AN INVESTIGATION INTO THE ELASTICITY OF MARGINAL UTILITY

by

Mateus Libardi de Carvalho

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

Department of Economics

Birmingham Business School

College of Social Sciences

University of Birmingham

September 2017

Abstract

Prioritizing public investments is arguably one of the most important and complex tasks Governments face. In this thesis, I contribute to such a task by examining the Elasticity of the Marginal Utility (EMU). This parameter is central to the determination of the Social Discount Rate, which is the discount rate used for Cost-Benefit Analysis in the public sector. I estimate the EMU using an unprecedentedly large dataset and test variants of the estimation technique which include National Insurance Contributions and Supernumerary Income. I also test the robustness of the estimates obtained. I further investigate the validity of the estimates by testing for the first time the key assumption underlying the estimation technique that the degree of progressivity of the income tax schedule represents society's inequality aversion. Next, I examine causality between tax progressivity and income inequality, which is a theme that emerges from testing the assumption mentioned. Finally, I estimate the EMU in different contexts, relating the estimated values and their context-sensitivity to psychological traits. Overall, the results suggest an EMU of 1.5 and that the estimation methodology implemented is acceptable. They also show bidirectional causality between progressivity and inequality, and that the EMU values vary significantly with psychological traits.

Acknowledgments

Firstly, I would like to thank the Brazilian Government for the funds provided during the four years of my PhD through the National Council of Science and Technology Coordination for the Improvement of Higher Education Personnel (CAPES).

Secondly, I would like to thank Prof. David Maddison and Prof. Jane Raymond for their patient supervision. David, thanks for the immense support and guidance, your careful comments and helpful assistance were essential. Jane, thanks for the insightful conversations and your readiness to help.

Thirdly, I would like to thank my family, mainly my wife Vanessa, my children Bonifácio, Manoel and Maria Isabel, my parents Elias and Rosa Angélica, and my sisters Sara and Rebeca.

Contents

List of Figures

Chapter 4

List of Tables

Appendix

Introduction to the Thesis

The pressure on public administrations to evaluate investments in a way that maximizes social welfare grows as societies organize around centralized Governments, which largely take on the complex responsibility for societal planning. The Social Discount Rate (SDR) (often estimated by the Ramsey rule) is essential for such activity, given it sets the discount rate used to assess public projects via Cost-Benefit Analysis (CBA), the main economic tool for prioritizing investments.[1] The fact discounting of future events has been depicted as “one of the most critical problems in all of economics” (Weitzman 2001: 260) gives an idea of the magnitude of the matter.

The SDR, already important for influencing the prioritization of investments with a large impact on society, becomes even more significant in times where policy makers increasingly have to deal with questions involving the very long run, as those regarding infra-structure and long-lived pollutants.

To illustrate the point consider an example related to climate change found in Gollier (2013). Using a discount rate of 5% Nordhaus (2008) estimated the Net Present Value (NPV) of future damages caused by the emission of one more tone of CO_2 as 8 dollars. In this scenario it would not be worthwhile to immediately use high technology projects to decrease carbon emissions, given it would cost more than 8 dollars for them to curb one tone of CO_2. On the other hand, using a discount rate of 1.4% Stern (2007) estimate the future damage of a tone of CO_2 emission as 85 dollars, in which case the immediate use of high technology projects would be acceptable.

The Ramsey rule gives the minimum return rate to compensate deferring consumption, and is composed of three parameters, one of which is the Elasticity of Marginal Utility (EMU or η).[2] The parameter reflects inequality aversion and makes reference to a normative value, i.e. to an ideal level of inequality aversion. It is clear that the quality of the SDR and therefore of investment decisions depend on the accuracy of the EMU estimation.

A relevant issue bearing on the accuracy of the EMU used in the determination of the SDR is the amount of long-run data used in the estimation process, given the discount rate is often applied

---

[1] In basic terms the CBA can be defined as a technique to compare projects by examining their benefits and costs scattered over time. It consists of measuring monetarily an investment’s benefits and costs (spread over time), bringing such values to their present value via an appropriate discount rate and then comparing the costs and benefits in order to see whether (and in what magnitude) one outweighs the other.

[2] Which in simple terms indicates the degree of satisfaction caused by a small increase in income (or consumption).

to evaluate long-run projects. If the parameter is derived with few observations it is more likely it will be subject to idiosyncrasies involving the events taking place at the time the observations refer to. On the other hand, if the EMU is obtained using long time series such specificities are dissolved, resulting in a more general and useful estimate.

Another factor bearing on the EMU estimation accuracy is the technique used to estimate the parameter, such that it is important to establish ways to identify reliable EMU estimation methods e.g. by testing alternative techniques' assumptions and the robustness of the resulting estimates. Nonetheless, it is important not to lose sight of the fact that the number of studies estimating the value of the EMU is not commensurate with the perceived importance of the parameter.

One of the methodologies occasionally used to estimate the EMU [the Equal Absolute Sacrifice Approach (EASA)] is based on data from income tax rates. The key assumption underlying the technique is that societal aversion to inequality is reflected in the Government's choice of income tax rates and the progressivity thereof.

In chapter 1 I contribute to the crucial task of providing an accurate estimate of the EMU as one of the three components defining the value of the SDR via Ramsey rule. I do that by estimating the parameter via EASA using an unprecedentedly large income tax dataset. I also test the adequacy of the utility function assumed in the EASA (which implies a constant EMU) by testing the constancy of the EMU across income levels, time and cultural background (countries). In addition, I test the robustness of the technique by comparing it's estimates with estimates obtained with other two different methodologies – each of them estimating the EMU in a different context (as risk aversion, intratemporal inequality aversion and intertemporal inequality aversion). These tests are carried out in order to determine the reliability of the EASA, given the impact different estimation techniques have on the EMU value.

In estimating the EMU via EASA and the technique's reliability, I systematically approach works using the estimation procedure and compare variants within the methodology in order to understand the extent to which these variants influence the estimated values. Although some studies have compared different approaches within the EASA, none has done it as systematically and as comprehensively as I do in chapter 1.

As mentioned above, the key assumption made in the EASA is that Governments express society's aversion to inequality through the degree of progressivity of the income tax schedules they set. After writing chapter 1, it drew my attention that, although such assumption is sometimes mentioned as one of the possible problems of the technique, there were no works (to the best my knowledge) actually testing it.

So, in chapter 2 I attempt to further validate the EMU estimates obtained in chapter 1 by testing the EASA underlying assumption that Governments set income tax rates in order to democratically represent societal inequality aversion. To do so, I investigate factors determining the EMU estimates obtained in chapter 1 (via EASA) allowing for the possibility that Governments take decisions without considering people's preferences, and thus that the EMU reflects administrations' inequality aversion only.

I take advantage of the fact the EMU estimated by the EASA can be interpreted as an index of tax progressivity and collect variables supposedly associated with tax progressivity to examine whether they are also associated with the EMU as measured by the EASA. It must be noticed, however, that my intention is not to analyse the determinants of tax progressivity. Rather, my interest is in the susceptibility of the EMU estimates of chapter 1 to variation due to factors having nothing to do with the underlying aversion to inequality of society.

The analysis includes variables describing the population of a country e.g. its demographic composition. The statistical significance of these variables could simply reflect that a society with a different demographic composition wants different things. On the other hand, variables whose statistical significance are harder to explain except for saying that Governments do indeed face constraints or have purposes apart from reflecting society's aversion to inequality, decrease the reliability of EMU estimates obtained by EASA.

The results suggest the dependency ratio,[3] Government stability, openness, population size and governmental expenditure with education are associated with the EASA measures of the EMU. Given the population size variable strongly affects the EMU estimates obtained by the EASA, my trust in the EASA is weakened somewhat, although not much. For in face of the evidence in favour of the technique found in chapter one and of the fact that all other variables considered in the concerned chapter do not strongly affect the estimates concerned, I think it is possible to regard EMU estimates obtained via EASA as reasonable measures of societal inequality aversion.

The causes and possible solutions to income inequality – which have been much debated lately – are themes that are closely related to the issue addressed in chapter two, in that if Governments set income tax schedules reflecting societal inequality aversion, one would expect tax progressivity to be used as a tool to curb actual income inequality. Such a use for tax

[3] Percentage of those with more than 65 and less than 14 in the population.

progressivity is often mentioned in political and economic discussions regarding inequality. The United Nations, for example, clearly points to "making the tax system more progressive" as one of the policies to be implemented to tackle inequalities[4] and Piketty (2014), when commenting on the results found by Piketty et al. (2011), claims only progressive taxation could put an end on the "skyrocketing executive pay", designated as one of the great causes of inequality.

Also politicians – Sanders, Corbyn and Mélenchon (in US, UK and French, respectively) are evident examples – clearly relate more progressive income taxes to less income inequality. In addition, standard optimal taxation theory regarding income tax, largely organized around the foundational work of Mirrlees (1971), points out that greater inequality tends to make the optimal tax more progressive (Mankiw et al., 2009).

Interestingly, few works test whether inequality really plays an important role in determining the degree of progressivity of income taxes. Many of them, instead, investigate the effectiveness of progressivity in decreasing inequality (e.g. Joumard et al., 2012; Cooper et al., 2015; Verbist and Figari, 2013; Duncan and Sabirianova Peter, 2012). However, if tax schedules are not used as a means to react to growing inequality levels the efficiency of progressivity is of little relevance in practice.

Thus, in chapter 3 I investigate the causal relationship between pre-tax income inequality and income tax progressivity.

It is remarkable that the few studies found examining the issue suggest that if Governments used progressivity to stop inequality before it is not the case anymore. Moreover, the only work (Scheve and Stasavage, 2016, referred to as SS) formally analysing causality between the variables (i.e. using statistical causality tests) found little evidence to support inequality is important to drive income tax progressivity. Their results suggest there is causality from progressivity to pre-tax inequality only.

It is important to stress that information on the phenomenon under scrutiny is highly important to address inequality issues, since if authorities setting income taxes are not sensitive to arguments involving actual inequality, progressivity will not be used to curb inequality even if it is an effective form to do so.

I build on the work aforementioned (SS, which was published whilst my study was ongoing) in that I also formally test causality between progressivity and inequality, but using different

[4] See United Nations Conference on Trade and Development: Tackling inequality through trade and development in the post-2015 development agenda (2014), page 12.

(arguably better) indexes of progressivity and income inequality, and different (again arguably better) causality tests.

Our results suggest there is causality both from progressivity to inequality and vice-versa, which contradicts SS's conclusions regarding the issue.

After comparing the EASA estimates with estimates from another 2 methodologies in chapter 1, I concluded the EMU does not vary across contexts, since each estimation method mentioned estimates the parameter in a different context and the estimated values obtained coincide. However, in chapter 1 I compare EMU estimates across contexts in a revealed-preference setting. In a stated-preferences setting not only the results may change, I can explain differences in the EMU across contexts by reference to psychological traits, which is interesting for a couple of reasons.

Therefore, in chapter 4 I test the EMU context-sensitivity in a stated-preferences setting and examine how psychological traits affect both EMU context-sensitivity and EMU value.

The research contributes to resolving a puzzle in the literature regarding the EMU context-sensitivity. Atkinson et al.'s (2009) results – obtained in a stated-preference setting – suggest the parameter changes significantly across contexts, while Groom and Maddison's (2013) results – obtained in a revealed-preference setting – suggest the opposite. None of them however, examine how EMU context-sensibility relates to psychological traits, which may account for some of the differences found in the results.

The analysis also contributes to the achievement of an EMU value possessing more relevance. To the extent that normative significance can be related to desirable psychological traits, it is possible to estimate more relevant EMU parameters if the sample which the estimation procedure is applied to is composed of individuals possessing desirable psychological traits. To explore such idea, however, it is necessary to define the traits associated with normatively desirable behaviour.

Two clear candidates for that, although others can certainly be included, are empathy and reflectivity. The first (which basically means the ability to put one's self in someone else's shoes) is directly associated with the notions of goodness and morality. Baron-Cohen (2011) defends what is commonly understood as evil can be better described as 'empathy erosion', and Rifkin (2009) claims an increase in global empathy is the only way humanity can survive 'global entropic collapse'. Irrespective of whether these perspectives are right (and they have been challenged), they underline the importance of the trait for policy makers. Reflectivity is important basically because of the potential harmfulness of cognitive biases in the context of social policies. The trait is responsible for making the mind "engaged", i.e. less superficial and "lazy" when facing problems. Thus, it is less likely that those highly reflective will be subject to cognitive biases. It can be understood, moreover, as the main aspect to be considered when investigating rationality (Stanovich, 2011).

In order to carry out the investigation proposed I apply an experiment containing 10 tasks. I use it to measure the EMU in three contexts (risk aversion, inequality aversion and time preference), and to measure participant's level of empathy and reflectiveness (it must be said I also consider other traits besides empathy and reflectiveness, but these two traits are the central ones).

The results show basically that the EMU context-sensitivity varies with the traits mentioned and that the EMU value varies mainly with empathy, suggesting the parameter is higher for more empathetic people.

References

Atkinson, G., Dietz, S., Helgeson, J., Hepburn, C., & Sælen, H. (2009). Siblings, Not Triplets: Social Preferences for Risk, Inequality and Time in Discounting Climate Change. Economics: The Open-Access, Open-Assessment E-Journal, 3, 2009-26.

Baron-Cohen, S. (2011). The Science of Evil. New York: Basic Books.

Cooper, D. H., Lutz, B. F., & Palumbo, M. G. (2015). The Role of Taxes in Mitigating Income Inequality Across the U.S. States.

Duncan, D., & Sabirianova Peter, K. (2012). Unequal Inequalities: Do Progressive Taxes Reduce Income Inequality? IZA Discussion Paper No. 6910.

Gollier, C. (2013). Pricing the Planet's Future. Oxfordshire: Princeton.

Groom, B., & Maddison, D. (2013). Non-identical Quadruplets: Four new estimates. Centre for Climate Change Economics and Policy, Working Paper n. 141 and Grantham Research Institute on Climate Change and the Environment, Working Paper n. 121.

Joumard, I., Pisu, M., & Bloch, D. (2012). Tackling Income Inequality The Role of Taxes and Transfers. OECD Journal: Economic Studies, published online first.

Mankiw, N. G., Weinzier M. C., & Danny F. Y. (2009). Optimal Taxation in Theory and Practice. Journal of Economic Perspectives, 23(4), 147-174.

Mirrlees, James A., (1971). An Exploration in the Theory of Optimal Income Taxation. Review of Economic Studies, 38, 175-208.

Nordhaus, W. D. (2008). A Question of Balance: Weighing the Options on Global Warming Policies. New Haven, CT: Yale University Press.

Piketty, T. (2014). Capital in the Twenty-First Century. Cambridge, Mass.: Harvard University Press.

Piketty, T., Saez, E., & Stantcheva, S. (2011). Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities. NBER Working Paper No. 17616.

Rifkin, J. (2009). The Empathic Civilization: The Race to Global Consciousness in a World in Crisis. London: Penguin.

Scheve, K., & Stasavage, D. (2016). Taxing the Rich: a History of Fiscal Fairness in the United States and Europe. Princeton: Princeton University Press.

Stanovich, K. (2011). Rationality and the Reflective Mind (First ed.). New York: Oxford University.

Stern, N. (2007). The Economics of Climate Change: the Stern Review. Cambridge: Cambridge University Press.

Verbist, G., Figari, F. (2013). The redistributive effect and progressivity of taxes revisited: An International Comparison across the European Union. GINI Discussion Paper 88.

Weitzman, M. L. (2001). Gamma Discounting. American Economic Review, 91(1), 260-271.

Chapter 1: Addressing the Equal Absolute Sacrifice Approach to Estimating the Elasticity of Marginal Utility Systematically

1.1. Introduction

The importance of how to evaluate public investments is straightforward since it directly affects society's resource allocation, and for such task discounting is indispensable. To illustrate the point consider two kinds of projects Governments are supposed to evaluate: one concerning the decrease of GHG emissions and another regarding small-size infrastructure improvements. The costs relative to both are incurred in the present and near future, but the time the benefits come vary greatly between them. The first has in view benefits that start to accrue in the distant future, whilst for the second the benefits come within a few years. To prioritize these projects it is necessary, among other things, to account for the time at which costs and benefits accrue, an exercise that requires a discount rate.

In this chapter I expect to contribute to this subject by estimating one of the compounding parameters - the elasticity of marginal utility (EMU or η) - of the Ramsey equation, which is widely used to determine the social discount rate (SDR), which in turn gives the minimum return rate a public project should yield in order to be undertaken. The EMU may also be used to derive welfare weights,[5] but in the present case the parameter estimation aims to contribute mainly to the public investment appraisal literature.

Cost-benefit analysis (CBA) is a relatively simple technique which enables public investments to be prioritized.[6] It basically consists of bringing (by an appropriate discount rate) all the investments impacts, which take place at different points in time, to present value equivalents. This procedure allows comparing in a straightforward way costs and benefits and identifying the

[5]Although the focus of the estimation here is not the use of η in weighting distributional impacts, it must be noticed that the coefficient essentially defines how much utility one derives from an infinitesimal variation in income, being therefore useful to establish the optimal way society should weight the income change (or welfare change) of the individuals or groups considered. So, if a project affects individuals possessing different income levels, and therefore different utility levels, the variation in social welfare induced by the project will be different depending on how much weight is given to the utility variation of the different parts, such that CBA may be carried out taking into consideration these weights to indicate the most valuable project in terms of welfare. Yet another use for η regarding inequality is the one concerning the measurement of actual income inequality by the procedure indicated in Atkinson (1970), which explicitly takes into account the social welfare function implied in ranking income distributions to measure inequality.

[6] The concept idea was first presented in Dupuit (1863) and further formalized by Marshall (1920) and Otto (1958).

most desirable projects. The efficiency of this technique, however, depends on the discount rate used (usually called social discount rate (SDR) when applied to public projects), and one of the most direct applications of the EMU parameter occurs in this context.

The conceptual justification of the CBA is given by the Kaldor-Hicks compensation idea (Kaldor, 1939, and Hicks, 1939) that basically states a project which can more than compensate those who lose with the implementation of it is a worthwhile undertaking. Nothing is said about whether the compensation does indeed take place or not, but just about the possibility of compensation.

In the literature there are basically four ways to get a defensible value for the social discount rate (Boardman et al., 2001): to set the SDR equal to the estimated Social Rate of Time Preference (SRTP) obtained via Ramsey rule, for example;[7] the Social Opportunity Cost (SOC); a weighted average of the rates given by the SRTP, SOC and the government's foreign borrowing rate; or the shadow price of capital. All these approaches to obtaining the SDR are very briefly commented upon in the next section, and the derivation of the SRTP via Ramsey rule is illustrated.

It is important to notice that methods to estimate the SDR apart from the SRTP (as estimated by the Ramsey equation) use information from financial markets. However, there are many different indices in such markets that could indicate the value of the SDR (Boardman et al., 2001), such that aggregating them all in order to estimate a unique discount rate becomes difficult. This fact clearly favours the use of the SRTP method. Moreover, tax revenues are arguably the main source of governmental resources (Pomerleau, 2014), implying that for the most part, governmental investments are funded with deferred consumption. Therefore, the SRTP method giving the return rate that compensates for consumption deferment also favours its use. Once the use of the STPR is established to derive the SDR, the importance of the EMU estimation is obvious to the public investment appraisal debate.[8]

There are a variety of different ways to approach the estimation of the EMU, and the result achieved is sensitive to the methodology employed. Some are based on stated preferences (via

---

[7] There are two ways to estimate the SRTP. One is derived from Ramsey's (1928) growth model and the other by looking at low risk securities in the market. Concerns relative to the second way of deriving the SRTP are that people systematically do not reflect inter-temporal preferences in the market and their individual preferences do not reflect what they think is best for society as a whole (Zhuang et al., 2007; Dasgrupta and Pearce, 1972).

[8] It also must be noticed that in a risk-free environment, with frictionless financial market, the SDR obtained by the STPR would be same as the one given by the financial market interest and the return on the marginal productive capital of the economy.

experiments and questionnaires), others on revealed preferences [via market data or tax data, like the Equal Absolute Sacrifice Approach (EASA) used herein], and others on ethical constraints mathematically represented imposed on specific models. All of them are presented in more detail in the course of the study.

Many estimates of the EMU are available, but with the dataset I use it is possible to provide more trustworthy estimates, since they take into consideration a long period of time and many countries. It also permits me to compare different variations of the EASA thereby contributing to the debate related to this approach. Moreover, the dataset permits to test some assumptions made by the EASA, such as the constancy of the EMU over time. To anticipate my results the calculated estimates clearly concentrate around 1.5.

To estimate the EMU with a large data is important given that, as argued by Groom and Maddison (2013), most public projects take many years to be materialise, such that a proper coefficient to be used in an SDR which intends to evaluate these projects should itself be based on many years of observation. Moreover, an SDR intended to consider intergenerational projects should take long run data into account due to the fact that different generations will have to keep on funding the investment in order to complete the undertaking.[9]

In the literature consulted regarding the estimate of EMU, only Van Dalen (1995) considered a larger dataset in terms of time (1830 to 1990) than the one presented here, using a life-cycle behavioural model. But this data is available only for the UK. Evans et al. (2005) and Kula (2004) are examples of studies that derived the parameter using large datasets (1963 to 2002 and 1965 to 1995, respectively), but still they do not cover as much time as the one used in the present study and involve observations from only one country (the UK and India, respectively).

The methodology that I use to estimate EMU requires data related to income tax rates. My interest in this approach is due to the large dataset on income tax data available to me, which permits me to estimate the parameter over the long run, for different generations and countries. More specifically a reasonably long time-series for income tax rates and hence EMU can be derived for nine different Western European countries using the dataset and the EASA approach. In what follows, the literature related to the EASA is comprehensively revised, such that all the

[9] Also insofar as the Ramsey formula depends on the growth rate this too has to be estimated over a significant period of time, such that the Ramsey formula has gone through considerable modifications due to for example uncertainty in the growth rate and autocorrelated growth rates. These modifications, however, highlight even more the importance of the EMU, as discussed in the next section.

works deriving the EMU by EASA are included. This has not been attempted so far, and it reveals several unresolved issues which I will attempt to address.

Lastly, I compare the EMU estimates obtained using the EASA with estimates obtained with other two revealed-preferences methodologies in order to validate the approach and check whether the value of 1.5 is consistent across the different estimation methods. One of the referred methods uses data on consumption growth rates and the other uses data on insurance to derive EMU estimates. The results indicate the value of EMU is broadly the same across the methodologies, which greatly reinforces the 1.5 value obtained via EASA. This is one of the most novel aspects of the chapter, given there are (to my knowledge) no other works comparing the EMU across different methodologies and countries with the amount of data considered by me.

In short, I organize and systematize the literature regarding the derivation of the EMU via EASA in that I classify the different estimation methods coexisting within the EASA framework and compare the parameter values obtained with each of them. I also validate the EASA comparing the EMU value obtained with it with estimates obtained with two other estimation methods. A further attempt to validate the EASA approach is deferred until the next chapter.

The remainder of the research is as follows. In the next section I make a theoretical review linking the parameter to be estimated to public investments and listing the most common ways to estimate it, including the EASA. In the next section I undertake a comprehensive literature review, describing and commenting on all the studies found employing EASA to obtain EMU estimates. Then I discuss the literature searched, trying to clarify the main ambiguities found in it and how they could be solved, providing a brief description of the dataset utilized just after. Next I present the results, giving details on the suggested EMU value. In section 7 I compare the estimates obtained with the EASA with estimates obtained with two other revealed-preferences methodologies. Section 8 concludes.

1.2. Theoretical review

In this section I give more details on the CBA, present the most widely used methods to obtain the SDR (which reveals the importance of the Ramsey equation in this context), illustrate how the Ramsey equation is derived and present the most common ways to obtain the η, including the EASA – the method to derive the EMU utilized in this study.

1.2.1. Cost-Benefit Analysis (CBA) and the Social Discount Rate (SDR)

The CBA is a tool that enables to prioritize investments and which is highly used in the public sector (see HM Treasury, 2003, as an example). It basically consists of estimating the benefits and costs of a project in monetary terms, through time (presuming the impacts occur at different points of time), and then converting them to their present values (PV) by applying a discount rate, making it possible to compare them even if they are scattered over time. The desirability of the project being evaluated can be judged in terms of the net present value (NPV) of the project.

Boardman et al. (2001) define 9 steps to undertake a CBA. The first is to identify the alternative projects relative to the counterfactual; the second is to decide whose benefits and costs are going to be considered; the third is to classify the project impacts as benefits or costs; the fourth is to predict how the impacts are going to be distributed over time; the fifth is to monetize the impacts; the sixth is to discount the benefits and costs to obtain their present values; the seventh is to calculate the net present value of each alternative; the eighth is to conduct sensitivity analysis for the alternatives, to see what is the effect of utilizing different values (for the impacts and discount rate, for example) in the calculations; and the ninth is to recommend one of the alternatives.

The procedure presented above clearly depends on the discount rate used to convert the costs and benefits to their PV equivalents (step 6). In the context of public sector investments evaluation, this discount rate is commonly called the Social Discount Rate (SDR), meaning the minimal real return rate an investment should yield. It must be stated that many projects are acutely sensitive to the choice of the SDR, mainly those whose benefits and costs are accrued in the long-run, like environment-related projects.

The SDR is (usually) obtained in five different ways. It may be equalized to the Social Rate of Time Preference (SRTP), which is described by the Ramsey equation, and basically indicates what the minimal future return an investment should give to make society willing to defer consumption. The equation is composed of two terms, the pure time preference parameter and a multiplication between the elasticity of the marginal utility (EMU) of consumption and the consumption growth parameters. The main criticism of this approach is that it does not take into consideration the effect of displacing investments undertaken by the private sector, being more appropriate to evaluate projects which are funded by deferred consumption solely (Baumol, 1968). On the other hand, there seems to be some agreement in the literature concerning SDR and the conclusion is that the SRTP is the most suitable to discount intergenerational projects (Zhuang et al., 2007).

The SDR can also be equalized to the marginal Social Opportunity Cost of capital (SOC). The basic idea underpinning this approach is that public investments mostly displaces private investment, and should therefore at least equal the rate of return given by private projects (Harberger, 1972). One of the ways to estimate the SOC is to search for the real pre-tax return on low risk corporate bonds, which can be made by, for example, looking at the monthly average yields of Mody's AAA corporate bonds (Boardman et al., 2001). Some shortcomings of this approach are that returns caused by negative externalities are not accounted for (Dasgupta et al., 1999), the fact that the risk structure faced by corporations is different from the one faced by Governments causes the discount rate to be overestimated (Arrow and Lind, 1997), and there is no fixed pool of investments, such that many public projects do not replace, but enhance and enable private investments (Morrison and Schwartz, 1996).

The SDR may also be equalized to the Governments borrowing rate, which is by many regarded as the rate reflecting the cost of financing the project (Boardman et al., 2001). A major criticism about this approach is that it does not take into account that in many countries governmental borrowing (even borrowing abroad) increases real domestic interest rates, which causes investments and consumption crowding out that is not accounted for (Boardman et al., 2001).

The other two approaches to obtain the SDR are the Weighted Average approach (WA) and the Shadow Price of Capital approach (SPC). The WA emerged as a result of different contributions (as Burgees, 1888 and Sandmo and Dreze, 1971), and is based on the idea of combining the rates obtained by the SRTP, the SOC and the governments foreign borrowing rate by averaging them, with the weights reflecting the proportion of funding coming from the different sources (consumptions deferment, private capital displacement and foreign borrowing). This approach, although considering the cost of deferring consumption and the private capital displacement, does not allow for the possibility of reinvestment of part of the public investment return, assuming implicitly it is all consumed. A complicating factor about the WA approach is that the weights obviously vary across projects, given that they are often funded in different ways, making it difficult for policy makers to follow this rule. Moreover, the critiques relative to the previous three approaches also apply to this one, since it combines them.

The SPC also makes use of the SRTP and takes into account the return a project may give to the private sector in the form of reinvestments, besides considering the private capital displacement caused by public investment. In this case the costs associated with the project are given by the decrease of current consumption and future consumption, the latter being caused by private capital displacement. And the benefits are given by the consumption of the return and by the reinvestment of it, which generates more consumption in the future. To illustrate the SPC I resort to the work of Zerbe and Dively (1994), where the net present value (NPV) calculated by the approach is given by

$$NPV = \sum_{t=0}^{N} \frac{B_t[g_b V + (1-g_b) - C_t[g_b V + (1-g_c)]}{(1+i)}, \tag{1.1}$$

where $B_t[g_b V + (1-g_b)]$ and $C_t[g_c V + (1-g_c)]$ are the consumption level equal to benefits and equal to costs at time t, respectively; g_b and g_c are the proportion of benefits that return to the private sector by reinvestment and the proportion of costs caused by displacement of private capital, respectively; i is the SRTP and V is the SPC parameter, which converts the benefits and costs caused by generation and displacement of private investments into consumption equivalents. Lyon (1990) proposes two alternative equations representing V:

$$V = \frac{r - sr}{i + d - sr}, \tag{1.2}$$

where r is the gross return on private investment, d is the rate of depreciation and s is the saving rate from r. Or

$$V = \frac{l - jl}{i - jl}, \tag{1.3}$$

where l is the return rate on private investment net of depreciation and j is the saving rate from j.

The shortcomings of this approach are that the SPC parameter may vary too much according to the parameters assumed in 1.2 and 1.3 (Lyon, 1990), and the resulting NPV varies according to the time it takes for the project to be completed (Harberger and Jenkins, 2002), which makes it quite difficult for policy makers to utilize this approach in their SDR estimations.

In practice, the SRTP and the SOC are the most used methods (mainly due to the complications related to the WA and SPC approaches), the SRTP preferred by developed countries (e.g., France, Italy, Spain, the UK and several US departments), whilst the SOC is preferred by the developing countries (Zhuang et al., 2007).

As seen above, besides being simple to apply and, in general, recognized as the best way to estimate the SDR associated with intergenerational projects, the SRTP is also used both in the WA and SPC approaches. Thus, I give special attention to this methodology for obtaining the SDR, illustrating its derivation (via Ramsey equation) below.

1.2.2. The Ramsey equation and the SRTP

In this section I illustrate how the Ramsey equation can be derived in a simple two-period context (based on Gollier, 2013), showing the relation between this formula and the EMU. Considering a two-period (current, represented by the subscript 0, and future, represented by the subscript 1) model and a utility function based on current and future consumption amounts, one can argue that the minimum return an investment [in which the costs are given in the present (0) and the return in the future (1)] should yield to be profitably made is given by

$$\Delta U = -\varepsilon U c_0 + R\varepsilon U c_1, \Delta U = 0, \tag{1.4}$$

where ε is the per capita investment cost, R is the gross interest rate returned by the project in period 1 (both are assumed to be small), U_{c0} is the derivative of the utility with respect to the current level of consumption and U_{c1} is the derivative of the utility with respect the future level of consumption. It means that the future return in terms of utility should at least compensate the loss of utility in the present. Rearranging it we have that

$$R = Uc_0/Uc_1. \tag{1.5}$$

Writing R in terms of an infinitely compounded interest rate we have that

$$R = e^{rt} = lnR = rt, \tag{1.6}$$

and then the infinitely compounded interest rate is

$$r = t^{-1}\ln(Uc_0/Uc_1). \tag{1.7}$$

Making the further assumption that the utility function is separable in time and discounted in a constant rate fashion (time consistency) we have that

$$U(c_0, c_1) = u(c_0) + u(c_1)e^{dt}, \tag{1.8}$$

where d is the rate of impatience through which the utility is discounted. Substituting equation 1.8 in 1.7 gives

$$r = t^{-1}\ln\left(\frac{u_0}{e^{dt}u_1}\right) = d - t^{-1}\ln(u_0/u_1). \tag{1.9}$$

Approximating u_1 by a first order Taylor expansion around c_0 gives

$$r = d - t^{-1}\ln(1 + \frac{c_1 + c_0}{c_0}\eta c_0), \tag{1.10}$$

where ηc_0 is the elasticity of marginal utility at c_0, given by

$$\eta c_0 = \frac{c_0 u''(c_0)}{u'(c_0)}. \tag{1.11}$$

Since

$$\ln\left(1 + \frac{c_1 - c_0}{c_0}\eta c_0\right) \cong -\frac{c_1 - c_0}{c_0}\eta c_0, \tag{1.12}$$

we have that

$$r \cong d + \frac{c_1 - c_0}{t c_0}\eta c_0, \tag{1.13}$$

which is the Ramsey equation. It shows that two factors should be considered in the choice of the SDR, the pure time preference, or impatience and the wealth effect. The latter is composed of two elements: the consumption growth and the EMU, which is interpreted in this context as inequality aversion. The Ramsey formula is commonly written as

$$r = d + \eta g, \tag{1.14}$$

where d is the pure rate of time preference, η is the elasticity of marginal utility of consumption (income) and g is the average rate of growth of per capita real consumption (income). The right hand side of equation 1.14 is conventionally called social time preference rate (STPR), and essentially measures the rate of return some investment should provide in order to preserve the welfare level negatively affected by consumption rescheduling.

It is also important to notice further theoretical developments have been made affecting the Ramsey rule that allow, for example, to account for growth rate uncertainty. In the case where growth is assumed to be i.i.d and to follow a normal distribution with variance σ^2, the extended Ramsey rule can for example be written as

$$r = d + \eta \bar{g} - 0.5\eta(\eta + 1)\sigma^2, \tag{1.15}$$

where $\bar{g}$ is the growth rate of expected consumption. Notice that besides the pure time preference rate d and ηg, which refers to the wealth effect,[10] the equation includes another term which reflects the 'prudence effect', i.e. the understanding that people tend to invest (save) more when future consumption is not certain. The extension of the Ramsey rule further highlights the importance of the EMU (η) for the SDR; besides reflecting inequality aversion it also appears in an additional component referred to as the 'prudence effect'.

[10] Notice that the term ηg reflects the idea that if g is expected to be high people will be more willing to invest only if the investment return is higher, which gives the intuition as to why η is interpreted as a parameter indicating inequality aversion.

1.2.3. Methods to estimate η

Having established the importance of this parameter in this section I briefly explain the main methods found in the literature to derive EMU. I also provide some remarks about the diverse possible interpretations of the parameter.

There are basically six methods to derive EMU: those eliciting inequality or risk aversion by experiments or surveys, those doing it by analysing lifetime consumption behaviour, demand for wants-independent goods or insurance data (thus collecting data based on individuals' revealed preferences), those doing it by observing governmental income tax schedules (looking at socially revealed values) and those doing it by choosing an η which satisfies some desirable ethical properties in an intergenerational context model (essentially deriving the EMU by an ethical thought experiment). Further details on all the methods cited are provided later.

It must noticed that in the usual model to investigate societal welfare (where the utility functions are assumed to be iso-elastic and the social welfare function is assumed to be utilitarian) η can be interpreted in three ways: as an indicator of intra-generational inequality aversion, intergenerational inequality aversion and risk aversion. It is the diversity of ways in which the EMU can be interpreted which opens the door to the wide variety of methods for estimating its value.

The literature is ambiguous about whether this representation of the elasticity of marginal utility is consistent. Atkinson et al. (2009), for example, sustain that people do not think of the three referred aspects as the same, whilst Groom and Maddison (2013) find, for UK data, that different methodologies (which result from different views on how to interpret η, as exemplified below) for estimating η generate statistically equivalent results, which can be interpreted as evidence that the three aspects should indeed be regarded as equal. Also there seems to be neuroscientific evidence suggesting a linkage between intergenerational and intra-generational inequality aversion (Da Silva et al., 2015), and Epstein and Zin (1989) disentangle inter-temporal substitution and risk aversion as generalizing the expected-utility model.[11]

[11] They basically derive a class of recursive utility function over inter-temporal consumption lotteries for which the von Neumann-Morgenstern expected utility function is just one of the possible cases. The time aggregator is assumed to be a linear homogeneous CES, and different certainty-equivalent functions (risk aggregator) relative to future utilities are considered. As a means of example regard it as $U_t(U_{t+1}) = [E_t U_{1+1}^{a}]^{1/a}$, where U_t is the utility at time t, E_t is the expected value conditional on information given at t, and a implies risk aversion. Once the U_t is given by $[(1-\beta)c_t^r + \beta\mu_t(U_{t+1})^r]^{1/r}$, where β gives the marginal rate of time preference and r implies the elasticity of inter-temporal substitution, it can be seen how risk aversion is disentangled from inter-temporal substitution.

The different approaches to estimate η reflect these varied ways of understanding the parameter. The estimation based on income tax rates, for example, assumes the progressivity of an egalitarian (in terms of equal absolute sacrifice, as discussed below) tax schedule reveals the degree of inequality aversion. This way of eliciting η clearly focuses on intra-generational inequality aversion, once tax schedules are basically income transfers amongst different persons at a given time and world state. The approach based on lifetime consumption behaviour measures the elasticity of inter-temporal substitution (EIS) which given some assumptions is equal to the reciprocal of EMU by maximizing an inter-temporal utility function. The fact the model is concerned with individual (not societal) utility maximization over time shows the coefficient in this case is interpreted primarily as an inter-temporal inequality index. Moreover, considering the expected utility framework, where individual utility is dependent on different states of the world, it is largely known that the relative risk aversion (RRA) measure (equal to η for an iso-elastic utility function) gives the degree of risk aversion, such that studies concerned with EMU estimation may instead interpret it as a parameter indicating risk aversion.

Below I detail four of the six methods to estimate the EMU. The other two are described in section 1.7, where I compare the EMU value estimated via EASA with EMU values obtained from analysing consumption growth rates and insurance data.

1.2.3.1. Survey methods

The survey method is exemplified by studies like Pirttila and Uusitalo (2008), Amiel et al. (1999) and Carlsson (2005), which use questions about wage distributions and/or of the leaky bucket type (addressed below) to elicit inequality aversion.[12] The first approach basically poses income distributions differing in mean and variance to be seen by the respondents, who have to choose a preferred one. The leaky bucket approach infers inequality aversion by asking questions bearing the idea of a transfer from a rich to a poor in which just part of the contribution reaches the latter.[13] The amount of loss undergone by the rich which is accepted by the respondent indicates how inequality averse her or she is. In a context where an iso-elastic social utility function is assumed, a person with an EMU equal to 2 would accept someone five times richer than another to make a sacrifice of 25 currency units to transfer just 1 to the poorer, for example (Cowell and Gardiner, 1999).

---

[12] There are other methods of eliciting η via experiments, as for example looking at self-reported wellbeing (Layard et al., 2008), or at risk aversion (Barsky et al., 1997). The two methods presented here, however, seem to be the most popular and direct way of measuring inequality aversion by surveys.

[13] See Appendix 1.A for an example of a leaky bucket type question.

The comparison between different income distributions can reveal the EMU if one assumes individuals rank the distributions according to a given criterion function which involves the EMU. So, for example, if after devising two distributions which should be hypothetically ranked equally by an individual with EMU equal to 2, the respondent comparing them is indifferent between the two, such choice can be interpreted as indicating that his EMU is indeed around 2. On the other hand, if he prefers the more egalitarian distribution, such choice can be seen as an indication that his EMU is larger than 2. By contrast, if he chooses the less equalitarian distribution, he can be seen as possessing an EMU smaller than 2.[14]

The main problem faced by this kind of research is that estimates vary across context and with the wording of survey questions,[15] something which casts doubt on any particular value obtained based on a given survey. Also, the representativeness of the participants (students in many cases or experts) is an issue, that may limit the relevance of the estimate obtained (see Carlsson et al., 2005 and Amiel et al., 1999).

1.2.3.2 Method based on demand for wants-independent goods

It is also possible to derive EMU by considering the demand for wants-independent goods. These are goods for which the marginal utility of consumption is independent of the amount of other goods consumed. So a utility function in which food and non-food goods are taken as wants-independent would imply the form

$$U = u_f(f) + u_n(n), \tag{1.16}$$

where f is the amount of food goods consumed and n is the non-food goods consumed. In this case the increase in utility caused by an increase in the amount of food consumed is independent from the consumption of all other goods. Given an additively separable utility function Frisch (1959) shows that

$$\eta = \frac{k_i(1-w_i k_i)}{\varepsilon_i}, \tag{1.17}$$

where k_i is the income elasticity of demand, w_i is the income share and ε_i is the own compensated elasticity for i.

---

[14] See Appendix 1.A for an example of this type of question.

[15] Pirttila and Uusitalo (2008) find a median inequality aversion below 0.5 using the leaky bucket type question, which is in accordance with Amiel et al. (1999), who use the same technique and get a median inequality aversion between 0.1 and 0.2. Using the income distribution comparison, Pirttila and Uusitalo (2008) find a median inequality aversion larger than 3, which is a result similar to the one found by Carlsson et al. (2005), who find a median inequality aversion between 1 and 2 using the same kind of question.

The most common assumption utilized for EMU estimation purposes is the one assuming wants independence between food and non-food goods. The next step is to estimate the demand for food[16] to obtain the parameters entering the Frisch equation.

The additive preference assumption is probably the factor posing most difficulties to the approach's acceptance. Some do not even consider the method in their reviews of estimates of EMU (e.g. Cowell and Gardiner, 1999 and Pearce and Ulph, 1995). On the other hand, some consider it is reasonable to assume that food enters the utility function in an additive way (as Fellner, 1967; Evans and Sezer, 2002; Kula, 2002 among others), and Groom and Maddison (2013) test the assumption and find it holds.

1.2.3.3. Method based on ethical criteria imposition

All the methods to measure η seen before are positive (as opposed to normative), in the sense that they look for a parameter consistent with people's behaviour. The problem with this kind of approach is policy makers are arguably more interested in working with a normative value for the EMU, i.e. in an inequality aversion coefficient which is more consistent with ethical requirements. To address this issue, Buchholz and Schumacher (2010) derive EMU by imposing ethical criteria in an inter-generational allocation context. What they basically do is to assume a two-generational model with a linear transformation curve and no pure time discount and impose ethical restrictions mathematically represented in the model (such as solidarity between generations, meaning that both generations should be better off if productivity increases). By proceeding in this way, they define a value of 2 for EMU.

This way to approach the question of the correct value of EMU has the merit of explicitly considering the normative nature of the issue. The ethical principle underpinning measurements of EMU based on positive approaches is that policy makers should use an estimate of EMU consistent with society's behaviour as revealed in markets (savings behaviour or wants-independent goods), via experiments or tax schedules (as discussed below). But if Governments are interested primarily in utilizing ethically defensible values of EMU, why would they accept any estimate based on a revealed preference approach?

[16] The most used models for estimating the demand for food are the CEM (constant elasticity model), AIDS (almost ideal demand system) and QAIDS (quadratic almost ideal demand system) models. For more detail on their use to estimate η see Evans et al. (2005).

One could however argue there is merit in the positive type of estimate presented, which is that Governments should represent its population, and therefore apply a number as representative of society as possible.

1.2.3.4. The EASA

The equal absolute sacrifice approach (EASA), the method used in this study, basically extracts EMU – interpreted in this context as an index of societal inequality aversion – from income tax progressivity data.

The assumptions needed to apply the estimation method are twofold. The first is that the income tax schedule considered is grounded on the Equal Absolute Sacrifice Principle (EASP), which basically states that taxes should be designed in such a way that all individuals would pay equally in terms of sacrifice. This principle was first stated by Mill (1848): "Equality of taxation, therefore, as a maxim of politics, mean equality of sacrifice". It can be expressed mathematically as[17]

$$U_i(Y_i) - U_i\big(Y_i - T(Y_i)\big) = c,\ \ i = 1, 2, 3, \dots, N, \tag{1.18}$$

where U_i is individual i's utility function, Y_i is individual i's income, T is the tax applicable to Y_i and c is a constant (and society is composed by N tax payers). A form for the utility function also has to be assumed, and usually it is the isoelastic form (all the studies reviewed in the literature review use an iso-elastic utility function). Thus, the presuppositions can be summarized in equations 1.18 above and 1.19 (representing an iso-elastic utility) presented below.

$$U_i = \frac{Y_i^{1-\eta} - 1}{1-\eta}. \tag{1.19}$$

Substituting 1.19 in 1.18 gives

$$\frac{Y_i^{1-\eta} - 1}{1-\eta} - \frac{\left[\big(Y_i - T(Y_i)\big)^{1-\eta} - 1\right]}{1-\eta} = c. \tag{1.20}$$

Differentiating 1.20 with respect to Y_i and isolating η yields

[17] Notice there is other principle related to equal sacrifice in taxation, which is the equal relative sacrifice principle (ERSP), given by $\frac{U(Y-T)}{U(Y)} = 1 - c$. Notice as well that this principle corresponds to using the EASP with $lnU(Y)$ as the utility function instead of $U(Y)$ (Young, 1990). Since there is no prior reason for us to use the log utility form in the EASP and the ERSP can be thought as a particular case of the EASP, just the EASP is considered in the study.

$$\eta = \frac{\ln(1-\frac{\partial T(Y_i)}{\partial Y_i})}{\ln(1-\frac{T(Y_i)}{Y_i})}, \tag{1.21}$$

where $\partial T(Y_i)/\partial Y_i$ is the marginal tax rate (MTR) and $T(Y_i)/Y_i$ is the average tax rate (ATR).

There are two different ways to estimate EMU by the model described above. The first is referred to as direct calculation, and consists on simply calculating the MTR and ATR at a given income level, thus obtaining EMU. The second is the regression method, consisting of running the following regression:

$$\log(1 - MTR_i) = \eta \log(1 - ATR_i) + \varepsilon_i, \tag{1.22}$$

where it is assumed that ε is normally distributed, has mean zero and constant variance, and i indicates a point of the income distribution at which the MTR and ATR are derived.

It must be noticed that in the present context the regression is just a means of finding the curve best fitting the terms $\log(1 - MTR_i)$ and $\log(1 - ATR_i)$, there being no special interpretation for the error term, such that it just indicates the deviation of the observed values from the curve.

The direct calculation is simpler, but considers just one income level and/or year, whilst the regression method potentially considers many income levels and/or years, such that although the latter is in general more desirable, it is more demanding in terms of the number of observations required. In this study I use both methods.

Direct calculation is used to obtain the estimates corresponding to different years, countries, income levels and tax units, whilst the regression approach is used to compare different ways of implementing the EASA methodology. For each year, country, income level and tax unit there is just one observation on the ATR and MTR, such that it is not possible to use the regression approach. On the other hand, once the ATR and MTR are calculated for many countries, or years, or income levels and so on, it is possible to use the regression approach (the study estimation procedure is further explained ahead in the chapter).

Some concerns related to the method are that it cannot be exactly known to what extent a tax income schedule follows the equal absolute sacrifice principle (EASP), since to do this it would be necessary to know the utility functions of the individuals. Clearly it is not possible to fully test the isoelastic utility function assumption, i.e. there is no way to provide a conclusive answer about people's utility function form, although indirect evidence may indicate some general features, as for example in relation to the constancy of the EMU. Some empirical evidence for the constancy of the elasticity of marginal utility is provided by Blue and Tweeten (1997) and

Evans (2005). Richter (1983), Vitaliano (1977) and Young (1987) find empirical support for the influence of the EASP in tax structures.

One should also be aware that other principles, such as work incentives (Spackman, 2004), also influence the formulation of income tax schedules, something which undermines the assumption of the EASP as the only motivation for setting income tax schedules.[18] I will return to this topic in the next chapter.

Finally, it must be noted that also the EASA is a revealed preference approach, being therefore not based upon normative criteria. This could be viewed as a disadvantage. However, one could regard the EASP as a normative principle to be followed by the Government: the Government ought to implement tax schedules based on this principle. If that happens then there should be a close correspondence between the curvature of the utility function and the progressivity of the income tax schedule. In other words, in such case a simple criterion to decide what countries set the tax schedule closer to the norm would be to consider how well $\log(1-MTR)$ and $\log(1-ATR)$ fit into a curve, as shown in equation 1.22. It would indicate the extent to which a country follows the EASP when setting income tax schedules.

1.3. Literature review on η estimates based on EASA

Before the literature review a brief description of how it was undertaken is presented. The review is intended to be as comprehensive as possible, listing and commenting on all works estimating η by EASA. Lastly, a short comment on the studies reviewed intends to further classify them.

1.3.1. Brief description of literature review undertaking

Since the current chapter aims to review works in which the EMU is estimated by EASA, and this parameter is often used for SDR or welfare weights derivation, the key words employed in the search were separated into three groups: the first comprises words related to the parameter's name, the second contains words referring to the context in which the parameter is required and the third relates to the methodological assumption underpinning the technique. These words are presented below.

Table 1.1. Key words categorization

Group 1	Group 2	Group 3
Elasticity of marginal utility	Social discount rate	EASA
Inequality aversion	Welfare weights	Estimation

[18] Da Costa and Pereira (2014) argue, however, that in an environment equal to the one found in Mirrlees (1971) and an income tax schedule following the EASP, inefficiency arises just at the top of the income distribution and for MTR above the ones seen nowadays in most developed countries.

Source: see text.

Using the key words above and Boolean operators the word combination by which the search was undertaken can be written as (“Elasticity of marginal utility" OR “Inequality aversion" OR “social discount rate" OR “welfare weight*") AND (sacrifi* OR approach OR estimat*).[19] The databases employed were EconLit, ProQuest and Web of Science, and the number of works retrieved in each database were 467, 332 and 61 respectively. Since many studies found in the EconLit and ProQuest search were irrelevant, the analysis had to be narrowed by diminishing the number of words contained in groups 1 and 3. Following this strategy just 17 works (from both databases) were kept for further investigation, and out of these, only 10 were used in the literature review.

The search undertaken in ProQuest did not present any novelty in relation to what was found in the EconLit database. In the analysis with the Web of Science database, on the other hand, 2 works reviewed in the literature review below were detected by a citation search involving 2 key articles. In addition, 2 more works reviewed in the literature review were found by looking at the references of papers obtained in the search described above. A summary of the bibliographic investigations carried out is presented below.

Table 1.2. Summary of literature review research

Database	Total of entries	Saved for further research	Used
EconLit	467	17	10
ProQuest	332	-	-
Web of Science	61	4	2
References	-	-	2

Source: see text.

1.3.2. Literature review

This section reviews works estimating EMU by EASA. All the studies found in the search described before are presented (in chronological order) and briefly discussed.

In the literature consulted for this work, the first author to explicitly use the EASA was Stern (1977). He used UK data for the tax year 1973/1974 and tax rates for families with two children, achieving results for EMU by regression. To calculate the average tax rate (ATR) he did not use supernumerary income, i.e. did not subtract subsistence income from the taxable income

[19] The asterisk indicates all different forms the word ends in must be tracked. So if one searches for estimat*, for example, words like estimate or estimation will be tracked. The operator OR indicates interchangeable names which should all be tracked out. So once η can be tracked by various names, such as elasticity of marginal utility or inequality aversion, both names are, in this case, assigned as interchangeable. The operator AND indicates conjunction, such that in the case of the present research, for example, elasticity of marginal utility must be tracked in conjunction with the words sacrifi* or in conjunction with the word approach and so on.

amount. Cowell and Gardiner (1999) seem to be the second to estimate EMU by the EASA.[20] They did so for the tax years of 1998/1999 and 1999/2000, using a single non-elderly man with no special circumstances tax rates (also for UK). They also used regression, but it is not clear how they calculated the ATR. In addition to the results obtained just with income tax, the parameter was also estimated with income tax and employee's National Insurance Contributions (NICs) [Stern (1977) uses income tax rates only]. Stern (1977) obtained a value of 1.97 for EMU and Cowell and Gardiner (1999) obtained a value of 1.43 and 1.41 (for the tax year of 1998/1999 and 1999/2000, respectively) for EMU when considering just income tax data and 1.29 and 1.28 (for the tax year of 1998/1999 and 1999/2000, respectively) when considering income tax and NICs data. Although Stern (1977) suggests weighting the data by number of tax payers in a given income category, none of the aforementioned works weighted the data.

We also find estimates of the EMU using EASA in Evans and Sezer (2004), where the main goal is indeed to establish a SDR grounded on the same theoretical framework (STRP) for different nations (Australia, France, Germany, Japan, UK and US) in order to yield a consistent criteria for public resource allocation in these countries.[21] Data was considered for the year 2001 and for single persons in full time employment with no dependants.[22] They use direct calculation to obtain the final values and NICs are included (this is subsequently referred to as all-in taxes).

The justification for including NICs is that the EMU in this case were revealed to be steadier and more plausible by prior experimentation. They do not inform, however, what and how such experimentations were applied, besides they leave aside the issue of whether only the "steadiness" of the results found is enough to justify the inclusion of NICs. Results for EMU are first presented for the average production wage (APW) in manufacturing industries. In this context, the outcomes achieved are in the interval of 1.3 to 1.6. The highest (1.6) is for Germany and the lowest (1.3) are for France and the US. They draw attention, moreover, to the fact that if a change in the tax rate comes just below (above) the APW level, the η derived will be overestimated (underestimated). This matter gives rise to another set of η estimates (one estimate for each country) calculated by averaging the results achieved for each different wage

[20] Young (1990) also derives η using the EASP and an iso-elastic utility function. The latter is, however, of a generic form, and the procedure employed to obtain the parameter is not the one described in section 1.2.3.4. In the study the η is given by regressing $\ln(T(Y))$ against $0.5\ln(Y(Y-T(Y)))$. It must be noticed that the derivation shown in the section referred is also valid for a generic iso-elastic utility function and that Young (1990) makes some further assumptions and approximations in order to derive η.
[21] The issue of adopting a consistent SDR among countries is further commented ahead.
[22] They follow Cowell and Gardiner (1999) in this aspect instead of Stern (1977) who considered families. To collect data for single person started indeed to be the case for virtually every work using EASA since Cowell and Gardiner (1999).

levels (proportional to APW, namely, 0.67xAPW, APW, 1.33xAPW, 1.67xAPW). In this second analysis the parameter was between 1.28 and 1.72 (for France and Australia respectively).

Evans and Sezers (2004) calculate η at different income levels and averaging the results does not seem to solve completely the problem of over (under) estimation, but just to soften it, since the tax rate change may come before or after the income levels they consider. Moreover, the resulting estimates are weakened in terms of representativeness, given there are fewer people in the income levels lower and higher than the APW.

Evans (2004) focuses on deriving a SDR for France and on the way he estimates η by EASA. The worry related to France is a result of the considerably high social discount rate adopted by the country in public investment appraisal. In the referred to work the regression approach is not considered appropriate because of the small sample size. It is also claimed that it does not make behavioural sense to suppose that the ATR causes the MTR, as it is implied in the methodology under consideration, such that he uses direct calculation, estimating η at different points of the income distribution. According to Stern (1977), however, the regression is just used to obtain a best fit, such that there is no suitable causality interpretation in this case.

The η estimates refer to the proportions of the APW in French manufacturing industries [the same ones used in Evans and Sezer (2004)]. The data utilized was for the year 2001 and single persons with full time employment and no dependants. To calculate the ATR supernumerary income was used. The justification given is that it would make more theoretical sense to apply the decreasing marginal utility of consumption concept (implied in the utility function assumed) to the income in excess of personal tax allowance. The authors supposed that the income amount under the allowance level would correspond to the basic subsistence level, under which it would not make sense to consider decreasing marginal utility of consumption.[23] The average η was 1.35 and the estimates for different income levels were between 1.3 and 1.4. The parameter calculated for the income level equal to 1.33xAPW (η = 1.56) was not considered since there was detected a change in the marginal tax rate just below this wage level, which caused the observed estimate to be overestimated.[24]

As pointed out before in the introduction, the EMU with respect to consumption is essential to the achievement of welfare weights, and in Evans et al. (2005) the effort to estimate η is a step

[23] This argument favouring the use of supernumerary income seems to have influenced other researchers regarding standard allowances.
[24] Note that the same method of averaging η across different income levels found in Evans and Sezer (2004) was adopted.

to obtain distributional regional welfare weights for the UK (England, Scotland, Wales and Northern Ireland), which are defined in the considered work as

$$w_j = (Y^*/Y_j)^\eta, \tag{1.23}$$

where w_j is the distributional weight for region j, Y_j is the per-capita income in region j and Y^* is the national per-capita income. Basically, the purpose in estimating these weights is to use them in CBA, thereby accounting for equity issues regarding regional income distribution.

The data is for personal tax of 2002/03 and the results are obtained via regression. Estimates are achieved with weighted (by number of tax payers or income amount) and non-weighted data, and the intercept is not assumed to be 0, so that before removing it, it's statistical significance is tested. This procedure is said to detect to some extent whether the model assumptions have empirical support. There is no comment, however, on how to interpret an intercept in the given context, and on exactly how the procedure tests the model. Another device used in the paper is to calculate the reciprocal of η in order to improve the causal relation underpinning the model, since according to the authors it makes more sense to assume that variation in the ATR depends on variation on MTR. As seen before, however, there is no point on thinking about causality in this context, once the regression is just a curve-fitting procedure. Two models were selected as being preferred. In the first the reciprocal of η is used, the data is weighted by income[25] and, after the intercept was shown to be statistically insignificant, it was removed, rendering an estimate of 1.63. In the second, η (=1.57) is estimated directly, the data is also weighted by income and there is no intercept (it was removed after displaying statistical insignificance). The models were preferred because the intercepts were not significant, the R^2 were higher and the Durbin-Watson (DW) statistic suggested less correlation in the error than for the other models. In addition, in the case of non-weighted data the intercept was significant, which was seen by the authors as empirical support for weighting data.

Evans and Sezer (2005) propose a SDR grounded on a consistent basis (STPR) for several EU members (19 countries), and on the way they estimate η by EASA. They stress the importance of consistency in SDR values for the EU in a context of expansion and development of a single European market, although without giving any further explanation of why consistency across countries is important. It seems reasonable to think, however, that large SDR disparities would preclude different countries from investing in a common project, once their considerations on the profitability of the project would differ.

[25] No specific reason or explanation is given to weight the observations by income.

The tax data used is for the year 2001 and single persons with no dependents. All-in tax rates were used (income tax rates and NICs), and the justification for this procedure is the same given in Evans and Sezer (2004). The estimations are obtained by direct calculation, first for the APW in manufacturing for each country and after for different levels of wage [same proportions of APW presented in Evans and Sezer (2004), mentioned before]. Using just the APW the results were between 1.11 and 1.81; averaging the η calculated for different income levels and for each country, the results were between 1.1 and 2. This second result seems much less representative, once using income levels different of the APW. It is not clear if they used supernumerary income for the ATR calculation, and some values of η related to specific levels of income and countries were considered out of line (it is, too different from the results obtained for other income levels in the same country) and then excluded from the average η calculation. However, it is not clearly stated what values were excluded, and no rigorous criteria for exclusion is provided.

In Evans (2005) the estimation of η is carried out for 20 OECD countries, and the result is close to 1.4 (for average η) for developed countries which supports the argument that the UK Treasury's choice of η (equal to 1) should be revised. The data utilized is for income tax rates on gross wage earning of single person with no dependants in 2002. Supernumerary income is used in the ATR calculation and its usage is justified in the same way as in Evans (2004). The author does not use all-in tax rates in this case, but tax-only data (does not include NICs). The reason given by the author is that the NICs reflect a different motivation in relation to income tax, so that the inclusion of the former could compromise the legitimacy of the model, once the EASA assumes the data reflects the EASP.[26] Because of the small sample and the already commented upon opinions of the author concerning the EASA causality inaccuracy, the regression approach is put away, and η is directly calculated for different levels of income (high, substantially above the APW in the manufacturing industries, and low, close to the APW) for each country.

After obtaining values for high and low income, the average is calculated for all the countries, as well as the differences between the parameters obtained with the high and low incomes (and also the squared differences). With these results at hand the author tests (by t-test and non-parametric Wilcoxon signed ranks test) if there is a significant statistical difference between the parameters calculated at low and high income. The tests support the constancy of η across income level.

[26] Although this justification is contradicted by the results obtained in Evans and Sezer (2004) and Evans and Sezer (2005) it seems it had some success in raising discussion about NICs inclusion.

The estimates obtained for the countries were also regressed in different ways, with two of them taken as preferred. In the first the regression was run as usual and in the second the parameter estimated was the reciprocal of η. As before, this second approach reveals a concern about the causality of the regression, which does not seem to be justified, given the regression should be seen just as a curve fitting procedure. The preferred specifications for both models give estimates close to 1.4.

Sezer (2006) estimates η by EASA to calculate welfare weights based on regional incomes for Turkey. He draws attention to the fact that sometimes EASA is the only method at hand to assess the parameter (which is the case for Turkey). It is a demonstration of the approach's importance, since occasionally it is the only one which can be applied. The data is for the year 2004 and η is derived by direct calculation. The ATR and MTR are obtained for the midpoint incomes of each taxable band, and supernumerary income is used. The result is an η of approximately 1.25 (excluding the first income group, for which the result is equal to unity, the parameter falls between 1.2 and 1.27). It is not clear if NICs were considered and if the data was collected for single persons.

Lopez (2008) estimates SDR by STPR for nine Latin American countries (Argentina, Bolivia, Brazil, Chile, Colombia, Honduras, Nicaragua, Mexico, Peru), using EASA to derive η. There are two main difficulties to apply the approach in these countries. The first is that, for some cases, the tax evasion is high, something which implies that the 'de jure' statutory tax rates are different from the 'de facto' paid tax rates. He addresses this issue relying on the 'de facto' tax rates. The second is the choice of income level at which to calculate the η. For this case he divides the income distribution in five groups based on points corresponding to 20, 40, 60 and 80 percent of the distribution. After addressing these matters, two methods (both by direct calculation) are used to obtain the estimates. In the first he averages the MTR and ATR calculated as showed below across the quintiles (i) and then calculates η with these average values, and in the second he estimates an η for each income level - for each country - at the income distribution percentages referred above and then averages the parameters derived for each corresponding income level. The MRT and ART for the first method are obtained by the following equations:

$$MRT_i = T_{i+1} - \frac{T_i}{Y_{i+1}} - Y_i, i = 1, 2, 3, 4, \tag{1.24}$$

and

$$ART_i = T_{i+1} + \frac{T_i}{Y_{i+1}} + Y_i, i = 1, 2, 3, 4, \tag{1.25}$$

where T_i and Yi are the average taxes and average per capita income level, respectively, for the i-th quintile. The averaged η obtained with the first method is 1.5, whilst with the second it is 1.6. He uses the parameters obtained with the first to derive the SDR of the countries considered, giving no clear explanation of why the first method was preferred. It is not clear if the tax rates are for single persons or families, if it was used supernumerary income and if NICs were included.

Percoco (2008) also estimates η by EASA to find an appropriate SDR for Italy (by the STPR approach). He used data for the year 2004, and claims to follow the methodology of Evans and Sezer (2004) and Evans (2004) regarding the EASA. By this we can infer that the data was also related to single person with no dependants and that supernumerary income was used. However, information concerning the data is not clearly stated. He utilizes direct calculation to estimate η, and it was done for different points of the income distribution [the same proportions of APW seen in Evans (2004)]. The average η estimated was equal to 1.347.

In Evans (2008) η is estimated by EASA (and other methods), and a short literature review about the cited approach comprising the works of Stern (1977), Cowell and Gardiner (1999), Evans (2005) and Evans and Sezer (2005) is presented. The data utilized for the estimation of η is for the year 2005/2006 and single person households (in the UK). The income level at which the parameter is obtained (by direct calculation) is the APW of adult, full-time workers. Supernumerary income is used based on the same rationale mentioned in Evans (2004), NICs are not included [the justification is the same one presented in Evans (2005)], and data is not weighted on grounds that when η is calculated to be included in the STPR equation, "it is the circumstances of the average household that count" (Evans, 2008).[27] The calculation result for η is 1.06.

To provide aggregated measures of inequality in individual well-being considering other dimensions than just income (income, health and education, to be more precise) and EMU heterogeneity across country, Aristei and Perugini (2010) estimate the parameter for 26 countries by the regression approach developed by Young (1990).[28] It was done using data for the year 2006, and a median regression approach to prevent distortions caused by outliers. It is not clear if the tax rates considered include NICs and if standard allowances are considered.

[27] It is interesting to note that it goes against the empirical support favouring weighting in Evans et al. (2005), and diminishes the validity of the interpretation advocated in Evans and Sezer (2004) that the across-income level average of the EMU is more valuable than the one obtained at the APW, once the representativeness of the parameters found out of the APW is smaller than for those at the APW.
[28] See footnote number 20.

Since ATR and MTR are not calculated, the income level at which they are derived is not an applicable concept in this context. The results for η are between 1.04 and 1.77 (for Denmark and Ireland respectively), with an average of 1.35.

In Evans and Kula (2011) η is estimated by EASA (direct calculation) with the purpose to achieve appropriate SDR and regional welfare weights for Cyprus. They used data for the year 2007 and, although it is not clear, it seems that the rates were related to single persons with no dependants. The ATR and MTR were calculated for the midpoints of the taxable income bands. Supernumerary income is utilized, and it is not clear if NICs were included. The results obtained for each income level were weighted by the number of taxpayers in the brackets corresponding to the income level concerned, and an average η of 1.33 and 1.08 were achieved for the Turkish (North) and Greek (South) regions of Cyprus respectively. Despite the different values, the author claims it is reasonable to use the same estimate of η (between 1 and 1.3) for both parts of the island.

With the purpose to "identify the appropriate value of a social discount rate to be used by the German Statutory Health Insurance for the economics evaluation of health technologies", Mareike and Jurgen (2012: 217) estimate η by EASA. They use data for 2007 and direct calculation, checking the differences arising from the inclusion of NICs. The income level at which the derivations are undertaken are proportions of the average income (wage). No weighting is considered, and the results, considering together the data with and without NICs, are between 1.25 and 1.80. It must be noticed that in the model including NICs the highest income level is ignored because they claim that above a contribution ceiling "the redistribution mechanism in terms of social insurance contributions does not work completely" (Mareike and Jurgen, 2012: 135). It is not clear if supernumerary income was used.

Groom and Maddison (2013) review the literature related to the η estimates for UK and also produce new evidence (by EASA and other methods) concerning the parameter. They utilize the regression approach and the direct calculation to derive the results. The data analysed are earnings liable to income taxation for the years 2000-1 through 2009-10 (except 2008-09) for single person with no special circumstances (used in the regression approach and part of the direct calculation approach) and tax rates from 1948 to 2007 for single person household (used in the direct calculation approach). They use all-in tax data (include NICs) and tax-only data, but clearly favour the former procedure claiming that on their view "whilst historically NICs embodied a contributory principle, this linkage has now all but disappeared" (Groom and Maddison, 2013: 9). The use of supernumerary income is also criticized by a demonstration in which the subsistence income, estimated by the tax allowance, is incorporated in the iso-elastic utility function, yielding an η definition where the parameter explicitly depends on the subsistence wage and goes to infinity if the allowance approaches the income. Another concern expressed by the authors about supernumerary income is relative to the reliability of the underlying assumption that tax allowance is a good estimate of the subsistence income level.

In the regression approach they develop models with and without NICs, with weighted (by number of tax payers in each earning category) and non-weighted data, and with year dummies to test the stability of the parameter over time. The ATR and MTR were calculated for the mean earning in each earning category, and the preferred model included NICs, used weighted data (thus representing, according to the authors, the population of income tax payers) and presented no dummies, since these were statistically insignificant, providing evidence in favour of the stability of η over time. The preferred result obtained for η is 1.515.

In the direct approach they calculate η for the APW in each year (with and without the NICs), but then the need to look at more historical data in order to derive a more reliable estimate is detected. At this point they use data reporting income tax rates since 1948 to 2007, calculating η (not weighting the data, but including NICs) at the APW for each year. The average η obtained was 1.45. The parameter calculated at +/- 20% of the APW was significantly different from the one obtained for the APW level, providing then evidence against the constancy of η across the income level. An AR (1) model was also run for η at the APW, and a mean of 1.57 with a 95% confidence interval of 1.09-2.07 was derived.

1.3.3. Comments on literature review and further classification

The works presented in the last section can be classified in many different ways, but since this study intends to compare different methodologies inside the EASA, I present a classification which focuses on some technical aspects of the different η estimation procedures undertaken throughout the studies considered.

The technical elements which seem to vary most from work to work are the use of all-in or tax-only data, the use of supernumerary income, the application of regression or direct calculation (or both), the utilization or not of weighted data, the income level and the tax years for which η is derived. Some issues related to some of these techniques and ways of assessing them will be further debated in the sections to come. Tables 1.3 and 1.4 below present a classification based on the aspects referred.

1.4. Brief discussion on literature review

In this section I sum up briefly the main points of ambiguity in the literature just presented and consider some possible ways of advancing the debate.

The two major issues which come up from reading studies using EASA to estimate η – apart from the little dataset used, which is already tackled in this work by the usage of a larger dataset – are the usage or not of NICs and SI in the calculations. Cowell and Gardiner (1999) are the first to compare estimates derived with and without NICs. The core of this controversy seems to be on how much NICs are different from income tax, which seems not to be sufficiently solved yet to form a consensus in one or other direction. In Evans (2004), the discussion about including a subsistence wage (estimated by the tax free allowance) in the η estimation procedure is posed, being based upon the claim it makes more sense to consider utility as a concave function of income just above a given subsistence income. Groom and Maddison (2013) further consider this issue, pointing out that in the derivation of η described on section 1.2.3.4 the subsistence income should be taken into consideration in the iso-elastic utility function for the estimation procedure to be consistent with the argument justifying it. Also in this topic there seems to be no consensus.

Table 1.3. Short literature categorization

Study	NICs	SI	IL	Years
Stern (1977)	No	No	NC	1973/74
Cowell and Gardiner (1999) 1	No	NC	NC	1998/99-1999/00
Cowell and Gardiner (1999) 2	Yes	NC	NC	1998/99-1999/00
Evans and Sezer (2004) 1	Yes	NC	APWmi	2001
Evans and Sezer (2004) 2	Yes	NC	APWmi prop.	2001
Evans (2004)	NC	Yes	APWmi prop.	2001
Evans et al. (2005)	Yes	No	NC	2002/03
Evans and Sezer (2005) 1	Yes	No	APWmi	2001
Evans and Sezer (2005) 2	Yes	No	APWmi prop.	2001
Evans (2005)	No	Yes	APWmi prop.	2002
Sezer (2006)	NC	Yes	MTB	2004
Lopez (2008) 1	NC	NC	QAI	NC
Lopez (2008) 2	NC	NC	IDP	NC
Percoco (2008)	NC	NC	APW prop.	2004
Evans (2008)	No	Yes	AW	2005/06
Aristei and Perugini (2010)	NC	NC	NA	2006
Evans and Kula (2011)	NC	Yes	MTB	2011
Mareike and Jrgen (2012) 1	Yes	NC	AW prop.	2007
Mareike and Jrgen (2012) 2	No	NC	AW prop.	2007
Groom and Maddison (2013) 1	Yes	No	NC	2000/01-2009/10
Groom and Maddison (2013) 2	Yes	No	APW	1948-2007
Groom and Maddison (2013) 3	Yes	No	APW	1948-2007

Notes: (1) SI=supernumerary income; (2) IL=income level; (3) NC=Not clear; (4) APWmi=average production wage in manufacturing industries; (5) APWmi prop.=APWmi proportions; (6) MTB=midpoint of each taxable band; (7) QAI=quintile average income; (8) IDP=income distribution percentages (20,40,60,80 percent); (9) AW=average income; (10) NA=not applicable; (11) The years in Groom and Maddison (2013) 1 exclude the tax year 2008/09; (12) Some studies develop many different models, so due to space constraint just the preferred ones are presented. Source: see text.

To be able to establish a preferred way of estimation as far as NICs and SI are concerned I consider the best fit of four different estimation approaches inside the EASA: estimation considering income taxes and NICs (all in taxes) and not considering SI; considering all in taxes and SI; not considering NICs and SI and finally not considering NICs and considering SI. In other words, I gather across country and year the (dependent and independent) variables required to apply a regression of the type shown in equation 1.22. Comparing the R-squared of the different methods enables me to evaluate which of them fits better the EuroPTax data (which forms the basis for my empirical analysis and which I will go on to discuss in detail shortly). This is a simple and I think reasonable way to come out with a preferred approach (or at least to see whether the issue of NICS and SI makes an appreciable difference and is therefore worth worrying about), given the regression shown on equation 1.22 is to be seen just as a curve-fitting exercise, as Stern (1977) argues.

Table 1.4. Continuing short literature categorization

Study	Regression/Calculation	Weighting	Result
Stern (1977)	Regression	None	1.97
Cowell and Gardiner (1999) 1	Regression	None	1.43-1.41
Cowell and Gardiner (1999) 2	Regression	None	1.29-1.28
Evans and Sezer (2004) 1	Calculation	None	1.3-1.6
Evans and Sezer (2004) 2	Calculation	None	1.28-1.72
Evans (2004)	Calculation	None	1.35
Evans et al. (2005)	Regression	IA	1.57
Evans and Sezer (2005) 1	Calculation	None	1.11-1.81
Evans and Sezer (2005) 2	Calculation	None	1.1-2
Evans (2005)	Calculation	None	1.4
Sezer (2006)	Calculation	None	1.25
Lopez (2008) 1	Calculation	None	1.5
Lopez (2008) 2	Calculation	None	1.6
Percoco (2008)	Calculation	None	1.347
Evans (2008)	Calculation	None	1.06
Aristei and Perugini (2010)	Regression	None	1.35
Evans and Kula (2011)	Calculation	NTP	1-1.3
Mareike and Jrgen (2012) 1	Calculation	None	1.25-1.60
Mareike and Jrgen (2012) 2	Calculation	None	1.47-1.80
Groom and Maddison (2013) 1	Regression	NTP	1.515
Groom and Maddison (2013) 2	Calculation	None	1.45
Groom and Maddison (2013) 3	AR(1)	None	1.57

Notes: (1) the information related to Cowell and Gardiner were inferred, not being clearly stated in the paper; (2) the results for Cowell and Gardiner are presented for the years 1998/99 and 1999/00 respectively; (3) 1.3- 1.6=between 1.3 and 1.6; (4) IA=data is weighted by income amount; (5) the result for Evans et al. (2005) is relative to the preferred model that estimates directly (not the reciprocal of); (6) NTP=data weighted by number of tax payers in the band; (7) Some studies develop many different models, so due to space constraint just the preferred ones are presented. Source: see text.

Anyone investigating the literature soon realizes there is some concern about the constancy of η over income levels, with the results regarding this issue ambiguous in the studies herein considered. Evans (2005), for example, tests the η constancy over income by comparing the η of 20 countries at low and high income levels, finding evidence for the constancy of η, whilst Groom and Maddison (2013), using a long series of η for the UK, test the same parameter property and find evidence supporting the variability of η over the dimension concerned.

Since the iso-elastic utility function assumed in the EASA implies that income changes over time do not affect η and that utility varies with income only, it seems sensible to expect the parameter is constant not just over income levels but also across time. If this expectation is not confirmed, the utility function underpinning the exercise would be proved to be not fully appropriate. If the parameter is found to vary over income and time, variation over time might be thought as a result of variation in income; on the other hand, if there is variation over time but no variation over income level, other variable(s) evolving trough time but income could be thought to affect the EMU. Groom and Maddison (2013) seem to be the only work in which the constancy of η across time is formally tested, but they do that observing just UK tax rates for some few years.

Furthermore, to see if the η obtained are consistent with the assumption that utility is dependent only on income, and therefore cultural factors, for example, may be left outside the analysis, I test the constancy of η across countries.

The issue related to the tax unit which data should be collected on seems to be fairly uncontested, since after Stern (1977), virtually all studies use single people as units instead of families. However, it must be noticed that no clear theoretical reason is found to justify this procedure, and since the EuroPTax offers the possibility to estimate η for different tax units, estimates obtained considering families (using the four combinations regarding NICs and SI referred above) will below be compared with those derived using single persons. The criterion for the comparison in this case is also the goodness of fit.

There is also some discussion in the literature relating to weighting the data used to derive η in the EASA. I will not address this potentially important issue mainly because the EuroPTax provides unfortunately no information about the number of persons at any given tax bracket or income range, precluding any comparison between weighted and non-weighted data.

1.5. Data description

The dataset used in the estimation of η is the result of a simulation model called EuroPTax, developed in a project undertaken by Lynch and Weingarten (2010). In the research report made by Lynch (2009) the objective of the project is stated as "to construct a history of the personal tax and social security contributions paid by individuals and households in Western European democracies, at all points on the income scale, since the 1950s". Moreover, it is declared that the interest is in studying the construction of tax systems and the changes to which they are subjected in order to finance public spending. The project, according to the report, could be used for four main reasons: to analyse the variation in the tax burden on individuals or households over time, to observe the differences in tax rates over the countries considered in the research, to classify the tax systems according their different characteristics and to compare the tax rates applied to richer and poorer workers across time. In the case of the present study the possibility to access and compare tax rates since the 1950s to 2007 for different countries in Europe is of great interest, since it permits me to establish a more trustworthy basis for the estimation of EMU.

The simulation was done basically by parameterizing tax rules found in the countries' legislation for all the periods considered and then applying these rules to hypothetical tax units. The research from which the data is collected consists, therefore, essentially on reconstructing the tax rules (tax systems for all countries and years) of the countries considered with respect to specific tax units. The EuroPTax permits one to establish the income of the individual/household (the APW is given in the EuroPTax spreadsheet), the marital status and the number of dependent children. Given this information, the simulation provides the local tax percentage and its absolute value, national tax percentage and absolute value, national plus local tax percentages and the absolute value, social security contribution (SSC/NICs) percentages and their absolute value, all-in taxes (including SSC) percentages and their absolute value. The countries included in the database are Belgium, Denmark, France, Germany, Ireland, Italy, Norway, Sweden and UK.

Some caveats must attached to the model. In line with the OECD, the EuroPTax includes standard allowances in the calculations, so that the tax paid by people who derive a large proportion of their income from capital assets (usually this is the case for high income earners) are probably inflated, since taxation rules are different in this circumstance. It basically means that the higher the income investigated the more overestimated the correspondent tax rate retrieved by EuroPTax will probably be.[29] Finally, one should notice that since the derivations are made based on tax rules, its reliability depends on whether the rules were in fact accomplished and people paid their taxes.

1.6. Results

In this section I estimate the EMU for the 9 countries considered in the EuroPTax using different methodologies associated with the EASA (as listed in section 1.4). In particular I am going to start by investigating the sensitivity of the EMU to the inclusion of NICS and SI. Then I am going to consider the constancy of the EMU over different portions of the income distribution, countries, periods of time and tax units. Finding that estimates of the EMU are sensitive to different

[29] It should not be a concern for this study, once the income levels looked at are not very high.

assumptions and that they vary over income, countries, time or tax units would make it difficult without further arguments to use these results in a policy setting.

1.6.1. Preferred approach concerning NICs and SI

As observed before, an issue in the literature related to the estimation of η by the EASA is the inclusion or not of NICs and SI in the calculations. To set a preferred approach I estimate the parameter in the four different combinations referred to in section 1.4, and look at the goodness of fit. In this study's context the goodness of fit measure is how well the actual data fits the EASA prediction that $\log(1 - MTR) = \eta * \log(1 - ATR)$. This gives evidence on what set of assumptions (each of which results in a different set of estimates) are better explained by the EASA.

There are obviously other ways to carry out comparisons among the different assumptions regarding NICs and SI. An example would be to examine the NICs/SI combinations in terms of which provides results possessing properties which are more similar to those assumed by the EASA (e.g. constancy of η across income). We find, however, that comparing the goodness of fit provides a simple way to carry out the referred comparison.

The estimates obtained without considering SI are obtained as I previously described. To include SI in the estimation process, however, I consider Groom and Maddison's (2013) remarks about the issue, in which they argue subsistence income should also be considered in the iso-elastic utility function assumed in the EASA when SI is to be accounted for. The utility function is then given by

$$U_i = \frac{(Y-\gamma)^{1-\eta}-1}{1-\eta}, \tag{1.26}$$

where γ is the subsistence income.

In this case, by a similar procedure to the one used to derive equations 1.21 and 1.22, η is given by

$$\eta = \left(\frac{\ln(1-MTR)}{\ln\left(1-\frac{T(Y)}{Y-\gamma}\right)}\right)\left(\frac{Y}{Y-\gamma}\right), \tag{1.27}$$

such that the corresponding regression is

$$\ln(1 - MTR) = \frac{\eta \ln\left(1-\frac{T(Y_i)}{Y_i-\gamma_i}\right)}{\frac{Y_i}{Y_i-\gamma_i}} + \varepsilon. \tag{1.28}$$

As in equation 1.22, there is no interpretation for the error term in this case, given the use of regression as a curve-fitting tool.

Another issue related to the parameters obtained incorporating SI is the way one estimates the subsistence income to be used in the calculations. Given the EuroPTax directly returns the income tax rate at a given wage, I estimated the tax-free income by entering a guess for wages and looking at the tax rate retrieved. If this is equal or close to 0% (between 0% and 1% to be more precise), the corresponding income was considered tax free. The tax rate considered in the procedure did not include NICs.

For some few years (and countries) the EuroPTax did not return reasonable rates to the income levels entered according to the procedure just described,[30] such that the means of the previous and next years were taken to complement the data. For Denmark, France and Belgium, however, it happened for rather many years,[31] such that I run a regression with income at the APW as the independent variable and the tax-free income available as the dependent variable for each country. This regression[32] gives tax free income as a proportion of income at the APW, which is then used to complement the tax-free income data related to the three countries cited. To be more specific, the proportion referred to is applied to the APW of the years for which the subsistence income is missing, providing thus an estimation of the subsistence income for such years.[33]

Table 1.5 shows the R^2 and η estimates for each of the NICs/SI combinations used. There is little difference amongst them in terms of R^2, but the one considering NICs and not considering SI has a slight advantage in terms of curve-fitting.[34] Figures in appendix 1.B show how the direct-calculation estimates (as calculated by each approach) evolve through time, respectively.

---

[30] Mainly because of missing information in the EuroPTax, which was complemented by the average of the rates obtained for neighbour years, and because in some cases the rates never reach 0% or reach it just when the corresponding income is very low, clearly underestimating the subsistence income.

[31] For Denmark and France for many years one had to pay income taxes no matter how much he earned according to the EuroPtax (the amount "payed" was probably compensated with transfers) and for Belgium the EuroPTax does not show the corresponding tax rate when income gets too low.

[32] Algebraically specified as $APW_i = pSI_i + \varepsilon_i$, where i are the years for which the SI was obtained normally, and p is the proportion of the APW understood as corresponding to the SI, which is applied to the APW of the remaining years in order to derive a subsistence income relative to them.

[33] Notice it is another case in which there is no concern with causality, being the regression a curve-fitting exercise only. The R^2 for the referred regressions were high, being 0.88 the lowest one found (for Denmark). The APW proportions were 0.14, 0.6 and 0.27, for Denmark, France and Belgium, respectively.

[34] It is important to have in mind that since the regressions are concerned with curve-fitting only no assumptions need to be made regarding the errors and model specification.

Remember, however, that in the present analysis (comparing NICs/SI approaches within the EASA) I use the regression approach only.

Table 1.5. R-squared for each approach applied for single persons

Approach	R-squared	EMU	95% rob. C. I.	F-stat.	N. Obs.
All in/no SI	0.9551	1.515 (0.016) [95.4]**	1.48 - 1.55	9108**	452
All in/SI	0.9484	1.395 (0.016) [89.4]**	1.37 - 1.43	7996**	452
No NICs/no SI	0.9424	1.653 (0.022) [76.1]**	1.61 - 1.7	5790**	452
No NICs/SI	0.9433	1.589 (0.021) [75.6]**	1.55 - 1.63	5717**	452

*Source: author calculations. Notes: (1) Robust standard errors are used to derive the SEs; (2) Standard Errors are shown between parenthesis and t-statistics between brackets; (3) ** indicates the relevant coefficient is significant at 1%, * at 5% and ° at 10%.*

1.6.2. The constancy of the EMU across income level analysis

As seen before, η is assumed to be constant across income levels in the EASA. To test if the estimates obtained possess this property, I apply a chi-squared test of heterogeneity using the EMU estimates and variances derived from regressing $\ln(1-MTR)$ against $\ln(1-ATR)$, as shown in equation 1.22.

I derive the MTR and ATR for all the nine countries at four different income levels (0.67XAPW, APW, 1.33xAPW and 1.67xAPW)[35] and run one regression per income level with the data for all countries stacked. The All in/no SI combination is used given it is the most appropriate according to the previous analysis. Thus with an η estimate per income level and their respective standard errors it is possible to test the parameter's constancy across income level.

The test (chi-squared test of heterogeneity) is based on Cochrans (1954) Q statistic, given by

$$Q = \sum w_i (T_i - T)^2, \tag{1.29}$$

where T_i are the parameters whose homogeneity are being tested, w_i are the reciprocal of their variances and T is given by

$$T = \frac{\sum w_i T_i}{\sum w_i}. \tag{1.30}$$

Under the null hypothesis of homogeneity, which can be written as

[35] Same proportions of the APW used by Evans and Sezer (2004), Evans (2004), Evans and Sezer (2005) and Percoco (2008).

$$\eta_j = \eta_k, \forall j \neq k, \tag{1.31}$$

where j and k are different income levels, the statistic follows a chi-squared distribution with $k-1$ degrees of freedom.

In this case the null was rejected,[36] providing evidence for the non-constancy of η across income levels. However, the estimates obtained per each income level and the confidence intervals suggest that although the differences are statistically significant they are very small (see table 1.6).

Table 1.6. Constancy of η across income levels

Income level	η (EMU)	95% rob. C. I.	F-stat.	N. Obs
APW	1.515 (0.016) [95.4]**	1.483 - 1.546	9108**	452
0.67APW	1.482 (0.016) [91.7]**	1.451 - 1.514	8403**	452
1.33APW	1.611 (0.018) [90.1]**	1.578 - 1.643	8276**	452
1.67APW	1.575 (0.026) [59.3]**	1.532 - 1.617	3521**	447[37]

*Source: author calculations. Notes: (1) Robust standard errors are used to derive the SEs; (2) Standard Errors are shown between parenthesis and t-statistics between brackets; (3) ** indicates the relevant coefficient is significant at 1%, * at 5% and ° at 10%.*

1.6.3. The constancy of the EMU across countries analysis

By testing if η varies across countries we can find evidence on whether societal inequality aversion changes according to cultural norms and whether it is reasonable to suggest the same parameter value applies to different Countries. The estimates being homogeneous over the set of nations considered gives rise to an argument for adopting just one η value for all the referred countries, given it could facilitate international investment decisions amongst them.

One can reasonably argue that differences from countries may result from differences in income levels, once the APW in different countries may actually correspond to different wage levels. Because the previous test indicates variation across income levels is small, however, it seems appropriate to disregard η variation across income and assume potential differences across countries are mainly due to elements other than income level, such as culture.

[36] Heterogeneity chi-squared = 32.65 (d.f. = 3) p = 0.000.

[37] The number of observations decreased for the 1.67xAPW income level due to the fact that the MTR obtained for 4 observations for Belgium and 1 observation for Germany were larger than 1, rendering negative values for $\ln(1 - MTR)$.

Here the chi-squared test of heterogeneity was once more applied. The means and variances of the η time series obtained via direct calculation at the APW for the nine countries and the EMU parameters and standard errors derived from running regression 1.22 were tested. The null can be written as

$$\bar{\eta}_i = \bar{\eta}_j, \forall i \neq j, \tag{1.32}$$

where i and j are countries, $\bar{\eta}$ is the mean value (or the EMU derived via regression 1.22) of η over 1950s to 2007.

Table 1.7. Constancy across countries

	ES (average)		**ES (reg)**	
Country	η	SE	η	Robust SE
UK	1.45	0.02	1.37	0.016
Sweden	1.63	0.07	1.68	0.071
Norway	1.45	0.03	1.47	0.037
Italy	1.42	0.02	1.46	0.020
Ireland	1.89	0.06	1.87	0.059
Germany	1.44	0.02	1.45	0.017
France	1.48	0.02	1.44	0.024
Denmark	1.58	0.08	1.45	0.033
Belgium	1.45	0.04	1.48	0.035
Chi-square Homogeneity	53.50** (0.000)	-	86.57** (0.000)	-

*Source: see text. Notes: (1) Standard errors in parenthesis; (2) ** indicates the relevant coefficient is significant at 1%, * at 5% and ° at 10%.*

The result suggests η varies across countries, which implies it would not be reasonable for all the nine countries considered to apply the same η value. Whilst this procedure would arguably facilitate international joint investment decisions amongst governments, apparently the countries should adopt different EMU values to account for their particular characteristics. The results also suggest that societal evaluation of inequality varies with cultural factors.

Nonetheless, here again the EMU differences across countries are, though statistically significant, actually quite small in absolute terms. The highest estimate is 1.89 for Ireland (average) and the lowest is 1.37 for the UK (regression).

1.6.4. The constancy of the EMU across time analysis

Another characteristic of the utility function assumed in the EASA is that is does not include time in its representation. To check if the results obtained are consistent with this assumption, I test the constancy of η over time.

To undertake this task I use the nonparametric test of trend across ordered groups proposed by Cuzick (1985). To apply the test, the k groups $n_i (i = 1, 2, \ldots, k)$ in which the data is separated must be given scores l_i, which usually correspond to the ordering (e.g. the first group receives score 1, the second 2 and so on). Then the data is ranked and the rank sum correspondent to each group (R_i) is found.

The test is then based on the statistic T given by

$$T = \sum_{i=1}^{k} l_i R_i. \tag{1.33}$$

Under the null hypothesis of no trend, the expectation of T is given by

$$E(T) = 0.5(N + 1)L, \tag{1.34}$$

and its standard error is given by

$$s(e) = \sqrt{\frac{N+1}{12}} (N \sum_{i=1}^{k} l_i^2 n_i L^2), \tag{1.35}$$

where $N = \sum n_i$ and $L = \sum_{i=1}^{k} l_i n_i$.

In this case T is asymptotically normal, such that the z statistic given by $T - E(T)/se(T)$ is also normal under the null.[38]

For the present purposes the η time series considered for each country and NICs/SI combination were calculated at the APW and were gathered per decade.[39] The data corresponding to the first decade was scored as 1, the second decade as 2 and so on.

Table 1.8 shows the results of the test applied. For the combinations "no NICs/no SI" and "no NICs/SI", the no trend hypothesis was accepted for four countries, whilst for the other two combinations it was accepted only for two countries. Therefore, whilst the combination "all in/no SI" seems to be the best by the best fit criteria, the ones not considering NICs seem to result in estimates possessing the desirable property of being constant over time for a larger number of countries.

In any case, the results suggest the EMU is not constant over time, and that it is important that future works develop a model capable of incorporating η's time variation.

---

[38] Which can be written as $\eta_i = \eta_j, \forall i \neq j$, where i and j are periods corresponding to different decades within a country.

[39] I also applied the test with the data gathered per approximately five and fifteen years, and the results did not change significantly (Appendix 1.C).

1.6.5. The constancy of the EMU across tax unit analysis

It is reasonable to think the η estimates derived by the EASA are sensitive to the tax unit chosen in the analysis, but there seems to be little discussion in the literature about which would be the best one to pick, although virtually all studies after Stern(1977) take single persons as the tax unit. To advance the debate in this area I calculate η taking families with two dependent children as tax units for all the NICs/SI combinations (with 100% of income coming from the head of household, at the APW). I test whether the results derived with the two distinct tax units are statistically different, what is the preferred combination in terms of the best curve-fitting for the estimates obtained with families and compare the parameters derived with single and families tax units by the best curve-fitting criterion.

Table 1.8. Test of the null hypothesis that η is constant across time

Country	all in/no SI	all in/SI	no NICs/no SI	no NICs/SI
UK	0	0	0	0
SWEEDEN	0	0	0	0
NORWAY	0	0	0	0
ITALY	0	0.025	0.39	0.94
IRELAND	0	0	0.34	0.12
GERMANY	0.99	0.25	0.9	0.72
FRANCE	0	0	0	0
DENMARK	0	0	0.02	0
BELGIUM	0.24	0.43	0.4	0.58

Note: The numbers shown in the table correspond to the p-values.

Before presenting the results, I note that the number of observations dropped considerably for the combinations taking families as tax units and SI into account,[40] since for many years the subsistence wage was too close to or larger than the APW.[41]

A paired t-test was applied to compare the average of the time series related to each country and combination (at the APW) for single and families tax units. For the "all in/no SI" combination, only for Sweden was the null hypothesis of homogeneity not rejected.[42] For the "all in/SI" combination the null was not rejected for the UK, Sweden, Germany, Denmark and Belgium. For the "no NICs/no SI" combination the null was rejected for all countries and for the "no NICs/SI" combination the null was not rejected for all countries except for the UK. The results seem to indicate that the "no NICs/SI" combination is the most robust one regarding variation over tax units (See table 1.9).

[40] France was not considered in the concerned combination because just one observation was left.

[41] The subsistence wage of a family was roughly estimated as three times the one of a single person. It is assumed one partner spends as much as the other and a child spend half as much as an adult. This is obviously not intended to be exact, but just to give an idea of a family's SI. Notice also that this procedure makes it difficult to derive η at income levels lower than the APW.

[42] All tests were applied at 5% significance level.

Table 1.10 shows the R-squared obtained by running regressions of the type shown in equations 1.22 and 1.28 using families as tax units. The combination "all in/no SI" once again fits better than the others. Comparing the r-squared for this preferred method over single persons and families tax units it can be seen that in terms of best fitting the usage of single persons is slightly superior.

Table 1.9. T-test comparison between mean η obtained with single and family tax units by country and NICs/SI combination (null hypothesis of homogeneity)

Country	all in/no SI	all in/SI	no NICs/no SI	no NICs/SI
	(t-statistics)	(t-statistics)	(t-statistics)	(t-statistics)
UK	6.83**	-1	3.06**	-2.44*
SWEEDEN	1.93°	-0.3	2.52*	-1.5
NORWAY	3.82**	2.6*	5.06**	-1.67
ITALY	7.82**	2.95**	7**	0.8
IRELAND	2.12*	2.23*	2.45*	1.11
GERMANY	2.09*	0.02	2.03*	-1.93°
FRANCE	-9.11**	-	5.21**	-
DENMARK	7.14**	1.63	7.86**	-1.05
BELGIUM	7.79**	2.47°	7.68**	-1.06

*Source: see text. Notes: (1) For the combination "all in/SI" there is very few data for France and Belgium; mainly for France, for which the test was not carried out once just one η observations was valid. The basic problem in the case of France was that for families the subsistence incomes calculated were in general larger than the APW. In the case of Belgium the problem was that in many cases the tax to be paid were larger than the difference between the SI and the APW, resulting on a $T(Y)/(Y-\gamma)$ larger than 1, making it impossible to calculate $ln(\frac{1-T(Y)}{Y-\gamma})$;(2) Also for the combination no NICs/SI just one observation for France, considering families, was valid; (3) Degrees of freedom vary across country and combination; (4) ** indicates the relevant coefficient is significant at 1%, * at 5% and ° at 10%.*

Table 1.10. R-squared for each approach applied for single persons and families

Approach	R^2 for single	R^2 for families	EMU (single)	EMU (fam.)	N. Obs. (fam)
All in/no SI	0.9551	0.9369	1.515 (0.016) [95.4]**	1.65 (0.022) [74.1]** {5496}**	428
All in/SI	0.9484	0.8169	1.395 (0.016) [89.4]**	1.04 (0.062) [16.8]** {281}**	276
NoNICs/no SI	0.9424	0.9142	1.653 (0.022) [76.1]**	1.85 (0.032) [57.8]** {3339}**	408
NoNICs/SI	0.9433	0.8796	1.589 (0.021) [75.6]**	1.54 (0.038) [40.8]** {1663}**	317

*Source: author calculations. Notes: (1) Robust standard errors are used to derive the SEs; (2) Standard Errors are shown between parenthesis and t-statistics between brackets; (3) F-statistics for families are between {}; (4) F-statistics and N. Obs. for single people are presented in table 1.5; (5) ** indicates the relevant coefficient is significant at 1%, * at 5% and ° at 10%.*

1.6.6. Suggestion of which estimate of η is to be used in Ramsey equation

Considering the graphs showing the η estimates in the different NICs/SI combinations, it is possible to see that there is a concentration around 1.5 (see figures on Appendix 1.B). The η estimated in the regressions run to test the best fitting combination (with single persons as tax units) fell in the interval 1.39 - 1.65. The value of 1.51 (with standard error of 0.016) was obtained by the preferred combination (considering NICs and not considering SI).

Using families as tax units, the estimates obtained in the regressions run to test the best fitting combination fell between 1.045 and 1.85, being the one derived by the preferred combination (again the "all in/no SI" combination, see table 1.10) equal to 1.65 (standard deviation of 0.022).

Given the estimates obtained using single persons as tax units fitted better the actual data and are largely used in the literature, I give more weight to the results derived from such tax units, pointing thus to an η of 1.5 as the preferred value.

Table 1.11. η value for the different countries considered

Countries	η for single (average)	η for single (regression)
UK	1.45	1.37
Sweden	1.63	1.68
Norway	1.45	1.47
Italy	1.42	1.46
Ireland	1.9	1.87
Germany	1.44	1.45
France	1.48	1.44
Denmark	1.58	1.45
Belgium	1.45	1.48

Source: see text. Notes: (1) The term 'average' in the table corresponds to the average EMU obtained by direct calculation across the years considered for the individual countries.

It is quite remarkable that the results cluster so tightly around 1.5. It strongly substantiates such a value as the preferred one. Given different values of η were found for each country under scrutiny, however, I present table 1.11 with all the resulting parameters obtained per country with the best NICs/SI combination (all in and no SI) with regards to single persons as tax units and measured at the APW. The estimates presented are obtained both by averaging the individual parameters estimated via direct calculation for each country across all years and by running the kind of regression shown on equation 1.22, in which the MTR and the ATR are clustered across years for each country.

It is also worthwhile noticing that considering a heterogeneity test, the EMU value of 1.5 proposed (obtained with the "all in/no SI" combination and single people as tax units) is not heterogeneous to the EMU value derived from averaging the series of EMU estimates found in the works listed in the literature review (see table 1.4) – i.e. other EMU values dominating the literature the last 30 years.[43] Nonetheless, considering a simple t-test (and thus not accounting for the variance regarding the EMU series in the literature review), the EMU value proposed is

[43] Heterogeneity chi-squared (H0 = homogeneity) = 2.63 (d.f. = 1), p = 0.105 (the average for the EMU series estimated by works listed in the literature review is 1,446 and the correspondent SEM is 0,039). The midpoint EMU value was considered for works that estimated EMU value ranges.

significantly different from the average obtained from the literature review series (t-stat.=4.341, p<0.01). These results show the parameters cannot be regarded as heterogeneous because of the variance of the EMU values across works, and suggest that although the proposed value is somewhat different from the one found in the literature regarding the estimation of the EMU via EASA, it mostly endorses the value indicated by such literature.

1.7. Comparing EASA estimates with estimates obtained from other methodologies based on revealed preferences

In this section, in an attempt to validate them, I compare the EMU estimates obtained with the EASA with EMU estimates obtained with two other methods: in the first the parameters are drawn from insurance data and in the second from consumption data. After giving more details about the relevance of the exercise carried out herein I explain in more depth the two methodologies employed. Following that I describe the dataset used in both the insurance and the consumption based approaches, and then finally present and compare the estimates.

It is quite uncommon to find works comparing the estimates of the EMU provided by different techniques. It is also the case that the three different methodologies used herein (EASA, insurance and consumption) refer to the estimation of the EMU in the three different contexts it can be estimated in (each corresponding to a different interpretation of the EMU, as discussed earlier) – intra-temporal aversion to inequality (EASA), intertemporal aversion to inequality (method using consumption growth rates) and risk (method using insurance data).

It is important to note that Atkinson et al. (2009) suggested, in a stated preference/survey based setting, that EMU estimates obtained within the three different contexts referred produce different results, while Groom and Maddison (2013), in a revealed preference based setting, derive results suggesting that EMU estimates derived in the three different contexts do not differ significantly. Given that I test for the constancy of the EMU estimates across such contexts (in a revealed preference setting), the evidence obtained herein can help to shed light on this puzzle. Moreover, the tests undertaken in this section mirror to some extent the exercise carried out in the final chapter, in which I compare EMU estimates derived from the three referred contexts in a stated preference setting and relate the obtained values with psychological traits.

If the three different methodologies give similar results, I will conclude it suggests the EMU value obtained via EASA is strongly substantiated and that estimating the parameter as intratemporal inequality aversion is not significantly different from estimating it as risk aversion or as intertemporal inequality aversion. On the other hand, if the methodologies produce different results two possible conclusions can be taken. Or it weakens the legitimacy of the EASA as a technique to estimate the EMU, or estimating the parameter as intratemporal inequality aversion is too different from estimating it as risk aversion or intertemporal inequality aversion.

Importantly, the comparison across different methodologies is made for the same countries rather than for different countries (in order to exclude the possibility that differences are due to country-specific cultural factors).

1.7.1. Method of estimating EMU based on insurance data

Szpiro (1986) derives relative risk aversion (RRA) estimates (equivalent to EMU) for the United States using data on wealth (wealth of households, non-profit organisations, the Government and the net foreign balance) and property and liability insurance. Below we present his methodology. Notice the ubiquity of insurance makes this a natural opportunity to calculate the EMU (or as it is known in this context relative risk aversion).

Consider an individual with wealth W which is at risk of being lost with probability q. He can insure part I of W against such risk by paying a premium π, which includes expected claims and proportional loading fee. Thus $\pi = qI(1+\lambda)$, where λ represents the loading factor. In this case the individual's expected utility function is given by:

$$EU = (1-q)U(W-\pi) + qU(W-\pi-D), \tag{1.36}$$

where D is the non-insured part of W. Expanding the right hand side of the equation using a two term Taylor series around W gives

$$EU = U - \pi U' + \left(\frac{\pi^2}{2}\right)U'' - DqU' + \pi DqU'' + \left(\frac{D^2}{2}\right)qU''. \tag{1.37}$$

To find the optimal amount insured he differentiates the equation with respect to W, sets the result equal to 0 and divide both sides by $-1/U'$. Rearranging the terms of the resulting equation it yields the uninsured part of W (D):

$$D = \frac{\lambda}{a(W)}\frac{1}{1-q(1-\lambda^2)} - qW(1+\lambda)\frac{\lambda}{1-q(1-\lambda^2)}, \tag{1.38}$$

where $a(W)$ denotes the Pratt-Arrow measure of Absolute Risk Aversion (ARA). If we assume q is small the equation can be written as

$$D = \lambda/r, \tag{1.39}$$

such that we can write the insured amount I as

$$I = W - \lambda/r. \tag{1.40}$$

Given that

$$a(W) = r/W, \tag{1.41}$$

where r represents a constant RRA, we can describe I as

$$I = W - \lambda W/r. \tag{1.42}$$

Since I is not observable, Szpiro (1986) assumes $I = Q/q$, where Q represents total claims and q (loss probability) is constant. Incorporating these assumptions in the equation above yields

$$Q = qW + nW\lambda. \tag{1.43}$$

Both q and n can be thus estimated, and r is given by

$$r = -q/n. \tag{1.44}$$

Using this methodology Szpiro (1986) obtain EMU estimates between 1 and 2 for the United States, and concludes the RRA is constant with respect to wealth.

1.7.2. Method based on lifetime consumption behaviour

To use consumption data to derive the EMU behaviour regarding lifetime consumption is taken into consideration and the estimation process is based on the Euler equation, with the elasticity of inter-temporal substitution taken as the reciprocal of η. This last equality is based on assumptions about the equation transforming within-period to period-specific (iso-elastic) utility function.

To illustrate the procedure, consider an inter-temporal utility function of the form

$$U = \sum_{t=0}^{t=\infty} \beta^t u(c_t), \tag{1.45}$$

where β is the discount factor specified by $(1+\delta)^{-t}$ and $u(c_t)$ is utility at time t.

Maximizing this welfare equation assuming an iso-elastic utility function $(u(c_t) = (c_t^{1-\eta} - 1)/(1-\eta))$ and a riskless inter-temporal wealth constraint of the form

$$A_{t+1}A_t = r_t A_t + Y_t C_t \tag{1.46}$$

(where A_t indicates the amount of assets in period t and r_t is the interest rate at t), it is possible to derive an Euler equation of the form

$$ln(c_t) - ln(c_{t-1}) = \eta^{-1}\, ln(1+\delta) + \eta^{-1} ln(1+r_t), \tag{1.47}$$

which, using a Taylor approximation for $(1 + r_t)$, can be reduced to an empirical specification of

$$\delta\, ln(c_t) = a + br_t + v_t, \tag{1.48}$$

where b is the reciprocal of η, a is a constant and v_t is an error term. More detail on this methodology and its derivation can be found on Hall (1978), Altonji (1986) and Blundell et al. (1994).

1.7.3. Data

As explained above, the method to obtain the EMU using the consumption data consists on running a simple linear regression on per capita consumption growth rates against real interest rates. The EMU is given by the reciprocal of the estimated coefficient. Table 1.12 presents summary statistics for per capita consumption growth rates for non-durable goods, and table 1.13 presents summary statistics for real interest rates.

The two variables are derived from information on annual (at current and previous year prices) consumption of non-durable goods (including services), on population and on short-interest rates.[44] The period span observed varies across countries; for per capita consumption growth rates the oldest observation is for 1960 and the newest for 2014. For real interest rates the oldest observation is for 1970 and the newest is for 2015.

Table 1.12. Summary statistics for per capita consumption growth rates for non-durable goods

	Mean	Sdev	Median	Max	Min	NbObs
All countries	0.01	0.02	0.01	0.06	-0.04	261
UK	0.01	0.02	0.01	0.03	-0.04	19
Sweden	0.01	0.01	0.01	0.03	0	20
Norway	0.02	0.01	0.02	0.05	-0.02	44
Italy	0	0.02	0	0.03	-0.03	18
Ireland	0.02	0.03	0.02	0.06	-0.04	19
Germany	0.01	0.01	0.01	0.02	-0.01	22
France	0.02	0.01	0.02	0.05	-0.01	55
Denmark	0.01	0.02	0.01	0.06	-0.03	47
Belgium	0.01	0.01	0	0.03	-0.01	17

Source: see text.

Table 1.13. Summary statistics real interest rates

	Mean	Sdev	Median	Max	Min	NbObs
All countries	0.02	0.03	0.01	0.1	-0.03	238
UK	0.01	0.02	0.02	0.04	-0.03	20
Sweden	0.01	0.02	0.01	0.06	-0.02	22
Norway	0.03	0.03	0.03	0.09	-0.01	37
Italy	0.01	0.02	0	0.05	-0.02	19
Ireland	0	0.03	0	0.08	-0.03	20
Germany	0.01	0.02	0.02	0.05	-0.01	24
France	0.02	0.03	0.02	0.07	-0.03	46

[44] This data was obtained from OECD database. The derived non-durable consumption growth accounts for population growth.

Denmark	0.02	0.03	0.02	0.1	-0.02	30
Belgium	0	0.01	0.01	0.03	-0.02	20

Source: see text.

Szpiro's (1986) method to obtain the EMU using insurance data, on the other hand, consists of regressing total claims against wealth and wealth times the loading factor. The EMU is given by minus the coefficient regarding wealth divided by the coefficient regarding wealth times the loading factor. In this study wealth is represented by the GDP; we consider premiums and claims for all forms of insurance (including health) but not life insurance.[45] Table 1.14, 1.15 and 1.16 present the summary statistics for the GDP, total claims and loading factor,[46] respectively.

Table 1.14. Summary statistics for GDP (in millions)

	Mean	Sdev	Median	Max	Min	NbObs
All countries	1794851	1021994	1791726	4181103	156144	108
UK	1588589	169758.2	1568346	1870693	1304874	12
Sweden	3461161	411419.9	3453797	4181103	2805115	12
Norway	2587282	450477.1	2597735	3140371	1781981	12
Italy	1585317	60992.3	1607075	1642444	1448363	12
Ireland	184269	25385.82	177980.9	255815.1	156144	12
Germany	2610368	242613.1	2570900	3032820	2270620	12
France	1983127	146841.7	1997166	2181064	1710760	12
Denmark	1793632	155680.4	1806198	2027171	1506001	12
Belgium	359915	35636.77	359583.5	410351	298711	12

Source: see text.

Table 1.15. Summary statistics for total claims

	Mean	Sdev	Median	Max	Min	NbObs
All countries	34788.5	22570.45	33091	91347	1585	97
UK	33009.9	4520.46	32098.68	43179.23	27145.2	10
Sweden	54779.6	7727.51	54616	63840	36141	11
Norway	28745.4	5122.74	29392	35784.13	21235	11
Italy	27482.1	2744.85	27214	33645	22882	11
Ireland	2095.6	401.91	2049	2750	1585	10
Germany	78672.1	7521.49	77088.45	91347	67952.32	11
France	39736.1	4262.28	40653	46570	34301	11
Denmark	39166.7	5626.1	40750.06	46193	29732	11
Belgium	6275.49	958.23	6163.25	8732.11	5146.79	11

Source: see text.

Table 1.16. Summary statistics for loading factor

	Mean	Sdev	Median	Max	Min	NbObs
All countries	0.39	0.24	0.42	0.97	-0.34	97
UK	0.55	0.19	0.54	0.97	0.22	10
Sweden	0.14	0.32	0.11	0.94	-0.34	11
Norway	0.49	0.11	0.51	0.66	0.31	11
Italy	0.28	0.12	0.3	0.46	0.04	11
Ireland	0.61	0.28	0.65	0.94	0.14	10
Germany	0.13	0.04	0.12	0.2	0.06	11

[45] The data is from the European insurance industry database.
[46] The loading factor is given by the premium amount not claimed divided by the total claims amount.

France	0.52	0.05	0.51	0.6	0.43	11
Denmark	0.34	0.14	0.34	0.6	0	11
Belgium	0.52	0.17	0.56	0.64	0.06	11

Source: see text.

1.7.4. Results

Table 1.17 presents the EMU(η) estimates and standard errors for all the methods considered in this section. Notice the standard errors (SE) for the estimates derived with the Euler equation and insurance data [Szpiro's (1986) approach] are obtained using the delta method and using bootstrapping. Once again the EMU obtained with the EASA are derived both by averaging the estimates obtained with direct calculation across years per country and by regressing $\ln(1 - MTR)$ against $\ln(1 - ATR)$ per country.

Table 1.17. EMU estimates and standard errors obtained with Equal Sacrifice, Euler equation and Szpiro's (1986) approaches

	ES (average)		ES (reg)		Euler equation (cons.)			Insurance		
Country	η	SE	η	Rob. SE	η	SE (d)	SE (bs)	η	SE (d)	SE (bs)
UK	1.45	0.02	1.37	0.02	2.22	0.77	0.6	2.02	0.19	3.69
Sweden	1.63	0.07	1.68	0.07	9.69	8.72	188	3.67	1.77	77.7
Norway	1.45	0.03	1.47	0.04	4.98	1.74	47	3.28	1.24	44.4
Italy	1.42	0.02	1.46	0.02	2.35	0.96	9	1.63	0.22	13.0
Ireland	1.89	0.06	1.87	0.06	-2.6	1.82	113	2.37	0.67	28.2
Germany	1.44	0.02	1.45	0.02	6.12	4.35	198	2.05	0.64	58.5
France	1.48	0.02	1.44	0.02	47.6	153	522	1.63	0.20	0.35
Denmark	1.58	0.08	1.45	0.03	16.7	20.8	338	4.15	2.52	74.8
Belgium	1.45	0.04	1.48	0.03	2.68	1.14	15	1.79	0.08	0.37

Source: see text. Notes: 1) ES stands for Equal Sacrifice; 2) (d) and (bs) indicate whether the standard error was derived using the delta method or bootstrapping.

To compare the estimates we use Meta-Analysis (MA). The purpose of the MA is to produce an overall best estimate from different estimates of the same parameter whilst having some regard to the fact that competing estimates may have differing degrees of precision so that some should carry more weight. A distinction can also be drawn between those studies that assume that the only possible difference between estimates is due to random sampling rather than differences due to the nature of the methodology employed. The latter are known as random effects estimates. MA has seen significant use in economics in recent years (e.g Groom and Maddison, 2013; Smith and Pattanayak, 2002). Table 1.18, 1.19 and 1.20 present the meta-value, chi-square statistics for homogeneity test and 95% confidence intervals across country and methodology for the EMU estimates.

Note that the Euler equation and the Szpiro (1986) methods do not provide standard errors for the estimates, such that we derive them using both the delta method and bootstrapping.

In table 1.18 the standard errors regarding the estimates obtained with the Euler equation and insurance data are derived by bootstrapping. In table 1.19 the standard errors regarding the estimates obtained with the two referred-to methodologies are derived with the delta method. In both tables the EASA estimates (and standard errors) presented are the ones obtained via regression. Table 1.20 shows the same type of results considering bootstrapping standard errors for the estimates derived with all methodologies.

Table 1.18. EMU meta-values, 95% confidence interval and homogeneity tests across countries and methodologies (bootstrapping SEs for estimates from Euler equation and insurance data)

Country	EMU (ES_reg)	EMU (Euler)	EMU (Insurance)	Meta-value	IC (95%)	Chi-square Homogeneity
UK	1.3750	2.225036	2.021844	1.376	1.337-1.414	1.84 (0.398)
Sweden	1.68258	9.698102	3.675867	1.683	1.541-1.824	0.00 (0.999)
Norway	1.47883	4.984873	3.284117	1.479	1.406-1.552	0.01 (0.996)
Italy	1.46639	2.353571	1.638616	1.466	1.429-1.504	0.01 (0.995)
Ireland	1.87333	-2.67771	2.379725	1.873	1.757-1.989	0.00 (0.999)
Germany	1.4542	6.127598	2.053863	1.454	1.420-1.488	0.00 (1.000)
France	1.44288	47.69287	1.635152	1.444	1.402-1.485	0.29 (0.863)
Denmark	1.45745	16.75575	4.159995	1.457	1.364-1.551	0.00 (0.998)
Belgium	1.48685	2.685955	1.797955	1.490	1.411-1.570	0.68 (0.713)
Meta-value	1.448	2.227	1.714			
IC (95%)	1.432-1.464	0.994-3.460	1.205-2.222			
Chi-square Homogeneity	86.57** (0.000)	0.02 (1.000)	0.11 (1.000)			

*Source: see text. Notes: 1) The number in parenthesis correspond to p-values; 2) ** indicates the relevant coefficient is significant at 1%, * at 5% and ° at 10%.*

Table 1.19. EMU meta-values, 95% confidence interval and homogeneity tests across countries and methodologies (delta SEs for estimates from Euler equation and insurance data)

Country	EMU (ES_reg)	EMU (Euler)	EMU (Insurance)	Meta-value	IC (95%)	Chi-square Homogeneity
UK	1.3750	2.225036	2.021844	1.382	1.344-1.420	12.07** (0.002)
Sweden	1.68258	9.698102	3.675867	1.686	1.545-1.828	2.10 (0.350)
Norway	1.47883	4.984873	3.284117	1.482	1.409-1.555	6.11* (0.047)
Italy	1.46639	2.353571	1.638616	1.468	1.430-1.506	1.42 (0.492)
Ireland	1.87333	-2.67771	2.379725	1.872	1.757-1.988	6.79* (0.034)
Germany	1.4542	6.127598	2.053863	1.455	1.420-1.489	2.01 (0.367)
France	1.44288	47.69287	1.635152	1.445	1.403-1.487	0.98 (0.611)
Denmark	1.45745	16.75575	4.159995	1.458	1.365-1.552	1.68 (0.431)
Belgium	1.48685	2.685955	1.797955	1.549	1.477-1.620	12.37** (0.002)

Meta-value	1.448	2.287	1.810			
IC (95%)	1.432-1.464	1.334-3.240	1.678-1.941			
Chi-square Homogeneity	86.57** (0.000)	11.98 (0.152)	6.75 (0.564)			

*Source: see text. Notes: 1) The number in parenthesis correspond to p-values; 2) ** indicates the relevant coefficient is significant at 1%, * at 5% and ° at 10%.*

Table 1.20. EMU meta-values, 95% confidence interval and homogeneity tests across countries and methodologies (bootstrapping SEs for all estimation techniques)

Country	EMU (ES_reg)	EMU (Euler)	EMU (Insurance)	Meta-value	IC (95%)	Chi-square Homogeneity
UK	1.3750	2.225036	2.021844	1.376	1.344-1.407	1.84 (0.398)
Sweden	1.68258	9.698102	3.675867	1.683	1.546-1.820	0.00 (0.99)
Norway	1.47883	4.984873	3.284117	1.479	1.407-1.550	0.01 (0.996)
Italy	1.46639	2.353571	1.638616	1.466	1.427-1.506	0.01 (0.995)
Ireland	1.87333	-2.67771	2.379725	1.873	1.756-1.990	0.00 (0.99)
Germany	1.4542	6.127598	2.053863	1.454	1.420-1.489	0 (1)
France	1.44288	47.69287	1.635152	1.444	1.398-1.489	0.29 (0.863)
Denmark	1.45745	16.75575	4.159995	1.457	1.393-1.522	0.00 (0.98)
Belgium	1.48685	2.685955	1.797955	1.489	1.422-1.557	0.68 (0.713)
Meta-value	1.448	2.287	1.810			
IC (95%)	1.432-1.465	1.334-3.240	1.678-1.941			
Chi-square Homogeneity	85.79 (0.00)	11.98 (0.152)	6.75 (0.564)			

*Source: see text. Notes: 1) The number in parenthesis correspond to p-values; 2) ** indicates the relevant coefficient is significant at 1%, * at 5% and ° at 10%.*

In all the tables the EMU estimates obtained with the Euler equation and insurance data are homogeneous across countries and the EASA estimates vary across countries. For tables 1.18 and 1.20, there is no variation across estimation technique considering all countries. In table 1.19 the estimates for the UK, Norway, Ireland and Belgium vary across technique.

No definitive conclusion about variation across countries and methodology can be derived from these results. They show different conclusions about EMU variation across countries are reached depending on what methodology is used and different conclusions about variation across methodologies are reached depending on what countries are considered.

However, it is striking that for virtually all countries the meta-values across techniques is close to 1.5, which is the overall EMU value suggested in section 1.6.6, which considers EMU estimates obtained by the EASA. It strongly suggests the EASA-derived value is sound. It also makes one wonder why others (e.g. Atkinson et al., 2009) using survey techniques have reached such different conclusions about the value of EMU estimates derived in different contexts (this is a question we will visit in the final chapter).

The fact the meta-values do not differ significantly across methodologies suggests that the standard welfare economics account of the nature of the EMU as a general inequality aversion parameter – i.e. as an intra-temporal, inter-temporal and across-states-of-nature (risk) inequality aversion parameter – is reasonably accurate. Put simply, the basic intuition why the different methods result in non-significantly different estimated values is that people in general tend to treat the three types of inequality aversion similarly. Yet in other words, the same individual in different points of time or different states of nature is modelled as many persons by the decision-taker when redistributive choices are considered. In chapter 4 I go back to this topic.

1.8. Conclusion

In this chapter I estimate the Elasticity of Marginal Utility parameter (EMU or η) using the Equal Absolute Sacrifice Approach (EASA) for 9 European countries from the 50s to 2007 (one estimate per year). This methodology assumes Governments set income tax schedules based on the Equal Absolute Sacrifice Principle (EASP), which in short determines all individuals paying income tax should have the same loss of utility. Given this assumption the EMU can be obtained from the Average Tax Rate (ATR) and the Marginal Tax Rates (MTR) relative to the income tax schedule set by the Government.

The ATR and MTR for each country and year is obtained from the EroPTax dataset. It simulates the tax rate and social contributions paid by individuals or families given their income, based on the income tax legislation in force.

I use 4 main variations of the income tax data obtained to calculate the parameter. The first regards National Insurance Contributions (NICs) as income tax and the Average Production Wage (APW) as the income level at which the ATR and MTR are calculated (all in/no SI). The second also regards NICs as income tax, but does not consider the APW as it is for the calculation basis; instead it includes an estimated Supernumerary Income (SI) in the basic income level (all in/SI). The third and fourth only differ from the first two in that they do not regard NICs as income tax (no NICs/no SI and no NICs/SI).

After calculating the parameters I perform a series of tests intended to reveal which of the 4 data combination is the most appropriate and whether the estimates are constant across income levels, countries, time and tax unity definition (single and families). The first test consists of comparing the R^2 of regressions run with the different income tax data considering all countries and years. Each regression reveals an EMU as the slope coefficient. There is no econometric interpretations for these regressions, instead they are just a best-fit exercise, which justifies comparing the R^2 to decide what combination is the best. The results indicate the "all in/no SI" combination is the most appropriate.

To test both the constancy of the EMU across income and the EMU constancy across countries I use the chi-squared test of heterogeneity. To test for EMU variation across time I use Cuzick's (1985) nonparametric test of trend across ordered groups, and to test for variation across tax unity definition (single and families) I use paired t-tests. We also compare the R^2 of regressions using family and single people as tax units to determine the most appropriate tax unit to be used. Parameters estimated with single people rendered a better R^2, thus leading to the conclusion that considering such a tax unit is more appropriate.

In so doing I analyse the literature regarding the estimation of the EMU via EASA and approach it systematically, in a way which has not been done before. In other words, although there are many works using the EASA approach to derive the EMU, I am the first to identify what are the key aspects of the estimation procedure differing across previous works and compare them in order to reveal whether different assumptions have a significant impact on the value of the parameter.

The results generally show that the parameter is statistically speaking not constant across income, countries, time and tax unit definition. However, the fact the parameters differ very little across income levels in absolute terms suggest the assumption of iso-elastic utility function is reasonable for the analysis. Moreover, the minor differences of the parameters estimates across countries also suggests national idiosyncrasies (at least with regards to the 9 countries considered) have little importance in shaping the EMU's value.

We suggest a value of 1.5 for the EMU. This value is obtained by the EASA regression method using the ATR and MTR of all countries for all years, the most appropriate combination (all in/no SI) and the most appropriate tax units (single people).

Finally, we compare the EMU estimated by EASA with EMU estimates obtained with different methodologies, namely the Euler equation and Szpiro's (1986) technique. The first uses data on consumption growth rate for non-durable goods and derives the parameter as an intertemporal inequality aversion index. The second uses data on insurance and derive the parameter as a risk aversion parameter. The results show the EMU estimates are consistent across estimation methods, which substantiates the EMU's suggested value of 1.5.

Moreover, the fact the three methods implemented derive the parameter in different contexts makes it even more interesting. It seems to confirm the findings of Groom and Maddison (2013) who also suggest a EMU value close to 1.5 and present results suggesting that estimates of the EMU derived in different contexts produce values that do not vary significantly (but only for the United Kingdom).

References

Altonji, J. G. (1986). Intertemporal Substitution in Labor Supply: Evidence from Micro Data. Journal of Political Economy, 94(3), S176-S215.

Amiel, Y., Creedy, J., & Hurn, S. (1999). Measuring Attitudes Towards Inequality. Scand. J. of Economics, 101(1), 83-96.

Aristei, D., & Perugini, C. (2010). Preferences for Redistribution and Inequality in Well-Being Across Europe. Journal of Policy Modeling, 32, 176-195.

Arrow, K. J., & Lind, R. C. (1970). Uncertainty and the Evaluation of Public Investment Decisions. The American Economic Review, 60(3), 364-378.

Atkinson, G., Dietz, S., Helgeson, J., Hepburn, C., & Sælen, H. (2009). Siblings, Not Triplets: Social Preferences for Risk, Inequality and Time in Discounting Climate Change. Economics: The Open-Access, Open-Assessment E-Journal, 3, 2009-26.

Barsky, R. B., Juster F. T., Kimball M. S., & Shapiro M. D. (1997). Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study. Quarterly Journal of Economics 112(2), 537–579.

Blue, E.N., & Tweeten, C. (1997). "The estimation of marginal utility of income for application to agricultural policy analysis", Journal of Agricultural Economics, 16, 155-169.

Blundell, R., Browning, M., & Meghir, C. (1994). Consumer Demand and the Life-Cycle Allocation of Household Expenditures. Review of Economic Studies, 61, 57-80.

Boardman, A., Greenberg, D., Vining, A., & Weimer, D. (2001). Cost-Benefit Analysis: Concepts and Practice. Upper Saddle River: Pearson Education.

Buchholz, W., & Schumacher, J. (2010). Discounting and welfare analysis over time: Choosing the eta. European Journalnof Political Economy , 26, 372-385.

Carlsson, F., Daruvala, D., & Johansson-Stenman, O. (2005). Are People Inequality-Averse, or Just Risk-Averse? Economica, 72, 375–396.

Costa, C. E., & Pereira, T. (2013). On the Efficiency of Equal Sacrifice Income Tax Schedules. European Economic Review, 70, 399-418.

Cowell, F A., & Gardiner, K (1999) Welfare Weights. STICERD, London School of Economics.

Da Silva, S., Matsushita, R., & De Carvalho, M. (2015). Prosocial People Take Better Care of Their Own Future Well-Being. Open Access Library Journal, 2, e2181.

Dasgrupta, A. K., & Pearce D. W. (1972). Cost-Benefit Analysis: Theory and Practice (First ed.). London: Macmillan.

Dasgupta, P. S., Maler, K. G., & Barrett, S. (1999). Intergenerational Equity, Social Discount Rates and Global Warming, in Portney, P. R. and Weyant, J. P. Discounting and Intergenerational Equity. Washington, DC: Resources for the Future, 51-78.

Dupuit, J. (1848). Theoretical and Practical Studies on the Movement of Running Water (Second ed.). Paris: Dunod.

Epstein, L. G., & Zin S. E. (1989). Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: a Theoretical Framework. Econometrica 57(4), 937– 968.

Evans D. J., & Sezer H. (2002). A Time Preference Measure of the Social Discount Rate for the UK. Applied Economics 34, 1925-1934.

Evans, D. J., & Sezer, H. (2004). Social discount rates for six major countries. Applied Economics Letters, 11, 557–560.

Evans D. J., & Sezer H. (2005). Social Discount Rates for Member Countries of the European Union. Journal of Economic Studies, 32(1), 47-59.

Evans D. J. (2004). The Elevated Status of the Elasticity of Marginal Utility. Applied Economics Letters, 11, 443-447.

Evans, D. J. (2005). The Elasticity of Marginal Utility of Consumption: Estimates for 20 OECD Countries. Fiscal Studies, 26(2), 197–224.

Evans, D. J. (2008). "The marginal social valuation of income for the UK", Journal of Economic Studies, 35(1), 26-43.

Evans, D. J., Kula, E., & Sezer, H. (2005). Regional welfare weights for the UK: England, Scotland, Wales and Northern Ireland. Regional Studies, 39(7), 923-937.

Fellner, W. (1967), "Operational utility: the theoretical background and a measurement". In: Fellner, W. Ten Economic Studies in the Tradition of Irving Fisher, John Wiley & Sons, New York, NY, pp. 39-75.

Gollier, C. (2013). Pricing the Planet's Future. Oxfordshire: Princeton.

Groom, B., & Maddison, D. (2013). Non-identical Quadruplets: Four new estimates. Centre for Climate Change Economics and Policy, Working Paper n. 141 and Grantham Research Institute on Climate Change and the Environment, Working Paper n. 121.

Hall, R. E. (1978). Stochastic Implications of the Life Cycle - Permanent Income Hypothesis: Theory and Evidence. Journal of Political Economy, 86(6), 971-987.

Harberger, A. (1972). Project Evaluation: Collected Papers (First ed.). Chicago: The University of Chicago Press.

Harberger, A., & G. Jenkins. (2002). Cost–Benefit Analysis for Investment Decisions. Queen's University, Canada.

Hicks, J. (1939). The Foundations of Welfare Economics. Economic Journal, 49, 696-712.

Kaldor, N. (1939). Welfare Propositions of Economics and Interpersonal Comparison of Utility. Economic Journal, 49(195), 549-552.

Kula, E. (2002). Regional Welfare Weights in Investment Appraisal - The Case of India. The Journal of Regional Analysis & Policy, 32(1), 99-114.

Kula, E. (2004). Estimation of a Social Rate of Interest for India. Journal of Agricultural Economics, 55(1), 91-99.

Layard, R., Nickell, S., & Mayraz, G. (2008). The Marginal Utility of Income, 92, 1846–1857.

Lopez, H. (2008). The social discount rate : estimates for nine Latin American countries. Policy Research working paper ; no. WPS 4639. Washington, DC: World Bank.

Lynch, F. (2009). Who pays for the State? The evolution of personal taxation in post-war Europe: Full Research Report ESRC End of Award Report, RES-000-22-1408. Swindon: ESRC.

Lynch, F., & Weingarten, N. (2010). EuroPTax. Who Pays for the State? The Evolution of Personal Taxation in Postwar Europe, 1958-2007. [data collection]. UK Data Service. SN: 6358, http://doi.org/10.5255/UKDA-SN-6358-1.

Lyon, R. M. (1990). Federal Discount Rate Policy, the Shadow Price of Capital, and Challenges for Reforms. Journal of Environmental Economics and Management, 18, S-29-S-50.

Mareike, S., & Jurgen, J. (2012). Towards a Social Discount Rate for the Economic Evaluation of Health Technologies in Germany: an Exploratory Analysis.

Marshall, A. (1920). Principles of Economics. London: Macmillan.

Mirrlees, J. A. (1971). An Exploration in the Theory of Optimum Income Taxation. The Review of Economic Studies, 38(2), 175-208.

Morrison, C. J., & Schwartz, A. E. (1996). State Infrastructure and Productive Performance. The American Economic Review, 86(5), 1095-1111.

Otto, E. (1958). Water Resource Development: The Economics of Project Evaluation. Cambridge: Harvard University Press.

Pearce, D., & Ulph, D. (1995). A Social Discount Rate for the United Kingdom. CSERGE Working Paper GEC 95-01.

Percoco, M. (2008). A social discount rate for Italy, Applied Economics Letters, 15(1), 73-77.

Pirttila, J., & Uusitalo, R. (2008). A 'Leaky Bucket' in the Real World: Estimating Inequality Aversion using Survey Data. Economica, 77, 60-76.

Pomerleau, K. (2014). Sources of Government Revenue in the OECD, 2014. Tax Foundation, Fiscal Fact, no. 443.

Ramsey, F. (1928). A Mathematical Theory of Saving. Economic Journal, 38(152), 543-559.

Richter, W. F. (1983). From ability to pay to concepts of equal sacrifice. Journal of Public Economics, 20, 211-29.

Sezer, H. (2006) "Regional welfare weights for Turkey", Journal of Economic Studies, 33(5), 357-368.

Smith, V. K., & Pattanayak, S. K. (2002). Is meta-analysis a Noah's ark for non-market valuation? Environmental and Resource Economics, 22(1-2), 271-296.

Spackman, M. (2004) Time Discounting and the Cost of Capital in Government Fiscal Studies, 25(4), 467-518.

Stern, N. (1977) Welfare weights and the elasticity of the marginal valuation of income. In: Artis, M. and Nobay, R. Proceedings of the Aute Edinburgh Meeting of April 1976. Basil Blackwell.

Szpiro, G. G. (1986). Measuring Risk Aversion: An Alternative Approach. The Review of Economics and Statistics, 68(1), 156-159.

Van Dalen, H. (1995). Intertemporal Substitution in War and Peace: Evidence from the United Kingdom, 1830-1990 Journal of Macroeconomics, 17(3), 447-469.

Vitaliano, D. F. (1977), "The tax sacrifice rules under alternative definitions of progressivity", Public Finance Quarterly, 5, 489-94.

Young, H. P. (1990). Progressive Taxation and Equal Sacrifice. The American Economic Review, 80(1), 253-266.

Young, H. P. (1987). Progressive taxation and the equal sacrifice principle. Journal of Public Economics, 32, 203-12.

Zerbe, R. & Dively, D. (1994). Benefit Cost Analysis in Theory and Practice. New York: Harper Collins.

Zhuang, J., Liang, Z., Lin, T., & De Guzman, F. (2007). Theory and Practice in the Choice of Social Discount Rate for Cost-Benefit Analysis: A Survey. ERD Working Paper, no. 94.

Appendix 1.A – Examples of survey questionnaires

Leaky bucket type question in Pirttila and Uusitalo (2008) for η = 1:[47]

What is your opinion of the following reform proposal?

The taxation of all high-income earners, whose disposable income exceeds 3300 per month, is increased. The money is spent for the benefit of those low-income earners whose disposable income is less than 800 per month.

The high-income earners can, however, react to the tax increase by reducing their work effort, and part of the money goes to administrative expenses. Therefore, for each 100 paid by the high-income earners, only 25 can be spent for the benefit of low-income earners.

Are you, nevertheless, in favour of this proposal?

1. Yes
2. No
3. Cannot say

Income distribution comparison type of question in Pirttila and Uusitalo (2008) for η = 1:[48]

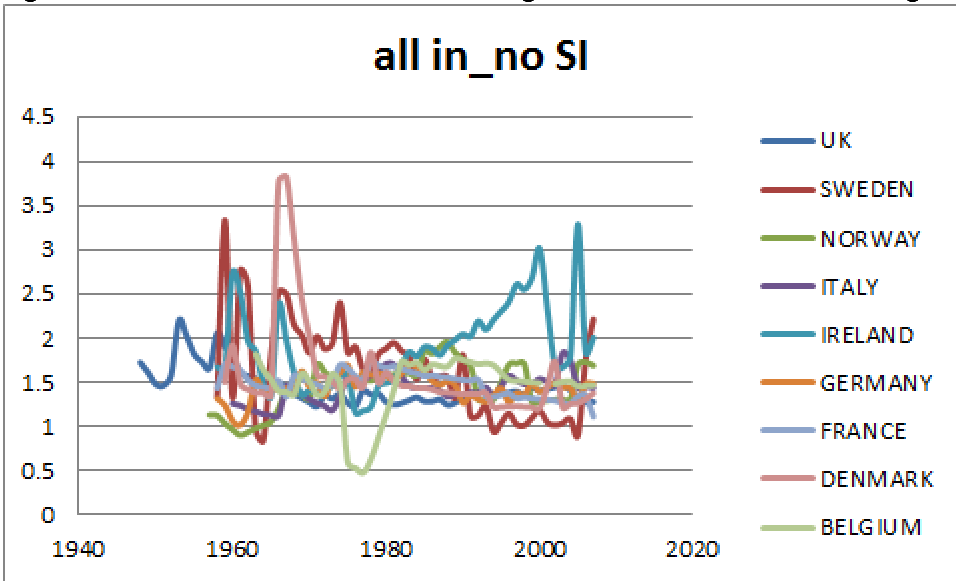
Let us imagine that in wage negotiations two different alternatives are considered. Which of the following do you prefer?

1. If all employees are ordered from the lowest-income earner to the highest-income earner, someone belonging to the lowest decile earns £1570 in a month, a person with average income earns £2340 and a person belonging to the highest decile earns £3480.
2. Income differences rise and the average income is increased, so that the low income earner gets £1280 per month, the person with average income gets £2580 and the high-income earner £5190.
3. Cannot say.

[47] For other η values the amount paid by high earners which is lost (which in this case is 75) is different.
[48] For other values of η the lowest decile, the average and the highest decile change.

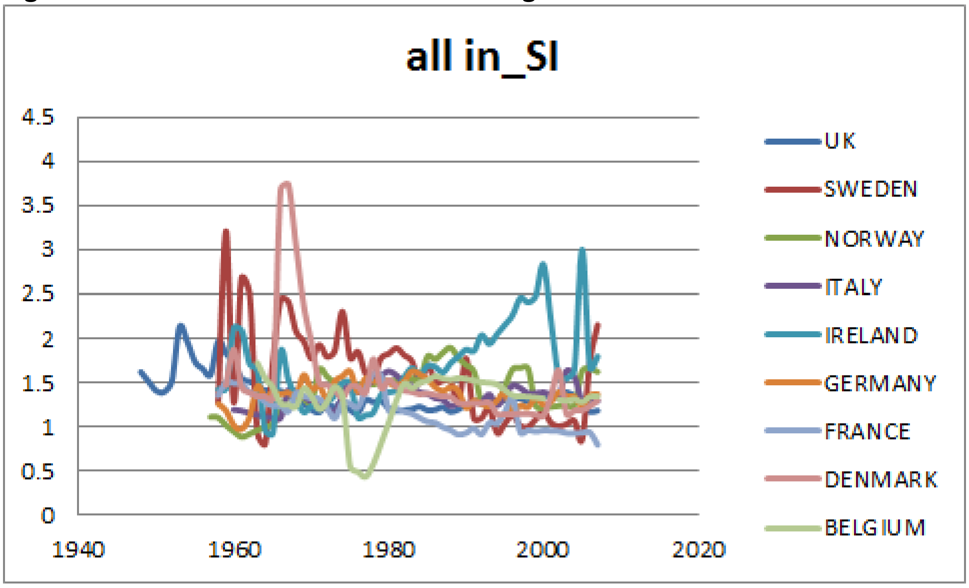
Appendix 1.B – EMU estimates Graph

Figure 1B.1. Estimates obtained considering all in taxes and not considering SI



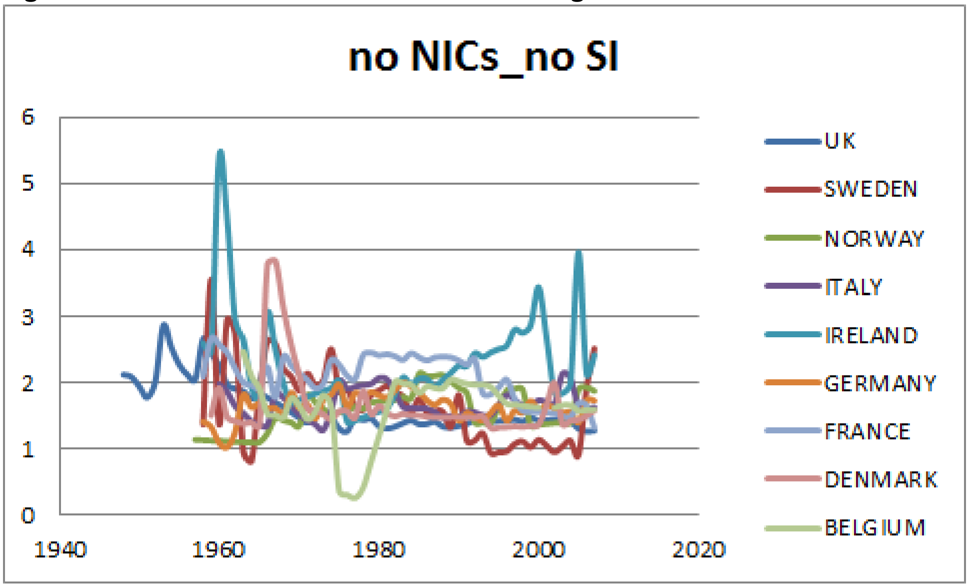
Source: see text

Figure 1B.2. Estimates obtained considering NICs and SI



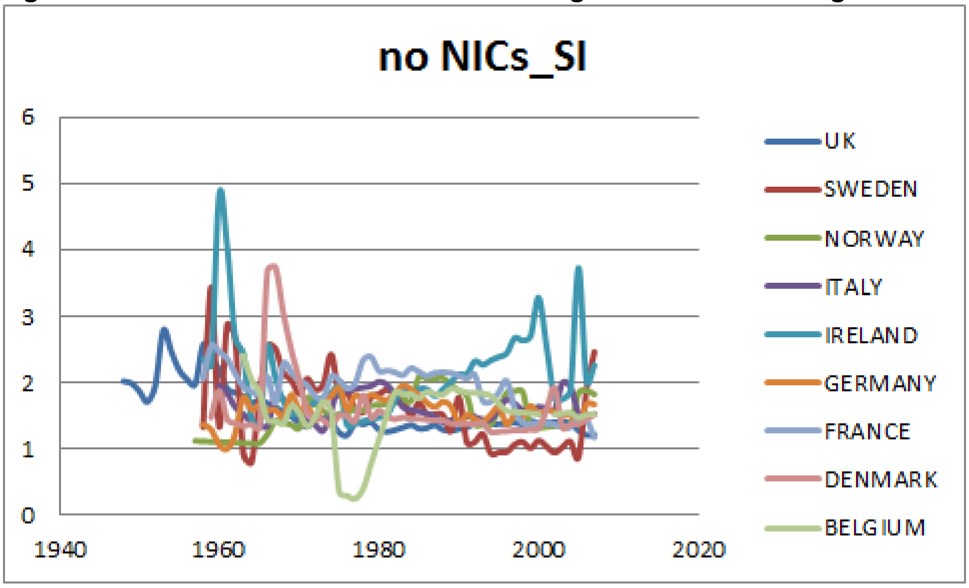
Source: see text

Figure 1B.3. Estimates obtained not considering NICs and SI



Source: see text

Figure 1B.4. Estimates obtained not considering NICs and considering SI



Source: see text

Appendix 1.C – EMU constancy across time (tables)

Table 1C.1. Constancy of η across time (data separated per 5 years)

Country	all in/no SI	all in/SI	no NICs/no SI	no NICs/SI
UK	0	0	0	0
SWEEDEN	0	0	0	0
NORWAY	0	0	0	0
ITALY	0	0.015	0.3	0.9
IRELAND	0	0	0.34	0.13
GERMANY	0.8	0.375	0.87	0.95
FRANCE	0	0	0	0
DENMARK	0	0	0.01	0
BELGIUM	0.26	0.5	0.46	0.66

Note: The numbers shown in the table correspond to the p-values of the trend z test.

Table 1C.2. Constancy of η across time (data separated per 15 years)

Country	all in/no SI	all in/SI	no NICs/no SI	no NICs/SI
UK	0	0	0	0
SWEEDEN	0	0	0	0
NORWAY	0	0	0	0
ITALY	0	0	0.12	0.45
IRELAND	0	0	0.22	0.06
GERMANY	0.85	0.23	0.92	0.76
FRANCE	0	0	0	0
DENMARK	0	0	0	0
BELGIUM	0.22	0.47	0.37	0.53

Note: The numbers shown in the table correspond to the p-values of the trend z test.

Chapter 2: The Determinants of the Elasticity of Marginal Utility as Estimated by the Equal Absolute Sacrifice Approach

2.1. Introduction

In this chapter I provide the first analysis (to the best of my knowledge) on the determinants of the EMU as estimated by the EASA. In the EASA context the parameter is seen as a measure of societal inequality aversion through the progressivity of the income tax schedule. It is implied that the Governments' inequality aversion represents the general population's inequality aversion (in a democratic country).

However, as others have notice (e.g. Spackman , 2004), there are other aspects that affect Governments' decisions to set income tax schedules apart from representing its population's aversion to inequality.

Herein I allow for a more realistic look at the EMU – one in which it is formed not merely through society's aversion to intratemporal inequality but which also takes account of other factors that might impinge on the Government's desire or ability to set tax rates in order to reflect societal aversion to inequality. These might include features of the democratic process and certain economic constraints e.g. tax competition and concerns about the impact of high taxes on labour market incentives. These were pushed into the background of the analysis performed in chapter 1. To the extent that certain factors influence the value of the EMU it may be less appropriate to view it as reflecting solely societal aversion to inequality. At the same time, the influence of the general public on governmental decision-taking concerning taxes is not disregarded, but seen as one of the factors driving tax policies.

In other words, I attempt to determine to some degree the extent to which other factors besides the Government's presumed desire to democratically address society's aversion to intratemporal inequality impinge on the setting of income tax rates. In fact, there is already a literature on the determinants of income tax progressivity (although there aren't many papers), and the estimate of EMU from the EASA can be interpreted as another measure of income tax progressivity. So, I try to explain my series for the EMU as revealed by the EASA using the same variables that a handful of others have used to explain tax progressivity.

The research is best interpreted as an attempt to validate the EASA approach to estimating EMU rather than another analysis of the determinants of income tax progressivity. I am interested in investigating the susceptibility of the estimates of EMU to factors having nothing to do with the underlying aversion to inequality of society.

The empirical analysis includes variables describing the population of a country e.g. its demographic composition. The statistical significance of these variables could simply reflect that a society with a different demographic composition wants different things. The statistical significance of other variables is however harder to explain except for saying that Governments do indeed face constraints or have goals of their own quite apart from reflecting society's aversion to inequality. To the extent that these variables are significant I might be less inclined to believe that the results obtained in chapter 1 really are measures of EMU.

Notice I am the first to investigate whether estimates of EMU from the EASA can actually be explained by political constraints and institutional features. In so doing I hope to be less naive than others who popularised the EASA technique and have simply accepted these estimates of EMU at face value.

I use the EMU estimates obtained via direct calculation in the previous chapter (per country and year) and carry out an empirical analysis by implementing a standard fixed effects model. Although the explanatory variables hypothesized to be associated with the EMU are drawn from those expected to be related to tax progressivity, as mentioned above, I am not going to limit myself necessarily to including just these variables.

The results show the dependency ratio, governmental stability, openness, population and governmental expenditure with education are associated with measured EMU.

The structure of this chapter goes as follows: firstly I make explicit why the EMU estimated by the EASA can be interpreted as a measure of income tax progressivity. Next I turn to the literature review regarding the main variables affecting income tax progressivity. Then I give more details about the data used and the empirical strategy. In section 5 I present the results and discuss them and in section I 6 conclude the chapter.

2.2. The EMU (η) as an index of tax progressivity

The EMU estimated via EASA is a measure of the progressivity of the income tax schedule. To see it consider that a given income tax schedule can be regarded as progressive at income level y if $\frac{dATR(y)}{dy} > 0$, where ATR (at y) stands for the Average Tax Rate, given by $T(y)/y$, where $T(y)$ is the tax burden at y.

By the definition of local income tax progression given above it can be shown that the EMU estimates obtained via EASA are indices of income tax progression at the APW, as demonstrated below (Seidl, 2009):

$$\frac{dATR(y)}{dy} = \frac{d}{dy}\frac{T(y)}{y} = \frac{T\prime(y)y - T(y)}{y^2} = \frac{MTR(y) - ATR(y)}{y} > 0 \Leftrightarrow MTR(y) > ATR(y) \Leftrightarrow$$
$$\frac{\ln(1 - MTR(y))}{\ln(1 - ATR(y))} = \eta > 1, \qquad (2.1)$$

where MTR (at y) stands for Marginal Tax Rate which can be expressed as $T'(y)$.

Once we have established the relationship between EMU estimates obtained by the EASA and indices of income tax progressivity, we can search among variables determining income tax progressivity in order to increase our understanding on the determinants of the measured EMU. That is done in the next section.

2.3. Literature review on determinants of income tax progressivity

In order to locate relevant studies related to the topic under investigation (the determinants of income tax progressivity) I used boolean operators and different databases. First I searched in EconLit with following search terms: "determinant* AND income tax (schedule* OR system* OR structure* OR progress*)."[49] In this occasion 37 works were retrieved, but just one work was selected to further investigation. Another search was undertaken on Google using variants of the phrase "determinants of income tax schedule" and "determinants of tax progressivity". In that case 10 works were kept for further investigation. Also, the references of works related to the application of the EASA were analysed, and 4 works were selected from further investigation. More than three other searches were undertaken in EconLit, using the words: "income tax* AND progress* AND measure*", "determinant* AND tax* AND progress" and "tax progress* measurement". In total 329 works were retrieved and 32 were kept for further investigation.

ProQuest was also used along with the search words: "determinant* AND income tax AND progress*" and "tax progress* measurement". Just five new works were kept for further investigation. The references of the works selected were also examined and more 23 works were separated for further investigation. From all works kept for further analysis, 10 were relevant and thus used in the following literature review (which is surprisingly few given the topicality of

[49] Remember the asterisk means that the form of the word from the asterisk on can change. So the command progress* would cause the dataset to search for words like progressiveness, progression and so on. The operator OR indicates words that must be tracked interchangeably and the operator AND indicates words to be tracked in conjunction.

the subject). Moreover, I examined the bibliography of the works selected to see whether additional works were left out.[50]

Amongst the works retrieved and analysed, two resemble closely the analysis undertaken – given the way they measure Personal Income Tax (PIT) progressivity, as discussed later on. They are Egger et al. (n.d.) and Fletcher and Murray (2008) (hereafter FM). The first considers 209 countries in the period of 1980 to 2009 and examines the determinants of income tax progressivity considering six different indicators of progressivity, but focussing on two of them: the difference in average tax rates between ten times the average wage and a wage of zero and the difference in marginal tax rates between ten times the average wage and a wage of zero. For these indices they employ generalized linear models for non-negative data (Poisson pseudo-maximum likelihood model and a negative binomial model), since the data generated by the indices is non-negative.

FM consider US states in 2003. Aspects of the PIT which can indicate progressivity are used such as the top bracket rate, the number of brackets, and the income threshold for the first bracket. Tobit regression analysis is used to investigate the issue given all indices are said to be subject to some censoring.

Table 2.1 shows the variables considered as possible determinants of to the PIT progressivity indexes and their expected signs in the aforementioned studies. I list all variables considered, even the non-significant ones.[51]

Table 2.1. Factors affecting PIT progressivity in Egger et al. (n.d.) and Fletcher and Murray (2008)

Variables	Works	Expected signs
Labour supply elasticity	E*	Negative
Marginal social utility of income*	E	Positive
Population	E	Positive
Unemployment rates	E	Negative
Inflation	E	Positive
Dependent ratio*	E	Positive
Openness*	E	Positive
Income inequality*	E	Positive
Ratio of pop with sec. educ.*	E	Negative
Political orientation: right	E	Negative
Political orientation: left	E	Negative
Political stability	E	Positive
Political process participation	E	Positive
Unitary or federalist structure	E	Positive
Average gross wage	E	Positive
Clothing exemption	FM*	Positive

[50] Many studies investigating tax progressivity focus on Governments' tax mix choice instead of determinants of income tax progressivity. Since such topic does not contribute to the one investigated here these works are not considered.

[51] I focus on Egger et al. (n.d.) results table and concluding section to present the expected signs relative to such work and on the results regarding top bracket rate to present the expected signs regarding FM.

Welfare benefits	FM	Positive
Sales tax share of revenue	FM	Positive
PIT share of revenue	FM	Positive
Years since PIT adoption	FM	Positive
Democrat governor	FM	Negative
Same party dominates state government	FM	Positive
Federal aid	FM	Positive
County income ratio	FM	Negative
Percentage deducting state and/or local taxes	FM	Positive
Population density	FM	Positive
Median income	FM	Negative
Percentage black	FM	Negative
Percentage older than 65	FM	Negative
Percentage in poverty	FM	Negative
Percentage younger than 18	FM	Positive
Percentage homeowner	FM	Negative

*Source: see text. *See notes. Notes: (1) E stands for Egger et al. (n.d.); (2) The marginal social utility of income is measured by "the share of unemployment insurance benefits a worker receives relative to her last gross earnings" (Egger et al. n.d.); (3) Dependent ratio is the sum of the shares of the population over 64 and below 14 in the overall population; (4) The level of openness is measured by the sum of exports and imports divided by the GDP; (5) The inequality aversion is given by the gross Gini indices; (6) "Ratio of pop with sec. educ." stands for ratio of population with secondary education; (7) The variables shown for Fletcher and Murray (2008) are just those found to be significantly related to one of the PIT aspects considered as indicator of the degree of progressivity.*

Since FM investigate the PIT progressivity for US states, some variables ["Federal aid" (to capture the effect of decreasing the cost to provide assistance for poor households), "Percentage deducting state and/or local taxes" (to capture the effect of decreasing the collection cost) "Democrat governors" (to capture government preference for redistribution), "Same party dominates state government" (to capture degree of political stability) and "County income ratio" (to capture income distribution)] are not directly applicable to this study, but nevertheless address aspects that can be considered in a multi-country analysis, as the present case, through different indices.

There is some overlap between the variables chosen by the two studies, such that just those which are expected to best capture the factors concerned should be used. I give preference to the variables used in Egger et al., n.d..[52] This is particularly the case for the variables intended to capture income inequality. For example Egger et al. (n.d.) uses the Gini and average gross wage[53] as indicators of inequality whilst FM uses both the "County income ratio"[54] and the median income for each state. In my case the Gini index is more appropriate, since it is usually calculated for countries. By contrast, the variables "Dependent ratio"[55] in Egger et al. (n.d.), and

[52] Due to the fact they use countries as the analysis unit.
[53] To justify average gross wages as an index of income inequality Egger et al. (n.d.) cite evidence (Piketty and Saez, 2006) suggesting that in the last decades in developed economies, increases in income are accompanied by increases in (pre-tax) inequality.
[54] It refers to the ratio between the highest- and lowest-county per capita income in each state.
[55] It is given by the number of those aged more or 64 plus those aged less or 14 divided by the rest of the population.

“Percentage older than 65” and “Percentage younger than 18” in FM are almost identical in terms of capturing the effect of population age structure on progressivity.

I classify the variables used in the different works considered with the intention to select those which were significantly related to progressivity at least twice[56]. But whilst I am including variables used in these and other studies examining PIT progressivity, I am open to using other variables too because my index is not viewed merely as a measure of progressivity, but also a measure of societal aversion to inequality. Egger et al. (n.d.), FM and the other studies included in the review all have a more narrow view of the work they were doing.

Egger et al.’s (n.d.) criteria to choose variables potentially affecting PIT progressivity is based on the modern theory of optimal taxation and on positive theories of Government behaviour.[57] From the first follows two basic predictions (see Appendix 2.A): 1) the optimal MTR is smaller the larger the average (compensated) labour supply elasticity with respect to tax of the overall population, and 2) the optimal MTR is larger the greater is the covariance between the marginal social benefit of income to a worker and his gross income. The inclusion of the variables “Labour supply elasticity”[58] and “Marginal social utility of income”[59] come directly from theory.

Other variables, such as the Gini coefficients, average gross wage, unemployment, openness, inflation, secondary education level and “Dependent ratio”, are included due their hypothetical relation to income inequality, which was expected in Egger et al. (n.d.) to influence the progressivity for the same reasons the “Marginal social utility of income” would do (i.e. because of Governments or people redistribution preferences).

Based on positive (as opposed to normative) views of Government behaviour Egger et al. (n.d.) pick two political variables. “Political orientation”, which is chosen based on the general perception that left-wing parties would be prone to increase the progressivity of a given PIT schedule [although they recognize that earlier empirical evidence is ambiguous about the subject (e.g. Alt and Lassen, 2006)] and “Political process participation”, which is selected based on the median voter hypothesis, according to which when inequality causes average income to

[56] The fact a variable is found to be significantly related to progressivity by (at least) two different studies substantiates the importance of the variable via replication.

[57] This is, theories regarding how Governments actually behaves, as opposed to theories on how Government should behave (normative theories).

[58] Given by $\frac{\ln(labour\ force\ participation\ at\ year\ t)-\ln(labour\ force\ participation\ at\ year\ t-1)}{\ln(net\ of\ tax\ wage\ at\ t-1)-\ln(net\ of\ tax\ wage\ at\ t-2)}$.

[59] Given by “the share of unemployment insurance benefits a worker receives relative to her last gross earnings (Egger et al., n.d.).

be larger than the median voter's income, the latter prefer a more progressive tax (Meltzer and Richard, 1981).

The basic point of Meltzer and Richard (1981) is that if there is enough participation in the political process, the median voter will bear inequality up to the point where his income is not below the average income. After that point, a rational median voter votes in favour of a more progressive income tax schedule, given it benefits him directly. The other political variables ("Political stability", and "Unitary or federalist structure) are picked with no explicit priors. The first seeks to capture concentration of power and the second political stability.

FM's approach to selectíng variables possibly determining PIT progressivity is different from Egger et al. (n.d.), and more similar to most of the other works considered in the remainder of this literature review. The theoretical basis for the empirical model in this case is found in the work of Hettich and Winer (1999) (who present what is also known as the probabilistic voting model), where the tax system is modelled as a result of government rational action towards the maximization of expected support in terms of votes. The basic prediction of the model is that Governments will tax less those who are more likely to participate in the electoral process (see Appendix 2.B for more details about the probabilistic voting model). Notice that this presents a very different view of the Government: trying to maximise its support rather than reflect societal preferences for inequality aversion.

The variables "Percentage black", "Percentage in poverty" and "Percentage older than 65" are chosen based on the fact that in the US black, poorer and younger people are expected to be less prone to vote than elders, white and richer individuals. It causes the political costs of levying tax from elders, white and wealthy to be higher than the political cost related to taxing black, poor and young people.[60]

The collection of data on "Clothing exemption" and "Welfare benefits" aims to test if redistribution policies act as substitutes or complements to tax progressivity. More specifically, they are used to test whether States in which these exemptions and benefits are largely implemented tend to have less progressive income taxes (in this case suggesting the programs serve as substitutes) or more progressive income taxes (in this case suggesting the programs serve as complements).

[60] Assuming the groups have the same tax base elasticity (i.e. the change in their taxable base given a change in the tax rate applicable to them is the same). However, the difference in tax rate is even larger if we assume the tax base elasticity of elder, wealthy and white people is higher, which is probably the case.

“Sales tax share of revenue” and “PIT share of revenue” capture the State´s reliance on one of the two referred fiscal tools. Hettich and Winer (1999) argue that larger reliance on any given tax increases the opposition towards it, thus increasing the political costs related to making it regressive. Therefore, larger reliance on PIT is hypothesized to be related to more progressive income tax schedules, while larger reliance on sales tax is hypothesized to be related to more regressive income tax schedules.

”Years since PIT adoption” is intended to capture the influence of the historical sentiments related to PIT schedules when they were first formulated. The hypothesis underpinning it is that States which adopted a more progressive (regressive) PIT at the beginning may have more progressive (regressive) taxes currently, indicating people become accustomed to particular tax schedules [they follow Morgan (1994) in this respect].

“Democrat governor” and “Same party dominates state government” are intended to reflect the political composition of the State. Although it is not clearly stated how these variables would influence tax structure, they can be claimed to express the governments ideology and the degree of political competition (how influential a party is in the State, and therefore how unpopular the measures it takes can be without diminishing significantly its power), respectively. “Population density” is reported as a simple control. No further explanation on how it could affect tax structure is given. Also, no direct explanation is given for the inclusion of “Percentage homeowner”.

The literature review undertaken identified many other works investigating the determinants of tax progressivity, but these others considered the general tax system, unlike the two presented before, which took into account just PIT progressivity. It is important to note however, that given some studies choose indicators of progressivity associated with PIT (e.g. top income tax rates) to measure the progressivity of the whole tax system (including sales taxes, property taxes, etc.), one can regard them as measuring PIT progressivity directly – although their intention was broader. This is the case for Galli (2002) and Scheve and Stasavage (2010). In my description of these works just the most important (statistically significant)[61] variables affecting tax structure are considered.

Galli (2002) finds that per capita income, government expenditures on transfers and subsidies, the degree of openness, EU membership, the coalition variable (measuring the power dispersion in the government) and government instability are related to tax progressivity, which is

[61] At 5% confidence level.

measured by the statutory top marginal tax rates on personal income. The author applied fixed effects regressions on a panel of 16 OECD countries between 1965 to 1995.[62]

In Galli (2002) the variables of interest were divided in three groups, the economic, the interest groups and the politico-institutional variables. According to the author, the first set would mainly measure the need for redistribution of income, the second would measure the impact of different voter groups on progressivity and the last would reflect governmental and political aspects which can increase or decrease tax progressivity.

Per capita income, government expenditures on transfers and subsidies, the degree of openness and EU membership are economic variables. The first is included more as an indirect driver, since it is assumed to impact governmental spending; the second is thus justified on grounds of complementarity / substitutability between government expenditure and tax policy. Openness and EU membership are variables included to control for the impact economic integration may have on PIT progressivity.[63] Both openness and integration increase the cost of progressivity, given they increase people's flexibility regarding their country of residence. In other words, they are expected to increase the elasticity of taxable activities (Slemrod, 2000).

The variables "coalition" (political competition) and "government instability" are institutional-political. The hypothesis as to how the first relates to progressivity is that if power is dispersed in the government (coalition government), there will be probably more debts and spending, given the difficulty to balance the different interests of the coalition (Roubini and Sachs, 1989). The author then argues that since it is politically easier to raise funds by augmenting tax burden than by diminishing expenditures, it can potentially influence progressivity. For government instability the prior is similar. It is argued that since a strong Government is needed to cut public spending, unstable Governments are more likely to present a deficit bias (Haan and Sturm, 1994). Thus, in case austerity measures are needed such Governments would be prone to raise taxes to finance spending. Notice an income tax system intended to raise more revenue may be more or less progressive in character, but progressivity is likely to be impacted.[64]

[62] The regressions were weighed least square models with White heteroscedastic consistent variance and covariance matrix.
[63] The study makes no reference to the possible influence of the EU on income tax rates. The variable just serves as an indicator of the degree of economics integration.
[64] Although these arguments are more directly related to the whole tax system, they are indirectly linked to income tax progressivity.

Scheve and Stasavage (2010)[65] analyse how war mobilization may affect tax progressivity using a diff-in-diff approach. They compare countries which entered in World War I (WWI) and those which did not. They find that war mobilization, GDP per capita, male universal suffrage and revenue to GDP affect tax progressivity. The latter is measured by the highest marginal income tax rate in the income tax schedule.

The study's main hypothesis is that mobilization for war creates an atmosphere where a new social pact is established, according to which the population in general agrees to fight whilst the rich agree to bear a larger tax burden. To test the hypothesis, they use a dummy variable indicating countries participating and not participating in WWI. Also included are a variable indicating the percentage of seats in the legislature occupied by leftist parties (left seat share; which was not significant), country fixed effects and time trends or year fixed effects. In all specifications only the "left seat share" variable was not significant.

The dummy indicating countries and years in which there was male universal suffrage is intended to address the hypothesis that progressivity depends on the degree of political rights possessed by the population. Revenue to GDP (ratio of central government revenue to GDP), on the other hand, indicates whether Governments altered progressivity due to revenue needs – again it is not clear whether this requires greater or lesser degrees of progressivity. No clear prior is given as to why GDP per capita is included.

Foster (2013) finds that ideology, ethnic congruence between poor and non-poor, degree of tension between ethnic groups, poverty, sales activity and neighbours' progressivity influence tax progressivity. The tax progressivity indices used in the work are the ratio between the ATR faced by the top income quintile and the bottom income quintile, the ratio between the ATR in the top and in the middle quintile, and the ratio between the ATR in the middle quintile and in the bottom quintile. His analysis is confined to US states.

It must be noticed that the measures of progressivity appear to refer to the whole tax burden,[66] such that the variables thought to affect progressivity in the study may not apply to PIT progressivity in the same manner. In any case, it seems there is no reason to expect all the factors considered to influence just tax systems as a whole and not PIT progressivity.

[65] This paper is further developed by the authors in their book Scheve and Stasavage (2016), which is the main reference for the next chapter.

[66] The work is not clear about whether the ATR refers to income tax or tax burden as a whole, but the overall context of the study reveals they refer to the tax system as a whole.

His explanation for hypothesizing that ideology affects tax progressivity comes from the studies of Fong (2001) and Alesina and La Ferrara (2005), where the preference for redistribution is said to be related to conceptions of fairness, which are, according to the author, based upon established ideologies. The choice of the variable "ethnic congruence between poor and non-poor" is based on studies undertaken by social psychologists and political scientists which indicate that people tend to be more generous to those more similar to them ethnically, racially, linguistically, etc. (Fong and Luttmer, 2011, give a literature review on the subject). Notice this one again drives a wedge between the EMU of individuals and the EMU as revealed by social choices. The degree of tension between ethnic groups, in turn, is measured based on Roch and Rushton (2008), where there is also evidence that the this variable affects support for redistributive fiscal policies.

The sales activity index (the ratio between a state's per-capita retail sales and the national average) variable is intended to capture the importance of tourism in the different states, given it is an important way to export consumption taxes. The neighbour State's progressivity is measured by their weighted average progressivity.[67] It may influence progressivity due to the possibility for individuals to move to other States in order to flee high tax rates (which is similar to the opportunities created by greater integration in the EU considered by Galli, 2002). No theoretical priors are given to justify the inclusion of the poverty variable. It is included solely based on the fact it is included on other studies investigating the "political economy of subnational tax and spending policies". Finally, they use Generalized Methods of Moments (GMM) model to perform the analysis.

In a forerunner to their 2008 paper Fletcher and Murray (2006) argue that time since income tax adoption is related to tax progressivity, which is measured as the ratio of the tax burden share for those in the bottom 75 percent of the top quintile relative to the burden share of those in the bottom quintile. Chernick (2005), based on Hettich and Winer (1988), hypothesize that variables which affect the political costs of taxation and/or the elasticity of the tax base for different income groups should be considered as potentially affecting tax structure. Based on that he shows the existence of correlation between tax exportability, Republican (or Democrat) Party dominance, neighbour state progressivity, income inequality and tax progressivity. In this case progressivity is measured by the ratio of the tax burden in the top quintile to the tax burden in the lowest one. Both studies are relative to US states.

[67] They use two different weighing matrixes: in one the population of the neighbour states is considered and in the other it is not.

Note that in both Fletcher and Murray (2006) and Chernick (2005), as well as in other works related to the determinants of tax progressivity in the US, the question of neighbouring States' tax schedules affecting tax progressivity is raised. The literature is not conclusive about this point, but Fletcher and Murray (2006) and Fletcher and Murray (2008), who study this issue more closely, did not find evidence supporting this possibility. Chernick (2005) uses instrumental variables[68] to obtain his results, while Fletcher and Murray (2006) use OLS and OLS with State fixed effects.[69]

Morgan (1994) points that newspaper circulation (as a measure of how informed population is) and level of conservatism affect tax progressivity. Jacobs and Waldman (1983) find that small business percent, income inequality and percent of blacks in the population are related to tax progressiveness. They argue the results suggest that states are hostile to the interests of the less well-off specially if they are part of a discriminated-against minority, that small businessman are more likely to condemn redistributive policies, and that in States where there is more income inequality the tax system does less redistributive damage.[70]

Both Morgan (1994) and Waldman (1983) use the progressivity measure given by Phares (1973), which is basically the coefficient b of the regression expressed by

$$ETD_i = a + bX_i, \tag{2.2}$$

where ETD_i is the effective tax rate in the ith income class and X_i is the average income in the ith income class. The analysis units are US states for both papers, and both use multiple regression techniques to obtain their results.

2.4. Data and empirical strategy

In this section I provide details on the variables used in my study into the determinants of aversion to inequality as measured by the EASA and present the empirical strategy followed. The variables identified as being used (and being statistically significant)[71] more than once are shown in table 2.2. From the 12 variables displayed in it, only 7 are used in my investigation.

[68] He argues that tax exportability is endogenous to tax structure. The author refers to Feldstein and Metcalf (1987) for the identification strategy utilized.

[69] Fletcher and Murray (2006) use different techniques (standard spatial autoregressive models and geographically weighed regression models) to test correlation between a state tax progressivity and its neighbours'.

[70] They add that their results do not allow them to infer that states use tax progressivity to decrease income inequality, for the overall structure of the state taxes is in general regressive.

[71] The statistical significance gives evidence on the relevance of the variable to the topic. Just for Egger et al. (n.d.) and FM I do not look at statistical significance, but to all variables chosen. I do that due to the fact their works are more similar to the present analysis (mainly Egger et al., n.d.).

Below I provide more details on why I do not include 5 of them and instead include others which are not in the table (as noted I view the analysis as not merely dealing with the determinants of tax progressivity but primarily societal inequality aversion).

Although we have an index of "political process participation" for all the countries considered,[72] it does not vary significantly between countries, such that it is not included.[73] We only possess data on "poverty" for a few years (and not for all countries considered),[74] so such a variable is not included in the analysis. Moreover, there are other factors (such as GDP per capita, inflation, unemployment) included which can to some extent indicate a country's poverty level. The "Neighbour progressivity" variable is not included since Fletcher and Murray (2006) find weak evidence in support of it.[75] If there is little evidence to support such a relation in the US, where there is high mobility, it can be assumed the variable is of little importance in an international context. In addition, my data is not for contiguous countries. The variable "time since PIT adoption" is not included because all countries had already adopted PIT at the start of the period under consideration. We also do not include the variable "percentage black" because it is more related to the US context.

Given the importance of income inequality in the analysis (used 3 times in the works retrieved), we in the same fashion as in Egger et al. (n.d.) include indirect measures. They are unemployment, average gross wage, the population size, inflation and Government's expenditure with education (a proxy for the population educational level). We also include a direct measure (in this case GINI coefficients). The first 4, the GINI, the dependency ratio and openness are hypothesized to be positively correlated with inequality, while the expenditure with education is hypothesized to be negatively correlated with it. These expectations are based on Egger et al. (n.d.).

Another reason for including governmental expenditure on education is that the level of state funded education in society is closely related to societal fairness in that it can be argued that a similar level of education across individuals provides them with similar initial opportunities. In

[72] It consists of the political rights ratings obtained from Freedom House.

[73] The highest ratings are given to all countries and all years, except for Italy, between 1976 and 1979 (then it is given the second best ratings) and Ireland in 1975 (it is given the second highest rating).

[74] From 2004, to be more specific. We looked at the World Bank database and the Eurostat database. All works including this variable collect data on US states.

[75] They mainly use spatial regression techniques to check Chernick (2005) results, but also analyse this issue more deeply than Foster (2013), which besides using techniques which are similar to Chernick (2005) regarding this topic, does not mention their evidence contradicting the latter.

other words, since expenditure in education can be thought as a means to achieve fairness in the same way as tax progressivity, these two policies can be understood as substitutes.

To summarise, the variables to be included in the analysis are: GDP per capita, openness, centralization of power, government stability, ideology, income inequality, dependency ratio, unemployment, average gross wage, population size, inflation and governmental expenditure with education. Table 2.3 presents the definitions and sources of the variables considered, and table 2.4 presents the summary statistics regarding the same variables.

Table 2.2. Variables used more than once (regarding works included in the literature review)

Variables	Times used	Works
GDP per capita (used)*	2	Galli (2002); SS (2010)*
Openness (used)	2	Galli (2002); Egger et al. (n.d.)
Centralization of power (used)	3	Galli (2002); FM (2008); Egger et al. (n.d.) [76]
Government stability (used)	2	Galli (2002); Egger (n.d.)
Political process participation	2	SS (2010); Egger et al. (n.d.)[77]
Ideology (used)	4	Foster (2013); Chernick (2005); Egger et al. (n.d.); FM (2008)
Poverty	2	Foster (2013); FM (2008)
Neighbour progressivity	2	Foster (2013); Chernick (2005)
Time since PIT adoption	2	FM (2006); FM (2008)
Income inequality[78] (used)	3	Egger et al. (n.d.); FM (2008); Chernick (2005);
Percentage of black	2	FM (2008); JW (1983)*
Dependent ratio (used)	2	Egger et al. (n.d.); FM (2008)

*Source: see text. *See notes. Notes: (1) SS (2010) stands for Scheve and Stasavage (2010); (2) JW (1983) stands for Jacobs and Waldman (1983); (3) (used) indicates the variables was used in the analysis.*

Table 2.3. Acronym, definition and source of the (independent) variables considered

Acronym	Definition	Source
GDPPERCAP	GDP per capita (constant 2005 US$)	World Bank Indicators
DEPRATIO	Dependency ratio=(Population > 65 + pop. < 14)/(total pop.)	World Bank Indicators

[76] Galli (2002) captures it by the coalition variable; FM (2008) captures it by the "same party dominates state government" variable and Egger et al. (n.d.) captures it by the "unitary of federalist" variable.
[77] SS (2010) captures it by the "male universal suffrage".
[78] Which in this work (as in Egger et al., n.d.) is associated to GINI coefficients, unemployment rate, average gross wage, population size, dependency ratio, openness, inflation and education.

GINI_gross	Ginni coefficients before taxes	WIDER[79]
GOVSTAB	Government stability: Percent of veto players[80] who drop from the government in any given year	WBDPI[81]
IDEOL	Ideology: Right=1;[82] Left=3;[83] Center=2[84]	Freedom House
OPENESS	Openness= (Exports + Imports)/(GDP)	World Bank Indicators
LNPOP	Log of population	World Bank Indicators
EDUCEXPEND	Govern expenditure with education (% of GDP)	World Bank Indicators
GOVDISP	Does party of executive control all relevant houses? (yes=1; no=0)	WBDPI
UNEMP	Unemployment (% of total labour force)	World Bank Indicators
LNAPW	Average gross wage: log of APW	EuroPTax
INFL	GDP deflator (annual %)	World Bank Indicators

Source: see text.

Table 2.4. Summary statistics of the independent variables

Variables	Mean	Std. dev	Median	Min	Max	NbObs
GDPPERCAP	28285	11111.22	26796	9010	69095	409
DEPRATIO	0.3530	0.02389205	0.3520	0.3022	0.4237	429
GINI_gross	41.47	6.31671	41.60	19.90	54.70	407
GOVSTAB	0.1592	0.2734169	0	0	1	284
IDEOL	1.898	0.916495	2	1	3	295
OPENESS	66.08	29.74537	59.44	24.95	178.25	419
LNPOP	16.49	1.192187	16.08	14.85	18.23	402
EDUCEXPEND	5.361	1.106591	5.287	2.907	8.215	301
GOVDISP	0.2068	0.4056843	0	0	1	295
UNEMP	7.763	2.84584	7.773	2.493	15.775	153
LNAPW	10.914	2.364954	10.676	5.858	17.501	459
INFL	5.514	4.397078	4.307	-1.695	25.789	403

Source: see text.

First we apply a standard fixed effects model with time trend. It can thus be represented by

$$\eta_{it} = \alpha_i + \beta X_{it} + \lambda T + u_{it}, \tag{2.3}$$

where η_{it} represents the EMU estimates for country i at period t as revealed by the EASA using the all in / no SI series for a single person, X_{it} is the time-variant regressor matrix consisting of

[79] World Income Inequality Database (UNU-WIDER).
[80] In presidential systems, veto players are defined as the president and the largest party in the legislature. In parliamentary systems, the veto players are defined as the PM and the three largest government parties.
[81] World Bank Database of Political Institutions.
[82] "Parties that are defined as conservative, Christian democratic, or right-wing." (WBDPI guide)
[83] "Parties that are defined as communist, socialist, social democratic, or left-wing." (WBDPI guide)
[84] "Parties that are defined as centrist or when party position can best be described as centrist (e.g. party advocates strengthening private enterprise in a social-liberal context). *Not* described as centrist if competing factions "average out" to a centrist position (e.g. a party of "right-wing Muslims and Beijing-oriented Marxists")." (WBDPI guide)

the explanatory variables presented above, λ is the coefficient regarding the time trend T, β is the coefficient vector of interest, and u is the error term.

The fixed effect model is justified given the results of an F test for individual effects, which tests whether a fixed effects model is more appropriate than a OLS pooled model, and a Hausman test, which tests whether a fixed effects or random effects model is more appropriate. In both cases the fixed effects model was found the more appropriate model.[85]

After running this regression we remove a non-significant variable which caused the number of observations in the regression to increase (the variable removed is UNEMP) and run it again. After that it was not necessary to exclude any further non-significant variables to increase the number of observations considered, so that this is the specification I focus on in the section below. We also run the F test for individual effects and Hausman test for this specification, and again the results indicate the fixed effects model is the most appropriate.[86]

Notice the works considered in the literature review interpret the variables discussed above as determinants of the degree of progressivity of income tax schedules, such that there is a causality direction underlying the analyses. Following them I interpret the variables considered as causes of the EMU, once the latter is a measure of the degree of progressivity.

Although to investigate such causality assumption among the variables would be another research in itself, endogeneity tests for all statistically significant variables are presented on Appendix 2.C.

In the next section then, I make hypotheses explaining the correlations (interpreted as causalities) between the significant variables referred to above and the EMU. In the conclusion I discuss the implications of such correlations to the reliability of the EMU as estimated by the EASA.

2.5. Results and discussion

The results of the regression analysis described in the last section are presented in table 2.5. According to the results there are 5 variables that are significantly correlated with the EMU (as measured by EASA): DEPRATIO, GOVSTAB, OPENESS, LNPOP, EDUCEXPEND. The dependent

[85] F test for individual effects: F = 6.4237, p-value = 5.156e-07 (alternative hypothesis: significant effects); Hausman Test: chisq = 39.626, p-value = 0.0001587 (alternative hypothesis: random effects model is inconsistent).
[86] F test for individual effects: F = 10.287, p-value = 2.223e-12 (alternative hypothesis: significant effects); Hausman Test: chisq = 59.473, p-value = 2.816e-08 (alternative hypothesis: random effects model is inconsistent).

variable is interpreted as an index of governmental inequality aversion (as reflected by the progressivity of the income tax schedule) and is not seen as necessarily reflecting people's inequality aversion.

To have a better understanding of why DEPRATIO affects the EMU negatively, we run the same regression substituting it by the percentage of people with more than 65 years (ABOVE65) and the percentage of people with less than 14 years (UPTO14) in the population.[87] Both variables are significant and affect the EMU negatively.[88]

One way to explain the result concerning the ABOVE65 variable is by the Hettich and Winer (1988) probabilistic voting model. Specifically it is known that older people are usually both more likely to vote and more likely to be richer, which increases the political cost to fight inequality, leading Governments to appear less inequality averse.

It is more difficult to interpret the UPTO14 variable because the extent to which people vote based on how proposed policies affect children is unclear. However, we can hypothesize that parents (the share of which is assumed to grow with the percentage of those under 14 years) behave similarly to the elders, which would explain the negative correlation between UPTO14 and the EMU.

An alternative interpretation is that older people and parents (again assuming a larger share of people under 14 in the population entails a larger number of parents) tend to be less inequality averse themselves, and therefore would favour less inequality averse Governments. Notice that in this case governmental policies would reflect people's preferences. Moreover, the explanation given in the previous paragraph is not inconsistent with the one discussed in this paragraph. This is, it is possible that DEPRATIO affects Governments' tax policies through a desire to maintain political support and also a wish to reflect the preferences of the populace.

The GOVSTAB variable affects the EMU positively. To explain this correlation we can again refer to Hettich and Winer's (1988) theory. If for a Government, being inequality averse implies a higher political cost (due to the larger political influence of wealthy people), as suggested by the mentioned theory, the result is compatible with the idea that stable Governments are in better condition to pay such cost without being dismantled.

[87] Notice DEPRATIO=ABOVE65+UPTO14.

[88] The coefficient regarding the variable ABOVE65 was -9.0002e-02 (standard error of 2.8384e-02 and significant at 1%). The coefficient regarding the variable UPTO14 was -6.9308e-02 (standard error of 1.7836e-02 and significant at 0.1%). For the referred regression, the Hausman test and the F test for individual effects also point the fixed effects model as the most appropriate.

On the other hand, it is also possible to conjecture that political stability per se causes people to become more inequality averse, which is reflected in the kind of administration elected. It is reasonable to assume in times of instability people become more selfish. Again, the two different explanations regarding GOVSTAB are not incompatible.

Openness is positively related to the EMU, which is consistent with the Egger et al. (n.d.) finding that openness is positively correlated with income tax progressivity. They explain this correlation basically by resorting to the idea that openness is related to income inequality and thus would indirectly reveal the effect of income inequality on progressivity. This is, it suggests inequality aversion increases with actual inequality.[89] It is also important to notice this result is at odds with Galli's (2002) finding that openness is negatively correlated with progressivity (which supports the hypothesis openness increases the cost of setting progressive taxes by encouraging workers mobility; workers can escape high taxes by changing the country of residence).

Therefore, there is a tension between two possible effects of openness on progressivity. In one hand it can increase income inequality and so inequality aversion, which leads people to vote for more progressive schedules. On the other hand, it can increase the cost of progressivity due to higher workers mobility, which may drive administrations to reduce the level of progressivity. In the present study the results suggest the first effect is stronger. Moreover, notice the interpretation based on Egger et al. (n.d.) seems to imply Governments reflect people's preferences, while the interpretations based on Galli (2002) does not make such an assumption.

Yet another explanation is that by increasing workers mobility openness may cause some countries to become less sensitive to tax competition (which decreases the cost of implementing egalitarian-driven policies) instead of more sensitive to it as hypothesized by Galli (2002). For at the same time workers can flee high taxes by changing countries, they can enter a country for reasons not related to taxes (e.g. political liberty, safety, jobs opportunities, etc.) irrespective of paying higher taxes.

In the same way as for openness, Egger et al. (n.d.) justify the inclusion of the LNPOP variable by referring to its probable positive relationship with income inequality. The nature of this linkage is however completely unexplained. In their case they find progressivity rises with LNPOP, which is consistent with the understanding that inequality raises progressivity. In the present study I instead find a negative correlation between LNPOP and the EMU.

---

[89] Notice the GINI is also positively correlated with the EMU, although it is not statistically significant.

It can be hypothesized that to live in a populous country have an effect on individuals' inequality aversion by, for example, decreasing their 'sense of community' (McMillan and Chavis, 1986).[90] Also, it is possible that population size affects progressivity because it promotes remoteness of the central Government from the people, and more remote Governments care less about inequality. Only the first argument (based on McMillan and Chavis, 1986) assumes Governments reflect people's preferences. The second argument suggests that this does not happen, at least in large countries.

Finally, the variable EDUCEXPEND is negatively correlated with the EMU. If we assume, as Egger et al. (n.d.) do that education affects negatively income inequality, then we can explain the result above by referring to this indirect association. This is, an increase in EDUCEXPEND affects the EMU negatively given its negative effect on income inequality.

Another possibility is that a better educated population is simply less inequality averse.

A third argument is to say that state spending on education and progressive PIT are substitutes. The idea of spending on substitutes or complements to tax progressivity is particularly discussed in FM. If spending on education is high this therefore guarantees a certain level of fairness within society, making the kind of compensation brought by progressivity unnecessary. It is here also possible to draw a parallel between the results concerning expenditure on education and the Scheve and Stasavage (2010) results regarding military conscription. They suggest progressivity increased during WWI to compensate for conscription to the military from poorer households and restore a certain level of fairness to society.

Table 2.5. Determinants of the EMU (estimated by the EASA)

Variables	EMU
TT (time trend)	3.7391e-04 (6.3416e-03)
GDPPERCAP	1.9686e-06 (7.5405e-06)
DEPRATIO	-7.0998 ** (1.7717)
GINI_gross	1.2355e-03 (5.3087e-03)
GOVSTAB	1.0577e-01 * (5.2950e-02)
IDEOL	1.0615e-02 (1.8062e-02)
OPENESS	7.4046e-03 ** (1.6494e-03)
LNPOP	-2.4927 * (9.6494e-01)

[90] McMillan and Chavis (1986) define it as "a feeling that members have of belonging, a feeling that members matter to one another and to the group, and a shared faith that members' needs will be met through their commitment to be together."

EDUCEXPEND	-9.4362e-02 ** (2.6701e-02)
GOVDISP	6.5051e-03 (2.6701e-02)
LNAPW	-2.0328e-02 (1.4447e-02)
INFL	6.7081e-03 (6.4790e-03)
R^2 (within)/(between)/(overall)	0.3363/0.2901/0.1114
F-statistic	10.2607 **
Nb. Obs.	264

*Source: see text. Notes: (1) Standard errors in parenthesis; (2) ** indicates the relevant coefficient is significant at 1%, * at 5% and ° at 10%.*

2.6. Conclusion

In this chapter I investigated the determinants of the EMU as estimated by the EASA. In such context the estimates are supposed to reflect societal intratemporal inequality aversion by the progressivity of the income tax schedule. In the study carried out I took a more realistic point of view regarding the parameter in that I considered the possibility that there are factors other than societal inequality aversion that affect Governments' willingness and capacity to set income taxes schedules. Discovering that estimates of the EMU derived from the EASA are largely explained by macroeconomic circumstances and institutional features rather than the characteristics of the populace would undermine the validity of the estimates of EMU in chapter 1.

To get insight into variables affecting the EMU I reviewed the literature regarding the determinants of tax progressivity and selected variables which were significantly correlated to tax progressivity more than once. Then I ran a standard fixed effects model regression with time trend to detect those within this group associated with the EMU. The results show the dependency ratio, governmental stability, openness, population and governmental expenditure with education are associated with the EMU.

Two possible interpretations were provided to explain each variable's correlation with the EMU. One focuses on aspects that are not related to societal inequality aversion and other which focus on aspects that affect societal inequality aversion and may through some channel affect the parameter. In other words, some interpretations see the EMU as indicating governmental inequality aversion (as expressed by the degree of progressivity of income tax schedules) as something separated from societal inequality aversion, and others see the EMU as indicating governmental inequality aversion as a reflection of societal inequality aversion. Notice the interpretations are mutually non-excluding and only well-established democracies are included in the analysis. Also notice that, although I think it is important to show the influence of the considered variables on EMU can be explained differently, I do not assume both explanations are equally convincing.

The dependency ratio (as well as the percentage of those above 65 and under 14 years in the population) was found to be negatively correlated with the EMU. The first possible interpretation (which takes governmental inequality aversion as distinct from societal inequality aversion) is that since elder people are more likely to vote and more likely to be richer, the political cost of being more inequality averse for the Government (i.e. setting more progressive income tax schedules) is higher. We can make the same hypothesis for parents (i.e. that parents are more likely to vote and to be richer), which can explain why the percentage of those below 14 years (which is assumed to grow with the percentage of parents) is also negatively correlated with the EMU. Another explanation (which regards governmental inequality as an expression of societal inequality) is that elder and parents are less inequality averse, which would increase the probability that a less inequality averse Government is chosen.

Government stability was found to be positively correlated with the EMU. A possible explanation is that it is more costly for less stable Governments to oppose inequality. Yet another hypothesis is that stability itself causes people to be more inequality averse. Openness was found to be positively correlated with the EMU. An explanation is that it affects the parameter indirectly by increasing income inequality (Egger et al., n.d.), which would then impact inequality aversion positively. Another possible explanation is that by increasing workers mobility openness makes some countries less sensible to tax competition, which decreases the cost of adopting egalitarian-driven policies.

Population size was negatively correlated to the EMU. One possible explanation for this result is that living in a populous country may have some effect on one's inequality aversion via, for example, their 'sense of community' (McMillan and Chavis, 1986), which would impact governmental policies. The alternative explanation is that the governments of large countries are too remote to implement societal preferences for intratemporal equality.

Finally, governmental expenditure with education was also found to be negatively correlated to the EMU. One explanation for such result is that there tends to be less inequality among more educated populations, such that assuming more inequality leads to more aversion to inequality and Governments reflect people's preferences it is possible to say that more education spending leads to less egalitarian policies. It is also possible to interpret this result by saying education and tax progressivity are substitutes in the pursuit of societal fairness. This is, education spending can be seen as an alternative way in which the Government achieves distributional goals – an alternative (substitute) to more progressive taxes. This is the same argument used by Scheve and Stasavage (2010) – a key paper underpinning the next chapter – albeit concerning military conscription rather than education.

Although it is possible to explain – without contradicting the EASA tenets – how the measured EMU might change in accordance with demographic factors (i.e. dependency ratio), it is harder to explain – in the same way – why the EMU estimates should be affected by political and economic factors. To the extent that they are affected by these factors (i.e. Government stability, openness, population size[91] and expenditure with education) the estimates of EMU provided by the EASA are compromised.

The results regarding GDP per capita and the APW confirm the hypothesis expressed by the iso-elastic utility function assumed in the EASA that the EMU (as societal inequality aversion) is constant across income level. The Government stability, openness and education spending variables have little (although significant) influence on the EMU, which also in a sense substantiates the EASA estimates – since these factors' effect on the EMU are more easily explained by their influence on governmental decision-taking. On the other hand, the size of the population variable (LNPOP) is notably significant in terms of its influence on the EMU (one standard deviation in the LNPOP variable causes the EMU to vary about seven times its standard deviation).[92]

The influence of population size on EMU weakens my belief in the EASA technique somewhat but not much, given that the little influence of the other statistically significant variables capturing political and economic factors and the statistical insignificance of GDP per capita and wages (APW) attest the validity of the technique. Moreover, the estimates of EMU from the EASA do not vary much and they did seem to match estimates of EMU obtained using other quite different techniques even if these estimates refer to other contexts (as shown in chapter 1).[93]

[91] Here such variable is understood more as a political than a demographic aspect, given it is viewed as more likely to affect how Governments approach inequality than to affect how inequality averse people are.

[92] In what follows I describe how much (in terms of standard deviations) the variables considered in my study cause the EMU to vary: 1sdDEPRATIO – 0.42sd(i.e. 0.168)EMU; 1sdGOVSTAB – 0.0675sd(i.e. 0.027)EMU; 1sdOPENNESS – 0.5sd(i.e. 0.2)EMU; 1sdLNPOP – 7sd(i.e. 2.8)EMU; 1sdEDUCEXP – 0.25sd(i.e. 0.1)EMU.

[93] It also must be noticed that when robust SEs are used just OPENESS and EDUCEXPEND are statistically significant, and their influence on the EMU is not significant in terms of magnitude. It is also the case when robust SEs are used with time dummies instead of time trend. These results reinforce the conclusion that the EASA is a reasonable technique to measure the EMU.

References

Alesina, Alberto, and Eliana La Ferrara. 2005. ‘‘Preferences for Redistribution in the Land of Opportunities.’’ Journal of Public Economics 89, 897-931.

Alt, J. E. & Lassen, D. D. (2006). Fiscal transparency, political parties and debt in OECD countries. European Economic Review, 50, 1403-1439.

Atkinson, A. B., & Stiglitz, J. E. (1980). Lectures on Public Economics. Maidenhead: McGraw-Hill.

Chernick, H. (2005). On the Determinants of Subnational Tax Progressivity in the U.S. National Tax Journal, 58(1), 93-112.

Dixit, A., & Sandno, A. (1977). Some Simplified Formulae for Optimal Income Taxation. The Scandinavian Journal of Economics, 79(4), 417-423.

Egger, P., Radulescu, D., & Rees, R. (n.d.). The Determinants of Personal Income Tax Progressivity Around the Globe.

Feldstein, M., & Metcalf, G. (1987) The Effect of Federal Tax Deductibility on State and Local Taxes and Spending. Journal of Political Economy, 95(4), 710-36.

Fletcher, J. M., & Murray, M. N. (2006). Competition over the tax base in the state sales tax. Public Finance Review, 34(3), 258-281.

Fletcher, J. M., & Murray, M. N. (2008). What Factors Influence the Structure of the State Income Tax? Public Finance Review, 36(4), 475-496.

Fong, C. (2001). Social Preferences, Self-interest, and the Demand for Redistribution. Journal of Public Economics, 82, 225-246.

Fong, C., & Luttmer, E. F. P. (2011). ‘‘Do Fairness and Race Matter in Generosity? Evidence from a Nationally Representative Charity Experiment. Journal of Public Economics, 95, 372-94.

Foster, J. M. (2013). Voter Ideology, Economic Factors, and State and Local Tax Progressivity. Public Finance Review, 41(2), 177-202.

Galli, E. (2002). Tax Progressivity and Economic and Political Determinants: an Empirical Analysis for the OECD Countries. SIEP XIV conference.

Gollier, C. (2013). Pricing the Planet's Future. Oxfordshire: Princeton.

Haan, J., & Sturm, J. (1994). Political and Institutional Determinants of fiscal policy in the European Community. Public Choice, 80, 157-172.

Hettich, W., & Winer S. L. (1988). Economic and Political Foundations of Tax Structure. American Economic Review, 78(4), 701–12.

Hettich, W., & Winer, S., L., (1999). Democratic Choice and Taxation: A Theoretical and Empirical Analysis. Cambridge: Cambridge University Press.

Jacobs, D., & Waldman, D. (1983). Toward a Fiscal Sociology: Determinants of Tax Regressivity in the American States. Social Science Quarterly, 550-565.

McMillan, D.W. & Chavis, D.M. (1986). Sense of community: A definition and theory. Journal of Community Psychology, 14(1), 6-23.

Meltzer, A. H., & Richard, S. F. (1981). A rational theory of the size of the government. Journal of Political Economy, 89, 914-927.

Morgan, D., R. (1994). Tax Equity in the American States: A Multivariate Analysis. Social Science Quarterly, 75, 510-523.

Phares, D. (1973). State-Local Tax Equity: an Empirical Analysis of the Fifty States. Lexington, Mass.: Lexington Books.

Piketty, T., & Saez, E. (2006). The evolution of top incomes: A historical and international perspective. NBER Working Paper 11955, Cambridge, MA.

Roch, C. H., and Rushton, M. (2008). Racial Context and Voting Over Taxes: Evidence from a Referendum in Alabama. Public Finance Review, 36, 614-634.

Roubini, N., & Sachs, J.D. (1989). Political and Economic Determinants of Budget Deficits in the Industrial Democracies. European Economic Review, 33, 903-938.

Scheve, K., & Stasavage, D. (2010). The Conscription of Wealth: Mass Warfare and the Demand for Progressive Taxation. International Organization, 64, 529-561.

Scheve, K., & Stasavage, D. (2016). Taxing the Rich: a History of Fiscal Fairness in the United States and Europe. Princeton: Princeton University Press.

Seidl, C. (2009). Measurement of Tax Progression with Different Income Distributions.

Sheshinski, E. (1972). The Optimal Linear Income-tax. The Review of Economics Studies, 39(3), 297-302.

Slemrod, J. & Bakija, J. (2000). Does Growing Inequality Reduce Tax Progressivity? Should It? NBER working paper, n.7576.

Appendix 2.A – Optimal taxation theory (brief review)

In the usual literature about optimal taxation, the individual utility depends both on (post-tax) income and labour supplied $[U(Y,L)]$. Sheshinski (1972) works within such framework and assumes a linear tax system.[94] This is represented by

$$T(Z) = tZ - G, \tag{2A.1}$$

where Z corresponds to income earned and is given by $Z = wL$ (w being the wage rate), t represents the tax rate and G a guaranteed minimum income. Y is given by

$$Y = Z - T = (1-t)wL + G. \tag{2A.2}$$

The revenue constraint is given by

$$G + R_0 = t\int_l^\infty wLdF, \tag{2A.3}$$

where l is a critical wage rate in the sense that $L(w) > 0$ for $w > l$, R_0 is governmental net revenue, $F(w)$ represents the cumulative ability distribution and the government maximizes a welfare function as follows

$$\int_l^\infty \Psi(U)dF, \tag{2A.4}$$

where $\Psi(U)$ is concave and increasing with U.[95] The resulting Lagrangean can be written as

$$L = \int_l^\infty [\Psi + \lambda(twL - G - R_0)]dF. \tag{2A.5}$$

The maximization first order condition with respect to G is

$$\int_t^\infty [\Psi\frac{\partial U}{\partial G} + \lambda(tw\frac{\partial L}{\partial G} - 1)]dF = 0 \tag{2A.6}$$

and

$$\int_t^\infty [\Psi\frac{\partial U}{\partial t} + \lambda(wL - tw\frac{\partial L}{\partial t})]dF = 0 \tag{2A.7}$$

with respect to t.

---

[94] Sheshinski (1972) treats utility as a function of consumption and labour supply. I follow Atkinson and Stiglitzs (1980) presentation where utility is shown as a function of income and labour.

[95] The concavity degree of Ψ in this context indicates the governmental inequality aversion, or governmental attitude toward redistribution.

By working on these conditions Dixit and Sandmo (1977) put the optimal tax rate problem in a more straightforward formulation, showed as

$$\frac{t}{1-t} = \frac{-cov[b,Z]}{\int Z\epsilon_{LL}dF}, \tag{2A.8}$$

where b is the net social marginal valuation of income in terms of revenue, i.e., a measure of the social benefit of transferring 1 currency unit (CU) (in terms of government revenue) allowing for the marginal tax paid on receiving this extra 1 CU (Atkinson and Stiglitz, 1980)[96] and ϵ_{LL} is the compensated labour supply elasticity.[97]

Then in the case considered the optimal tax depends on the (compensated) elasticity of labour supply and on the covariance between marginal social valuation of income and income.

[96] Algebraically it is given by $b = \Psi\left(\frac{\alpha}{\lambda}\right) + tw(\frac{\partial L}{\partial M})$, where α is the individual marginal utility of income and M is income (Atkinson and Stiglitzs, 1980).

[97] Atkinson and Stiglitz (1980) also present a conclusion regarding the lump-sum amount G given by $b = 1$, meaning that G must be adjusted up to the point where the average b is 1.

Appendix 2.B – The probabilistic voting model (brief review)

Hettich and Winer (1988 and 1999) are concerned with a theory capable of explaining the actual choices of governments towards a given tax structure. For this they propose an objective function of the form

$$EV = \sum_{h=1}^{H} \pi_h = \sum_{h=1}^{H} f_h(I_h), \tag{2B.1}$$

where EV stands for Expected Vote and is relative to a specific political party, H is the total number of individuals and the subscript h refers to one individual; π_h is the probability of an individual h to vote for the party concerned and is equal to a function of the difference (I_h) between the benefits of taxation for individual h [$b_h(G)$], which depend on the services received by a pure public good G, and the opposition to taxation [$c_h(v_h)$], which depends on the loss in full income v_h,[98] such that

$$\pi_h = f_h\big(b_h(G), c_h(v_h)\big) = b_h(G) - c_h(v_h) = I_h. \tag{2B.2}$$

The tax applied to a given individual (T_h) is given by a tax rate (t_h) multiplied by the individuals taxable activity (B_h). Assuming there is no administrative cost for the government to know voters response to taxation, that individuals see no relation between G and their tax burden and that they do not care about how others are affected by taxes and benefits, the governments problem is to choose a tax rate to each voter in order to maximize equation 2B.1 subject to the following budget constraint

$$G = \sum_{h=1}^{H} t_h B_h. \tag{2B.3}$$

The first order conditions include

$$\frac{\partial f_h/\partial I_h . \partial c_h/\partial t_h}{B_h(1+\varepsilon_h)} = \lambda, \tag{2B.4}$$

where ε_h is the elasticity of B_h with respect to t_h and λ is the Lagrange multiplier relative to government budget constraint,[99] and

$$\sum_{h=1}^{H} \frac{\partial b_h}{\partial G} - \lambda = 0. \tag{2B.5}$$

The interpretation of equation 2B.4 is basically that the marginal political cost (numerator) of raising one currency unit of revenue by increasing tax burden (denominator) must be equal

[98] The loss in full income (v_h) is given by $v_h = T_h + d_h$ where d_h is the welfare cost of taxation for the hth voter.
[99] Notice that $B_h(1 + \varepsilon_h) = \partial T_h/\partial t_h$.

across the taxpayers, or that the optimum tax structure is the one which minimizes the total political costs for a certain revenue required. Equation 2B.5 complements 2B.4 basically pointing that the size of government (revenue collected and spent) will be adjusted up to the point where the marginal political benefit of spending one more currency unit on public services equals the common marginal political cost (Hettich and Winer, 1999).[100]

By equation 2B.4 one can conclude that if two groups or persons present the same political cost (tax base elasticity) related to increasing marginally their tax burden but different tax base elasticity (political cost), the Government would adjust the latter by varying tax rates up to the point where both equal to λ. Assume marginal revenue levied is decreasing in tax burden. Then, if a group presents higher marginal political cost (and same base elasticity) than the other, its tax burden would be lowered (raising thus the marginal revenue to be obtained by increasing tax rate) in order for its marginal revenue to reach the point where equation 2B.4 is satisfied. In the case where both group present the same political cost but one of them has higher base elasticity, the latter would be subject to lower tax rates, once in its case the marginal revenue decreases faster.

[100] The model can be extended to encompass different activities, to consider the choice of the number of tax brackets and special provisions, but once we are dealing just with taxes levied on income and the relation between number of brackets and progressivity is ambiguous, we do not go further.

Appendix 2.C – Endogeneity tests for statistically significant variables

I test endogeneity for the statistically significant variables (i.e. the variables shown to be significant on table 2.5) in two ways. In the first I run 5 IV fixed-effects panel regressions using all the variables presented on table 2.5. For each regression I test endogeneity[101] for one significant variable [using the lagged (1) variable of the variable under investigation as an instrument]. The results are shown on table 2C.1. In the second I run one IV fixed-effects panel regression using the same variables shown on table 2.5 and test endogeneity for all 5 significant variables as a group (again using lagged variables as instruments). The results are also shown on table 2C.2.

Table 2C.1. Endogeneity test for each significant variable and for all significant variables as a group

Individual variables	**Chi-squared (1) stat. (H0=exogenous)**
DEPRATIO	0.011
GOVSTAB	3.003°
OPENESS	0.402
LNPOP	6.906**
EDUCEXPEND	0.005
Grouped variables	**Chi-squared (5) stat. (H0=exogenous)**
-	10.412°

*Source: see text. Notes: (1) Robust covariance matrices are used throughout theanalysis ; (2) ** indicates the relevant coefficient is significant at 1%, * at 5% and ° at 10%.*

The overall conclusion to be taken from table 2C.1 results is that although LNPOP is endogenous, the significant variables are exogenous as a group. However, it must be noticed that despite what the results say it is hardly likely to be the case that LNPOP is endogenous.

[101] More specifically, I apply the "difference-in-Sargan" test for endogeneity.

Chapter 3: The Causality Relationship Between Income Inequality and the Income Tax Progressivity

3.1. Introduction

There have been many works investigating the redistributive effects of taxes (e.g. Echevarria, 2012; Stanovnik and Verbic, 2014; Cooper et al., 2015; Joumard et al., 2012; Verbist and Figari, 2013; Kakwani, 1977; Musgrave et al., 1948).[102] In general, these studies imply one of the uses of tax progressivity, and more specifically income tax progressivity, can or should be to tackle pre-tax income inequality, confining themselves to the examination of how well the mechanism works, or would work.

This view on how PIT and income inequality relate to one another is currently the subject of much political rhetoric. Sanders, Corbyn and Melenchon (candidates in the US, UK and French most recent presidential elections, respectively), for example, have both called for highly progressive income taxes as a central plank of their campaigns. In Sanders' campaign website[103] it is stated that "the issue of wealth and income inequality is the great moral issue of our time, it is the great economic issue of our time, and it is the great political issue of our time." The first topic proposed to solve the 'issue' is "demanding that the wealthy and large corporations pay their fair share in taxes." In context, it means more progressive taxes.

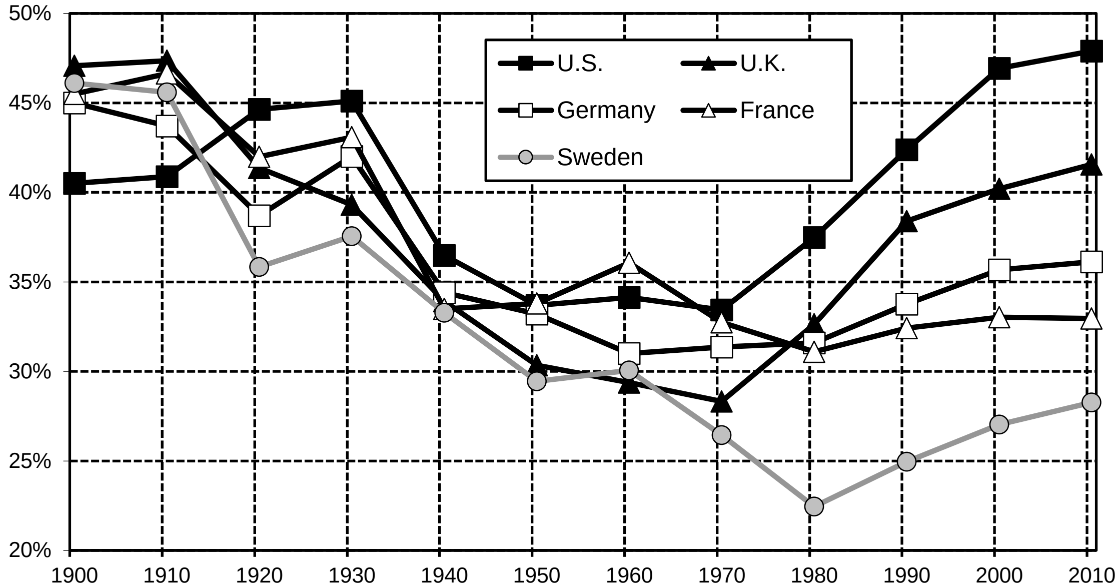
Despite all of the rhetoric (and the fact that all of the above political campaigns were narrowly unsuccessful), few works have examined what factors cause administrations to propose the tax schedules that they do and thus whether income inequality is even an issue considered by tax policy proposers. As far back as 1980, Atkinson and Stiglitz (1980: 293) declared in their classic textbook "redistributive impact … has been analysed in terms of the effect of a specified set of tax and expenditures, but no attempt has been made to examine how particular types of policy came to be adopted." A review of the literature reveals that little has changed since then in terms of studies concerning the determinants of income taxes.

Despite the lack of empirical evidence, some recent works (Scheve and Stasavage, 2016; Zolt, 2009; Piketty, 2014) suggest that if ever the mechanism whereby increases in inequality were met by a more progressive income tax schedule worked it has now stopped working. Not

[102] These works can be loosely classified in two groups: empirical studies measuring the phenomenon (e.g. by examining inequality in post- and pre-tax income distributions) and theoretical works presenting models in which alterations in parameters regarding the level of tax progression affect income distribution.
[103] https://berniesanders.com/issues/income-and-wealth-inequality/.

because income tax progressivity has no redistributive power, but because Governments do not seem to take it into consideration to set income tax schedules (or have for some reason stopped doing so). The only statistical evidence we have on causality (Scheve and Stasavage, 2016) suggests that indeed Governments do not react to growing inequality by using PIT progressivity. At the same time the tendency towards greater income inequality (figure 3.1, taken from Piketty, 2014, illustrates this point) has raised much concern, which brings up questions regarding the best and feasible instruments to deal with it.

Figure 3.1. The top decile income share



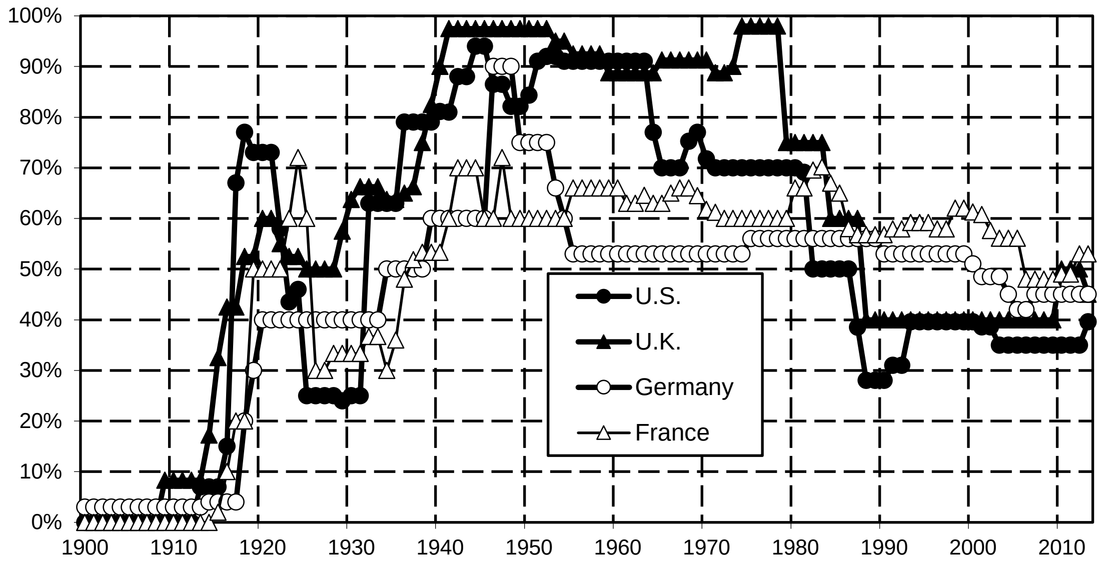
Source: Piketty (2014).

An interesting example of how income tax is often seen as a means to tackle inequality can be found in Piketty's (2014) celebrated book. In a section dedicated to the analysis of income tax he uses historical data on top income tax rates[104] to demonstrate how the UK and the US were the first countries to use income tax to prevent high levels of income inequality, but stopped using it in such way in the 1980s, possibly (he argues) due to the sensation other countries were catching-up with them economically. The argument is basically that when a Government sets income tax rates at 70% or 80% (and the US and UK Governments were the first to do that in a sustained way until about the 80s; see figures 3.2 also taken from Piketty, 2014) its primary intention cannot be to increase revenue, and thus the objective must be to decrease inequality. He goes on to describe progressive taxes as a "relatively liberal method for reducing inequality" (Piketty, 2014: 505), given they do not involve "absolute prohibition or expropriation". He argues the US should go back to use income taxes to address inequality.

[104] He also uses top rates applied to inheritances, but herein we focus on income taxes.

Piketty sees with worry that income taxes are no longer being used to fight inequality in the US, arguing democracy itself is at risk of being captured by a few due to such inaction. As he puts it "the history of the progressive tax over the twentieth century suggests that the risk of a drift toward oligarchy is real and gives little reason for optimism about where the United States is headed" (Piketty, 2014: 514). In other words, he fears large inequality levels because, among other things, it is becoming possible for the very rich to acquire enough power to impose their will over the majority's. Therefore, the issue investigated in this work can be of great importance not just economically, but also politically.

Figure 3.2. Top income tax rates, 1900-2013



Source: Piketty (2014).

Piketty (2014) argues that, taxation is "perhaps the most important of all political issues", and "at the heart of every major political upheaval lies a fiscal revolution", such that it is crucial to understand its determinants, among which are the determinants of income tax schedules. However, there are remarkably few empirical works in this respect and just one (to the best of my knowledge) formally investigating causality between PITP and inequality, which stresses the relevance of the present analysis in the debate regarding not just the determinants of income taxation, but also regarding the determination of taxations systems in general.

Given the recent importance attached to this issue it is no surprise that two prominent US academics have recently taken up the challenge to investigate causality between income inequality and tax progressivity. Scheve and Stasavage (2016) (henceforth SS) deepen the analysis done in Scheve and Stasavage (2010) (which is reviewed in chapter 2) and statistically test causality between inequality and tax progressivity. They are to my knowledge the only ones to examine the matter in such a way, so to some extent we build on their work (although it must be mentioned that my research was already underway when SS published their book, which is testimony to the interest in this area).

SS's main objective is to establish whether Governments do in fact consider pre-tax income inequality in order to define income tax schedules, and although this is the issue that concerns us most, there are also reasons to believe PITP systematically affects pre-tax income inequality too. To see this notice there are many possible reactions to income taxation, such as substituting work hours for leisure, evading taxes, reclassifying earnings or exerting less effort in work (which can be seen as a form to exercise the work-for-leisure substitution), all of which can potentially affect pre-tax income inequality as a result of income taxes.

As income taxes decrease disposable income from work, the cost of leisure also decreases. Then by the 'substitution effect' one can anticipate an increase on demand for leisure. By the 'income effect', on the other hand, one can predict demand for leisure decreases (assuming leisure is a normal good) (Ehrenberg and Smith, 2012). Given there is no reason to believe these two effects balance each other and that the strength of the effects can vary across individuals at different points of the income distribution, it is reasonable to expect changes in income taxes to affect labour supply and therefore pre-tax income inequality. Tax evasion and avoidance (which includes reclassification of earnings) have been much studied mainly due to the fact they are widely practised (Fisher, 2014; House of Lords, 2013; Sikka and Willmott, 2013; degl'Innocenti and Rablen, 2016; Alm and Finlay, 2012). Since there is no reason to believe these practices (by which people pay less tax than they should according to the law) are used equally across people with different income levels, it is also reasonable to expect they also affect pre-income inequality.

In their investigation of factors causing tax progressivity, SS give 3 main reasons as to why democracies would resort to progressivity to stop inequality. The first is that the more inequality increases the more voters would find it is in their interest to tax more heavily the rich given their higher ability to pay. The second stems from the idea inequality of income reflects inequality of opportunity, such that more progressive taxation (among other things like public spending on education, health, etc.) would serve as a way to compensate for such inequalities. The third is related to the possibility an elite would disproportionately influence the political system.

Given these hypotheses as to why there could be a causal relationship between the two variables, they test it by using top marginal rates of income taxation to measure progressivity and top gross income shares to measure income inequality. Their yearly data covers 20 countries (all democracies) and goes from 1900 to 2010. They obtain their progressivity index from consulting the countries' legislations and their income inequality indices are based on income tax returns.[105]

They use Granger non-causality tests with country fixed effects and time effects (period fixed effects or common time trends) in their analysis.[106] For both dependent and independent variables they use running five-year average values throughout their study. Their investigation is restricted to a one-lag analysis and there is no consideration given to the variables' stationarity.[107]

They conclude top marginal rates Granger cause gross income inequality, but surprisingly find little evidence to support the reverse causality. It suggests progressivity has indeed effects on the pre-tax income distribution, but the pre-tax income distribution does not drive progressivity. This is, although PITP can impact gross inequality, that is not why it exists. Such a conclusion is consistent with their hypothesis that compensation for luck and privileges (mainly those granted by Governments), like the ones that influenced mass war mobilization for the first and second World Wars, and not income inequality, is the main factor causing progressivity to vary.

To be more specific, SS make the point that although it is largely agreed income taxes should be designed based on the principle of 'treating citizens as equals', the meaning of equality in such context is far from consensual. Nonetheless, they present the 'ability to pay' doctrine and the 'compensatory theory' as the main criteria through which an income tax schedule can be defined as treating citizens equally or not. The first criterion basically defines a tax schedule is treating citizens as equals when people subjected to it are taxed according to their ability to pay, which results in progressive schedules. However, it is clear that if income earned is proportional to productivity and effort, it is not fair to tax the rich more heavily.

The second criterion implies income tax schedules treat citizens as equals when they compensate for privileges that caused some to accrue more wealth than others. Thus, if it is assumed luck and having good contacts, for instance, play an important role in accumulating riches, progressive taxes would be justified. SS argue throughout their book that this principle is

---

105 The top income shares are from The World Top Incomes Database, a work led by Tony Atkinson, Thomas Piketty and Emmanuel Saez. The top marginal rates are obtained by SS and assistants.

106 They also make another analysis related to testing correlation between PITP and income inequality. Since it does not address causality between top income shares and top rates we do not consider it herein. Moreover, although they find no correlation between the variables, other works (e.g. Roine et al., 2009; Galli, 2002; Egger et al., n.d.) find it, which further justifies an analysis of causality.

107 However, they notice the results of the tests are qualitatively similar when they "use the first value of the five-year period for the dependent and independent variables", when they "use the first value of the dependent variable and its lag and the mean value of the independent variable", and when they "estimate the models using annual data with two-period lags".

the most effective to impel administrations to set progressive schedules. If that is the case, however, the primarily goal of imposing more progressive income taxes cannot be to react to inequality, but to compensate for privileges. Their argument in favour of the 'compensatory theory' receives support from testing the causality between the degree of tax progression of the income schedule and income inequality.

I assert that my work improves SS' causality analysis in many different ways. In first place, we use the liability progression (LP) and residual progression (RP) (both first presented in Musgrave and Thin, 1948) measures of progressivity at the Average Production Wage (APW). Both are traditional measures of progressivity, which are calculated at the APW[108] and give a much more accurate account of progressivity than the top marginal rates used by SS.

Secondly, my data includes local tax and Social Security Contributions (SSC) in addition to just national taxes, as in SS. Although they claimed to have included local taxes (which is discussed in more detail below), they do not consider SSC in their published analysis. Given SSCs can be regarded as very similar, if not essentially the same, as income taxes (Reed and Dixon, 2005; Adam and Loutzenhiser, 2007), it is arguably very important to include them in taxation.[109]

Thirdly, I use the GINI index to measure inequality, which is also more considerate of the whole income distribution than the top income shares used by SS.

Fourthly, given that I use time series to perform the causality tests, I address standard stationarity issues (which SS do not do despite the fact their analysis involves time series). I also carry out the tests for many lags (not just for one as in SS). Although they add time effects (common time trend or year dummies) to their analysis, which may alleviate or eliminate spurious correlation, they do not test the effectiveness of this measure, such that it is not certain whether their variables are stationary. Needless to say, stationarity is central in a Granger non-causality analysis, which is the base of my (and their) statistical testing.

Fifthly, and perhaps most importantly of all besides the Granger non-causality test, we use the Granger non-causality in heterogeneous panels test (Dumitrescu and Hurlin, 2012; henceforth DH test), whose alternative hypothesis identifies causality for a subgroup of the panels (the null

[108] One of the best approximations for most people's income, since labour income is usually the main component of income and other elements of income are harder to observe.
[109] To illustrate the point consider taxation in the UK in 2007 at the APW (single people) according to our dataset (EuroPtax). More than 30% of the total income taxes paid are due to NICs (SSC). In France, it is almost 50%. In Norway the communal tax rate (11.71%) for the APW (2007 and single people) is higher than the national tax rate (10.03%), and in Denmark the difference between local and national income tax rates (in the same conditions) is striking. It is more than 25% for the local rate and about 5% for the national rate.

hypothesis of non-causality refers to all panels) and allows for the testing parameters (the autoregressive parameters and the regression coefficients slopes) to vary across them. More details are given in the sections to come, but the basic advantages just outlined render this test far more general than the usual Granger non-causality test, which makes it more relevant when one wishes to make wider assertions of the type made by SS.

Using the conventional Granger non-causality test we find evidence suggesting causality from income tax progressivity to pre-tax income inequality only, similarly to SS's results. However, when we use DH test we find evidence suggesting causality in both directions. This shows two things: there is a behavioural response from workers because tax progressivity alters gross earnings, and Governments do indeed change tax progressivity as a result of gross earnings inequality. The fact that the conventional causality tests do not show causality from inequality to progressivity but the DH tests do is most plausibly because different countries have different decision-making processes regarding income inequality.

It is also valuable to notice that besides being a measure of inequality aversion, the Elasticity of Marginal Utility (EMU or η) estimated in chapter 1 (by the EASA) can be regarded as a measure of progressivity (as discussed in chapter 2). It implies the level of progression in the income tax reflects the Government's and society's degree of inequality aversion, which in turn suggests causality from pre-tax income inequality to PITP. This is, implies that Governments systematically react to inequality in pre-tax incomes when setting income tax schedules. Thus, the causality pattern found confirms the intuition of the EMU estimates derived in chapter 1.

Notice to test the causal relation referred above it makes more sense (like in SS) to examine gross instead of net income distributions; net income is by definition being affected by progressivity. In other words, if Governments are to react to income inequality with more progressive income tax schedules, then it is necessary that the inequality be in the pre-tax income distribution, once the post-tax income distribution will reflect the progressivity of the income tax implemented.

In the next section I review the literature regarding causality between the two variables of interest. In section 3 I give more details on the measures used to describe income inequality and tax progressivity, underlining their advantages and disadvantages compared with those used in SS. In section 4 I describe my dataset and contrast it with that of SS. In the fifth section I present the results and discuss them. The sixth section concludes.

3.2. Literature review

For the literature review I made two searches on EconLit and also examined the relevant references cited by SS. I look for works that, as in SS and the present study, aim to formally investigate (i.e. by using statistical causality tests) causality between gross income inequality and income tax progressivity.[110] To begin with however I present more details on SS' study and then turn to the papers identified by the reviews carried out with EconLit and the SS references.

The overall objective in SS's work is to investigate the main factor driving income tax progressivity. The first thing they do is to present some common ideas about its determinants (among which is the one that "democracies tax the rich when inequality is high"). Then they focus on the most common and persuasive arguments to obtain political support to implement more progressive income tax schedules: the ability to pay doctrine and the compensatory theory. Both of them are criteria based on which a given income tax schedule can be regarded as 'treating citizens as equals' or not. They are key matters to be considered when deciding the tax system to be implemented.

The ability to pay doctrine basically claims it is fair that those who possess more, being thus able to pay more, be subject to a higher tax burden. The idea that Governments should react to income inequality with more progressive income tax follows directly from it. The compensatory theory, on the other hand, is based upon the opinion that progressive tax is fair when used to compensate privileges obtained by a class, specially privileges provided by the Government.

A possible argument opposing progressivity when based on the ability to pay doctrine is that it is not plausible to oblige someone to pay more based only on the fact he earns more, especially if earnings are understood as proportional to effort or contribution to society. Moreover, one can argue it is not certain that transferring income from rich to poor increases overall happiness as stated by standard utilitarian theory, for it is not possible to know how much utility is lost or earned by the different agents.

Regarding the compensatory argument, it is often claimed the imbalance caused by the Government in favour of or to the detriment of different groups in society should be solved instead of compensated with income taxes. It is also worth mentioning that perhaps the most salient argument in economics against tax progressivity, which states it degenerates the mechanism of incentives in the economy rendering it inefficient, is applicable to any increase in

[110] Notice works presented at chapter 2, which show correlation between income inequality and progressivity (see table 2.2) can also be regarded as contributing to this analysis. Mainly Egger et al. (n.d.), which explicitly consider pre-tax income inequality.

progressivity, no matter whether it is based on the ability to pay doctrine or on the compensatory argument.[111]

The main hypothesis empirically tested in SS' study is that mass war mobilizations as occurred in World War I (WWI) was the leading driver of income tax progressivity. The rich agreed to pay higher taxes whilst the sons of the poor went to war. They believe it is also a way to empirically test the efficiency of compensatory arguments in comparison with ability-to-pay arguments in pushing income tax progressivity since general conscription gave raise to very powerful compensatory arguments.

Their results suggest that indeed compensatory arguments are more effective. They argue rich people were less likely to be conscripted,[112] and were in many cases earning more than usual due to the war, such that the situation was clearly propitious to the use of compensatory arguments, which indeed led Governments to tax the wealthy much more heavily than before. Moreover, general conscription referred only to the need for labour, such that the necessity for capital in the war effort was unattended. It again gave room for compensatory arguments demanding for the 'conscription of capital' in order to compensate for the 'conscription of labour' already taking place.

They also include in their study analyses in which they examine whether democracy or worries about growth (economic efficiency) can be regarded as the main driver of income tax progressivity. According to their results none can. Concerning their analysis examining the association between the institution of democratic states (identified with the establishment of male universal suffrage) and progressivity, they compare the periods at which countries established universal suffrage and adopted progressive taxes to see whether those democratizing first are also the first to set progressive income tax schedules. They find it is not the case, concluding democratization is not the central factor leading to more progressive taxation. It is important to notice that in this analysis they give more weight to progressivity in inheritance tax, given it takes less administrative capacity (which can vary much across countries

[111] Taxes (including income taxes) generate distortions in the economy by changing the way prices are perceived by agents (consumers and producers). It precludes Pareto optimum equilibria and generates social welfare loss. The only kind of taxes causing no distortions and social welfare loss are lump-sum transfers, but their application is politically and financially unfeasible. They are highly unpopular and since there is no proportion between the amount to pay and ability to pay it would be difficult to raise as much revenue as usual tax systems do.

[112] To be more specific, they claim rich were probably more apt to get service exemption. One of the reasons for that, they argue, has to do with the fact that the possibility of getting the referred service exemption was related with being aged, which is associated with higher wealth levels.

and influence the results) to tax inheritance than to tax income. Their conclusion, however, refers to progressiveness in general and therefore includes income taxation.

Concerning their analysis of whether worries about economic growth affect tax policies, they examined if Governments experiencing economic recession or slow down cut taxes in order to reverse the situation. They found no evidence of that, concluding it also does not play a central role in determining the degree of progressivity of tax systems. They also point out their goal is not to assess the soundness of the theories asserting there is a relationship between taxes and economic efficiency, but just to evaluate their influence on actual tax policies.

They also approach the issue of whether the movements in progressivity are related to elites capturing the states. They once again conclude there is not enough evidence to establish whether this factor is the main driver of progressivity. In the analysis concerning the topic they observe tax schedules in countries in which political campaigns are financed privately and countries where private financing is limited. Given they fail to find a relationship between how campaigns are financed and progressivity they conclude it is also not a central factor for the topic under scrutiny.

The part of SS's work which interests us most is the one in which they test causality between pre-tax income inequality and tax progressivity with the intention to check whether the idea that progressivity is caused by pre-tax income inequality is sound. They make use of the standard Granger non-causality test in order to examine the issue. Progressivity is captured by top marginal income tax rates and income inequality is captured by top pre-tax income shares (i.e. the income share of the 1% and 0.01% most rich individuals).

To be more specific about the methodology, they implement the following model:

$$T_{i,t} = \alpha + \beta_1 T_{i,t-1} + \beta_2 I_{i,t-1} + \eta_i + \theta_t + \varepsilon_{i,t} \tag{3.1}$$

$$I_{i,t} = \alpha + \beta_1 I_{i,t-1} + \beta_2 T_{i,t-1} + \eta_i + \theta_t + \varepsilon_{i,t}, \tag{3.2}$$

where T represents top marginal rates and I represents top income shares.[113] Notice besides individual fixed effect parameters (η_i) they include period fixed effect parameters (θ_t) (or common time trends, although it is not shown in the equation)[114] and a constant (α) which despite being included in SS is necessarily omitted in a model with fixed effects.

[113] As noted in the introduction they use 5-years means.

[114] They include two specifications regarding time effects in the analysis: one with time trend and the other with period fixed effects parameters.

The η_i and θ_t (or common time trends) have the objective to capture effects of potential causal variables not included in the model. Although they may help to alleviate such problems there is no analysis of the variables' stationarity, which raises concerns regarding spurious correlation between them. Also notice their analysis, for the most part, includes just 1 lag order.[115]

Their results suggest there is causality from progressivity to pre-tax income inequality, but not from inequality to progressivity. It is consistent with their hypothesis that income inequality is not central to explain movements on the income tax progressivity level, and thus arguments relying on the first are not likely to affect the latter. The findings suggest debates organized around income inequality are politically too weak to affect Governments' decisions regarding income tax schedules.

Below we argue, among other things that considering more appropriate progressivity and inequality measures and using a more general causality test that allows for parameters to vary across panels yields results suggesting there is causality in both directions (i.e. from progressivity to inequality and the reverse). It supports the idea that the "inequality argument" is more effective than portrayed by SS.

SS cite three works which also analyse association between the concerned variables. None of them however, use statistical causality tests to approach the issue. Two of them (Sokoloff and Zolt, 2006 and Zolt, 2009) investigate possible effects of income inequality on tax schedules and the other (Roine et al., 2009) focus on the determinants of income inequality, being top tax rates one of the explicative variables included in the analysis. Below I give more details and comment on the works referred. After I proceed to my independent literature review undertaken using EconLit.[116]

Sokoloff and Zolt (2006) compare tax systems (but in general, not just income tax) in Latin America and North America (United States and Canada) in order to get insight into whether gross income inequality (which is very high in Latin America) influences how tax systems are built.

They find there is not much difference between North and South American countries' tax regimes at the national level, but at the local level they find more regressive tax schedules in Latin America. Such findings suggest it is important to consider local taxes in addition to just

[115] They run other regressions and obtain the same qualitative result (see introduction). The model presented corresponds to their main specification.

[116] They are presented in chronological order and then commented on. I must note Sokoloff and Zolt (2006) and Zolt (2009) are published in law journals, while Roine et al. (2009) is published in an economics journal.

national taxes when investigating the relation between inequality and progressivity: something that SS did not do. This is, tax systems might look very different in terms of regressivity depending on whether local taxes are included in the investigation or not. This substantiates my decision to use a dataset which includes not only income taxes but quasi-income taxes.

Zolt (2009) use long run data on United States local and state taxes to investigate, among other things, the link between the level of taxation and income inequality. They do that mainly by examining correlation between income inequality (measured by GINI coefficients and top income shares) and property taxes in the different periods considered (from the Colonial Period to the Civil War, from Civil War to World War II and Post World War II). The relation under consideration changed considerably along the years, being noticeable at the beginning of the analysis (from about the American Civil War), but non-existent at the end of the 20th century.

One of the possible explanations the author gives for the noted change in the relation between taxing patterns and inequality over the years has to do with empirical evidence (he refers to Rhode and Strumpf, 2003) supporting the view that higher heterogeneity (of individual public good preferences and public good provisions) within communities leads to convergence in their taxing and spending policies (irrespective of the inequality level). This understanding suggests (to the extent we can transpose the view from communities to countries) the degree of heterogeneity in a country's population may affect how inequality and taxes relate in a country. The analysis is consistent with Piketty's (2014) description of how the US and the UK used income taxes to address inequality from the period between World Wars to the 1980s referred in the introduction.

Finally, Roine et al. (2009) investigate determinants of gross income inequality by running dynamic first differences and first differenced GLS models with a panel of 16 countries over the twentieth century. The tax progressivity and income inequality indices used are reminiscent of those considered in SS's (top marginal statutory tax rates and top income shares). They admittedly do not empirically test causality, but assume progressivity is one of the factors driving income inequality.

The difference between their inequality index and that of SS is that they examine shares of different top income percentiles[117]. Regarding progressivity they use two measures. The first is a combination of top rates and rates paid by those with income equal to five times the GDP per

[117] They investigate shares at three different income distribution percentiles: the top 1%, the top 10% to 1% (next nine percentiles) and the bottom 90% (residual share received by the 90% at the bottom of the distribution).

capita; the second consists of the full set of statutory rates (it is not available for all countries). The reason to use the combination referred in the first measure instead of just top rates is that top statutory marginal rates have been, in practice, applied in varying degrees on top incomes both across countries and across years within a country. They find that progressivity reduces top income shares.

Both Sokoloff and Zolt (2006) and Roine et al. (2009) highlight some drawbacks of SS's tax progressivity index (national top marginal statutory tax rates). The first suggests that local taxes are an important factor to examine when studying the relationship addressed herein and the second unveils the problem that top marginal rates are often not binding at high income levels (i.e. high earners have more means to pay less than it is stipulated by the referred rates).

The two issues regarding SS's tax progressivity index raised below are addressed in the present research, given we include local taxes in our analysis and also calculate the two progressivity indexes at the APW, a much lower income level than the ones referring to the top marginal rates but one of relevance to many more people.

Now I turn to the literature review undertaken in EconLit. In the first research (done last time on February of 2017) the following search terms were used: TX[118] "tax progressivity" AND TX "income inequality" AND TX causality. Since no works were retrieved I searched again using just the first two terms: TX "tax progressivity" AND TX "income inequality". Fourteen articles in English were retrieved. None of them use statistical tests to examine causality between inequality and progressivity, but 2[119] investigate the relationship under investigation herein, such that they are kept and reviewed.

The works retrieved but not used in the literature review were excluded due to the following: assuming the relationship under investigation without further analysing it; examining variables which both inequality and progressivity are hypothesized to affect; just comparing methods to measure the distributional effects of tax schedules without addressing the issue considered herein; comparing progressivity across countries using different measures of progressivity and inequality (which are needed to derive the progressivity measures used) again without addressing the issue of how the variables relate; restricting their analysis to the correlation between tax progressivity and net income inequality.

[118] TX indicates the term was searched through the works whole text.
[119] Roine et al. (2009), one of the works cited by SS (the only one published in an economics journal), was retrieved but has already been presented.

Now I give more detail on each of the 2 works found to be of some relevance and then comment on them.

Sarte (1997) assumes a Ramsey model which admits heterogeneity across individuals' constant rate of impatience. It allows for a non-degenerated income distribution when agents are faced with progressive income tax schedules (which is not possible in the standard Ramsey context).[120] They assume a tax schedule based on the Equal Absolute Sacrifice Principle (EASP)[121] and perform numerical simulations to investigate the effects of tax policies on income distribution.

Their results suggest two main conclusions regarding the relationship between inequality and progressivity. The first is that in the model progressivity, as measured by the relative tax burden borne by the different income classes, is positively correlated with income inequality due to behavioural responses of wealthy agents and the increase of interest rates, which affect those at the bottom of the income distribution that borrow. The second refers to the long time it takes to tax policies to affect income distribution; it occurs mainly to the heterogeneous responses to such policies.

It is important to notice in Sarte (1997) the results showing tax schedule changes may take a long time to affect income inequality underlines the importance of including many lags in an investigation into causality between progressivity and inequality.

Thoresen (2004) analyses the reasons that led the Norwegian tax schedules to become less progressive in the 90s (between 1991 and 1999), focusing on the potential influence of tax reforms that took place in the period. In doing so they test the hypothesis that the referred reforms led to behavioural responses that caused pre-tax distributions to become more unequal thus rendering the tax system less progressive.[122]

The referred test is based on the elasticity of gross income with respect to net-of-tax rate (one minus the marginal tax rate), which are estimated by two-stage least square regressions with panel data.[123] Their dataset consists of income tax returns of 1,500 individuals for the period

---

[120] As claimed by Ramsey (1928) and confirmed by Becker (1980), such variability across individuals in the standard Ramsey environment leads to all resources to be possessed by a single household (the most patient) due to the agents' inability to adjust their rates of impatience to the interest rate.
[121] See chapter 1 for more details on the EASP.
[122] Since they use global measures of progressivity, which compare post and pre-tax income inequality in order to assess tax progression, it is not surprising that they associate larger pre-tax inequality to less progressivity. Nonetheless they continue to analyse such relation by testing whether changes in tax policies induced behavioural responses that affected pre-tax inequality, and in doing that they shed light on one of the main effect transmission mechanisms from tax progressivity to income inequality.
[123] The model considered in the study is represented by
$\log\left(\frac{X_2}{X_1}\right) = k + \eta \log\left[\frac{1-\tau_2}{1-\tau_1}\right] + \gamma R + \varepsilon,$

considered (9 years). They find the elasticity estimated ranges from 0 to 0.2, which suggests tax progressivity has some influence on pre-tax inequality due to behavioural responses to fiscal policy, though the magnitude is small.

Both works find a relationship between inequality and progressivity, which supports their hypothesis of causality from the second to the first based on the idea that behavioural responses to tax policies affect pre-tax income distribution. In the present study we also hypothesize there is indeed such a behavioural response, but instead of just examining the correlation we test temporal causality between the variables. Moreover, we also hypothesize (and test) Governments reaction to pre-tax income inequality by increasing tax progressivity.

3.3. Methodology

In this section I contrast mine and SS's methods to measure gross income inequality and progressivity and give more details on the causality tests used in both works, also contrasting them in order to underline how the present study adds to SS's work. I also explain the procedure employed to test stationarity and finally give details on the implementation of the causality tests.

3.3.1. Measurements

SS make two analyses in order to verify whether the use of top rates is adequate to represent progressivity in general. In both they rely on the full schedules of statutory income tax rates for 6 of the 20 countries (France, Germany, New Zealand, Sweden, UK and US) for 7 or 5 years (at 25-years interval).[124]

In the first analysis they examine whether top rates are non-decreasing with income. For all countries and years (with the exception of New Zealand in 1925) that is the case, which is interpreted as indicating higher top rates in general represent higher progressivity. In the second they examine how top rates relate to a more general index of progressivity which consists of subtracting the marginal tax rate relative to low incomes and the marginal tax rate relative to high incomes. They find a high correlation 0.93 between the two variables. Therefore, both analyses, it is claimed, suggest the use of top rates is satisfactory.

---

where 1 and 2 represent subsequent periods, X stands for pre-tax income, τ stands for marginal tax rate, R represents other explanatory variables and η the elasticity of pre-tax income with respect to net-of-tax rate.

[124] It is seven years (1875, 1900, 1925, 1950, 1975, 2000 and 2010) for the UK and Sweden, 6 years (1900 to 2010) for New Zealand and 5 years (1925 to 2010) for the US, Germany and France.

However, the fact that the marginal rates grow with income does not mean an increase in the top rate would represent higher progressivity, given a schedule can become flatter at the same time top marginal rates increase (or stagnate). As an example, take the UK in 2010 and 1925. Although the top rates are virtually the same, the 2010 schedule is much flatter. As the authors themselves notice "there is important variation in the rate of increase of the statutory marginal rates among cases with similar top marginal income taxes", which "can matter for how the income tax system influences the income distribution, and it is not captured by the top statutory rate measure" (SS, 2016, n.8 of chapter 3).

Moreover, although the difference between marginal rates at the top and bottom of the income distribution are more representative of the general progressivity than top marginal rates, it is quite inexact in many instances, given two tax schedules can be very similar in the extremes, but very different in the middle. A good example is France in 1950 and 1925. Both schedules have virtually the same top and bottom marginal tax rate, but the schedule in 1925 was visibly more progressive.

In order to obtain a PITP index more representative of that experienced by taxpayers in general, we calculate the Residual Progression (RP) and Liability Progression (LP) progressivity indexes (Musgrave and Thin, 1948) at the APW. At such a point the indexes reflect how progressive the income tax schedule is for most people. Thus, although both are local progressivity measures (i.e. do not give information on the whole tax schedule) when calculated at a representative point as the APW they become representative themselves.[125]

Top marginal income tax rates indicate progressivity not because it shows the average tax rate increases with income, but because in general an increase in the top marginal rate is not succeeded by a proportional (or higher) increase in the lower marginal tax rates. The LP and RP progression measure, on the other hand, show (for a given income level) whether the average income tax rate grows with income, which is the standard definition of progressivity.

Now we briefly consider how the two measures are defined and how they relate to the stated progressivity definition. According to the LP measure of PITP (Musgrave and Thin, 1948) an income tax schedule can be defined as progressive if

$$(dT(y)/T(y))/(dy/y) > 1, \qquad (3.3)$$

[125] See Appendix 3.E for a brief discussion of global progressivity measures.

where $(dT(y)/T(y))/(dy/y)$ is the elasticity of the tax raised with respect to income (y represents a given income level and $T(y)$ the statutory tax rate at y). Simple manipulation yields

$$LP(y) = \frac{MTR}{ATR} > 1, \tag{3.4}$$

where MTR is the marginal tax rate and ATR is the average tax rate. Notice that the higher the MTR is with respect to the ATR, the more progressive the income tax schedule is considered (at the income level concerned) according to the index.[126]

Another well known progressivity index is the RP measure, also proposed by Musgrave and Thin (1948), which is given by the elasticity of the post-tax income with respect to pre-tax income. In this case progressivity is characterized by

$$RP(y) = \frac{d(y-T(y))/(y-T(y))}{dy/y} < 1, \tag{3.5}$$

Which after simple manipulation yields

$$(1 - MTR(y))/(1 - ATR(y)) < 1. \tag{3.6}$$

To see that both indexes stem from the progressivity definition that a tax function can be regarded as progressive when the ATR increases with income ($dATR(y)/dy > 0$), we present below an extended version of the Seidl (2009) proof shown in chapter 2.

$$\frac{dATR(y)}{dy} = \frac{d}{dy}\frac{T(y)}{y} = \frac{T(y)y-T(y)}{y^2} = \frac{MTR(y)-ATR(y)}{y} > 0 \Leftrightarrow MTR(y) > ATR(y) \Leftrightarrow \frac{MTR(y)}{ATR(y)} =$$
$$LP(y) > 1 \Leftrightarrow 1 - MTR(y) < 1 - ATR(y) \Leftrightarrow \frac{1-MTR(y)}{1-ATR(y)} = RP(y) < 1. \tag{3.7}$$

Notice thus that progressivity in the RP and LP measure imply $dATR/dy > 0$.

Another important addition regarding the progressivity measures is that we include national and local statutory tax rates, besides the Social Security Contributions (SSC),[127] in the calculations relative to the indexes. SS examine the issue of whether it makes much difference to include local taxes and find it does not, although they do not give much detail on the tests performed with tax data including local taxes.[128] I, however, find (as pointed in the introduction) that the SSC has considerable importance in income taxation and include it in the analysis, which adds to

[126] Notice the LP is very similar to the EMU (η) calculated in chapter 1, where we have that $\eta = \ln(1 - MTR)/\ln(1 - ATR)$. It shows clearly that the η can also be interpreted as a measure of income tax progressivity.
[127] Referred as National Insurance Contributions (NIC) in the UK and in chapter 1.
[128] They collect local taxes assuming the subject lives in the largest city of the country, and then including the local taxes relative to such place.

the accuracy of the progressivity measures calculated. Moreover, our dataset considers tax systems subtleties related to marital status and number of dependent children.[129]

Concerning gross income inequality, SS use top income shares, more specifically the shares of the top 1% and 0.01%, while we use the GINI index. They do not analyse how their income inequality index relate to other indices, probably leaning on the fact the famous analysis by Piketty (2014) relies on top income shares (both works use the same database regarding top 1% and top 0.01% income shares).

Piketty (2014) dedicates a short section explaining why he uses top income shares instead of indices like GINI, which claim to indicate in one number all inequality information contained in the whole income distribution. He basically argues that the GINI is an over simplification, given that inequality at different income levels have difference economic and political meanings, which should be analysed separately. Two other aspects he points to are that the GINI does not differentiate between income coming from labour and capital and give an abstract view of inequality, not allowing people to realize its meaning in their own experience.

Although these are legitimate criticisms, the fact remains that indexes like the GINI give (at least theoretically, given it is a global measure) a better description of inequality in the whole income distribution. To be sure they are simplifications, but nonetheless able in my opinion to convey more information about inequality than top income shares.

Regarding the issue of not dividing income between earned by labour or capital, although that is an important issue in Piketty's (2014) analysis it is not relevant in the present study. We want to know whether the relation between PITP and gross income inequality irrespective of whether taxation and inequality refer to income earned by labour or capital.

The GINI coefficient is given by

$$G = \frac{1}{2n^2\bar{x}} \sum_{i=1}^{n} \sum_{j=1}^{n} \left| x_i - x_j \right|, \tag{3.8}$$

where x_i is individual's income, n is the number of people and $\bar{x}$ is the mean value of incomes. The index is better understood with reference to the Lorenz curve, which plots the cumulative percentage of income against the cumulative percentage of population. For any given income distribution, the index is equal to twice the area between the diagonal (which represents perfect equality) and its Lorenz curve. It is scale independent, so a proportional change in all individual's

[129] We examine tax rates assuming single individuals.

income does not change the coefficient value and it satisfies the Pigou-Dalton principle of transfers, according to which any transfer from a poor to a rich person increases inequality.[130]

3.3.2. Causality testing

In the present study the term causality refers to the concept of 'Granger causality', which basically means there is causality between two variables when past values of one significantly affect the current values of the other. There are two basic assumptions upon which the concept of Granger causality is predicated: that past or present cause the future, but the reverse is not possible and that the causing variable has unique information regarding the caused variable (Granger, 1980). If a candidate causing variable contains redundant information, then it is not clear it is causing, giving there is another variable providing the same information regarding the dependent variable.

Bi-variate Granger non-causality tests in panel data context are computed by using regressions of the following kind:

$$y_{i,t} = \alpha_i + \sum_{k=1}^{k} \gamma_k \, y_{i,t-k} + \sum_{k=1}^{k} \beta_k \, x_{i,t-k} + \varepsilon_{i,t}, \tag{3.9}$$

where i represents the panel, t represents time period, k represents lag order, γ and β are parameters to be estimated (auto-regressive and slope parameters, respectivelly), and ε represents the error. In the case of the equation above causality from x to y is being tested.

To implement the test both variables must be stationary. Once the regression is run, the null hypothesis of non-causality is not rejected if and only if the joint lagged variables of the independent variable is not significant. The test has been applied to several contexts. Foresti (2006) and Mahdavi and Sohrabian (1991), for example, use it to examine causality between economic growth and stock prices. Shyamal and Rabindra (2004) investigate causality between energy consumption and economic growth; Kónya (2006) analyse causality between exports and growth and Lin and Ali (2009) investigate causality between economic inequality and military spending. All use the causality testing technique described above.

Dumitrescu and Hurlin (2012) develop a panel causality test based on the panel Granger non-causality test referred above (the Granger non-causality in heterogeneous panels, referred here as DH) which has two important features to the present study: a) it allows the testing parameters to vary across panels, which makes for a more general causality test; b) the alternative

[130] Yet another reason for us to use the GINI is that for this index we have data covering all the countries considered, while we have no data for Belgium concerning top income shares.

hypothesis divides the panels in two groups, for one of which there is causality between the panels.

Notice the first feature of the DH test makes it possible to detect causality even if the regression specification is different for the countries, and the second allows us to test whether non-causality is general or valid just for a subgroup of countries. Unfortunately, if the alternative hypothesis is not rejected the test does not give information on the panels in which causality is present. Although quite new the test has already been applied in various contexts. Liddle and Messinis (2015), for example, use it to investigate causality between urbanization and economic growth, whereas Paramati et al. (2016) apply it to examine the effect of foreign direct investment and stock market growth on clean energy use and Salahuddin et al. (2016) utilize it to analyse the effects of Internet usage and economic growth on CO_2 emissions.[131]

The test is based on the following model

$$y_{i,t} = \alpha_i + \sum_{k=1}^{k} \gamma_i^k \, y_{i,t-k} + \sum_{k=1}^{k} \beta_i^k \, x_{i,t-k} + \varepsilon_{i,t}, \tag{3.10}$$

which is importantly different to the one assumed in the Granger non-causality test in that it allows the testing and autoregressive parameters to vary across the panels. Assuming the two variables[132] are stationary and that the cross-sections are independent the following statistic follows a standard asymptotic normal distribution under the null:

$$Z = \sqrt{\frac{N}{2K}}(W - K), \tag{3.11}$$

where N is the number of panels, K is the lag order and W is the average of the usual Granger non-causality test calculated for each panel. Weighing the statistics for unbalanced panels renders

$$\bar{Z} = \sqrt{N}\left[W - K \times \frac{(T_i - 2k - 1)}{(T_i - 2K - 3)}\right] \times [2K \times \frac{(T_i - 2K - 1)^2 \times (T_i - K - 3)}{(T_i - 2K - 1)^2 \times (T_i - 2K - 5)}], \tag{3.12}$$

where T_i represents the periods specific to each country.

[131] Many of these works (e.g. Paramati et al., 2016 and Salahuddin et al., 2016), as the present study, use Eviews to implement the test. A version of Lopez and Weber (2017) working paper presenting the STATA command xtgcause to run the test reported there is an error on the Eviews code, but after some further investigation it was shown there is no error in the Eviews code (http://forums.eviews.com/viewtopic.php?t=18349).

[132] It is designed for bivariate models.

The standard DH test assumes that the lag length is the same across all the panels. If different lag lengths are used across panels there is an adjustment for Z. However, there is no guidance regarding the optimal lag length in Dumitrescu and Hurlin (2012).

Finally, we can state the null hypothesis for the DH test as indicating no Granger causality from the independent to the dependent variables for all panels considered. The alternative hypothesis, on the other hand, indicates the existence of Granger causality from the independent to the dependent variable for at least one of the panels considered.

The null can thus be algebraically represented as $x_i \neg\rightarrow y_i$, where x represents the independent variable, y the dependent variable, $i = 1, 2, \dots, N$ the panels under investigation and the arrow stands for Granger causality from x to y. On the other hand, the alternative hypothesis can be represented by $x_t \rightarrow y_t$, where t stands for a non-empty subset of i.

3.4. Data description

In this section I present the summary statistics concerning the three variables considered in the research and give details on their sources.

Table 3.1 shows the summary statistics for the GINI, the RP and the LP variables. Both the LP and RP were calculated using income tax data obtained from the EuroPTax dataset. The GINI coefficients were obtained from the UNU-WIDER World Income Inequality Database (WIDER).

Table 3.1. Summary statistics for the GINI, the RP and the LP

Var	Mean	Stdev	Median	Max	Min	NbObs[133]
GINI	41.47	6.32	41.6	54.7	19.9	407
RP	0.84	0.11	0.87	1.23	0.43	453
LP	1.3830	0.28	1.34	2.78	0.52	453

Source: see text.

With the EuroPTax database it is possible to obtain the national and local income tax rate plus the SSC rate for any income level;[134] moreover it gives the APW for each country and year considered. With such information it is simple to calculate the RP and the LP. The EuroPTax provides data from 9 European countries from the 50s or 60s (varies across countries) to 2007. The countries are: UK, Sweden, Norway, Ireland, Italy, Germany, France, Denmark and Belgium.[135]

133 Therefore for a lag order of 1 the causality tests count with 398 observations.

134 However, the higher the income level entered the less trustworthy the income tax returned is.

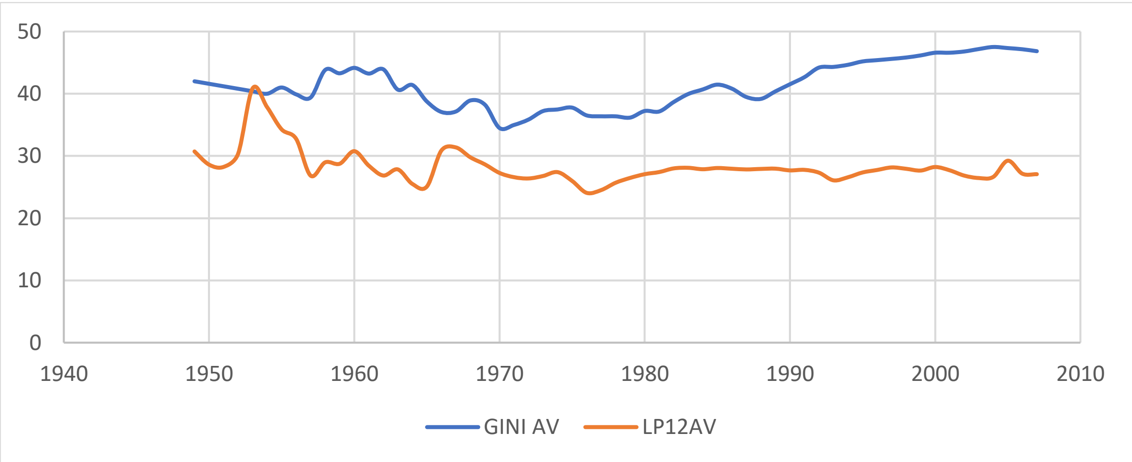
135 See chapter 1 for more details on the EuroPTax.

The WIDER dataset provides GINI coefficients from many studies for all the nine countries covered by the EuroPTax dataset and many others. The period covered varies across countries, being 1974 to 2007 the shortest coverage (for Italy). In several years there was more than one GINI value available so an average of them was used in the study.

Moreover, although we just used the GINI coefficients derived from studies using pre-tax income, there is some variation in the income definition used by these studies; for example, some use gross income and others gross monetary income, the difference being regarding the inclusion or not of in-kind incomes.[136] Once the dataset does not provide observations for all years we use linear interpolation to complete the series.

In order to provide an easy way to visualize the data we present below the GINI and the LP and the RP (in augmented scales in order to improve visualization) as averaged across countries (figures 3.3 and 3.4) and including the series for all countries together (figures 3.5 and 3.6) (charts considering only one country are shown in Appendix 3.B).

Figure 3.3. The GINI coefficients and the LP index

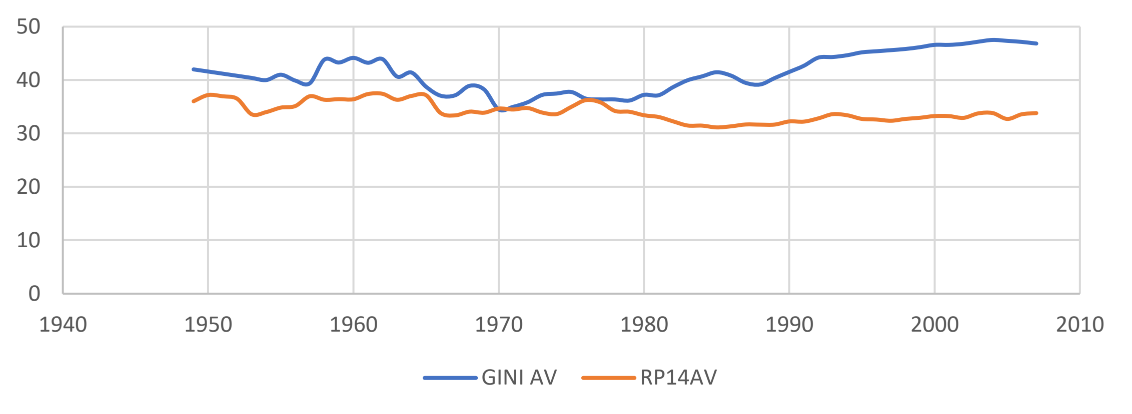


Source: see text. Notes: 1) The GINI coefficients from years 1949 to 1956 are regarding the UK alone; for 1957 it corresponds to the average between the UK and Norway; for 1958 onwards it corresponds to the average across the UK, Sweden and Norway; in 1960 Germany enters the series; in 1962 France enters the series; in 1963 Denmark enters the series; in 1969 Belgium enters the series; in 1973 Ireland enters the series and finally on 1974 Italy enters the series. 2) The LP index, from 1949 to 1956 it concerns to the UK alone; in 1957 Norway enters the series; in 1958 Sweden, Ireland, Germany and France enter the series; in 1959 Denmark enters the series; in 1960 Italy enters the series and

[136] The income definitions included in the study are the following (they go as in the WIDER guide version 3.0b; see the guide for details regarding each definition): gross income; monetary gross income; market income, factor income and primary income; taxable income; 'income…' In cases where the definition is given as 'income…' there is no information about whether the income is pre or post-tax. However, given that it occurs just 27 times from 240 and in many cases (14 times) the coefficient is complemented by other estimates (obtained with pre-tax income) for the same year and country, we decided to include these observations.

finally in 1963 Belgium enters the series. 3) We multiply the LP index by 12 in order to make it easier to visualize how the two variables behave.

Figure 3.4. The GINI coefficients and the RP index

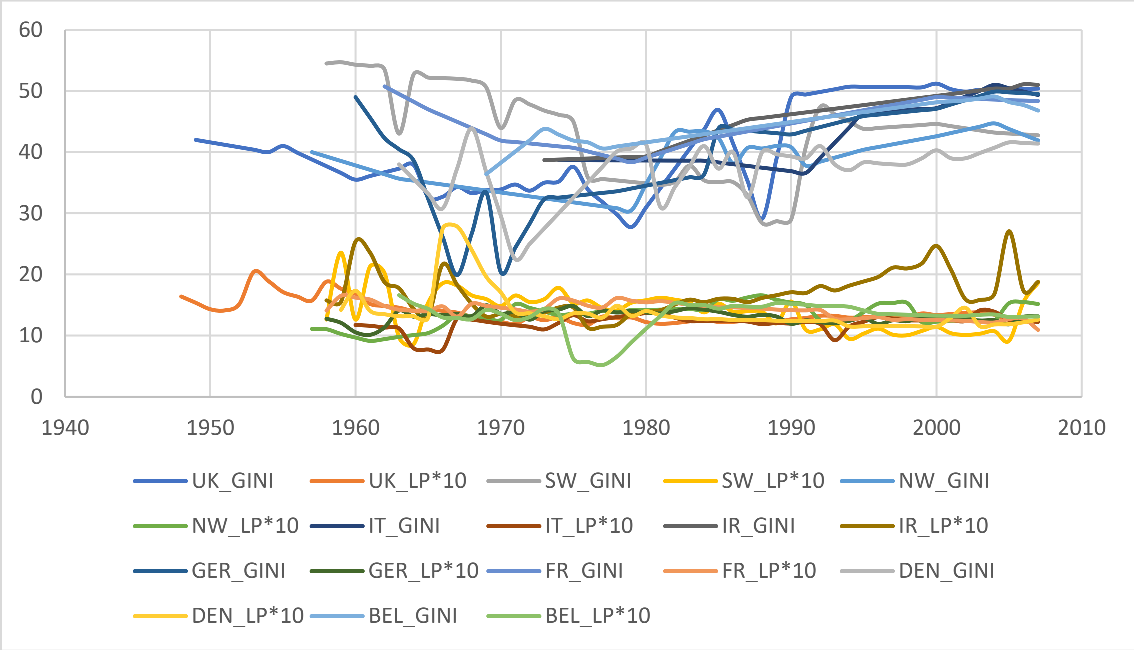


Source: see text. Notes: 1) The GINI coefficients from years 1949 to 1956 are regarding the UK alone; for 1957 it corresponds to the average between the UK and Norway; for 1958 onwards it corresponds to the average across the UK, Sweden and Norway; in 1960 Germany enters the series; in 1962 France enters the series; in 1963 Denmark enters the series; in 1969 Belgium enters the series; in 1973 Ireland enters the series and finally on 1974 Italy enters the series. 2) The RP index, from 1949 to 1956 it concerns to the UK alone; in 1957 Norway enters the series; in 1958 Sweden, Ireland, Germany and France enter the series; in 1959 Denmark enters the series; in 1960 Italy enters the series and finally in 1963 Belgium enters the series. 3) We multiply the RP index by 14 in order to make it easier to visualize how the two variables behave.

In charts (figures) 3.3 and 3.4 we see all the three series tend to fall for a period and then resume growth. In the cases of the RP and the LP the series stabilizes after rising and in the case of the GINI coefficients it continues to grow. Both the LP and GINI variables stop falling around the middle 70s; since then the GINI coefficients keep on an upward trend, whilst the LP grows up until to the early 80s and then stabilizes. On the other hand the LP index retakes growth very timidly around the middle 80s, stabilizing in the middle 90s.

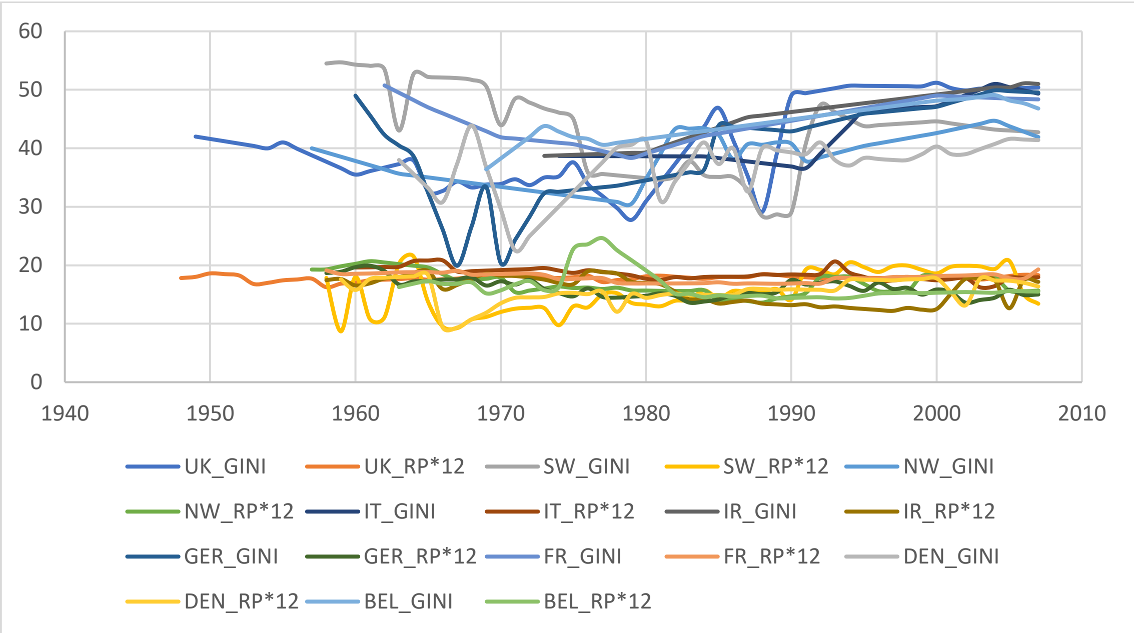
Although no conclusion can be taken from such chart analysis, the progressivity indexes and the GINI seem to be related mainly before the middle to late 80s, which is reminiscent of Zolt's (2009) finding that correlation between taxes and inequality gets weaker in the end of the 20th century.

Figure 3.5. The GINI and the LP for the countries separately



Source: see text. Note: The LP is multiplied by 10 in order to facilitate visualization

Figure 3.6. The GINI and the LP for the countries separately



Source: see text. Note: The RP is multiplied by 12 in order to facilitate visualization.

In charts (figures) 3.5 and 3.6, on the other hand, we see that although the overall tendency of the GINI showed in figures 3.3 and 3.4 is confirmed, the pattern displayed by the LP and RP is different. In both cases the variables vary around a central value (10 for the RP and just over 10 for the LP). However, there are countries that clearly stand out from the general pattern in certain periods (in the case of the LP, see Denmark in the middle 60s, Belgium in the middle 70s and Ireland from the 90s on).

Finally, the analysis of the individual charts (see Appendix 3.B) shows there are significant ways in which the GINI (e.g. Sweden and Germany) and the progressivity indices (e.g. in the UK and Ireland) vary, which stresses the importance of allowing the countries to have different functional forms in order to capture causality, as the DH test does.

3.5. Results and discussion

In this section I present the results for the stationarity tests and causality tests.

3.5.1. Stationarity tests

We investigate whether the LP, RP and GINI variables are stationary by using 5 panel unit root tests: Levin, Lin and Chu (2002); Im, Pesaran and Shin (2003); Fisher-type tests using ADF and PP tests (Maddala and Wu, 1999 and Choi, 2001), and Hadri (2000).

The first two and Hadri's (2000) test make use of common unit root process, i.e. assume the persistence parameter does not vary across cross-sections. The others allow for such variability. For all tests the null hypothesis is of non-stationarity, with the exception of Hadri's (2000), which could not be used due to the variables being highly autocorrelated. Below I give more details on how the two groups of tests differ and on why the results from the Hadri test were not considered.

Consider the following Auto-Regressive (AR) panel process:

$$y_{i,t} = \rho_i y_{i,t-1} + X_{i,t}\delta_i + \varepsilon_{i,t}, \tag{3.13}$$

where $X_{i,t}$ represents exogenous variables, ρ_i the persistence parameters and the errors $\varepsilon_{i,t}$ are independent. If $|\rho_i| < 1$, $y_{i,t}$ is stationary; if $|\rho_i| = 1$, on the other hand, it contains a unit root and cannot be regarded as stationary.

Some tests (those assuming common unit root process) assume the persistence parameters are the same across panels, so $\rho_i = \rho$, while others allow the parameter to vary across the countries.

Regarding the Hadri test simulations have shown the test is subject to size distortion when the variable concerned is highly autocorrelated. To be more specific, the simulations show it overrejects the null hypothesis of stationarity. Given all three variables considered in the study present this characteristic,[137] we conclude the Hadri test is not appropriate.[138]

Finally, we use the Akaike criterion to obtain the optimal lag orders required to implement all the stationarity tests. Table 12 presents the stationarity results regarding the LP, RP and GINI. Given almost all of the tests (the only exception is the PP test, regarding the GINI coefficient) confirm the variables are stationary, we can proceed to implementing both the Granger non-causality and the DH tests with the variables in levels.

Table 3.2. Stationarity tests for the GINI, RP and LP (H0=not stationary)

Variables	Levin, Lin and Chu	Im, Pesaran and Shin	ADF	PP
RP	-3.62565** (0.0001)	-3.40985** (0.0003)	41.8882** (0.0011)	38.9564** (0.0029)
LP	-2.89127** (0.0019)	-2.70695** (0.0034)	38.6354** (0.0032)	49.4620** (0.0001)
GINI	-6.02306** (0.0000)	-2.19309* (0.0142)	35.7096** (0.0077)	17.2298 (0.5074)

*Source: see text. Notes: The numbers in parenthesis are probability values. ** indicates the relevant coefficient is significant at 1%; * at 5%.*

3.5.2. Causality tests

For the analysis done with the standard Granger non-causality test we included up to 10 lag orders. The optimal lag length is 1 for both the analysis considering the LP and RP.[139] Table 3.3 shows the results for causality between the LP and the GINI variables.

The results strongly suggest there is no causality from income inequality to tax progressivity, but on the other hand present weak evidence (see 5, 6 and 7 lag orders) of reverse causality. Such conclusion is essentially the same as SS's. Table 3.4 presents the results regarding causality between the RP and the GINI coefficients. They are very similar to the results displayed in table 3.3 thus suggesting the same conclusion stated previously.

Table 3.3. Results for causality between the LP and the GINI using standard Granger non-causality test

Null hypothesis	Lag order	F-statistics	N. Obs.
GINI does not Granger Cause LP	1	0.98034 (0.3227)	398
LP does not Granger Cause GINI	1	0.20310 (0.6525)	398
GINI does not Granger Cause LP	2	0.28035 (0.7557)	389

[137] See Appendix 3.A to see the autocorrelation tests regarding the variables.
[138] We think it is important to give some explanation about the issue concerning the Hadri test because the standard procedure to test stationarity is to have some tests whose null hypothesis indicates stationarity and others whose null indicates non-stationarity.
[139] To obtain the optimal lag length we consider the Akaike Information Criterion (AIC).

LP does not Granger Cause GINI	2	0.48327 (0.6171)	389
GINI does not Granger Cause LP	3	0.65224 (0.5820)	380
LP does not Granger Cause GINI	3	1.08832 (0.3539)	380
GINI does not Granger Cause LP	4	1.09774 (0.3574)	371
LP does not Granger Cause GINI	4	1.52989 (0.1929)	371
GINI does not Granger Cause LP	5	1.40444 (0.2219)	362
LP does not Granger Cause GINI	5	4.55849** (0.0005)	362
GINI does not Granger Cause LP	6	1.88638° (0.0824)	353
LP does not Granger Cause GINI	6	3.17945** (0.0048)	353
GINI does not Granger Cause LP	7	1.66832 (0.1159)	344
LP does not Granger Cause GINI	7	2.23595* (0.0311)	344
GINI does not Granger Cause LP	8	0.94468 (0.4797)	335
LP does not Granger Cause GINI	8	1.90040 (0.0593)	335
GINI does not Granger Cause LP	9	0.39049 (0.9393)	326
LP does not Granger Cause GINI	9	1.17378 (0.3112)	326
GINI does not Granger Cause LP	10	0.38927 (0.9509)	317
LP does not Granger Cause GINI	10	1.14431 (0.3288)	317

*The numbers in parenthesis are probability values. ** indicates the relevant coefficient is significant at 1%; * at 5%.*

Table 3.4. Results for causality between the RP and the GINI using standard Granger non-causality test

Null hypothesis	**Lag order**	**F-statistics**	**N. Obs.**
GINI does not Granger Cause RP	1	0.36413 (0.5466)	398
RP does not Granger Cause GINI	1	0.06513 (0.7987)	398
GINI does not Granger Cause RP	2	1.75644 (0.1740)	389
RP does not Granger Cause GINI	2	0.19641 (0.8218)	389
GINI does not Granger Cause RP	3	1.20503 (0.3077)	380
RP does not Granger Cause GINI	3	0.33918 (0.7970)	380
GINI does not Granger Cause RP	4	1.64335 (0.1628)	371
RP does not Granger Cause GINI	4	0.55051 (0.6988)	371

GINI does not Granger Cause RP	5	1.45877 (0.2027)	362
RP does not Granger Cause GINI	5	4.75264** (0.0003)	362
GINI does not Granger Cause RP	6	1.45986 (0.1912)	353
RP does not Granger Cause GINI	6	3.27383** (0.0038)	353
GINI does not Granger Cause RP	7	1.05404 (0.3932)	344
RP does not Granger Cause GINI	7	2.00353° (0.0541)	344
GINI does not Granger Cause RP	8	0.60643 (0.7724)	335
RP does not Granger Cause GINI	8	1.68342 (0.1015)	335
GINI does not Granger Cause RP	9	0.16878 (0.9969)	326
RP does not Granger Cause GINI	9	1.20461 (0.2914)	326
GINI does not Granger Cause RP	10	0.13932 (0.9992)	317
RP does not Granger Cause GINI	10	1.29096 (0.2347)	317

*The numbers in parenthesis are probability values. ** indicates the relevant coefficient is significant at 1%; * at 5%.*

Given the standard non-causality test null-hypothesis is regarding all countries, the final conclusion derived from the testing is also regarding all countries, and it is consistent with SS's findings. Although there is no evidence of causality from inequality to progressivity we can reject the null of non-causality in the other direction (although there is not much evidence to support it at the optimal lag length of 1). In other words, the results basically suggest Governments do not react to income inequality with more progressive income tax schedules, and provide some equivocal evidence that tax progressivity affects pre-tax income inequality. Thus, following the basic Granger causality testing procedures our findings are at odds with the idea that one of the main uses of income taxes is to fight income inequality.

Now we do the same analysis using the DH test instead of the standard Granger non-causality. Table 3.5 shows the results for the LP and the GINI. This time we consider just 9 lag orders because it is the maximum possible for the DH test given the number of observations available.

The results show there is a striking difference between using the standard Granger and the DH causality test, namely that bidirectional causality is observed at almost all lag orders when the DH technique is implemented. It strongly suggests (the only exceptions occur with 8 and 9 lag orders, presumably because the reduced number of observations) a group of countries (at least one) within the nine uses tax progressivity as a tool to combat income inequality, i.e. income inequality causes income tax to become more progressive.

Table 3.6 presents the results for the RP and the GINI (also in this case the maximum lag order is 9). They confirm the hypothesis just discussed, although less strongly (lag orders 1 and 2 show causality just from RP to GINI, while lag orders 7, 8 and 9 show causality in the opposite direction).

Table 3.5. Results for causality between the LP and the GINI using the DH causality test

Null hypothesis	Lag order	W-statistics	Zbar-statistics
GINI does not Granger Cause LP	1	3.60539	4.93657** (8.E-07)
LP does not Granger Cause GINI	1	3.67146	5.06437** (4.E-07)
GINI does not Granger Cause LP	2	4.75149	3.52490** (0.0004)
LP does not Granger Cause GINI	2	4.35398	2.99329** (0.0028)
GINI does not Granger Cause LP	3	10.7278	8.01411** (1.E-15)
LP does not Granger Cause GINI	3	8.28067	5.41192** (6.E-08)
GINI does not Granger Cause LP	4	11.2274	6.19194** (6.E-10)
LP does not Granger Cause GINI	4	10.9529	5.94718** (3.E-09)
GINI does not Granger Cause LP	5	18.3880	9.95097** (0.0000)
LP does not Granger Cause GINI	5	14.6830	7.11207** (1.E-12)
GINI does not Granger Cause LP	6	30.7995	16.0914** (0.0000)
LP does not Granger Cause GINI	6	12.0999	3.67988** (0.0002)
GINI does not Granger Cause LP	7	16.5873	5.03138** (5.E-07)
LP does not Granger Cause GINI	7	11.9540	2.38684* (0.0170)
GINI does not Granger Cause LP	8	16.2051	3.36024** (0.0008)
LP does not Granger Cause GINI	8	11.3394	1.05639 (0.2908)
GINI does not Granger Cause LP	9	21.1481	3.49531** (0.0005)
LP does not Granger Cause GINI	9	10.6966	-0.01993 (0.9841)

*The numbers in parenthesis are probability values. ** indicates the relevant coefficient is significant at 1%; * at 5%.*

Table 3.6. Results for causality between the RP and the GINI using the DH causality test

Null hypothesis	Lag order	W-statistics	Zbar-statistics
GINI does not Granger Cause RP	1	1.75012	1.34821 (0.1776)
RP does not Granger Cause GINI	1	10.4149	18.1072** (0.0000)
GINI does not Granger Cause RP	2	3.17962	1.42277 (0.1548)

RP does not Granger Cause GINI	2	3.95963	2.46590* (0.0137)
GINI does not Granger Cause RP	3	8.37511	5.51234** (4.E-08)
RP does not Granger Cause GINI	3	7.13914	4.19808** (3.E-05)
GINI does not Granger Cause RP	4	11.7504	6.65839** (3.E-11)
RP does not Granger Cause GINI	4	9.84256	4.95704** (7.E-07)
GINI does not Granger Cause RP	5	22.6581	13.2229** (0.0000)
RP does not Granger Cause GINI	5	11.7747	4.88366** (1.E-06)
GINI does not Granger Cause RP	6	44.0337	24.8754** (0.0000)
RP does not Granger Cause GINI	6	10.6335	2.70664** (0.0068)
GINI does not Granger Cause RP	7	16.6151	5.04725** (4.E-07)
RP does not Granger Cause GINI	7	9.34525	0.89786 (0.3693)
GINI does not Granger Cause RP	8	14.6153	2.60751** (0.0091)
RP does not Granger Cause GINI	8	9.11899	0.00508 (0.9959)
GINI does not Granger Cause RP	9	41.4124	10.3110** (0.0000)
RP does not Granger Cause GINI	9	10.1178	-0.21459 (0.8301)

*The numbers in parenthesis are probability values. ** indicates the relevant coefficient is significant at 1%; * at 5%.*

Thus, the general conclusion based on these results is that we cannot rule out causality from PITP to pre-tax income inequality (that countries use income tax to tackle income inequality) when the causality-testing technique used allows coefficients to vary across panels. The different test results are probably due to the arguably different ways that countries react to inequality. Such variation is not detected by the standard Granger causality, but can be grasped to some extent by the DH causality test, in which auto-regressive and slope coefficients vary with country.

There are several works investigating factors determining inequality aversion, and therefore pointing out variables potentially causing countries to treat inequality differently. Standard public choice economics, for example, considers current income position as key to the degree of inequality (e.g. Romer, 1975, Meltzer and Richards, 1981), whereas expected future income is pointed out as an important factor by works like Hirschman (1973) and Bénabou and OK (2001). Hirschman (1973), for example, argues future income expectance may cause countries developing fast to be less inequality averse, given in such countries there tends to be more people expecting their income to be increasing in the future.

Mobility experience – i.e. positive or negative past experiences concerning one's career – can also be pointed out as a potential cause of the degree of inequality aversion (e.g. see Alesina and Ferrara, 2001), given its relationship with earnings expectance. More specifically, earning expectations tend to be higher if past experiences regarding earnings are positive. Also, Piketty (1995) points out that mobility experience is likely to be related to beliefs regarding the costs of redistribution. This is, those who had positive experiences regarding their careers tend to believe the income distribution depends more on effort than individuals who had negative experiences. Piketty (1995)'s argument stresses the importance of perceptions of the determinants of income generation (e.g. hard work vs luck for example) to the determination of the degree of income inequality. Also, perception of actual income inequality and legacy of communist ideology (see Suhrcke, 2001) are examples of factors potentially causing countries to treat inequality differently.

The results also suggest that the EMU parameters estimated in chapter one are based on reasonably solid grounds (although they simplify the way in which income taxation is conducted) by substantiating the intuition implicit in the referred estimates that the degree of progressivity reflects inequality aversion. Moreover, it implies it is not accurate to portrait political rhetoric stressing problems related to income inequality as a way to push more progressive income tax schedules as being poorly effective.

Notice I am not saying that compensatory arguments[140] (described by SS as the most powerful explanation for increased progressivity) are less effective than the inequality one. I just point out that the evidence provided here finds the arguments favouring progressivity based on income inequality as more effective than in SS's study. In other words, the evidence suggests Governments in general understand income tax progressivity should be used to combat income inequality and use it with such a purpose. However, as Piketty (2014) and Zolt (2009) point out, the intensity with which it is used for the aforementioned purpose varies over time (and probably over country).

[140] This is, the idea that Governments tax the rich more heavily (increase progressivity) when it is clear they have privileges which are not granted to the rest of the population as a way to compensate the general public.

To check the robustness of SS's findings we run standard Granger non-causality tests and DH causality tests using SS's dataset.[141] In this analysis the index of pre-tax income inequality are the top 1% and 0.01% income shares and the index of tax progressivity is the top marginal income tax rates.[142] The top rates data cover 20 countries from 1900 to 2010; the top 1% income share data covers 18 countries and the top 0.01% income share data covers 14 countries (in both cases the period covered varies across countries). All results regarding the analysis are shown in Appendix 3.C.

Before applying the causality tests, we check whether the variables are stationary at level. The results show we cannot reject this hypothesis, so we run the causality tests with the variables in levels. We must notice that in the same way with the RP, LP and GINI, the top marginal rates and top income shares are highly autocorrelated, which makes the Hadri (2000) stationarity test unreliable.

The standard Granger non-causality tests results show bidirectional causality between top rates and both top 1% and 0.01% income shares at most lag lengths (we consider up to 10 lags). At the optimal lag length we have bidirectional causality between top rates and the top 1% and causality from top 0.01% income share. Such findings are indeed very different from SS's results, also based on Granger non-causality, which suggests their findings regarding the issue are somewhat fragile.

The DH causality tests also show similar results; for almost all lag lengths there is causality between income inequality and progressivity.[143] All these results confirm the findings presented earlier, also suggesting there is bidirectional causality between inequality and progressivity.

We also analyse the SS dataset (top rates and top income shares) and the dataset for this study (LP, RP and GINI) in the same fashion as SS. This is, we include time fixed effects and time trends to control for variables potentially affecting both income inequality and progressivity. The analysis is carried out for annual data a lag length of 2. (The results for this analysis are presented on Appendix 3.D).

To be more specific, the model implemented is the same one reviewed in the literature review regarding SS, but instead of one lag we include 2, and instead of using 5-years mean values we

[141] It is basically an illustration of the importance of the DH methodology and a demonstration of the fact that the same findings are found across different datasets.
[142] The top marginal rates used by SS are available at http://press.princeton.edu/titles/10674.html and the top income shares are obtained from the World Wealth and Income Database.
[143] The only exceptions are at 1 and 2 lags for 0.01% top income shares and at 3 lags for 1% top income shares.

use annual data. Although this is not SS's main specification, it is very similar to the specification they implement.

The results regarding top rates and top 1% shares suggest there is causality from progressivity to income inequality only, as SS defend. However, the results obtained with top rates and top 0.01% shares and with the GINI and the RP suggest bidirectional causality and causality from inequality to progressivity respectively (no causality was found with the GINI and the LP indexes). These results demonstrate that even with the conventional Granger causality testing methodology the presence or absence of causality and also its direction depend greatly on the precise measures of inequality and progressivity employed. This stands in contrast to the robust results obtained using the DH procedure.

I also run endogeneity tests between the GINI, the LP and the RP. I consider IV fixed effects panel regressions with time trend or time dummies, and use lagged variables for each considered variable as instrument. In all cases the endogeneity hypothesis was rejected. See Appendix 3.F for more details.

3.6. Conclusion

In this study we investigate causality between gross income inequality and tax progressivity. Income tax, something at the core of both economics and politics, is often thought of as one of the most powerful tools to counter income inequality, which in turn is often seen as a threat capable of undermining countries' political and economic stability. Simultaneously tax progressivity can, according to standard economics theory, affect pre-tax income inequality via behavioural responses of the individuals affected by changes in the level of progressivity.

Given the universal nature of income taxes there are strikingly few empirical works investigating income tax determinants; and the present investigation contributes to the literature by providing further work in the area. One of the main contributions consists of providing evidence which contradicts the only other substantive piece of research which suggests income inequality is of little relevance in determining the level of progression of income tax schedules.

By shedding light on the causality issue we also contribute to future debates regarding Government action towards income inequality. According to the results obtained the call to use political rhetoric involving income inequality in order to increase progressivity reveals itself effective.

Moreover, the causality relation from inequality to progressivity is supportive of the EASA method used in chapter 1, given that aversion to societal inequality is meant to lead to changes in tax progressivity. Therefore, this chapter provides evidence supporting the basic intuition of the method used to obtain an estimate of EMU in chapter 1.

Notice, however, that I am not saying the validity of the estimates derived in chapter 1 depend on the present analysis. It seems to me possible that there could be no causal influence and still the degree of tax progressivity could reveal something about inequality aversion. What I say is that since the EASA assume inequality aversion and progressivity move together and it makes more sense to think that inequality aversion causes progressivity, the causality relation supported by the evidence shown herein provides in turn some support for the EASA and EMU estimates obtained by it.

Since I found just SS using statistical causality tests to address the issue I built on their work mainly by using more precise indexes of tax progressivity and gross income inequality, by checking issues related to stationarity that were overlooked in their analysis and finally by implementing more appropriate causality tests.

The tax progressivity measures used in this study (the LP and RP calculated at the APW income level) are both more representative given the income level at which they are calculated are more representative of the population's income level and both indices stem directly from the standard tax progression definition. GINI coefficients were used to measure income inequality. Compared to the top income shares used by SS, they provide more information on inequality throughout the whole income distribution. An issue with regards to using the GINI, however, is that some observations are obtained via interpolation.

Part of the standard procedure to implement bivariate Granger non-causality is to test whether the variables considered are stationary. However, although SS use the standard causality test they overlook the stationarity testing of the variables indexing income inequality and tax progressivity. Thus another way we contribute to the understanding of causality between the two variables is by investigating stationarity of the variables involved. Our results suggest the inequality and progressivity measures used in the study are stationary in levels.

Besides the Granger non-causality test we also resort to Dumitrescu and Hurlin's (2012) Granger non-causality test in heterogenous panels test (DH test). This last test allows autoregressive and slope coefficients to vary across panels and is hence able to identify causality within a group of panels, thus providing more detailed information on causality between the variables.

The Granger non-causality tests results suggest a conclusion very similar to the one given by SS, i.e. that although tax progressivity affects income inequality, governments do not tend to use the first as a tool to address the second. Put simply, we find evidence that tax progressivity affects income inequality, but do not find evidence of causality in the reverse direction.

The results regarding the DH test, on the other hand, show that gross income inequality affects tax progressivity. It is consistent with the hypothesis that countries use income tax as a means to address inequality. Some Governments (although perhaps not all) do indeed change tax progressivity as a result of gross earnings inequality, but there is also a behavioural response from workers because tax progressivity alters gross earnings. The fact that the conventional causality tests show no causality but the DH tests do is probably because different countries have different decision-making processes some of which are slow and unresponsive.

I also must also note that contrary to what SS report when I analyse their data using conventional Granger tests I find bidirectional causality. I do not replicate perfectly their analysis, but because I use their data and the same causality testing technique, the same qualitative results were expected. This is a major surprise given that when I use conventional Granger tests and my dataset I find unidirectional causality. Nonetheless, it further substantiates the importance of using the DH methodology from the outset. Applying this methodology the results are unambiguous for both datasets.

In summary, the results from the study suggest at least some Governments take inequality into consideration when proposing income tax schedules and thus that the political rhetoric describing income tax as a useful instrument to tackle the dangers of income inequality cannot be rejected as inefficient. It does not mean, however, that action via income taxes is instantaneously taken to avoid inequality, and that the latter is the only factor determining the form of tax schedules. The determination of such fiscal policy is certainly complex and, although the fact progressivity is not rising as fast as before, results imply inequality play a significant role in determining income tax progressivity.

References

Adam, S., & Loutzenhiser, G. (2007). Integrating Income Tax and National Insurance: An Interim Report. The Institute of for Fiscal Studies, Working Paper 21.

Alesina, A. and E. La Ferrara (2005). Preferences for Redistribution in the Land of Opportunities. Journal of Public Economics, 89, 897-931.

Alm, J., & Finlay, K. (2012). Who Benefits from Tax Evasion? Tulane Economics Working Paper Series.

Atkinson, A. B. (1970). On the Measurement of Inequality. Journal of Economic Theory, 2(3), 244-63.

Atkinson, A. B., & Stiglitz, J. E. (1980). Lectures on Public Economics. Maidenhead: McGraw-Hill.

Becker, G. (1980). On the long mn steady state in a simple dynamic model of equilibrium with heterogenous agents. Quarterly Journal of Economics 95, 375-382.

Bénabou, R. and E. Ok (2001). Social Mobility and the Demand for Redistribution: the POUM Hypothesis. Quarterly Journal of Economics, 447-487.

Choi, I. (2001). "Unit Root Tests for Panel Data," Journal of International Money and Finance, 20, 249-272.

Cooper, D. H., Lutz, B. F., & Palumbo, M. G. (2015). The Role of Taxes in Mitigating Income Inequality Across the U.S. States.

degl'Innocenti, D. G., & Rablen, M. D. (2016). Income Tax Avoidance and Evasion: A Narrow Bracketing Approach. Public Finance Review, 1-23.

Dumitrescu, E., & Hurlin, C. (2012). Testing for Granger Non-causality in Heterogeneous Panels. Economic Modelling, 29, 1450-1460.

Echevarría, C. A. (2012). Income tax progressivity, physical capital, aggregate uncertainty and long-run growth in an OLG economy. Journal of Macroeconomics, 34, 955-974.

Egger, P., Radulescu, D., & Rees, R. (n.d.). The Determinants of Personal Income Tax Progressivity Around the Globe.

Ehrenberg, R. G., & Smith, R. S. (2012). Modern Labor Economics: Theory and Public Policy. Boston: Pearson.

Fisher, J. M. (2014). Fairer Shores: Tax Havens, Tax Avoidance, And Corporate Social Responsibility. Boston University Law Review, 94, 337-365.

Galli, E. (2002). Tax Progressivity and Economic and Political Determinants: an Empirical Analysis for the OECD Countries. SIEP XIV conference.

Granger, C. W. J. (1980). Testing for Causality: a Personal Viewpoint. Journal of Economic Dynamics and Control, 2, 329-352.

Hadri, K. (2000). Testing for Stationarity in Heterogeneous Panel Data. Econometric Journal, 3, 148-161.

Hirschman, A.O. (1973). The Changing Tolerance for Income Inequality in the Course of Economic Development, with a Mathematical Appendix by Michael Rothschild. Quarterly Journal of Economics, 87, 544-566.

House of Lords: Select Committee on Economic Affairs. (2013). Tackling Corporate Tax Avoidance in a Global Economy: is a New Approach Needed? Report.

Im, K. S., Pesaran M. H., & Shin Y. (2003). Testing for Unit Roots in Heterogeneous Panels. Journal of Econometrics, 115, 53-74.

Joumard, I., Pisu, M., & Bloch, D. (2012). Tackling Income Inequality The Role of Taxes and Transfers. OECD Journal: Economic Studies, published online first.

Kakwani, N. (1977). Measurement of tax progressivity: an international comparison. Economic Journal, 87, 71-80.

Kesselman, J. R., & Cheung, R. (2004). Tax Incidence, Progressivity, and Inequality in Canada. Canadian Tax Journal, 52(3), 709-789.

Kiefer, D. W. (1984). Distributional Tax Progressivity Indexes. National Tax Journal, 37(4), 497-513.

Levin, A., Lin C. F., & Chu C. (2002). Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties. Journal of Econometrics, 108, 1-24.

Liddle, B., & Messinis, G. (2015). Which comes first – urbanization or economic growth? Evidence from heterogeneous panel causality tests. Applied Economics Letters, 22(5), 349-355.

Maddala, G. S., & Wu S. (1999). A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test. Oxford Bulletin of Economics and Statistics, 61, 631-652.

Meltzer, A. and S. Richards (1981). A Rational Theory of the Size of Government. Journal of Political Economy, 89, 914-27.

Musgrave, R. A., & Thin, T. (1948). Income Tax Progression, 1929-48. Journal of Political Economy, 56, 498-514.

Piketty, T. (2014). Capital in the Twenty-First Century. Cambridge, Mass.: Harvard University Press.

Piketty, T., Saez, E., & Stantcheva, S. (2011). Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities. NBER Working Paper No. 17616.

Piketty, T. (1995). Social Mobility and Redistributive Politics. Quarterly Journal of Economics 110 (3), 551-584.

Ramsey, F. P. (1929). A Mathematical Theory of Saving. The Economic Journal, 38(152), 543-559.

Reed, H., & Dixon, M. (2005). National Insurance: Does it Have a Future? Public Policy Research, 102-110.

Reynolds M. O., & Smolensky, E. (1977). Public Expenditures, Taxes, and the Distribution of Income: The United States, 1950, 1961, 1970. New York: Academic Press.

Roine, J., Vlachos, J., & Waldenström, D. (2009). The long-run determinants of inequality: What can we learn from top income data? Journal of Public Economics, 93, 974-988.

Romer, T. (1975). Individual Welfare, Majority Voting and the Properties of a Linear Income Tax. Journal of Public Economics, 7, 163-188.

Sarte, P. G. (1997). Progressive Taxation and Income Inequality in Dynamic Competitive Equilibrium. Journal of Public Economics, 66, 145-171.

Scheve, K., & Stasavage, D. (2010). The Conscription of Wealth: Mass Warfare and the Demand for Progressive Taxation. International Organization, 64, 529-561.

Scheve, K., & Stasavage, D. (2016). Taxing the Rich: a History of Fiscal Fairness in the United States and Europe. Princeton: Princeton University Press.

Sikka, P., & Willmott, H. (2013). The tax avoidance industry: accountancy firms on the make. Critical perspectives on international business, 9(4), 415-443.

Sokoloff, K. L., & Zolt, E. M. (2006). Inequality and Taxation: Evidence from the Americas on How Inequality May Influence Tax Institutions. Tax Law Review, 59, 201-276.

Stanovnik, T., & Verbic, M. (2014). Personal income tax reforms and tax progressivity in Slovenia, 1991-2012. Financial Theory and Practice, 38 (4) 441-463.

Suhrcke, Marc (2001). Preferences for Inequality: East vs. West. Innocenti Working Paper No. 89. Florence: UNICEF Innocenti Research Centre.

Thoresen, T. O. (2004). Reduced Tax Progressivity in Norway in the Nineties: The Effect from Tax Changes. International Tax and Public Finance, 11, 487-506.

Verbist, G., Figari, F. (2013). The redistributive effect and progressivity of taxes revisited: An International Comparison across the European Union. GINI Discussion Paper 88.

Zolt, E. M. (2009). Inequality, Collective Action, and Taxing and Spending Patterns of State and Local Governments. Tax Law Review, 62(4), 445-504.

Appendix 3.A – Autocorrelation tests

Autocorrelation tests for the LP, RP and GINI:

LP:

Date: 11/04/16 Time: 11:04
Sample: 1948 2007
Included observations: 453

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
.|***** |	.|***** |	1	0.730	0.730	242.83	0.000
.|**** |	.|. |	2	0.559	0.057	385.71	0.000
.|*** |	.|. |	3	0.433	0.015	471.73	0.000
.|*** |	.|. |	4	0.364	0.062	532.47	0.000
.|** |	.|* |	5	0.351	0.113	589.18	0.000
.|*** |	.|* |	6	0.353	0.083	646.70	0.000
.|** |	.|. |	7	0.308	-0.041	690.51	0.000
.|** |	.|. |	8	0.260	-0.010	721.77	0.000
.|* |	.|. |	9	0.208	-0.016	741.92	0.000
.|* |	.|. |	10	0.163	-0.021	754.29	0.000
.|* |	.|. |	11	0.119	-0.044	760.89	0.000
.|* |	.|. |	12	0.097	-0.001	765.29	0.000
.|. |	.|. |	13	0.068	-0.029	767.43	0.000
.|. |	.|. |	14	0.048	-0.006	768.52	0.000
.|. |	.|. |	15	0.022	-0.027	768.75	0.000
.|. |	.|. |	16	-0.000	-0.013	768.75	0.000
.|. |	.|. |	17	-0.001	0.030	768.75	0.000
.|. |	.|. |	18	-0.010	-0.014	768.79	0.000
.|. |	.|. |	19	0.001	0.036	768.79	0.000
.|. |	.|. |	20	-0.003	-0.009	768.80	0.000
.|. |	.|. |	21	-0.008	0.001	768.83	0.000
.|. |	.|. |	22	-0.013	0.003	768.91	0.000
.|. |	.|. |	23	-0.019	-0.010	769.09	0.000
.|. |	.|. |	24	-0.012	0.019	769.16	0.000
.|. |	.|. |	25	-0.017	-0.023	769.29	0.000
.|. |	.|. |	26	-0.015	0.003	769.39	0.000
.|. |	.|. |	27	-0.030	-0.039	769.83	0.000
.|. |	.|. |	28	-0.022	0.031	770.06	0.000

RP:

Date: 11/04/16 Time: 11:06
Sample: 1948 2007
Included observations: 453

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
.|******|	.|******|	1	0.788	0.788	283.15	0.000
.|***** |	.|* |	2	0.658	0.098	480.97	0.000
.|**** |	.|* |	3	0.574	0.078	631.91	0.000
.|**** |	.|. |	4	0.501	0.023	747.34	0.000
.|*** |	.|. |	5	0.455	0.055	842.45	0.000
.|*** |	.|. |	6	0.430	0.069	927.81	0.000
.|*** |	.|. |	7	0.395	0.005	1000.0	0.000
.|*** |	.|. |	8	0.356	-0.008	1058.6	0.000
.|** |	.|. |	9	0.308	-0.035	1102.6	0.000

.|** |	.|. |	10	0.259	-0.031	1133.7	0.000
.|** |	.|. |	11	0.224	0.002	1157.1	0.000
.|* |	.|. |	12	0.195	-0.005	1175.0	0.000
.|* |	.|. |	13	0.160	-0.032	1187.0	0.000
.|* |	.|. |	14	0.132	-0.013	1195.1	0.000
.|* |	.|. |	15	0.098	-0.034	1199.6	0.000
.|. |	.|. |	16	0.060	-0.035	1201.4	0.000
.|. |	.|. |	17	0.044	0.020	1202.3	0.000
.|. |	.|. |	18	0.023	-0.023	1202.5	0.000
.|. |	.|. |	19	0.009	0.006	1202.6	0.000
.|. |	.|. |	20	-0.001	-0.002	1202.6	0.000
.|. |	.|. |	21	-0.021	-0.027	1202.8	0.000
.|. |	.|. |	22	-0.033	0.002	1203.3	0.000
.|. |	.|. |	23	-0.047	-0.016	1204.4	0.000
.|. |	.|. |	24	-0.050	0.017	1205.6	0.000
*|. |	.|. |	25	-0.074	-0.060	1208.2	0.000
*|. |	.|. |	26	-0.073	0.026	1210.8	0.000
*|. |	.|. |	27	-0.079	-0.014	1213.8	0.000
*|. |	.|. |	28	-0.067	0.043	1216.0	0.000

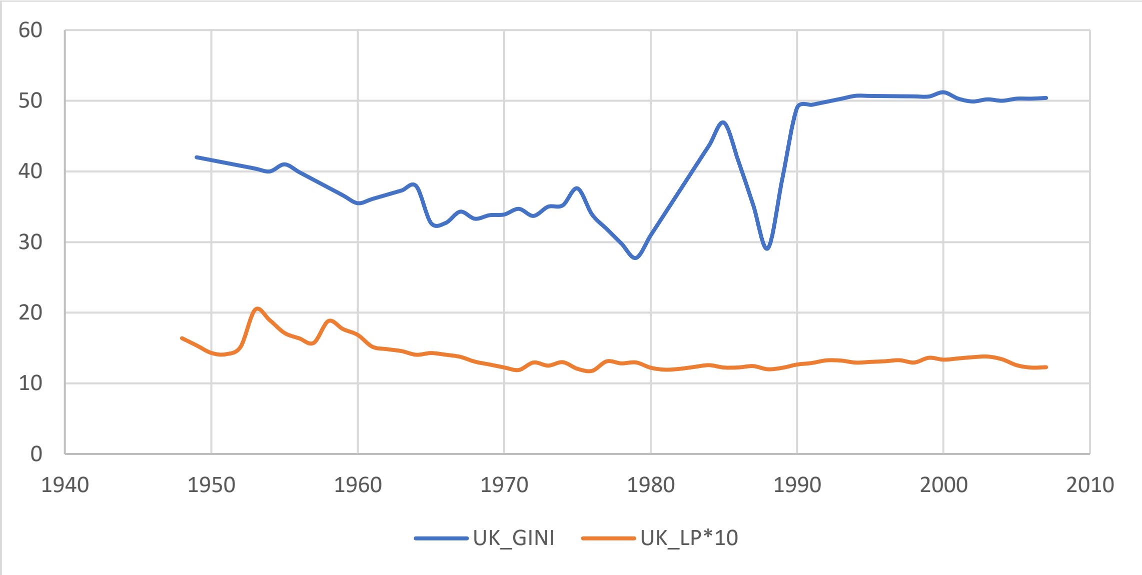
GINI:

Date: 11/04/16 Time: 11:08
Sample: 1948 2007
Included observations: 407

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
.|*******	.|*******	1	0.906	0.906	336.33	0.000
.|******|	*|. |	2	0.800	-0.115	599.15	0.000
.|***** |	.|. |	3	0.708	0.028	805.88	0.000
.|***** |	.|* |	4	0.642	0.074	975.87	0.000
.|**** |	.|. |	5	0.583	-0.005	1116.7	0.000
.|**** |	.|. |	6	0.539	0.056	1237.4	0.000
.|**** |	*|. |	7	0.486	-0.078	1335.6	0.000
.|*** |	.|. |	8	0.437	0.020	1415.4	0.000
.|*** |	.|. |	9	0.383	-0.062	1476.7	0.000
.|** |	*|. |	10	0.318	-0.090	1519.1	0.000
.|** |	.|. |	11	0.254	-0.029	1546.3	0.000
.|* |	*|. |	12	0.189	-0.080	1561.3	0.000
.|* |	.|. |	13	0.128	-0.028	1568.1	0.000
.|. |	.|. |	14	0.072	-0.039	1570.4	0.000
.|. |	.|. |	15	0.022	-0.030	1570.6	0.000
.|. |	.|. |	16	-0.015	0.031	1570.7	0.000
.|. |	*|. |	17	-0.056	-0.077	1572.0	0.000
*|. |	.|. |	18	-0.093	0.001	1575.7	0.000
*|. |	.|. |	19	-0.116	0.054	1581.5	0.000
*|. |	.|. |	20	-0.124	0.040	1588.1	0.000
*|. |	.|. |	21	-0.133	-0.013	1595.7	0.000
*|. |	.|. |	22	-0.146	-0.033	1604.9	0.000
*|. |	.|. |	23	-0.161	-0.002	1616.1	0.000
*|. |	.|. |	24	-0.168	0.016	1628.3	0.000

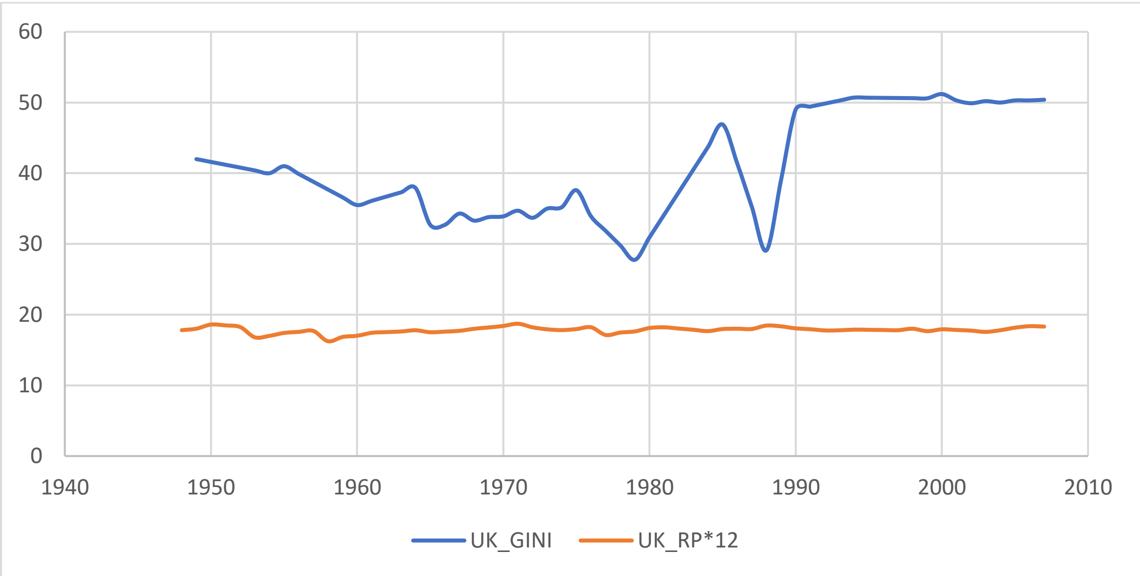
Appendix 3.B – GINI, LP and RP charts (each country separately)

Figure 3B.1. The GINI and the LP index for the UK



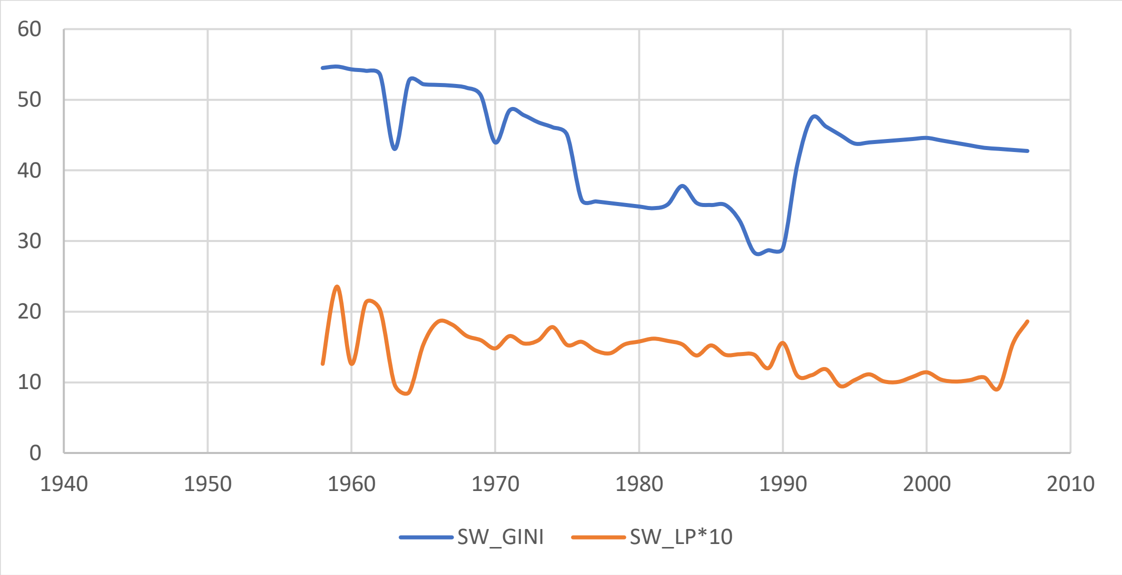
Source: see text. Note: The LP is multiplied by 10 in order to facilitate visualization

Figure 3B.2. The GINI and the RP index for the UK



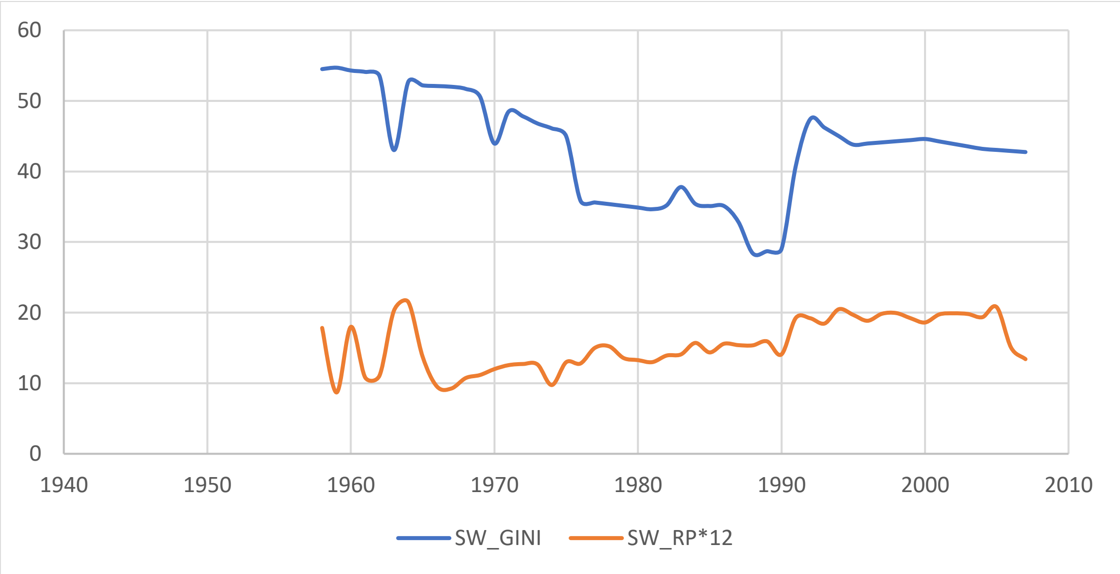
Source: see text. Note: The RP is multiplied by 12 in order to facilitate visualization

Figure 3B.3. The GINI and the LP index for Sweden



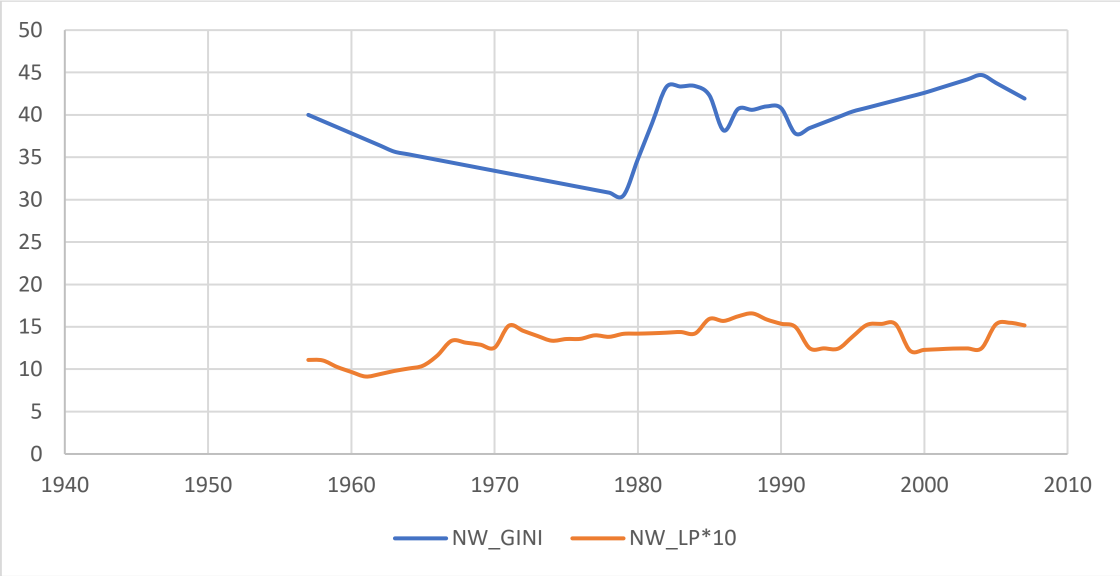
Source: see text. Note: The LP is multiplied by 10 in order to facilitate visualization

Figure 3B.4. The GINI and the RP index for Sweden



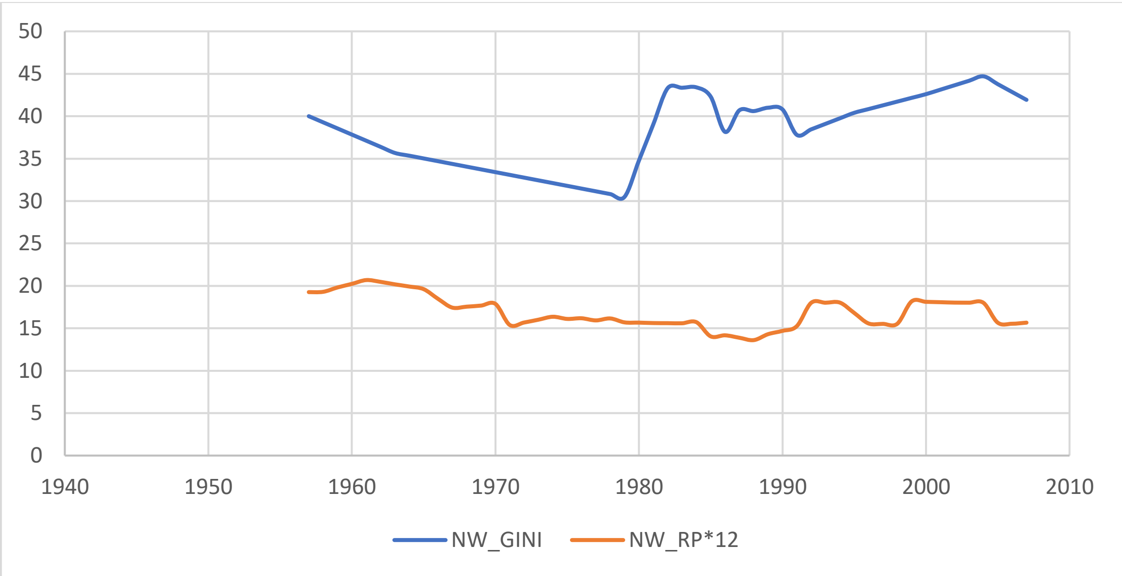
Source: see text. Note: The RP is multiplied by 12 in order to facilitate visualization

Figure 3B.5. The GINI and the LP index for Norway



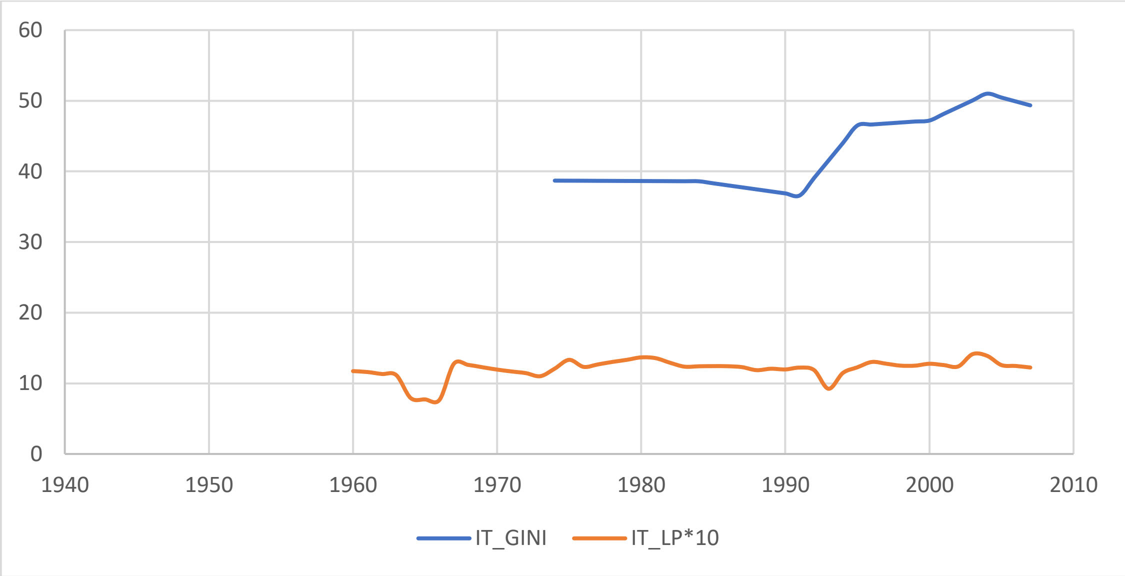
Source: see text. Note: The LP is multiplied by 10 in order to facilitate visualization

Figure 3B.6. The GINI and the RP index for Norway



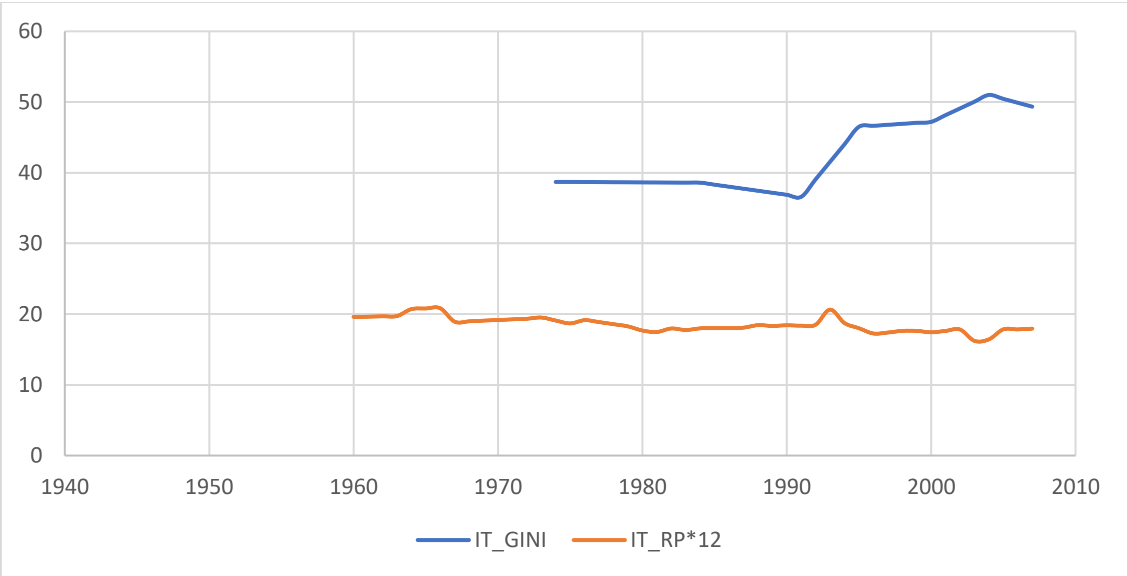
Source: see text. Note: The RP is multiplied by 12 in order to facilitate visualization

Figure 3B.7. The GINI and the LP index for Italy



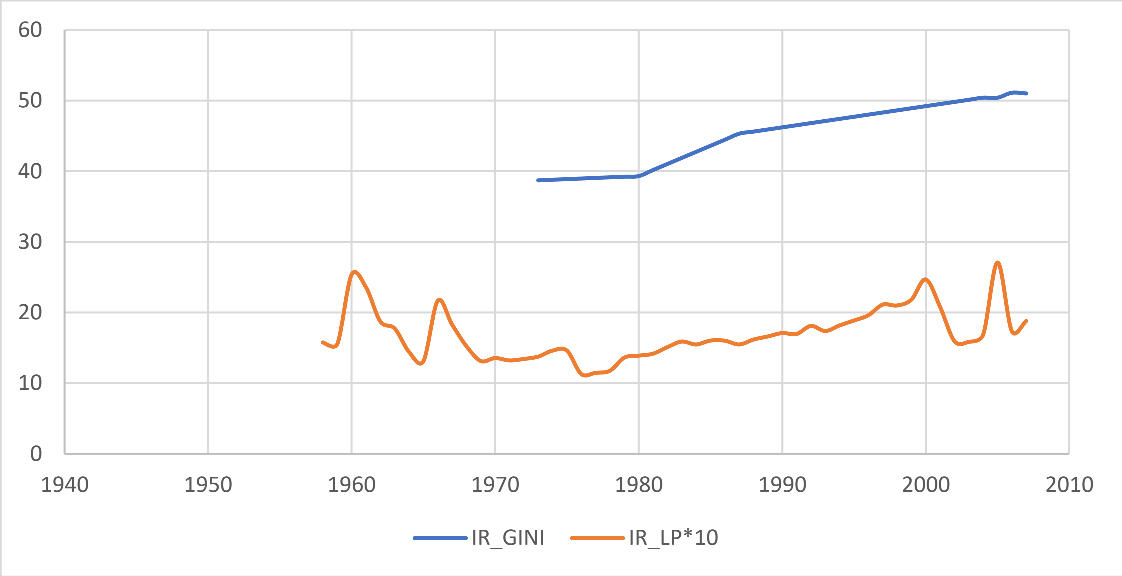
Source: see text. Note: The LP is multiplied by 10 in order to facilitate visualization

Figure 3B.8. The GINI and the RP index for Italy



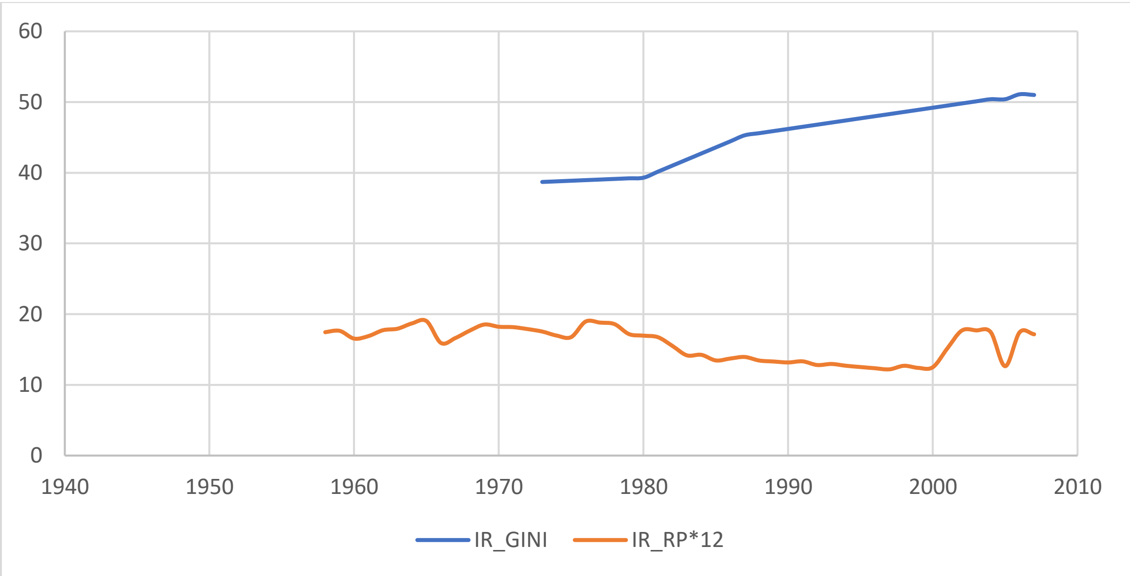
Source: see text. Note: The RP is multiplied by 12 in order to facilitate visualization

Figure 3B.9. The GINI and the LP index for Ireland



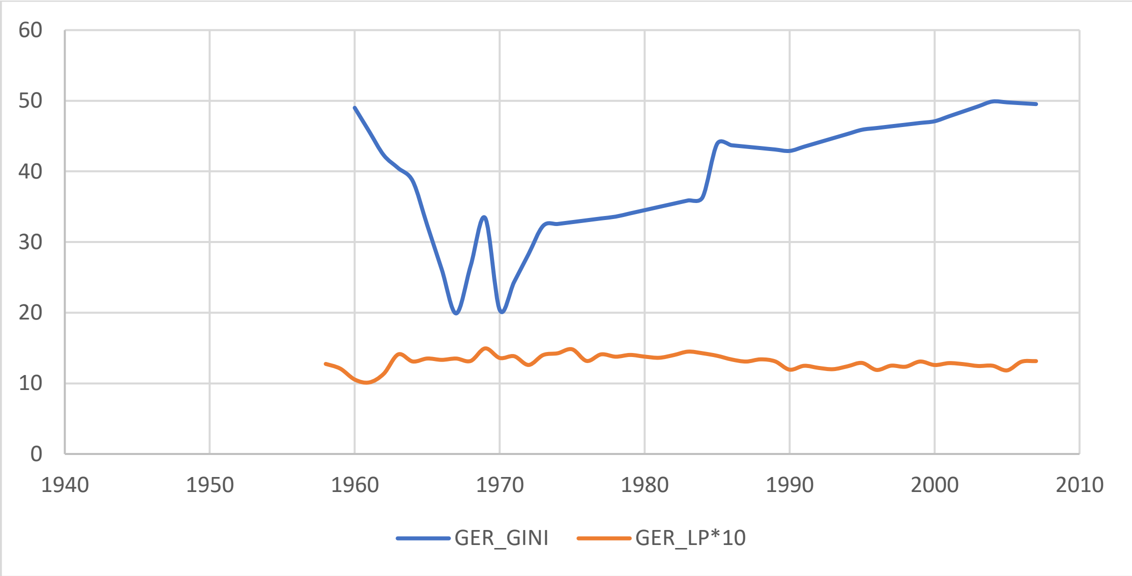
Source: see text. Note: The LP is multiplied by 10 in order to facilitate visualization

Figure 3B.10. The GINI and the RP index for Ireland



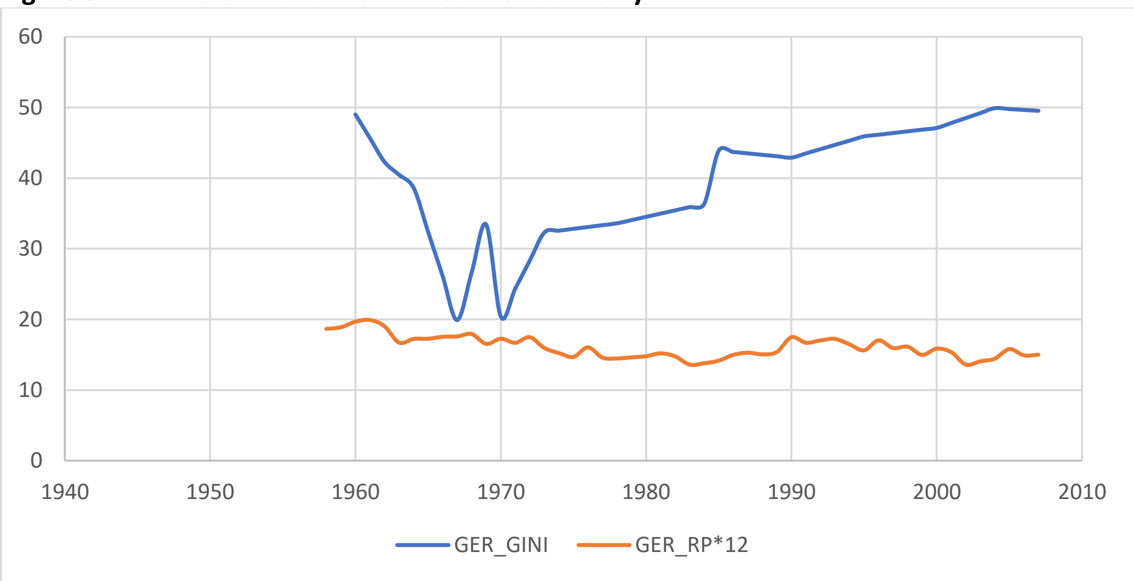
Source: see text. Note: The RP is multiplied by 12 in order to facilitate visualization

Figure 3B.11. The GINI and the LP index for Germany



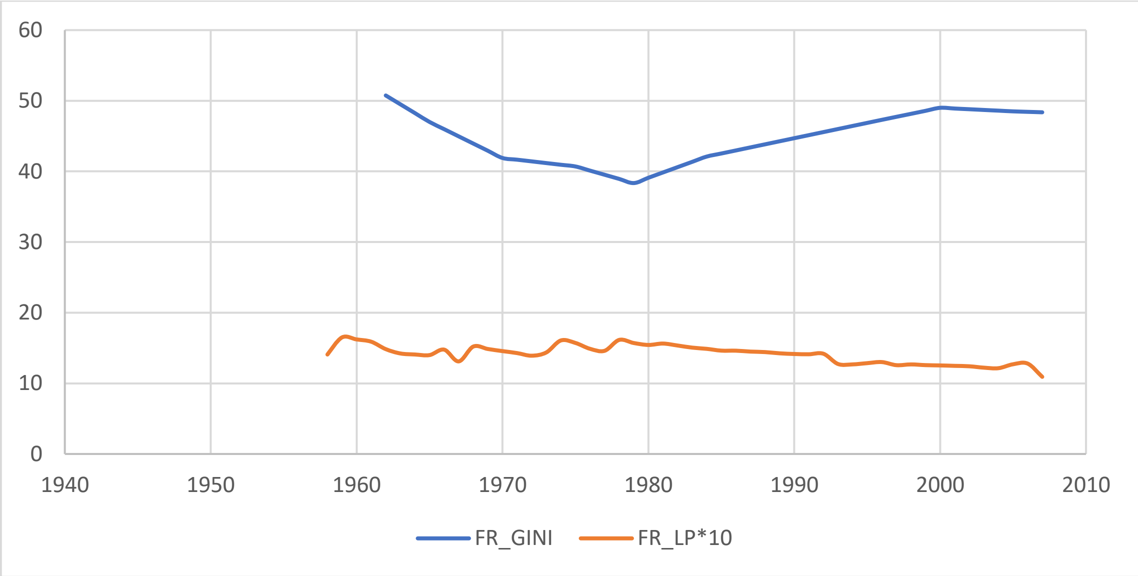
Source: see text. Note: The LP is multiplied by 10 in order to facilitate visualization

Figure 3B.12. The GINI and the RP index for Germany



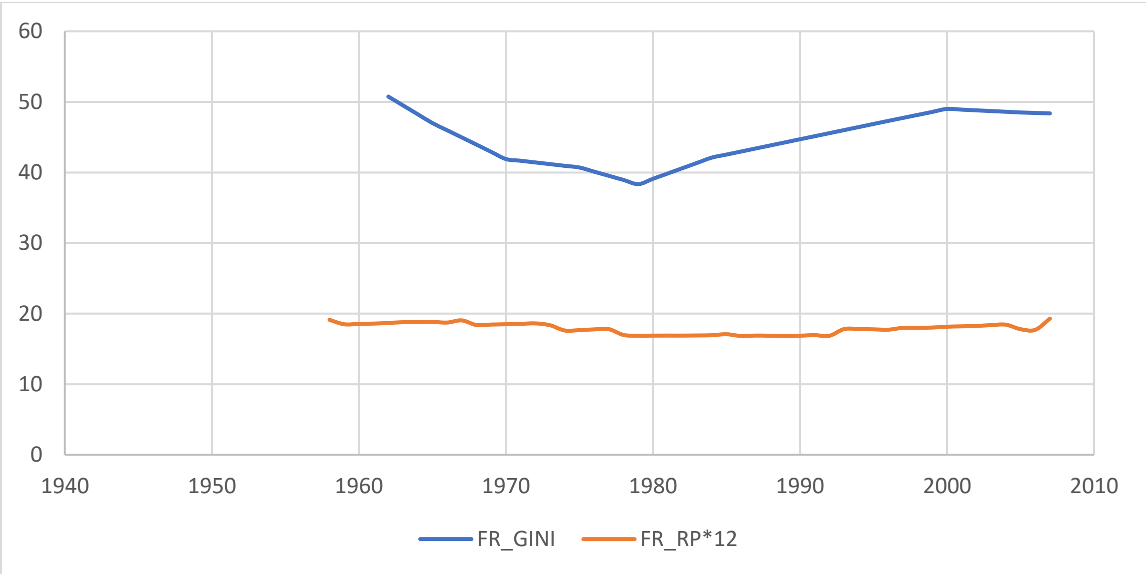
Source: see text. Note: The RP is multiplied by 12 in order to facilitate visualization

Figure 3B.13. The GINI and the LP index for France



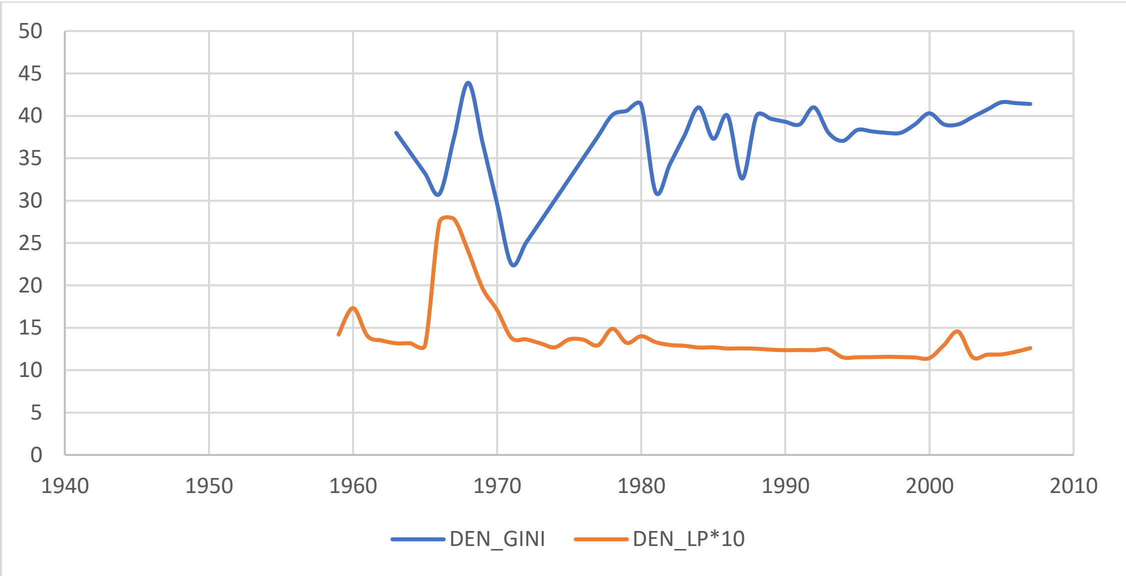
Source: see text. Note: The LP is multiplied by 10 in order to facilitate visualization

Figure 3B.14. The GINI and the RP index for France



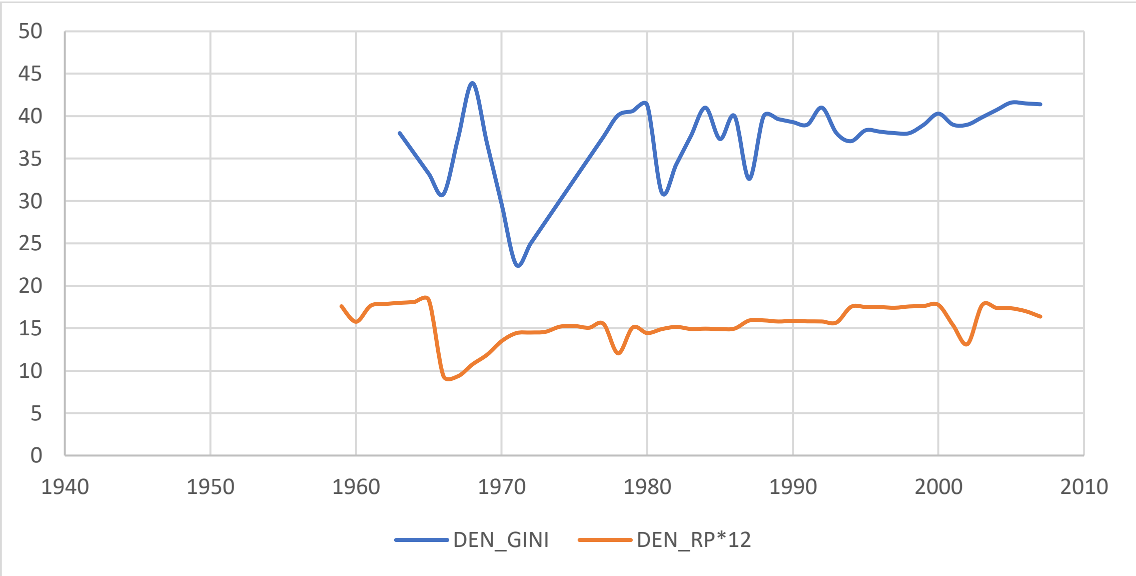
Source: see text. Note: The RP is multiplied by 12 in order to facilitate visualization

Figure 3B.15. The GINI and the LP index for Denmark



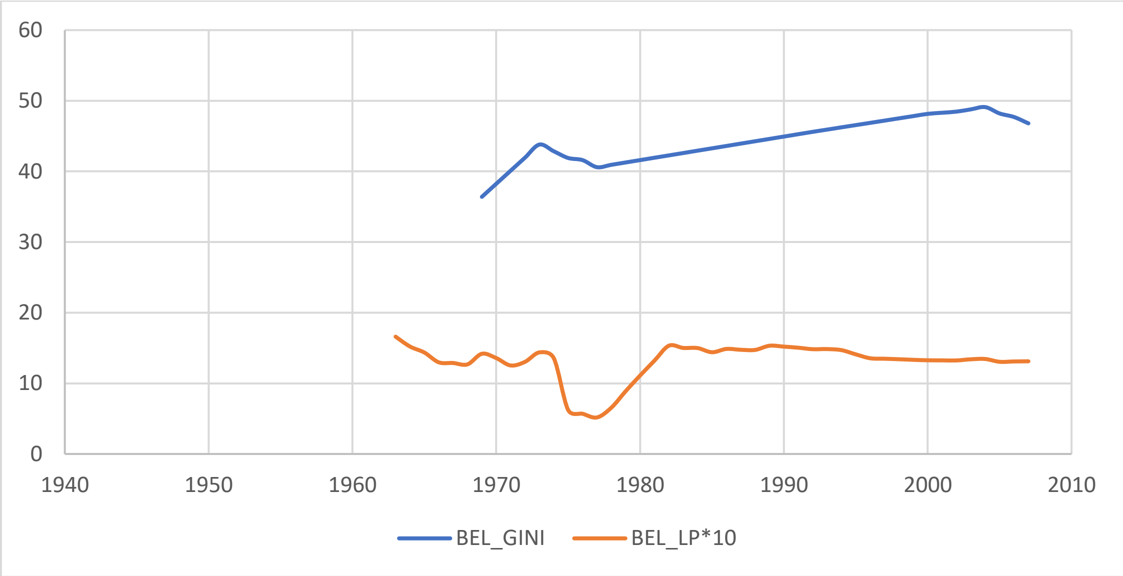
Source: see text. Note: The LP is multiplied by 10 in order to facilitate visualization

Figure 3B.16. The GINI and the RP index for Denmark



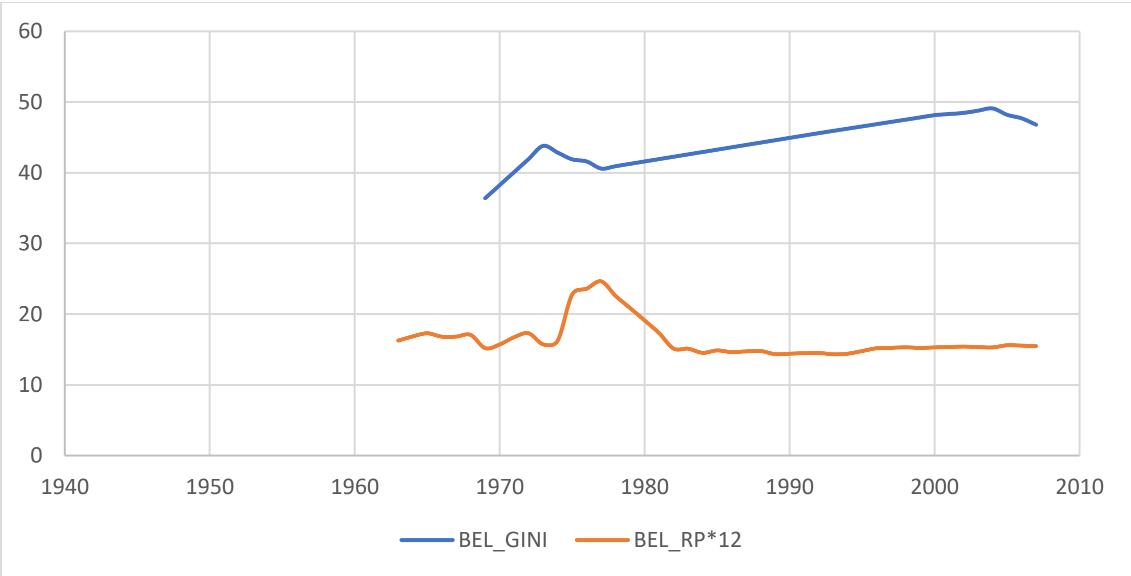
Source: see text. Note: The RP is multiplied by 12 in order to facilitate visualization

Figure 3B.17. The GINI and the LP index for Belgium



Source: see text. Note: The LP is multiplied by 10 in order to facilitate visualization

Figure 3B.18. The GINI and the RP index for Belgium



Source: see text. Note: The RP is multiplied by 12 in order to facilitate visualization

Appendix 3.C – Causality tests (SS's dataset)

Table 3C.1. Summary statistics for the top marginal tax rates and the top 1% and 0.01% income shares

Var	Mean	Stdev	Median	Max	Min	NbObs
Top rate	45.77050	22.98871	48.6268	97.50000	0	1988
Top 0.01%	1.268332	1.008856	0.850000	5.120000	0.17000	1220
Top 1%	10.53619	4.270170	9.440000	28.0400	3.49000	1595

Table 3C.2. Stationarity tests for the GINI, RP and LP (H0=not stationary)

Variables	Levin, Lin and Chu	Im, Pesaran and Shin	ADF	PP
Top rate	-2.30464 (0.0106)	-1.88709 (0.0296)	45.8127 (0.2437)	56.5099 (0.0434)
Top 0.01%	-4.42183 (0.0000)	-3.18556 (0.0007)	54.8348 (0.0018)	26.6215 (0.5389)
Top 1%	-4.37177 (0.0000)	-2.16772 (0.0151)	50.0830 (0.0595)	55.3272 (0.0207)

*Source: see text. Notes: The numbers in parenthesis are probability values. ** indicates the relevant coefficient is significant at 1%; * at 5%. The lag orders are obtained with the Akaike Information Criteria (AIC).*

Autocorrelation tests for the top marginal tax rates and the top 1% and 0.01% income shares:

TOP MARGINAL RATE:

TOP RATE
Date: 05/08/17 Time: 10:39
Sample: 1900 2010
Included observations: 1988

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
|*******	|*******	1	0.961	0.961	1839.3	0.000
|*******	| |	2	0.919	-0.060	3522.5	0.000
|******|	| |	3	0.878	-0.007	5060.0	0.000
|******|	| |	4	0.838	-0.017	6459.8	0.000
|******|	| |	5	0.797	-0.027	7727.0	0.000
|***** |	| |	6	0.760	0.024	8878.9	0.000
|***** |	| |	7	0.724	-0.009	9924.4	0.000
|***** |	| |	8	0.689	-0.006	10872.	0.000
|***** |	| |	9	0.657	0.025	11735.	0.000
|***** |	| |	10	0.626	-0.022	12518.	0.000
|**** |	| |	11	0.596	0.013	13230.	0.000
|**** |	| |	12	0.567	-0.026	13873.	0.000
|**** |	| |	13	0.535	-0.044	14447.	0.000
|**** |	| |	14	0.506	0.014	14959.	0.000
|*** |	| |	15	0.478	0.000	15417.	0.000
|*** |	| |	16	0.450	-0.010	15824.	0.000
|*** |	| |	17	0.426	0.019	16188.	0.000
|*** |	| |	18	0.401	-0.018	16511.	0.000
|*** |	| |	19	0.378	-0.005	16798.	0.000
|** |	| |	20	0.352	-0.049	17047.	0.000
|** |	| |	21	0.328	0.014	17263.	0.000
|** |	| |	22	0.303	-0.026	17448.	0.000
|** |	| |	23	0.279	-0.012	17605.	0.000
|** |	| |	24	0.253	-0.038	17734.	0.000
|** |	| |	25	0.225	-0.053	17836.	0.000

```
       |*     |              |     |        26    0.195   -0.044   17913.   0.000
       |*     |              |     |        27    0.164   -0.026   17967.   0.000
       |*     |              |     |        28    0.133   -0.026   18003.   0.000
       |*     |              |     |        29    0.104   -0.004   18025.   0.000
       |*     |              |     |        30    0.078    0.003   18037.   0.000
       |      |              |     |        31    0.052   -0.023   18042.   0.000
       |      |              |     |        32    0.026   -0.026   18044.   0.000
       |      |              |     |        33    0.003    0.016   18044.   0.000
       |      |              |     |        34   -0.018   -0.004   18044.   0.000
       |      |              |     |        35   -0.039   -0.027   18047.   0.000
       |      |              |     |        36   -0.060   -0.016   18055.   0.000
```

TOP 0.01%:

TOP0_01PERC
Date: 05/08/17   Time: 10:46
Sample: 1900 2010
Included observations: 1220

```
==========================================================================================
   Autocorrelation      Partial Correlation          AC       PAC      Q-Stat    Prob
==========================================================================================
       |*******              |*******         1    0.970    0.970    1149.9   0.000
       |*******             *|     |          2    0.929   -0.188    2206.2   0.000
       |******|             *|     |          3    0.882   -0.100    3159.4   0.000
       |******|              |     |          4    0.836    0.013    4015.6   0.000
       |******|              |     |          5    0.793    0.037    4786.9   0.000
       |***** |              |     |          6    0.753    0.011    5483.7   0.000
       |***** |              |     |          7    0.716   -0.002    6114.4   0.000
       |***** |              |     |          8    0.682   -0.002    6685.7   0.000
       |***** |              |     |          9    0.650    0.020    7205.1   0.000
       |****  |              |     |         10    0.617   -0.035    7674.9   0.000
       |****  |              |     |         11    0.585   -0.032    8096.5   0.000
       |****  |              |     |         12    0.549   -0.061    8468.6   0.000
       |****  |              |     |         13    0.513   -0.010    8793.5   0.000
       |***   |              |     |         14    0.477   -0.007    9074.8   0.000
       |***   |              |     |         15    0.443   -0.000    9317.2   0.000
       |***   |              |     |         16    0.410   -0.015    9525.0   0.000
       |***   |              |     |         17    0.379    0.005    9703.0   0.000
       |**    |              |     |         18    0.352    0.030    9856.8   0.000
       |**    |              |     |         19    0.325   -0.048    9987.9   0.000
       |**    |              |     |         20    0.300    0.019    10100.   0.000
       |**    |              |     |         21    0.274   -0.064    10193.   0.000
       |**    |              |     |         22    0.245   -0.027    10268.   0.000
       |**    |              |     |         23    0.217   -0.012    10327.   0.000
       |*     |              |     |         24    0.190    0.014    10372.   0.000
       |*     |              |     |         25    0.165   -0.005    10405.   0.000
       |*     |              |     |         26    0.141   -0.003    10430.   0.000
       |*     |              |     |         27    0.117   -0.049    10447.   0.000
       |*     |              |     |         28    0.093   -0.018    10458.   0.000
       |      |              |     |         29    0.070    0.001    10464.   0.000
       |      |              |     |         30    0.050    0.007    10467.   0.000
       |      |              |     |         31    0.030   -0.019    10469.   0.000
       |      |              |     |         32    0.010   -0.007    10469.   0.000
       |      |              |     |         33   -0.008   -0.009    10469.   0.000
       |      |              |     |         34   -0.026   -0.006    10470.   0.000
       |      |              |     |         35   -0.043   -0.013    10472.   0.000
       |      |              |     |         36   -0.060   -0.020    10477.   0.000
==========================================================================================
```

TOP 1%:

TOP1PERC
Date: 05/08/17 Time: 10:52
Sample: 1900 2010
Included observations: 1595

<table>
<tr><th>Autocorrelation</th><th>Partial Correlation</th><th></th><th>AC</th><th>PAC</th><th>Q-Stat</th><th>Prob</th></tr>
<tr><td>|*******</td><td>|*******</td><td>1</td><td>0.963</td><td>0.963</td><td>1482.1</td><td>0.000</td></tr>
<tr><td>|*******</td><td>*| |</td><td>2</td><td>0.920</td><td>-0.101</td><td>2835.9</td><td>0.000</td></tr>
<tr><td>|******|</td><td>| |</td><td>3</td><td>0.876</td><td>-0.037</td><td>4063.0</td><td>0.000</td></tr>
<tr><td>|******|</td><td>| |</td><td>4</td><td>0.832</td><td>-0.008</td><td>5172.0</td><td>0.000</td></tr>
<tr><td>|******|</td><td>| |</td><td>5</td><td>0.795</td><td>0.063</td><td>6185.0</td><td>0.000</td></tr>
<tr><td>|***** |</td><td>| |</td><td>6</td><td>0.759</td><td>-0.016</td><td>7109.5</td><td>0.000</td></tr>
<tr><td>|***** |</td><td>| |</td><td>7</td><td>0.726</td><td>0.009</td><td>7955.4</td><td>0.000</td></tr>
<tr><td>|***** |</td><td>| |</td><td>8</td><td>0.696</td><td>0.014</td><td>8732.1</td><td>0.000</td></tr>
<tr><td>|***** |</td><td>| |</td><td>9</td><td>0.667</td><td>0.005</td><td>9445.9</td><td>0.000</td></tr>
<tr><td>|***** |</td><td>| |</td><td>10</td><td>0.637</td><td>-0.027</td><td>10098.</td><td>0.000</td></tr>
<tr><td>|**** |</td><td>| |</td><td>11</td><td>0.608</td><td>-0.015</td><td>10692.</td><td>0.000</td></tr>
<tr><td>|**** |</td><td>| |</td><td>12</td><td>0.575</td><td>-0.056</td><td>11224.</td><td>0.000</td></tr>
<tr><td>|**** |</td><td>| |</td><td>13</td><td>0.539</td><td>-0.060</td><td>11691.</td><td>0.000</td></tr>
<tr><td>|**** |</td><td>| |</td><td>14</td><td>0.501</td><td>-0.030</td><td>12096.</td><td>0.000</td></tr>
<tr><td>|*** |</td><td>| |</td><td>15</td><td>0.464</td><td>-0.016</td><td>12443.</td><td>0.000</td></tr>
<tr><td>|*** |</td><td>| |</td><td>16</td><td>0.429</td><td>0.004</td><td>12741.</td><td>0.000</td></tr>
<tr><td>|*** |</td><td>| |</td><td>17</td><td>0.400</td><td>0.039</td><td>12999.</td><td>0.000</td></tr>
<tr><td>|*** |</td><td>| |</td><td>18</td><td>0.374</td><td>0.006</td><td>13225.</td><td>0.000</td></tr>
<tr><td>|** |</td><td>| |</td><td>19</td><td>0.348</td><td>-0.023</td><td>13421.</td><td>0.000</td></tr>
<tr><td>|** |</td><td>| |</td><td>20</td><td>0.325</td><td>0.009</td><td>13591.</td><td>0.000</td></tr>
<tr><td>|** |</td><td>| |</td><td>21</td><td>0.301</td><td>-0.020</td><td>13738.</td><td>0.000</td></tr>
<tr><td>|** |</td><td>| |</td><td>22</td><td>0.276</td><td>-0.017</td><td>13861.</td><td>0.000</td></tr>
<tr><td>|** |</td><td>| |</td><td>23</td><td>0.252</td><td>-0.013</td><td>13964.</td><td>0.000</td></tr>
<tr><td>|** |</td><td>| |</td><td>24</td><td>0.227</td><td>-0.016</td><td>14048.</td><td>0.000</td></tr>
<tr><td>|* |</td><td>| |</td><td>25</td><td>0.203</td><td>0.005</td><td>14115.</td><td>0.000</td></tr>
<tr><td>|* |</td><td>| |</td><td>26</td><td>0.180</td><td>-0.011</td><td>14168.</td><td>0.000</td></tr>
<tr><td>|* |</td><td>| |</td><td>27</td><td>0.156</td><td>-0.044</td><td>14207.</td><td>0.000</td></tr>
<tr><td>|* |</td><td>| |</td><td>28</td><td>0.131</td><td>-0.036</td><td>14235.</td><td>0.000</td></tr>
<tr><td>|* |</td><td>| |</td><td>29</td><td>0.106</td><td>-0.014</td><td>14253.</td><td>0.000</td></tr>
<tr><td>|* |</td><td>| |</td><td>30</td><td>0.082</td><td>-0.016</td><td>14264.</td><td>0.000</td></tr>
<tr><td>| |</td><td>| |</td><td>31</td><td>0.060</td><td>-0.002</td><td>14270.</td><td>0.000</td></tr>
<tr><td>| |</td><td>| |</td><td>32</td><td>0.040</td><td>0.000</td><td>14273.</td><td>0.000</td></tr>
<tr><td>| |</td><td>| |</td><td>33</td><td>0.020</td><td>-0.019</td><td>14274.</td><td>0.000</td></tr>
<tr><td>| |</td><td>| |</td><td>34</td><td>0.000</td><td>-0.017</td><td>14274.</td><td>0.000</td></tr>
<tr><td>| |</td><td>| |</td><td>35</td><td>-0.019</td><td>-0.003</td><td>14274.</td><td>0.000</td></tr>
<tr><td>| |</td><td>| |</td><td>36</td><td>-0.037</td><td>-0.001</td><td>14276.</td><td>0.000</td></tr>
</table>

Standard Granger non-causality tests:

Table 3C.3. Results for causality between top marginal rates and top 1% income share using standard Granger non-causality test

Null hypothesis	Lag order	F-statistics
TOP RATE does not Granger Cause TOP 1%	1	23.8126** (1.E-06)
TOP 1% does not Granger Cause TOP RATE	1	18.1390** (2.E-05)
TOP RATE does not Granger Cause TOP 1%	2	10.7138** (2.E-05)
TOP 1% does not Granger Cause TOP RATE	2	7.07633** (0.0009)

TOP RATE does not Granger Cause TOP 1%	3	6.57428** (0.0002)
TOP 1% does not Granger Cause TOP RATE	3	5.64375** (0.0008)
TOP RATE does not Granger Cause TOP 1%	4	5.00835** (0.0005)
TOP 1% does not Granger Cause TOP RATE	4	4.29473** (0.0019)
TOP RATE does not Granger Cause TOP 1%	5	5.59896** (4.E-05)
TOP 1% does not Granger Cause TOP RATE	5	3.08654** (0.0089)
TOP RATE does not Granger Cause TOP 1%	6	3.76034** (0.0010)
TOP 1% does not Granger Cause TOP RATE	6	4.38101** (0.0002)
TOP RATE does not Granger Cause TOP 1%	7	3.62304** (0.0007)
TOP 1% does not Granger Cause TOP RATE	7	4.42852** (7.E-05)
TOP RATE does not Granger Cause TOP 1%	8	3.47724** (0.0006)
TOP 1% does not Granger Cause TOP RATE	8	4.49789** (2.E-05)
TOP RATE does not Granger Cause TOP 1%	9	3.22957** (0.0007)
TOP 1% does not Granger Cause TOP RATE	9	4.26689** (2.E-05)
TOP RATE does not Granger Cause TOP 1%	10	3.23017** (0.0004)
TOP 1% does not Granger Cause TOP RATE	10	3.87584** (3.E-05)

*The numbers in parenthesis are probability values. ** indicates the relevant coefficient is significant at 1%; * at 5%. Optimal lag length = 5 (AIC).*

Table 3C.4. Results for causality between top marginal rates and top 0.01% income share using standard Granger non-causality test

Null hypothesis	Lag order	F-statistics
TOP 0.01% does not Granger Cause TOP RATE	1	17.1437** (4.E-05)
TOP RATE does not Granger Cause TOP 0.01%	1	18.1885** (2.E-05)
TOP 0.01% does not Granger Cause TOP RATE	2	9.91814** (5.E-05)
TOP RATE does not Granger Cause TOP 0.01%	2	7.41109** (0.0006)
TOP 0.01% does not Granger Cause TOP RATE	3	7.78331** (4.E-05)
TOP RATE does not Granger Cause TOP 0.01%	3	4.92525** (0.0021)
TOP 0.01% does not Granger Cause TOP RATE	4	7.31120** (8.E-06)
TOP RATE does not Granger Cause TOP 0.01%	4	3.92518** (0.0036)
TOP 0.01% does not Granger Cause TOP RATE	5	7.63271** (5.E-07)
TOP RATE does not Granger Cause TOP 0.01%	5	3.60951** (0.0030)

TOP 0.01% does not Granger Cause TOP RATE	6	6.65685** (6.E-07)
TOP RATE does not Granger Cause TOP 0.01%	6	2.16427* (0.0441)
TOP 0.01% does not Granger Cause TOP RATE	7	7.03519** (3.E-08)
TOP RATE does not Granger Cause TOP 0.01%	7	1.96126 (0.0573)
TOP 0.01% does not Granger Cause TOP RATE	8	6.86498** (8.E-09)
TOP RATE does not Granger Cause TOP 0.01%	8	1.55868 (0.1330)
TOP 0.01% does not Granger Cause TOP RATE	9	6.34137** (9.E-09)
TOP RATE does not Granger Cause TOP 0.01%	9	1.48005 (0.1504)
TOP 0.01% does not Granger Cause TOP RATE	10	5.71196** (2.E-08)
TOP RATE does not Granger Cause TOP 0.01%	10	1.60644 (0.0996)

*The numbers in parenthesis are probability values. ** indicates the relevant coefficient is significant at 1%; * at 5%. Optimal lag length = 9 (AIC).*

DH causality tests:

Table 3C.5. Results for causality between top marginal rates and top 1% income share using DH causality test

Null hypothesis	Lag order	F-statistics
TOP RATE does not Granger Cause TOP 1%	1	16.3260** (0.0000)
TOP 1% does not Granger Cause TOP RATE	1	9.90204** (0.0000)
TOP RATE does not Granger Cause TOP 1%	2	7.93697** (2.E-15)
TOP 1% does not Granger Cause TOP RATE	2	2.05150* (0.0402)
TOP RATE does not Granger Cause TOP 1%	3	6.24780** (4.E-10)
TOP 1% does not Granger Cause TOP RATE	3	1.82016 (0.0687)
TOP RATE does not Granger Cause TOP 1%	4	8.88536** (0.0000)
TOP 1% does not Granger Cause TOP RATE	4	2.43755* (0.0148)
TOP RATE does not Granger Cause TOP 1%	5	11.4366** (0.0000)
TOP 1% does not Granger Cause TOP RATE	5	3.75514** (0.0002)
TOP RATE does not Granger Cause TOP 1%	6	10.5955** (0.0000)
TOP 1% does not Granger Cause TOP RATE	6	4.63268** (4.E-06)
TOP RATE does not Granger Cause TOP 1%	7	9.45150** (0.0000)
TOP 1% does not Granger Cause TOP RATE	7	4.72847** (2.E-06)
TOP RATE does not Granger Cause TOP 1%	8	6.25225** (4.E-10)

TOP 1% does not Granger Cause TOP RATE	8	2.94917** (0.0032)

*The numbers in parenthesis are probability values. ** indicates the relevant coefficient is significant at 1%; * at 5%.*

Table 3C.6. Results for causality between top marginal rates and top 0.01% income share using DH causality test

Null hypothesis	Lag order	F-statistics
TOP RATE does not Granger Cause TOP 0.01%	1	22.9705** (0.0000)
TOP 0.01% does not Granger Cause TOP RATE	1	1.64213 (0.1006)
TOP RATE does not Granger Cause TOP 0.01%	2	15.8432** (0.0000)
TOP 0.01% does not Granger Cause TOP RATE	2	0.82810 (0.4076)
TOP RATE does not Granger Cause TOP 0.01%	3	12.2173** (0.0000)
TOP 0.01% does not Granger Cause TOP RATE	3	2.23058* (0.0257)
TOP RATE does not Granger Cause TOP 0.01%	4	11.9852** (0.0000)
TOP 0.01% does not Granger Cause TOP RATE	4	8.42073** (0.0000)
TOP RATE does not Granger Cause TOP 0.01%	5	9.03063** (0.0000)
TOP 0.01% does not Granger Cause TOP RATE	5	48.8441** (0.0000)
TOP RATE does not Granger Cause TOP 0.01%	6	8.61655** (0.0000)
TOP 0.01% does not Granger Cause TOP RATE	6	47.7446** (0.0000)
TOP RATE does not Granger Cause TOP 0.01%	7	10.2194** (0.0000)
TOP 0.01% does not Granger Cause TOP RATE	7	44.2808** (0.0000)
TOP RATE does not Granger Cause TOP 0.01%	8	10.8639** (0.0000)
TOP 0.01% does not Granger Cause TOP RATE	8	78.2729** (0.0000)
TOP RATE does not Granger Cause TOP 0.01%	9	6.01330** (2.E-09)
TOP 0.01% does not Granger Cause TOP RATE	9	113.608** (0.0000)
TOP RATE does not Granger Cause TOP 0.01%	10	3.47590** (0.0005)
TOP 0.01% does not Granger Cause TOP RATE	10	67.7275** (0.0000)

*The numbers in parenthesis are probability values. ** indicates the relevant coefficient is significant at 1%; * at 5%.*

Appendix 3.D – Causality tests with country fixed effects and time effects

Table 3D.1. Causality with time effects (top rates and top 1% income share)

	Top income tax rate		Income share of top 1%	
	(1)	**(2)**	**(3)**	**(4)**
Top Income Tax Rate$_{t-1}$	0.9575** (0.0268) 0.0000	1.0316** (0.0258) 0.0000	-0.00331 (0.0038) 0.3897	-0.011508** (0.003753) 0.0022
Top Income Tax Rate$_{t-2}$	-0.05162 (0.0267) 0.0541	-0.085** (0.0261) 0.0011	-0.00226 (0.0038) 0.5552	0.001540 (0.003791) 0.6846
Income Share of Top $1\%_{t-1}$	0.057863 (0.1843) 0.7536	-0.15340 (0.1734) 0.3765	0.9998** (0.0264) 0.0000	1.072961** (0.025164) 0.0000
Income Share of Top $1\%_{t-2}$	-0.16383 (0.1819) 0.3680	0.107241 (0.1700) 0.5284	-0.084** (0.0260) 0.0012	-0.154053* (0.024692) 0.0000
Common Time Trends	No	Yes	No	Yes
Period Fixed Effects	Yes	No	Yes	No
Country Fixed Effects	Yes	Yes	Yes	Yes
R-squared	0.958000	0.950565	0.973937	0.968591
Number of Observations	1513	1513	1507	1507

*The table reports robust standard errors in parentheses and p-values. ** indicates the relevant coefficient is significant at 1%; * at 5%.*

Table 3D.2. Causality with time effects (top rates and top 0.01% income share)

	Top income tax rate		Income share of top 1%	
	(1)	**(2)**	**(3)**	**(4)**
Top Income Tax Rate$_{t-1}$	0.9476** (0.0314) 0.0000	1.0106** (0.03013) 0.0000	-0.00075 (0.0008) 0.3656	-0.003509** (0.000886) 0.0001
Top Income Tax Rate$_{t-2}$	-0.04379 (0.0311) 0.1606	-0.079** (0.0301) 0.0088	-0.00025 (0.0008) 0.7608	0.001129 (0.000885) 0.2024
Income Share of Top $0.01\%_{t-1}$	-2.5553* (1.1637) 0.0283	-2.3611* (0.9656) 0.0146	1.1211** (0.0306) 0.0000	1.234223** (0.028159) 0.0000
Income Share of Top $0.01\%_{t-2}$	2.38628* (1.1494) 0.0381	2.19015* (0.9300) 0.0187	-0.188** (0.0303) 0.0000	-0.313522** (0.027116) 0.0000
Common Time Trends	No	Yes	No	Yes
Period Fixed Effects	Yes	No	Yes	No
Country Fixed Effects	Yes	Yes	Yes	Yes
R-squared	0.958421	0.948805	0.981667	0.972423
Number of Observations	1138	1138	1133	1133

*The table reports robust standard errors in parentheses and p-values. ** indicates the relevant coefficient is significant at 1%; * at 5%.*

Table 3D.3. Causality with time effects (GINI and LP)

	GINI		LP	
	(1)	**(2)**	**(3)**	**(4)**
$\mathbf{GINI}_{t-1}$	0.9647**	0.9686**	0.006103	0.003654

	(0.0546) 0.0000	(0.0511) 0.0000	(0.0039) 0.1250	(0.003785) 0.3351
$\mathbf{GINI}_{t-2}$	-0.1288* (0.0543) 0.0182	-0.129** (0.0494) 0.0090	-0.00492 (0.0039) 0.2126	-0.002234 (0.003660) 0.5419
$\mathbf{LP}_{t-1}$	0.503039 (0.7348) 0.4941	0.543020 (0.6740) 0.4210	0.6378** (0.0533) 0.0000	0.652452** (0.049877) 0.0000
$\mathbf{LP}_{t-2}$	-0.76066 (0.7284) 0.2972	-0.51024 (0.6671) 0.4449	0.082449 (0.0528) 0.1199	0.047091 (0.049371) 0.3408
Common Time Trends	No	Yes	No	Yes
Period Fixed Effects	Yes	No	Yes	No
Country Fixed Effects	Yes	Yes	Yes	Yes
R-squared	0.893550	0.871287	0.678532	0.596034
Number of Observations	389	389	389	389

*The table reports robust standard errors in parentheses and p-values. ** indicates the relevant coefficient is significant at 1%; * at 5%.*

Table 3D.4. Causality with time effects (GINI and RP)

	GINI		**RP**	
	(1)	**(2)**	**(3)**	**(4)**
$\mathbf{GINI}_{t-1}$	0.9678** (0.0547) 0.0000	0.9712** (0.0511) 0.0000	-0.004** (0.0015) 0.0066	-0.003111* (0.001487) 0.0370
$\mathbf{GINI}_{t-2}$	-0.1368* (0.0547) 0.0129	-0.133** (0.0494) 0.0075	0.002816 (0.0015) 0.0748	0.002579 (0.001437) 0.0734
$\mathbf{RP}_{t-1}$	-0.16538 (1.8425) 0.9285	0.009624 (1.6922) 0.9955	0.6607** (0.0530) 0.0000	0.672538** (0.049170) 0.0000
$\mathbf{RP}_{t-2}$	-0.28585 (1.8176) 0.8751	-0.59917 (1.6667) 0.7194	0.080204 (0.0523) 0.1263	0.064621 (0.048431) 0.1829
Common Time Trends	No	Yes	No	Yes
Period Fixed Effects	Yes	No	Yes	No
Country Fixed Effects	Yes	Yes	Yes	Yes
R-squared	0.893223	0.871123	0.723691	0.660238
Number of Observations	389	389	389	389

*The table reports robust standard errors in parentheses and p-values. ** indicates the relevant coefficient is significant at 1%; * at 5%.*

Appendix 3.E – Brief discussion on global progressivity measures

Global indexes, which differently from the RP and LP measures take into account the whole income distribution, are mostly based on two related procedures: they or compare pre and post-taxes income distribution and regard as progressive the tax structures which after applied lead to more equal income distributions, or measure how much the tax structure deviate from a proportional tax system (Kesselman and Cheung, 2004). The first procedure is more common and both enclose the idea that progressivity is the property of PIT systems of leading to a more equal income distribution.

As an example of the first kind (comparison between pre and post-taxes income distributions) consider Reynolds and Smolensky (1977) measure, given by

$$G\big(Y - T(Y)\big) - G(Y), \tag{3E.1}$$

where $G(.)$ is the Gini income inequality measure.

A very popular measure constructed on the basis of the second procedure (deviation from proportionality) is the Kakwani (1977) measure, given by

$$C(T) - G(Y), \tag{3E.2}$$

where $C(T)$ is the tax concentration coefficient and $G(Y)$ is the Gini coefficient.

The GINI coefficient basically measures the area between a diagonal line indicating complete equality (in a Cartesian plane where the X axis indicates the cumulative percentage of the population and the Y axis represents the cumulative percentage of income) and the Lorenz curve, whilst the tax concentration coefficient correspond to the area between the diagonal line and the tax concentration (TC) curve (indicating the percentage of total taxes paid in a Cartesian plane where the X axis indicates the cumulative percentage of population and the Y axis represents the cumulative percentage of tax). Thus the Kakwani (1977) index measures the area between the Lorenz and the TC curve, which corresponds to measuring deviation from proportionality once the TC curve corresponds to the Lorenz curve when the tax system is proportional.

The measure is positive for progressive, zero for proportional and negative for regressive tax systems. Some other progressivity measures consider different income inequality index (as the Atkinson (1970) measure) to compare post and pre-tax income distributions. An example would be Kiefer's (1984) measure, given by

$$I(Y) - I(Y - T(Y)), \tag{3E.3}$$

where $I(.)$ is an Atkinson (1970) inequality measure.

Appendix 3.F – Endogeneity tests

To test endogeneity[144] between GINI, LP and RP I run 8 fixed effects IV panel regressions. Besides varying in terms of the variables used, the regressions vary in terms of the type of time effects used (time trend or time dummies). Also, I use lagged variables of order 1 as the instrument concerning the supposedly endogenous variable. So, for example, in the first regression the GINI is regressed against the LP and time trend is used to capture time effects; the regressor tested is the LP and the instrument used is the first order lag of the LP variable. Table 3F.1 present all the results.

Table 3F.1. Endogeneity tests regarding the GINI, RP and LP variables

	Dependent variables							
	GINI	GINI	GINI	GINI	LP	LP	RP	RP
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
LP	4.28* (1.74)	3.58** (1.3)	-	-	-	-	-	-
RP	-	-	-6.7 (4.2)	-14 (3.5)	-	-	-	-
GINI	-	-	-	-	.01** (.00)	.01** (.00)	-.003** (.00)	-.008 (.00)**
Endog. Test [Chi-sq(1)]	.25	.13	.00	.33	.07	.57	.05	3.37°
Time Trend	YES	NO	YES	NO	YES	NO	YES	NO
Period FE	NO	YES	NO	YES	NO	YES	NO	YES
Country FE	YES	YES	YES	YES	YES	YES	YES	YES
R-squared	.24	.03	.24	.08	.06	.03	.03	.08
N. Obs.	405	405	405	405	398	398	398	398

*The numbers in parenthesis are robust standard errors. ** indicates the relevant coefficient is significant at 1%; * at 5% and ° at 10%.*

The results suggest there is no endogeneity between the variables considered, which supports the validity of the causality analysis.

[144] Again, I apply the "difference-in-Sargan" test for endogeneity.

Chapter 4: The Elasticity of Marginal Utility and Psychological Traits

4.1. Introduction

Many studies have been concerned with estimating the EMU parameter, given the importance it has to public investment decision-taking and economics in general. Most methods to estimate it use observations from the general population or from convenience samples of students without bothering to measure individuals' psychological traits, such that little attention has been given to whether and how the EMU is affected by psychological traits and to the bearings that sampling has on the normative significance of the parameter.

In this chapter, there are two main analyses. The first investigates the relationship between the parameter and socially desirable traits. Such research has bearings on the normative significance of the EMU to the extent that attitudes of those possessing socially desirable traits can arguably be regarded as having more normative significance themselves. In the second analysis, I investigate how the context-sensitivity of the EMU relates to such traits. That is, the objective is to see whether people with particular psychological traits provide estimates of the EMU which are constant across the different contexts in which the parameter can be estimated. It sheds light on the issue of whether the Standard Welfare Economics Model (SWEM) is an appropriate normative framework, since it assumes the EMU is constant across the referred contexts (risk aversion, inequality aversion and time preference).[145]

This context analysis can also help to enlighten a debate within the literature regarding the different results found by Atkinson et al. (2009) and Groom and Maddison (2013) concerning the EMU context-sensitivity. Using experimental methods the first set of authors obtained results suggesting the EMU varies across contexts, while using mostly revealed-preferences based techniques the second set of authors obtained results suggesting the EMU does not vary across contexts. Examining how psychological traits affect the EMU context-sensitivity may reveal, to some extent, why they achieved different results. It must also be said the analyses executed in my study are unprecedented (as shown in the literature review). It is probably so due to the lack of attention devoted to the issue of how to test normative significance in economics.

[145] The SWEM referred is given by $E(W) = \sum_{i=1}^{I} \sum_{t=0}^{T} \sum_{p} p_s \,(u_{its})(1+\delta)^{-t}$, where $E(W)$ stands for the (expected, i.e. relative to the expected utility theory) welfare, p_s is the probability of state of nature s occurring, δ is the utility discount rate, i indicates a given individual and t a certain period of time. Many works use such framework to analyse welfare in economics, and in it the EMU can be interpreted in the three different ways referred above (Atkinson et al., 2009).

Underlying the first analysis is the idea that attitudes observed in a group possessing socially desirable traits can and should be given more normative weight – i.e. can be thought as being closer to what should be done. If that is the case the EMU obtained from people possessing desirable traits will have more normative value. Also, if that is the case, a model describing the behaviour of those possessing desirable psychological traits has more normative significance. This is, if the EMU of those possessing desirable traits is less likely to vary across contexts, as predicted by the SWEM, it would suggest the model gives a good description of individuals supposedly acting ideally, which would therefore support the model's usage as a normative model (i.e. as capable of prescribing ideal decisions). That relates to the second analysis.

I call the idea that normative estimates should be obtained only from those displaying particular psychological traits 'sample framing'. It suggests that when normative considerations are to be derived from empirical observations one should consider to frame its sample in order for it to reflect not the general population, but people whose attitudes have, for some reason (e.g. their psychological traits, their knowledge, etc.), more normative weight regarding the particular topic under consideration. In short, if we are examining what should be done based on observations, we should not base our conclusions on observations of just anyone, but just on observations of those who, for some reason, are expected to meet a normative criteria.

Notice the SWEM has been largely used even though it is not certain it describes people's behaviour, because its use can be justified based on the argument that it reflects rational conduct and thus normative behaviour. It basically means that decisions can be taken based on results derived from the model because it is assumed it unveils – to a large degree – the ideal decision. It highlights the importance of testing the model's assumed normative significance, which is done in the 'context analysis'.

The sample framing idea itself and its use to estimate an EMU with more normative significance are inspired by Drupp et al.'s (2015) study. This study surveys prominent economists who are familiar with social discounting on the components of the Ramsey Rule, one of which being the EMU. It is the only work to my knowledge that investigates the value of the EMU according to experts (their survey text can be found on Appendix 4.C; see question 3).[146]

The estimate of EMU they obtain can arguably be seen as possessing more weight than values obtained from the general population (or from a convenience sample of students) based, amongst other things, on the respondents' knowledge about the issue. The fact that they are

[146] Weitzman (2001) also surveys amongst experts on the SDR, but not on its determinants, such that there are no considerations regarding the EMU.

experts however does not mean that their answers have great normative significance. In terms of psychological traits, it is plausible to assume that prominent economists, as most prominent scientists, are likely to be weakly empathetic (Baron-Cohen et al., 2001). Socially desirable traits (if we accept that empathy is a desirable trait) are not always correlated with expertise, which shows how my study and Drupp et al.'s (2015) study differ in terms of 'sample-fitting'.

The concept of 'socially desirable' can certainly change over time and place, and although some traits' social desirability might be found to be considered more or less constant historically, it is by no means my intention to enter into such a discussion. Rather, I will ground my choice of socially desirable traits on the importance currently attributed to some traits by the psychology and economics mainstream literature – mainly when it comes to policy makers. This is the basis for my focusing on empathy and reflectiveness throughout the analysis. Besides, these traits summarize well[147] what one usually would expect a social planner to be – rational[148] (reflective, not impulsive) and considerate of others' problems, i.e. to be genuinely willing to solve them (empathetic).[149] Later on I will provide a much more detailed description of what I mean by these two traits.

At this point it should be noticed that although throughout the study I refer to empathy as a socially desirable trait, it is not everyone's opinion that decisions made by empathetic people are always better. Bloom (2016), for example, claims "on balance, empathy is a negative in human affairs. It's not cholesterol. It's sugary soda, tempting and delicious and bad for us" Bloom (2016: 13). He argues basically that the focal (i.e. inability to refer to more than one person) and biased nature of empathy leads people to make bad decisions systematically. Such opinion contrasts frontally with others that place empathy as central to morality (e.g. Baron-Cohen, 2011; Rifkin, 2009).[150] Moreover, Bloom (2016) declares that besides criticising the weight given to empathy he intends "to make a case for the value of conscious, deliberative reasoning" Bloom (2016: 5). That is, he defends more weight being given to reflectiveness than to empathy (mainly when it comes to decisions taken for society).

---

[147] In the sense they potentially origin many other characteristics perceived as socially beneficial.

[148] The use of reflectiveness as rationality is based on Stanovich (2011), in which it is argued this trait defines rationality. Also based on that we give preference to reflectiveness over other cognitive abilities in the study.

[149] Notice they regard the social planner's capacity and willingness to solve other people's problems.

[150] It must be noticed that Bloom (2016) focuses on emotional empathy, i.e. empathy as the ability to feel what someone else is feeling. As such the empathy measure which is most related to his analysis in this study is the Empathy Quotient (EQ), and the one which is least related to it is the Reading the Mind in the Eyes (RME), given that it focus on cognitive empathy, i.e. the ability to know what someone else is feeling or thinking.

Such controversy reinforces the importance of empathy (and reflectiveness) as the key psychological traits. Notice however, I am not aiming to resolve the question of whether empathy is a desirable trait in this chapter, but I do want to answer the question as to whether the possession of empathy makes a difference.

I also consider two other traits social planners are expected to have: time consistency when choosing for others – which basically refers to whether the different selves (in different points of time) of the decision-taker disagree about decisions taken on behalf of others – and maturity (i.e. being in full possession of important neural faculties needed to take decisions, which in general takes place about after the early twenties). Regarding the first trait, besides being socially desirable for allowing rational planning, a time consistent social planner (which by definition chooses for others) is often assumed in economics analysis,[151] which adds interest in looking at the EMU of those possessing it.

Regarding maturity, I refer basically to a body of literature showing adolescents' brains (individuals around the age of 12 to the early 20s) are still developing and because of that they are more likely to change their behaviour and choices after they grow older, besides being also more likely to engage in anti-social behaviour. Regarding the first aspect neuroimaging studies provide evidence that adolescence is hallmarked by goal flexibility. During this period, cognitive engagement depends more on "the social and motivational salience of a goal" (Crone and Dahl, 2012: 645). Such aspect coupled with the developmental nature of adolescence justifies the placing of less weight on the choices of such young people. Such evidence regularly provides support for courts to consider adolescents less responsible for their actions (Steinberg, 2013).

Regarding the second aspect of immaturity (being prone to engage in anti-social behaviour), brain models point to a gap in cognitive control and affective processes as one of the main cause of antisocial behaviour amongst young people (Steinberg, 2008; Somerville et al., 2010; Ernst and Fudge, 2009).[152] The basic idea is that sensation-seeking increases substantially during this period while impulse control just improves gradually with age, rendering the gap mentioned before (Harden and Tucker-Drob, 2011). Moreover, it is important to notice that although these findings have been absorbed and understood by psychologists they are still to penetrate the thinking of economists.

---

[151] It is implicit, for example, in the SWEM, in which utility is discounted exponentially and therefore time consistently.

[152] Other researchers have started to suggest less emphasis should be given to cognitive control, while more attention should be directed to how it articulates with affective processing and social development (Pfeifer and Allen, 2012).

Empathy is measured in three different ways [via the Empathy Quotient (EQ), based on Baron-Cohen and Wheelwright (2004), the Read the Mind in the Eyes test (RME), based on Baron-Cohen et al. (2001a) and the Autism Spectrum Quotient (AQ), based on Baron-Cohen et al. (2001b)]. Reflectiveness is measured in two different ways [via the CRT, based on Frederick (2005) and the Model Free/Model Based (MFMB) test, based on Doll et al. (2012) and Eppinger et al. (2013)]. I will provide more extensive details on the traits measurements later on in the chapter. All of them focus on different aspects of the factors to be gauged. By measuring the traits from different angles I thus get insight into aspects of empathy and reflectiveness that are most relevant to the determination of EMU.[153]

The data used in this study is obtained experimentally. I present respondents with a survey through which they reveal, in three different ways, a range within which their EMU value is contained. The ways to derive the EMU intervals correspond to the contexts in which the same EMU value is assumed by the SWEM. That is, I estimate the EMU as risk aversion, inequality aversion and as the EIS. In the same survey I measure the traits considered previously (empathy, reflectivity, time consistency and maturity). There are 9 relevant tasks in total: 3 measuring the EMU (as risk aversion, inequality aversion and the EIS), 3 measuring empathy, 2 measuring reflectiveness, and one identifying those choosing time consistently for others.[154]

In my study most tasks are hypothetical (in that they do not explicitly motivate individuals by using economic incentives).[155] This is in line with other experimental studies measuring empathy (e.g. Baron-Cohen and Wheelwright, 2004; Baron-Cohen et al., 2001; Baron-Cohen et al., 2001); reflectiveness (e.g. Frederick, 2005; Toplak et al., 2011); risk aversion (Weber et al., 2002; Hanoch et al., 2006), inequality aversion (e.g. Amiel et al., 1999; Carlsson et al., 2005; Pirttila and Uusitalo, 2008) and the EIS (e.g. Barsky et al., 1997; Atkinson et al., 2009).[156] In the task relative to choosing time-consistently for others it is very difficult to incentivize given the fact they choose for other people. The same happens for other tasks, as for example the one

[153] Notice to measure any of the traits one has to strict their definition, such that an individual may be considered highly empathetic (reflective) according to one measure and weakly so according another. The measuring of them in different ways allows us to understand better in which sense empathy (reflectiveness) relate to the EMU.
[154] There is also a task measuring Working Memory capacity in the survey, but it is not used in this particular study. More details on the tasks, including parts which are not used in this study are provided in section 4.3.
[155] In the MFMB task most of the participants had the possibility to get an actual prize (more details on that are given in section 4.3). Participants (mostly students) receive money or course credits to take part.
[156] Atkinson et al. (2009), one of the central references for this study measures inequality aversion, risk aversion and the EIS using hypothetical questions.

measuring inequality aversion. In the questionnaires measuring empathy it is likewise not possible to incentivize.

I use interval regression techniques throughout. I divide the sample between highly versus lowly reflective and high versus lowly empathetic participants according to the different test scores and then investigate for each group whether the EMU measured in a given context (say risk aversion) is significantly different from the EMU measured in another (say EIS)[157]. I do that by including dummies in the interval regression specifications in a way that allows us to test for variation across contexts. I repeat the same strategy to determine whether the other two traits (being time consistent when choosing for others and maturity) have any influence on the issue.[158]

I use three strategies to determine how the EMU estimates vary across the different traits (EMU estimate analysis). In the first I divide the participants according to the relevant traits in the same fashion as in the context analysis and estimate the EMU for the different groups using interval regressions. This time the regression specification consists of just an intercept, allowing to obtain a point EMU estimate per group.

In the second I run 6 interval regressions, each containing an index of empathy and reflectiveness, and three dummy variables: one indicating time-consistency when choosing for others, another indicating maturity and another indicating the context the EMU (as in all cases the dependent variable) was estimated in. After this process, I run another interval regression containing just the variables that were significant at least once in the previous regressions. The third is the same as the second, but without the context dummies. The goal is to observe how the EMU varies in response to the psychological traits only.

The results for the 'context analysis' show more empathetic people's EMU tends to be more context-sensitive. By contrast more reflective people's EMU (as measured by the MFMB[159] task) tend to be less context-sensitive. Time-consistent people's EMU tend to be more context-sensitive and more mature (with more than 24 years) people's EMU tend to be less context-sensitive. These results suggest that although the SWEM is a not a good description of any group

[157] Notice there are three combinations to test in this analysis: whether the EMU as risk aversion is equal to the EMU as (intra-temporal) inequality aversion, whether the EMU as risk aversion is equal to the EMU as the EIS and whether the EMU as (intra-temporal) inequality aversion is equal to the EMU as the EIS.

[158] With regards to these other two traits we compare participants being time consistent (for others) as opposed to those not being time consistent and those older than 24 as opposed to those less than or equal to 24 years old.

[159] The MFMB index basically measures how goal-directed, as opposed to habit-directed, one behaves. Since to be goal-directed requires reflectiveness, it can be considered as a reflectiveness index. The measure is discussed in more detail on section 4.3.

related to the traits examined (in no case the EMU is the same in all contexts), it describes better people who are highly reflective, mature, weakly empathetic and do not choose time-consistently for others. The findings confirm the suspicion that estimates of EMU differ depending on psychological traits and that for some sorts of respondents the differences across contexts are much more pronounced than others.

Moreover, the finding that traits are partially responsible for variation in the EMU context sensitivity may help to explain the different results found by Atkinson et al. (2009) and Groom and Maddison (2013) concerning the issue: Atkinson et al. (2009) used a convenience sample of students.

The results show basically that more empathetic and time consistent people tend to have higher EMU estimates, which suggests the ideal EMU value tends to be higher than estimated by works drawing samples from the general population or students (i.e. not 'sample framing') – if one is prepared to accept empathy as a socially desirable trait concerning decisions for society.

In subsection 2 I review the main methods concerning the estimation of the EMU, arguing that estimating it experimentally is the methodology that best serves the objectives of the present analysis. In the same section I review the literature concerned with estimating experimentally the EMU in the three contexts referred above. In subsection 3 I describe the experiment run in the study, giving details on how the EMU is estimated in the three different contexts, on the measuring of the aforementioned traits and on the empirical strategy to approach people's EMU context-sensitivity and the effects of desirable traits on participants' EMU values. In subsection 4 I present and discuss the results. Subsection 5 concludes.

4.2. Literature review

This section is organized as follows. In the first subsection I briefly review the main methods to estimate the EMU and give the reasons as to why I focus on works estimating it experimentally. In the second subsection I search for works estimating the parameter (using any method) and relating the estimates to the personality traits considered herein. Since no such work is found doing that I give more details on Atkinson et al. (2009). Although it does not relate the EMU estimates to the psychological traits considered herein, it is the only work found in the search estimating the EMU experimentally in the three contexts considered in the present study.[160]

[160] Besides, it is a central work to the present investigation due to being the first (to knowledge) to test experimentally the EMU parameter equality across the three contexts as implicit in the standard economics welfare theory.

In the third to fifth subsections I describe key aspects of studies experimentally estimating the EMU in each of the three contexts. What I mostly look for in the works retrieved is whether they relate the estimates to any of the traits considered in this study or whether they investigate EMU variation across contexts. In subsection 6 I comment on the review undertaken.

4.2.1. Main methods to derive the EMU

Methods to estimate the EMU can be divided in those based on revealed preferences and stated preferences.[161] The ones belonging to the second group are found in works measuring the EMU experimentally as an inequality aversion parameter (Amiel et al., 1999; Carlsson et al., 2005; Pirttila and Uusitalo, 2008); as a risk aversion parameter (Barsky et al., 1997; Eckel and Grossman, 2002; Holt and Larry, 2002) or as given by the EIS (Barsky, et al., 1997 and Atkinson, et al., 2009). Yet another kind of work estimating the EMU by stated preferences are those reporting experts' opinions about the appropriate value of the coefficient.[162]

In the first group we have works inferring the EMU from income tax schedules (Stern, 1977; Cowell and Gardiner, 1999; Evans, 2005; Groom and Maddison 2013), works estimating the EMU (as an intertemporal inequality aversion parameter) from lifetime consumption behaviour as revealed in the market (Blundell et al.,1994; Attanasio and Weber, 1989 and Groom and Maddison, 2013), works based on the Frisch's (1959) formula, which holds for additively separable utility function (wants independence) (Evans, 2004; Evans et al., 2005; Halicioglu and Karatas, 2013), and works estimating the parameter from insurance data (Szpiro, 1986).

Empirical works estimating the EMU based on revealed preferences data do not generally include information that would enable one to measure the psychological characteristics of the individual; such characteristics are usually measured experimentally via games or questionnaires. Of course, such coordination is possible, but far more difficult to undertake than to obtain both the personality traits and the EMU estimates experimentally.

The difficulty just mentioned regarding EMU estimates based on revealed preference values makes it quite unlikely that works using such an approach can perform the same kind of analysis undertaken herein. Therefore, I focus on works estimating the EMU experimentally[163] in the literature review.

---

[161] Notice I refer to empirical works only, since it is not possible to measure psychological traits theoretically (i.e. using models).
[162] Just Drupp et al. (2015) have used this methodology so far (to knowledge).
[163] With 'experimentally' I refer basically to Harrison and List's (2004) taxonomy of field experiments: conventional lab experiments, artefactual field experiments, framed field experiments and natural field experiments.

4.2.2. Works estimating the EMU

A search was executed on June the 2nd (2016) with the following search terms: TX "elasticity of marginal utility" and 49 works were retrieved. I further refined them by picking just published articles in English, thus ending up with 33 works.[164] Of those none analyses the normative significance of the parameter based on participants' psychologies or whether such psychologies cause the parameter to vary across the different contexts described above. The only investigation found estimating the EMU experimentally was Atkinson et al. (2009), which tests whether in general people have the same EMU when it is measured as risk aversion, inequality aversion and EIS. Given this work's centrality to the analysis I also made a search on Web of Science database for works citing it. In total 17 works were retrieved, but none carried out the kind of analyses of relevance to my research.

Atkinson et al. (2009) elicits the EMU by three different experiments, each one looking at the EMU from a different perspective. The first takes it as a risk averse parameter, the second as an (intra-temporal) inequality aversion parameter and the third as an inter-temporal inequality aversion parameter (i.e. as the EIS).[165] They use a "convenience sample" of more than 3,000 participants whose age median is 27 years. It indicates young people are overrepresented, given the median age in the UK is 40 (Office of National Statistics, 2016). Given what the psychological evidence on the immaturity of people under 24 presented before, this is a potentially very important aspect of Atkinson et al. (2009) study. They do not consider differences in people's personality traits in their study.

They also investigate factors causing variation in the EMU estimates. Between all the variables considered just the result regarding age is relevant to the present study. It shows age has a small and positive effect on the EMU measured as inequality aversion.

The median EMU measured as risk aversion is in the interval 3-5 (that is also the modal group); the median EMU measured as inequality aversion was in the interval 2-3 (the modal group is above 7.5); for the EMU implied in the EIS the midpoint of the median respondent was 8.8.

[164] Below I am going to search for estimates of EMU in different contexts where the parameter is usually known by a different name e.g. inequality aversion, risk aversion, the elasticity of intertemporal substitution.

[165] The methods used to estimate risk aversion and the EIS are based on Barsky et al. (1997), and the method used to estimate inequality aversion is based on Carlsson et al. (2005) and Johansson-Stennman et al. (2002).

It is worth noticing that in measuring the EMU as the EIS based in Barsky et al. (1997) they select consistent answers,[166] such that in a way they give more weight to responses possessing a socially desirable trait. Nonetheless, there is no analysis on how the values they found compare with those not possessing such a trait or any explicit link with normative significance.

Given the threefold interpretation regarding the EMU when the SWEM applies, I turn now to reviewing works estimating the EMU experimentally in each of them.[167]

4.2.3. EMU as inequality aversion

To search for works estimating the EMU as inequality aversion I used the following search terms (on EconLit, June 1st 2016): TX "inequality avers*" AND (TX measure* OR TX experiment*). 288 works were retrieved. I further select them by picking just published articles in English (such a procedure was adopted in all searches), ending up with 182 works. From those, 5 were found estimating the EMU experimentally as inequality aversion. Most works in the search are not experimental or do not use the curvature of a hypothesized utility function (i.e. the EMU) as the inequality aversion measure, which puts them out of the area of interest of the present research.

The first (in chronological order) work investigating the EMU as inequality aversion is Amiel et al. (1999). They use 'leaky bucket' questions[168] and traditional utilitarian welfare function in order to elicit the EMU. They do that for three different utility functions: an isoelastic, a constant absolute risk aversion, and another one based on the Gini measure of inequality.[169] Four different groups of students participated in the experiment (the first with 41 respondents, the second with 37, the third with 272 and the fourth with 56). The median EMU obtained from the participant groups for each welfare equation ranged from 0.095 to 0.227 and the parameter estimates derived from fitting the different welfare functions over all responses and individuals ranged from 0.05 to 0.248. Comparing to other estimates the ones obtained by Amiel et al. (1999) are remarkably low.

---

[166] Given we use a similar methodology to measure the EMU as the EIS, more details on the selection of consistent answers are given in section 4.3.

[167] Notice I do not include Atkinson et al. (2009) in the review to follow once it was already resented in the present section.

[168] Basically, a question asking how much loss per donation unit one is willing to accept in a transfer from a rich to a poor.

[169] See chapter 1 for more details on the iso-elastic utility function. The constant absolute risk aversion (CARA) utility (exponential utility) is given by $1 - e^{-\alpha y}$, where y is income/consumption and α is the CARA. The one based on the Gini is given by $1 + \frac{1}{N} - \frac{2}{\mu N^2}\sum_{i=1}^{N}(N + 1 - i)y_i$, where the number of incomes in a given income distribution is ranked as follows: $i = y_1, < y_2 < \cdots < y_N$, where y is income.

The second is Johansson-Stenman et al. (2002), which also estimates the EMU as an inequality aversion parameter, but now by asking participants to choose among different income distributions. They assume a utilitarian welfare function with isoelastic utility and present two uniform income distributions to the respondents, one being more unequal than the other. Comparing how one chooses allows the authors to infer the EMU.[170] Since participants choose from behind a veil of ignorance[171] in the experiment the EMU is interpreted as both inequality and risk aversion.[172]

They recruited students (374 of them), and the median inequality/risk parameter (EMU) was between 2 and 3. They relate the estimates[173] with some of the participants' characteristics, finding left-wing voters are more inequality averse and business students less inequality averse than other students.

The third is Carlsson et al. (2003). They investigate the inequality aversion parameter of Indian students taking decisions from behind a veil of ignorance in the same fashion of Johansson-Stenman et al. (2002). The difference in terms of methodology is basically that in this study the societies to be chosen by the interviewees present right-angled triangular distributions instead of uniform distributions.[174]

364 students were interviewed, but just 338 were considered once 26 gave answers which were inconsistent with the assumptions made. The median EMU value was about 3. They relate the estimates to some characteristics of the respondents using interval regression, finding students' household income is negatively related to the EMU, respondent's own income is positively related to the EMU, those whose parents are left-wing voters are more inequality averse and that economics, technology and social science students are less inequality averse than the rest.

The fourth is Carlsson et al. (2005). They assume a general utility function presenting a constant parameter of individual inequality aversion which can be interpreted as the curvature of the utility function (in the same way as the EMU). Given such preferences they derive the inequality

[170] Indifference between the two distributions implies an EMU number, such that the choice of one implies someone's EMU is above or below such value.
[171] It means after choosing a given distribution there is, hypothetically, an equal probability the participant will be in any point of the distribution he chose.
[172] I classify the work as measuring inequality aversion due to the influence its method has on other works' methods to measure inequality aversion.
[173] They use an estimation of the relative risk premium when the respondent is indifferent between the two societies as the dependent variable of a OLS regression to investigate the determinants of inequality (risk) aversion.
[174] These are claimed to be more compatible with the Indian case.

aversion parameter by presenting the participants different societal income distributions in which there is a trade-off between mean income and inequality (standard deviation).

To run the experiment, they relied on a sample of students (324 of them). The median value for the EMU is in the interval 0.09-0.22. They also relate the estimates to the participants' characteristics using interval regression, finding women and left-wingers are more inequality averse, while technology and business students are less inequality averse.

Finally Pirttila and Uusitalo (2008) gauge the EMU by two methods and compare them. The first is the one used by Amiel et al. (1999) and the second is the one used by Carlsson et al. (2005), explained above. The study has 3,000 participants from general Finnish population. The median EMU as measured by Amiel et al.'s (1999) methodology was below 0.5, and the median EMU as measured by Carlsson et al.'s (2005) methodology was larger than 3.

4.2.4. EMU as risk aversion

Due to the great number of works estimating the EMU as risk aversion,[175] I further specify the search in this section by seeking works not just estimating it experimentally, but also relating the estimates to the psychological traits considered in this research. The following search terms were used on EconLit (on the 28th of December, 2016): TX "risk avers*" AND (TX empath* OR TX "cognitive ability" OR TX "time consisten*" OR TX age). 147 journal articles in English were retrieved. From those 4 were found to be estimating the EMU as risk aversion and relating the estimates to one of the traits considered in the study. 2 related them to age and 2 related them to reflectiveness.[176] Many works measuring risk aversion experimentally did not do it with the EMU, and many others used risk aversion together with other traits not considered in the study (e.g. alcoholism, divorce, tendency to cheat, etc.).

The first work is Harrison et al. (2007), in which the EMU as risk aversion is estimated for a representative sample of the Danish population and related to various of its socio-demographic

[175] The search terms TX "risk avers*" AND (TX measure* OR experiment*), for example, retrieved 1,164 journal articles in English.

[176] To further check whether there were works relating the EMU as risk aversion to the traits considered I disentangled the main search in five smaller searches. In the first I used the terms TX "risk avers*" AND TX empath* and no works were retrieved. In the second I used the terms TX "risk avers*" AND TX "time consisten*" and 18 works were retrieved, but none related risk aversion to how time consistently one chooses for others. In the third I used the terms TX "risk avers*" AND TX "cognitive ability" and 10 works were retrieved, from which 2 fitted within the category concerned in the study. In the fourth I used the terms TX "risk avers*" AND TX age, and 119 works were retrieved, from which 2 fitted within the category concerned in the study. In the fifth I used the terms TX "risk avers*" AND TX maturity, and 38 works were retrieved, but none of them referred to maturity as a psychological trait, but as a financial term (e.g. debt maturity structure). The fact the four works kept were the same found in the main search substantiates the accuracy of the latter.

characteristics including age. The method used to elicit the EMU is a modification of Holt and Larry's (2002) Multiple Price List (MPL)[177] (assuming participants' preferences are described by an iso-elastic utility function). The first modification is that participants answer four times the MPL with a different set of 4 prizes (money) each time; it allows them to test whether risk aversion varies with income for the sample. The second involves refining the estimates by asking participants to choose among lotteries lying between the preferred and not-preferred options implied in the option switches. The third intends to tackle the framing effect[178] by including two asymmetric frames in the experiment: one with probabilities of 0.3 (and 0.7), 0.5, 0.7, 0.8, 0.9 and 1 for the lotteries[179] (skewHI treatment) and other with probabilities of 0.1, 0.2, 0.3, 0.5, 0.7, and 1 (skewLO)[180]. The asymmetric frames allow to test for framing effects if people tend to switch in the middle the skewLO (skweHI) would skew the answers to be low (high) probability ones.

253 people across Denmark with ages between 19 to 75 years old participated. The mean CRRA (which is the same as the EMU) coefficient is equal to 0.67 (for the skewedLO and skewedHI the mean CRRA coefficient is 0.43 and 0.91 respectively).[181] The authors also conclude the coefficient does not vary significantly over income, and thus the iso-elastic utility assumption is appropriate for their investigation. Finally, they find support for an inverse correlation between age and risk aversion.

The second work is Hryshko et al. (2011). They look at answers given to the Panel Study of Income Dynamics (PSID) (US) (a general population survey) which can be used as data coming from an experiment of the same kind of the one done by Barsky et al. (1997) to elicit EMU.[182] About 5000 people participate. The questions ask basically whether the interviewee would accept to change his actual job for a job in which he has a 50% chance to get paid the double of what he receives currently and a 50% chance to get a wage cut of $(1-\lambda)\%$ (which varies).[183]

[177] Holt and Larry's (2002) MPL consists basically of 10 choices between two lotteries, say A and B. The difference between the 2 prizes (which keep constant through the 10 questions in both lotteries) at stake in A is smaller than in B and the probabilities to gain a given prize is the same in A and B for a given question. Given participants' underlying utility function an EMU interval can be obtained using the switching point from A to B.

[178] More specifically, the tendency to switch option in the middle of the table.

[179] As opposed to the probabilities of 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1 of the usual symmetric frame.

[180] Notice the asymmetric frames have 6 questions instead of 10.

[181] They use interval regressions to obtain point estimates.

[182] Therefore they do not apply the experiment themselves, but use the data coming from one (though the "experimenter" was not aware of the experiment itself).

[183] The cut regarding the first question is 33.33%; if the respondent accepts the new job he answers another question in which $1-\lambda=50\%$, if he accepts again $1-\lambda=75\%$ and then the sequence stops; If he does not accept $1-\lambda=50\%$ it also stops; if he does not accept $1-\lambda=33.33\%$ another question

According to the expected utility theory, if the participant accepts the new job, then $\frac{1}{2}U(2c) + \frac{1}{2}U(\lambda c) \geq U(c)$, where $U(c)$ is the utility of consumption c. Assuming $U(c)$ is iso-elastic, if the respondent is indifferent then $\lambda = (2 - 2^{1-\eta})^{1/(1-\eta)}$. With such structure, it is possible to derive EMU (η) intervals from the answers to the questions described.[184]

The median EMU interval was 2 to 3.76 (with midpoint of 2.88), and the EMU interval mode was above 7.53. The authors also relate the EMU to some of the participants characteristics, concluding older individuals are more risk averse.

Now I present works relating the EMU as risk aversion to reflectiveness (included in the broader area of cognitive ability).

The first is Taylor (2013). He estimates risk aversion using the Holt and Larry (2002) MPL referred above in two ways: just hypothetically and with real incentives, and uses maximum likelihood methods to investigate, among other things, the influence of reflectiveness, as measured by the Cognitive Reflectiveness Test (CRT),[185] on the EMU estimates.

He finds CRT scores relate to the level of risk aversion in the hypothetical design, such that more reflective participants are significantly less risk averse in the hypothetical setting.[186] Considering such influences together with the also-found influence of gender[187] on the EMU estimates, and knowing the CRT varies from 0 to 3, we can determine that the EMU varies from 0.237 to 1.1 according to the econometric specifications.

Andersson et al. (2016) measure risk aversion using two different MPLs. The basic difference is that in the first (MPL 1) the switching point indicating risk neutrality is the third option (located relatively "high up"), while in MPL 2 such point is located relatively "low down" (is the sixth option). They also measure reflectiveness using the CRT.

Using interval regressions, they find the CRRA parameter for subjects with average cognitive ability score (measured by a standard intelligence test called "IST 2000 R") is 0.34 for MPL 1 and 0.01 for MPL 2. The relation between CRT scores and the EMU is ambiguous (like the relation

---

is asked with $1 - \lambda = 20\%$; if he accepts the sequence stops, if he does not then he is presented with a question in which $1 - \lambda = 10\%$ and then the sequence stops. Thus there are 6 possible groups regarding the questions sequence: yes, yes, yes (those who accept three times); yes, yes, no; yes, no; no, yes; no, no, yes and no, no, no.

[184] Those who (don't) accept the new job for a given question will have an EMU smaller (higher) than the one given by $\lambda = (2 - 2^{1-\eta})^{1/(1-\eta)}$.

[185] The details regarding this reflectiveness measure are given on section 4.3.

[186] There is no influence of reflectiveness on the EMU in the real incentive setting.

[187] Women are found to be significantly more risk averse than man.

between the IST and the EMU); for the MPL 1 they find a negative relation and for the MPL 2 they find a positive relation. Such results suggest cognitive ability (including reflectivity) is not actually related to risk aversion, but to noise linked to the measuring of it, generating spurious correlation between risk aversion and reflectiveness (cognitive ability). To be more precise, they argue the fact that high reflectiveness is associated with less noisy estimates of risk aversion generates correlation between the variables.

4.2.5. EMU as the EIS

To find experimental works estimating the EMU as intertemporal inequality aversion (measured by the EIS) I used the following search terms on EconLit (on December the 28th, 2016): TX "elasticity of intertemporal substitution" AND (TX measure* OR TX experiment*). 15 journal articles in English were retrieved, and just one was found estimating the EMU as EIS experimentally.

One such work was Barsky et al. (1997). They assume participants (from the general public) behave in accordance to the Ramsey rule, then ask them to choose among different spending plans implying a consumption growth rate for a specified rate of return. By considering three responses they can derive a EMU range per participant.[188] They have 198 observations, but once they exclude answers inconsistent with utility maximization and also uninformative responses they end up with only 116 observations. Therefore, as in other works performing this type of selection, the results reflect the choice of the consistent participants, and are hence linked to a psychological trait, albeit in a very indirect way. The midpoint of the modal EMU interval implied in the EIS estimated by them was 8.7.

4.2.6. Comments on literature review on works eliciting the EMU experimentally

The first clear conclusion to take from the literature review is that the work in this chapter is quite unprecedented. No work was found relating the EMU to psychological traits in order to get insight into its normative significance, or examining how the EMU context-sensitivity relates to psychological traits in order to get insight into the SWEM normative significance. Moreover, no work was found trying to solve the controversy implicit in Atkinson et al. (2009) and Groom and Maddison (2013), or selecting a sample based upon psychological traits [rather than expertise as Drupp et al. (2015) do] to define a value for the EMU.

[188] Given the Ramsey rule $r = \delta + \eta g$, where r represents a return rate, δ the utility discount in the SWEM and g growth, and that in each answer the participant chooses a preferred rate of growth given a return rate, for 2 of such answers it is possible to derive a value for η and δ.

Although some works relate the estimated EMU to participants' characteristics, just the ones estimating risk aversion were found relating them to traits that are considered in my study (i.e. maturity and reflectiveness).

The works estimating the EMU as risk aversion that related it to age got different conclusions on how age affects risk aversion. Harrison et al. (2007) found age is inversely correlated with risk aversion, while Hryshko et al. (2011) found the opposite (both use samples taken from the general public). The first result is surprising in view of studies in psychology relating age and risk-taking behaviour (Steinberg, 2008; Steinberg and Scott, 2003). It suggests the relation between the EMU as risk aversion and maturity may not be so clear as between maturity and risk-taking behaviour as expressed in more general measurement forms (used in psychology studies). Atkinson et al. (2009) find no correlation between age and risk aversion further adding to the ambiguity over the role of age.

All works estimating the EMU as inequality aversion, except Pirttila and Uusitalo (2008), worked with samples consisting of young people (students). The fact that Pirttila and Uusitalo (2008) got a larger inequality aversion estimate than Carlsson et al. (2005) and Atkinson et al. (2009) (all using similar methods) is also consistent with Atkinson et al. (2009) finding that in the inequality aversion context the EMU is positively related to age.

The only work found estimating the EMU in the EIS context, apart from Atkinson et al. (2009), is Barsky et al. (1997), which samples from the general population, such that the fact it obtains an EIS similar to Atkinsons et al. (2009) (which samples among young people), is once again consistent with Atkinson et al. (2009) not finding any correlation between age and the EIS.

It is important to clarify that although the present study does not investigate age, but maturity, these are obviously related variables (maturity is a binary variable identifying mature and non-mature people) – at the same time, it is worthy to keep in mind they are not the same thing.

Two works were found relating the EMU as risk aversion to reflectivity (as measured by the CRT, only). Taylor (2013) finds a negative correlation between the EMU and CRT scores, while Andersson et al. (2016) find both positive and negative correlation between reflectiveness and the EMU, depending on how they measure risk aversion. Therefore, the relationship between risk aversion and the CRT is unclear.

In this chapter besides measuring reflectiveness using the CRT, we also measure it using the MFMB task. As discussed in more detail later on, the MFMB has not just the advantage of measuring reflectivity from a different perspective, but also of allowing for a more accurate measurement of the trait.

No work was found relating empathy and the EMU directly, but two (Johansson-Stenman et al., 2002; Carlsson et al., 2005) related the parameter measured as inequality aversion to political preferences. Given there is evidence suggesting left-wingers are more empathetic (Dodd et al., 2011; Mondak et al., 2010), we expect to find empathetic people are inequality averse.

Therefore, according to the works retrieved in the literature review my study is the first one relating the EMU (in each context) to how time-consistently one chooses for other people, to maturity (as such, not as age) and to empathy. It is also the first time one relates the EMU as inequality aversion and as the EIS to reflectiveness. Although these are not main contributions, they are useful to expand the understanding of how key traits relate to risk and inequality aversion and the EIS.

Many of the works retrieved (Carlsson et al., 2003; Carlsson et al., 2005; Harisson et al., 2007; Andersson et al., 2016) used interval regression in their analysis to relate traits to the EMU estimates. Harrison et al. (2007) also used such techniques to obtain EMU point estimates. It suggests that it is an established technique in the literature to perform the kind of analysis intended here. Like Harisson et al. (2007), we use it not just to relate the estimates to the traits, but also to obtains point estimates.

Finally, table 4.1 presents all works cited in the literature review with some key characteristics of each.

Table 4.1. Literature review

Study	EMU	Context	Traits considered[189]
Atkinson et al. (2009) 1	2-3	IA	Age
Amiel et al. (1999)	0.095-0.248	IA	None
Johansson -Stenman et al. (2003)	2-3	IA	None
Carlsson et al. (2003)	3	IA	None
Carlsson et al. (2005)	0.09-0.22	IA	None
Pirttila and Uusitalo (2008) 1	<0.5	IA	None
Pirttila and Uusitalo (2008) 2	>3	IA	None
Atkinson et al. (2009) 2	3-5	RA	Age
Harrison et al. (2007)	0.43-0.91	RA	Age
Hryshko et al. (2011)	2-3.76	RA	Age
Taylor (2013)	0.23-1.1	RA	CRT
Andersson et al. (2016)	0.01-0.34	RA	CRT
Atkinson et al. (2009) 3	8.8	EIS	Age
Barsky et al. (1997)	8.7	EIS	None

[189] By traits we mean the traits considered in this study.

Source: see text

4.3. Methodology

4.3.1. Brief experiment description

In this brief experiment description I address the participants, the apparatus used, the experimental design and the general procedure.

4.3.1.1. Participants

93 adults (age between 18 and 33 years, mean = 21 years, s.d. = 3.88 years) of whom 73 (78%) were females[190] and 17 (18%) were over 24[191] participated in exchange for £7 or course credits. They were recruited through the University of Birmingham's Psychology Research Participation Scheme and all reported good physical and mental health. Informed consent was obtained prior to testing.

4.3.1.2. Apparatus and experimental design

A computer [64-bit Operating System, 1920 by 1080 pixels and 80Hz monitor (ProLite B2483HS)] was used to implement 8 of the 10 tasks. All tests were programmed using Psychopy as the programming platform (written in Python). The computer controlled and timed the displays used in each task and recorded responses using millisecond accuracy via a keyboard and/or mouse. The remaining 2 tasks were applied with pen and paper. The participants were tested individually, in a quiet and well-lit room.

The participants are randomly divided in two groups: half were given the Risky Prospects (RP) questionnaire (discussed below) as the first task and the Distributive Alternatives (DA) questionnaire (also discussed below) as the eighth task; the other half were given the DA questionnaire first and the RP questionnaire as the eighth task.[192] In all other respects participants in both groups are treated equally.[193] The tasks are hypothetical.[194]

[190] For three participants, the gender response is not available.
[191] Among these 17, 8 (47%) were female. Among the 76 participants under 24, 62 (82%) were female.
[192] For the first 10 participants half responded just to the DA version and the other half just to the RP version.
[193] By applying the DA and RP alternately we intended to test for priming effect in the first 15 questions of the DA and RP questionnaires. However, given these questions were not used in the present research we do not pursue the priming issue anymore.
[194] With the exception of the MFMB task.

4.3.1.3. General procedure

The experimental session comprises 10 tasks. Below I give a general idea of how the experiment looks like as a whole. Details and further explanations are given in the sections to come.

The tests completed (in the order applied) are: 1) the DA or the RP questionnaires – both have 16 questions, and in each participants choose one among alternatives presented to them to establish their EMU; 2) the MFMB task – a game consisting basically of choosing between two options, with probability to obtain a hypothetical reward thereby establishing the participant's degree of reflectivity;[195] 3) the RME task – in which participants choose among four words the one best describing eye expression photos; the task consisted of 36 trials and establishes the respondent's empathy; 4) the Working Memory task (WM) – in which participants are told to remember a sequence of numbers seen on the screen and type them (in the reverse and regular order);[196] 5) the CRT – in which participants are asked to answer to three seemingly simple questions; there are no alternatives, they had to type their (numerical) answers. This test also measures respondents' reflectivity; 6) the EQ questionnaire – in which participants are asked to answer 60 questions using a 4-points agreement scale (definitely agree, slightly agree, slightly disagree, definitely disagree) thereby providing an alternative measure of empathy; 7) the AQ, which is equal to the sixth task in terms of procedure; the questions are obviously different and there are 50 of them instead of 60. This provides an alternative measure of empathy; 8) the DA or the RP questionnaire (see the first task);[197] 9) the Time Consistency (when choosing for other people) task (TC) – in which participants choose on behalf of other people they know nothing about among alternatives having to do with receiving a given amount of money at a given point in time or another larger amount of money at a more distant point in time; 10) the EIS task – in which participants are asked to choose, again in behalf of other people they know nothing about, among different retirement plans thereby revealing a third estimate of EMU.[198] Respondents typically completed these tasks in less than 90 minutes.

---

[195] Depending on the amount of hypothetical rewards obtained the participant could get a real reward. More details are provided below.
[196] The data collected in this task (WM) was not used in this study. Nonetheless we describe the task given it was part of the experiment.
[197] If the participant answered the RP questionnaire first task he would answer the DA questionnaire in the eighth and vice-versa.
[198] Notice the first 10 individuals were not submitted to tasks 8, 9 and 10 and subjects 11 to 32 answered to a slightly different task 10 due to improvements undertaken along the experiment implementation. For participants 11 to 32 the questionnaire had 4 cards, each containing from 3 to 5 plans. The difference in card number is regarding a test to verify whether participants are consistent with the assumptions made. In the beginning, just one card was used to such end; after participant 32 two cards were dedicated to it. The main reason for the change however is not related to the accuracy of the results, but to calculation convenience.

4.3.2. Tasks and measures

In this section I now explain in detail the strategy to estimate the EMU (in each context) and to measure the traits through the tasks referred above.

4.3.2.1. The RP and the DA questionnaires and the elicitation of the EMU as risk aversion [16th question of the RP (task 1 or 8)]

The RP or DA (first or eighth task) is based on Tversky and Kahneman (1979); Michaelson (2015); Eckel and Grossman (2002); and Atkinson et al. (2009). These tasks consist of responding to a questionnaire containing 16 questions. Just the last question measures risk aversion for those answering the RP whereas for those answering the DA it measures inequality aversion. Here I focus on the last question of the RP to explain how risk aversion is estimated. The next section deals with the last question of the DA, used to estimate the EMU as inequality aversion. The first fifteen questions involve choosing between two risky prospects (or distributive alternatives) and one (measuring the EMU) involves choosing between 6 risky prospects (or distributive alternatives).[199]

The task is self-paced and unspeeded; the questions and choices are presented simultaneously and remain on the screen until a response is given. The first 15 questions are presented in a random order for each participant (the 16th is always the last question). [200]

The 16th question of the RP is based on Eckel and Grossman (2002) whose work is not reviewed in the literature review because they do not relate their estimates to any of the traits considered in this study. We chose their method given its simplicity, efficiency (given the assumptions described below) and because it can be applied rapidly (given that the participants have many tasks to complete in the experiment). This question is used to measure the EMU as risk aversion. Assuming respondents act in accordance to the Expected Utility Theory (EUT) and their preference is represented by an iso-elastic (CRRA) utility function we can estimate the EMU using the following question, which consists of choosing among 6 lotteries.

Indicate among the six following gambles the one you would most like to participate in:

[199] As noticed before the first 15 questions of both the DA and RP were not used in this study, but were included in the description for being part of the experiment.

[200] Half participants were assigned to each of these random orders. The alternatives presented per question in the first 15 questions were also randomized in two different ways to avoid participants answering randomly to obtain very high (low) scores. Given these questions are not used in the study I do not give a detailed explanation regarding them. The whole RP and DA questionnaires can be visualized in Appendix 4.A.

a) 28 pounds for sure;

b) A 50% chance of 24 pounds and a 50% chance of 36 pounds;

c) A 50% chance of 20 pounds and a 50% chance of 44 pounds;

d) A 50% chance of 16 pounds and a 50% chance of 52 pounds;

e) A 50% chance of 12 pounds and a 50% chance of 60 pounds;

f) A 50% chance of 2 pounds and a 50% chance of 70 pounds.

To understand how it measures the EMU as inequality aversion, consider the case where someone chooses alternative b. If he is an expected CRRA utility maximizer we can represent the utility concerning the alternative as

$$U(b) = 0.5\left(\frac{24^{(1-\eta)}-1}{1-\eta}\right) + 0.5\left(\frac{36^{(1-\eta)}-1}{1-\eta}\right). \tag{4.1}$$

For him to choose it, it must be larger then $U(i)$, where $i = a, c, d, e, f$.[201] To obtain the EMU (η) one has to calibrate the parameter with numbers satisfying the inequality referred above (i.e. $U(b) > U(i), \forall\ i$). In the case of the example, it corresponds to numbers in the interval 1.16 - 3.46.

Notice in this case the respondent is assumed to be taking only his own utility function into consideration, such that the EMU obtained represents his own risk aversion. It corresponds to the SWEM with one period, one person and two states of nature.[202]

4.3.2.2. The elicitation of the EMU as (intra-temporal) inequality aversion [16th question of the DA (task 1 or 8)]

The last question of the DA questionnaire, which measures inequality aversion, is based on Atkinson et al. (2009) (which draws from Johansson-Stenman et al., 2002 and Carlsson et al., 2005). It goes as follows:

Imagine now that MANY OTHER PEOPLE WERE INCLUDED IN THE GROUP[203] and that they will be UNIFORMLY DISTRIBUTED BETWEEN A MAXIMUM AND A MINIMUM INCOME AMOUNT YOU HAVE TO CHOOSE. So, for example, if you choose a maximum of 100 and a minimum of 90

[201] This is, the other 5 alternatives. Notice the only difference between $U(b)$ and $U(i)$ is that the values entering the equation are different.
[202] See the introduction (first footnote).
[203] It refers to a group of people who will hypothetically be affected by the participant's choices (remember there are 15 questions to be answered before the one detailed here). The instructions for the task (available in Appendix 4.A) contains more details.

pounds, 10% of the people will receive 91 pounds, other 10% will receive 92, and so on up to 100 pounds. If you choose a maximum of 110 and a minimum of 85 pounds, 4% of the people will get 86 pounds, other 4% will get 87 pounds, and so on up to 110 pounds. Indicate among the 6 following distributions the one you would like the people in the group to participate in:

a) People will be uniformly distributed between 308 and 319 pounds;

b) People will be uniformly distributed between 264 and 396 pounds;

c) People will be uniformly distributed between 220 and 484 pounds;

d) People will be uniformly distributed between 176 and 572 pounds;

e) People will be uniformly distributed between 132 and 660 pounds;

f) People will be uniformly distributed between 22 and 770 pounds.

Assuming respondents have a utilitarian welfare function with an isoelastic social utility function as the criterion function by which they choose among different income distributions, a uniform income distribution would yield the following choice criterion function:

$$W = \int_{min}^{max} (\frac{y^{1-\eta}}{1-\eta}) \left(\frac{1}{max-min}\right) dy = \left(\frac{1}{(1-\eta)(2-\eta)}\right) \left(\frac{max^{2-\eta}-min^{2-\eta}}{max-min}\right), \quad (4.2)$$

where max and min are the maximum and the minimum income value in the concerned distribution, respectively. If one is indifferent between two alternatives presented in the question presented above (e.g. alternatives A and B of DA's question 16) we have:

$$\left(\frac{319^{2-\eta}-308^{2-\eta}}{319-318}\right) = \left(\frac{396^{2-\eta}-264^{2-\eta}}{396-264}\right). \quad (4.3)$$

From the above equality an EMU (η) range can be obtained, since the parameter satisfying the equality give the maximum EMU for those choosing the more unequal distribution (B in the example) and the minimum EMU for those choosing the more equal distribution.

In this case the EMU is not assumed to be a parameter from the utility function of the respondent, but from his social welfare function (SWF), i.e. from the utility of the representative agent he is choosing for. It is normative in the sense that it is the inequality aversion level the respondent thinks of as appropriate. The model assumed corresponds to the SWEM with one period and many (homogeneous) individuals (i.e. one representative agent). Also, notice that for the representative agent there are many possible states of nature, given he can be assigned to any point of the income distribution with the same probability.

4.3.2.3. Reflectiveness (tasks 2 and 5)

In this section I explain how I extract reflectiveness measures from tasks 2 and 5 (the MFMB and the CRT respectively) of the experiment.

The MFMB (task 2) (based on Doll et al., 2012; Eppinger et al., 2013). In this game the participants are presented with the following context: there are two planets (each with a different colour, say pink and golden) and two spaceships (A and B), one of which leads to one of the two planets (e.g. the pink planet) on 70% of trips (common transition) and to the other planet on 30% of trips (rare transition). The other spaceship proceeds in the same way but inverting the planets (i.e. it would take one to the golden planet 70% of the time and to the pink planet 30% of the time). Each planet yields a gem or a rock with a changing-over-time (independent Gaussian random walks between p = 0.25 to 0.75, s.d. = 0.025) probability. The player's objective is to maximize the number of gems they get by choosing between the spaceships in 200 trials. Half of participants in each questionnaire group are submitted to treatment 1, where spaceship A leads to the pink planet 70% of the time and the other half is submitted to treatment 2 where spaceship B leads to the pink planet 70% of the time.

The number yielded in the MFMB task indicates how Model Based (MB) a respondent is and comes from the frequency the respondent acted as a typical Model Based (MB) subject (i.e. fully considering the game context to obtain gems, as I discuss below) minus the frequency with which they acted as the typical Model Free (MF) subject (acting fully based on association between actions and effects, as I also discuss below).[204]

The typical MF subject chooses the same spaceship he did previously if it yielded him a gem and switches spaceship if the previous yielded him a stone. Notice that in this case there is no accounting for the probabilities referred in the game instructions. On the other hand, the typical MB subject takes into consideration the probabilities associated with a spaceship leading to one or other planet to obtain the maximum amount of gems.

So if a spaceship goes to its "common planet" (the one it reaches with 70% chance) and it yields a gem (a common reward event) the Typical MB subject (TMB) acts in the same way as the Typical MF (TMF) subject and chooses the same spaceship again (because it suggests the common planet is likely to yield gems); if the spaceship instead takes to its "rare planet" (the one it reaches with a 30% chance) and it yields a gem (a rare reward event) the TMB switches

[204] The frequencies referred are regarding the number of times the participants choose the same spaceship they chose in the previous trial (i.e. the probability to stay).

spaceship (because it suggests the planet most likely to give gems is the other one). If the spaceship chosen goes to its common planet and gives a stone (common non-reward event) the TMB also chooses the same way as the TMF subject and switches spaceship; and finally if the spaceship chosen goes to the rare planet which gives a stone (a rare non-reward event) the TMB (differently from the TMF) chooses the same spaceship again (because it suggests the planet to which the other spaceship goes more often is more likely to give a stone).

To sum up, the probability to be MF is calculated by the probability to stay (i.e. choose the same spaceship as the one chosen previously) when a common event (when the spaceship leads to the planet it does 70% of the time) is rewarded plus the probability to stay when a rare event is rewarded minus the probability to stay when a common event is not rewarded minus the probability to stay when a rare event is not rewarded. The probability to be MB is given by the probability to stay when a common event is rewarded plus the probability to stay when a rare event is not rewarded minus the probability to stay when a rare event is rewarded minus the probability to stay when a common event is not rewarded.

The subjects are given both written and oral instructions before starting, including a clear statement about the probabilities linking spaceships to planets. A sheet summarizing the task is kept with them during the selection process.[205] At the beginning of each trial the subjects view both spaceships and have to choose between them by using a mouse to position a cursor over the spaceship right-clicking on the chosen one. After choosing they see an image of the planet the spaceship has taken them to (for 2 seconds) and then an image of the yield they got (for 1 second). The task is self-paced and unspeeded.

This task is used to index participants' reflectiveness; the basic idea is that more reflective people will tend to take into consideration the current level of probability of finding a gem rather than just using the known, stable probability of which spaceship goes to which planet. A tendency to use information about the reward received on the just previous trial indicates using a cognitive model of the two-chain event (goal-directed behaviour) and is therefore indicative of reflectance. In contrast, less reflective participants will tend to choose the spaceship based on the stable probabilities predicting the outcome of the first choice (habit-directed behaviour) without regard to eventual consequences (actual reward outcome). Such behavioural

[205] Starting from participant 24 the participants were presented with the score (number of gems obtained) ranking of the previous participants. Those scoring more than the third, second and first in the ranking earned prizes (a small chocolate bar for those getting the third place, a big chocolate bar for those getting the second place and both bars for those getting the first place).

tendencies show low reflectance. The instructions given to the respondents can be found in appendix 4.B.

Model-based behaviour is related to reflectiveness in that those acting in such way are more likely to supress habitual responses in order to respond in a manner more compatible with their goals. In the mind dual-system approach such operation corresponds to being reflective, i.e. stopping autonomous processing and simulate in order to figure out the most appropriate manner to achieve one's goals.

<u>The CRT</u> (task 5) (Cognitive reflection test) (Frederick, 2005). In this task participants are asked to respond to three apparently simple questions. Each one of them is kept on the screen until an answer is typed and entered. The task is self-paced and unspeeded.

The three questions are tailored in a way that causes a wrong answer to come most readily to mind just after the reading of it. The basic idea is that a low reflective subject would indeed give such a (wrong) answer, whereas a highly reflective one would resist the impulse of rashly giving that answer, being therefore more likely to give the correct answer. The questions and the answers explained can be found in appendix 4.B.

This test has been widely used as the main measure of reflectiveness. The capacity to refrain from giving the intuitive but wrong responses corresponds to control over autonomous processing. It shows greater reflectiveness, given it is linked to the ability to supervise and interrupt intuitive thinking when appropriate. The reflectiveness index yielded by the task consists of the number of correct answers given by the participant (0 correct responses indicates lowest reflectiveness and 3 correct responses indicate highest reflectiveness).

However, some works (Welsh et al., n.d. and Hoppe and Kusterer, 2011) argue the CRT has predictive power just with regards to biases that arise from problems having a calculable correct response, whose solution requires analytical skills. They thus suggest the CRT has serious limitations to measure reflectiveness in general. Moreover, the vast majority of respondents give no correct responses, which decreases the resolution of the test. Because of these shortcomings, we chose to measure reflectiveness in the alternative way presented above (the MFMB) which is the preferred approach.

4.3.2.4. Empathy (tasks 3, 6 and 7)

In this section I explain how I extract empathy measures from tasks 3, 6 and 7 (the RME, the EQ and the AQ, respectively).

We use three different tasks to look at empathy. The reason is because it is a broad concept and as such it can be approached in different ways. We want to check whether those alternative approaches change the results. We then look at empathy as related to 'theory of mind' or cognitive empathy (i.e. being able to recognise others' feelings and intentions, with the RME task), as related to autism (with AQ questionnaire) and as related to emotional empathy (feeling what someone else feels), which is the most common sense given to the word (with the EQ questionnaire).

The RME (task 3) (based on Baron-Cohen et al., 2001). In this task for each of the 36 trials subjects are presented with a photograph of the eye region of different actors expressing different emotions or thoughts. Together with the picture, which keeps being visualized until an answer is given, there are four alternatives describing the eye expression, one of which is correct. The response is given by typing a keyboard key corresponding to the description chosen (the subjects are given a list of word definitions in case they are not familiar with some of the words describing the thoughts or emotions concerned). The task is self-paced and unspeeded.

The test is originally described to be an 'advanced theory of mind test', where 'theory of mind' stands for "the ability to attribute mental states to oneself or another person" (Premack and Woodruff, 1978 as cited in Baron-Cohen et al., 2001). The concept is also referred by other names and overlaps with empathy (Baron-Cohen et al., 2001). In this study, it is one of the ways used to measure empathy. The instructions for the task and an example of a trial can be seen in appendix 4.B.

The RME contains 36 questions, each of which can be right or wrong. The scale formed by counting the correct answers allows to measure empathy as revealed by the capacity to "tune in" to other people's mental states. It goes from 0 to 36 correct answers, where 0 means low empathy and 36 means high empathy.

The EQ (task 6) (based on Baron-Cohen and Wheelwright, 2004). In this task subjects are asked to respond to 60 questions using a four-points scale going from 'strongly agree' to 'strongly disagree'. The question and the scale appear simultaneously on the screen and remain in view until an answer is given (by pressing a key corresponding to the option chosen). The task is self-paced and unspeeded.

As the name suggests the questionnaire contains questions tailored to measure subjects' degree of empathy. The questionnaire can be found on appendix 4.B.[206] The minimum score is 0 and

[206] The questions were given in a fixed order for all subjects.

the maximum 80, where 0 represents the lowest level of empathy and 80 represents the highest level of empathy (the scoring process is shown in appendix 4.B).

The AQ (task 7) (Baron-Cohen et al., 2001). This task is virtually identical to the previous one in structure. The only difference is that the questions are different and there are 50 instead of 60 of them.

The questionnaire is used to determine the autism degree of the subjects, and the scale goes from 0 to 50, where 0 indicates low level of autism (equivalent to a high degree of empathy) and 50 indicates high level of autism (related to a low degree of empathy). The scoring process is shown in appendix 4.B.

In the same way as with the RME, although the AQ is not intended to measure empathy directly, it allows us to look at it from a different standpoint, given the strong relation existing between autism and empathy.

4.3.2.5. Time consistency when choosing for others (task 9)

In this section I describe the strategy to identify those choosing more time-consistently for others.

The TC (task 9) (based on Chark et al., 2015). In this task the participants are asked to choose 40 times for someone they have never and will never meet between an amount of money to be received on a specific day and another larger amount to be received on a later day. There are 4 sets of 10 questions. The first set of choices is between a sum of money received 'today' and a larger sum received '7 days from now' (the amounts of money received vary across the questions within and between each set); the second set of choices is between a sum received '2 days from now' and a larger sum received 9 days from now'; the third set of choices is between '31 days from now' and '38 days from now'; the final set of choices is between '301 from now' and '308 days from now'. Participants receive a sheet containing the instructions and the questions and respond by ticking with a pen the preferred answers. The task is self-paced and unspeeded.

The basic idea is that a time-consistent (exponential discounter) individual would keep their inter-temporal preference the same in the different questions set, e.g. if they prefer £100 today rather than £125 in '7 days from now' they would also prefer £100 in '301 days from now' than £125 in '308 days from now'. The questionnaire can be found in Appendix 4.A. It is important to keep in mind that this task is regarding choosing time consistently for others, not for themselves.

Intuitively speaking the definition of time-consistency relies on the idea that when faced with options between receiving a sum of money in a specific date and another larger sum on another more distant date the time-consistent agent would base his choice on the distance between the two dates, not on the distance between the dates and the present. In other words, the time-consistent agent is not present-biased, i.e. does not give more weight to sums to be received closer to the present.[207]

Mathematically time-consistency is modelled as exponential discounting, which can be represented as

$$d = {}^{1}\!/_{(1+r)^{t}}, \tag{4.4}$$

where d is the discount factor multiplying the amount to be discounted, r is the discount rate and t is the period the sum is to be received.

The task allows me to distinguish between hyperbolic discounters[208] and exponential discounters. Also by including trade-offs between values received in the present and in 7 days from the present (which is not in Chark et al., 2015) it allows me to detect quasi-hyperbolic discounters[209] (the questionnaire can be seen in appendix 4.A).[210]

The ones choosing in the same way in the four scenarios are regarded as choosing (time) consistently for others.

4.3.2.6. The elicitation of the EMU as the EIS (task 10)

The EIS test (Task 10) (Barsky et al., 1997 and Atkinson et al., 2009). In this task the participants are asked to choose, for a group of people they do not know, among different retirement plans. The context is as follows: the hypothetical group of individuals is composed of 50 year old people, and they (and their partners) are expected to live to be 80. There is no inflation, and their income after tax is guaranteed to be £3000 each month from age 50 to age 80. The respondents are presented with 5 cards, each of which contains from 2 to 5 retirement plans which consist of getting a loan before retirement and paying it off afterwards, i.e. spending more

---

207 Risk issues with receiving a sum in the future must be null.

208 Whose discount factor increases for adjacent periods and can be represented, for example as $\frac{1}{1+at}$, where a is a parameter regarding the curvature of the discount factor with respect to time.

209 The quasi-hyperbolic discounter is biased in the short-run but discounts consistently in the long-run, i.e. when the trade-offs occurs further in the future. Mathematically it can be represented as the following sequence of discounting factors $\{1, bd, \ldots, bd^{t}, \ldots\}$, where $0<b<1$.

210 Respondents 1 to 10 did not answer this questionnaire and respondents 11 to 24 answered the questionnaire as in Chark et al. (2015). Also the no-risk statement in the instruction (see appendix A) was put from participant 27 onwards.

before than after retirement, or saving before retirement and spending more after. Each plan implies a different interest rate and consumption growth, from which the EMU as inter-temporal inequality aversion can be estimated.

The task was administrated by giving the participants a sheet of paper containing both the instructions and the alternatives. The task is self-paced and unspeeded.

In Barsky et al. (1997) the respondents take decisions for themselves, and in Atkinson et al. (2009) they decide for the government. In my study by contrast they are asked to decide directly for others. In decisions to be taken for society as a whole like in Atkinson et al. (2009) empathy has a relativelly small role to play, given the impersonal nature of the question. The context we use is more appropriate for this study because it is more likely to reveal how empathy affects the EIS and at the same time keeps the respondent in the position of a social planner (i.e. choosing for others).

A social planner maximizing a welfare function (consistent with the SWEM) represented by $W = \sum_{t=0}^{t=\infty}(1+\delta)^{-t}U(C_t)$, where δ is the utility discount rate and $U(C_t)$ is a iso-elastic utility function, subject to an inter-temporal wealth constraint represented by

$$A_{t+1} - A_t = r_t A_t + Y_t - C_t, \tag{4.5}$$

where A stands for the amount of assets possessed, r is the rate of interest and Y is labour income, yields the Euler equation, which can be expressed as

$$\Delta \log(C_t) = s(r_t - \rho), \tag{4.6}$$

where s is the EIS ($1/_{\eta}$ or $1/_{EMU}$) and ρ yields information on the value of the subjective discount rate. Notice it contains four parameters, consumption growth, return rate, pure discount rate and η (the EMU). In each decision for each card in the questionnaire concerned there is a consumption growth and a return rate implicit, so with two choices it is possible to derive an implicit pure discount rate and an EMU value.

There are at most five choices per card. Each card assumes an interest rate and each choice within a card a range of growth rates, such that the preferred growth can be anywhere between the alternatives above and below the chosen option. Therefore, an EMU range is estimated considering the answers for the cards presented, the breath of which depends on the number of questions asked per card.

The procedure to obtain the intervals is as follows. I compare the answers given for all combinations of card answers[211] thereby getting three interval estimates, and then I consider the interval which is common to all of them.[212] Only responses consistent with the assumed model are considered. This basically means the chosen growth rate must be increasing with interest rate.[213]

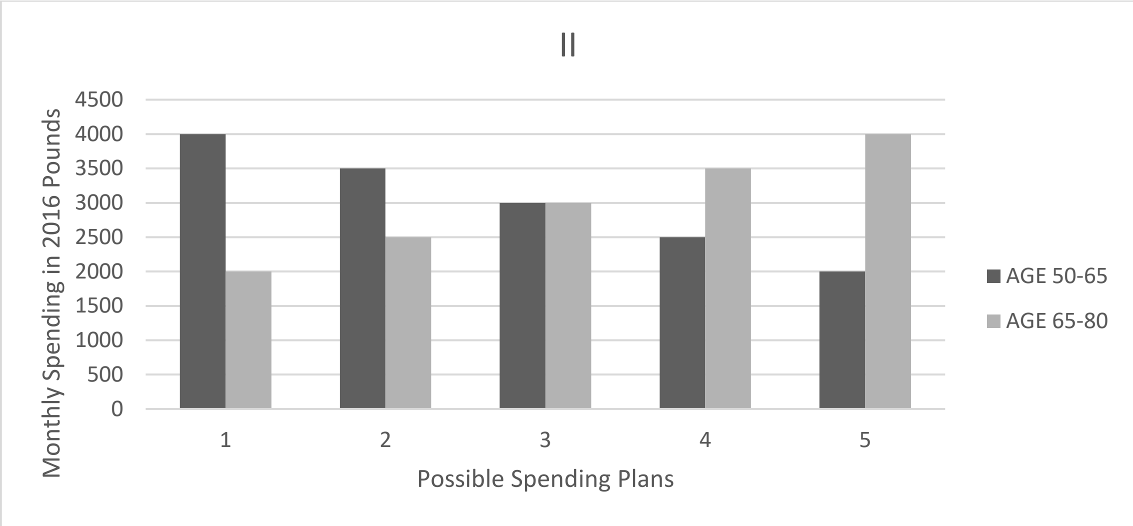
Consider the example given in figures 4.1 (second card) and 4.2 (fourth card) below. The card represented in figure 4.1 assumes an interest rate of 0%, while the one in figure 4.2 assumes an interest rate of 4.6%. Option 2 in figure 4.1 presents an annual decrease rate that ranges from 4.7% to 1.2% and option 4 in figure 4.2 presents an annual growth that ranges from 1.4% to 3.5%. Assuming the participant chooses option 2, for example, in the third card (presented in Appendix 4.B) then the consistent decrease rate for option 2 in figure 4.1 goes from 2.7% to 1.2%. With the information above we can substitute the interest rates and growth rates into the Euler equation in order to obtain an EMU range (which in the case above goes from 0.74 to 1.12).

[211] Which are 3, both for the first 22 participants (11 to 32) and for rest.

[212] When the intervals do not overlap, I consider the whole range covered by the three.

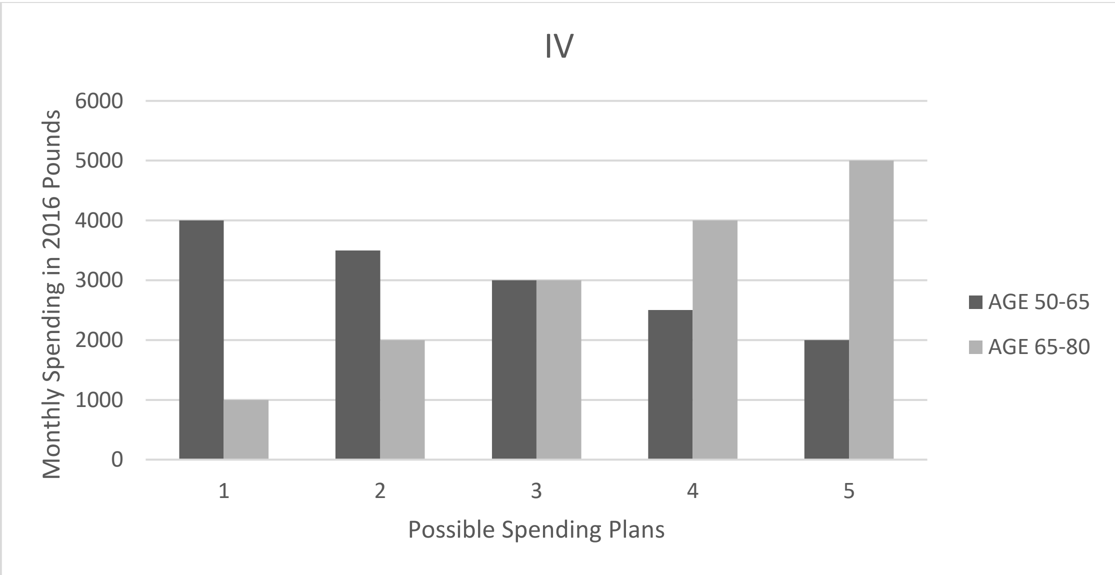
[213] There are also two (one for the first 22 participants) questions designed to check whether the participant understood the questionnaire.

Figure 4.1. Task 10 card example (0% interest rate)



Source: see text

Figure 4.2. Task 10 card example (4.6% interest rate)



Source: see text

As in the previous estimation procedure it is assumed that the respondents are maximizing a social welfare function, and are therefore are viewed as social planners taking decisions for others. As such the EMU estimated is not from their own utility function, but from the social utility function considered appropriate by the participant. The difference in this setting compared to the one regarding intra-temporal inequality aversion is that the social planner discounts others' utilities when they are realized in the future, or in other words, time is taken into consideration.

4.3.3. Data description

In this section I provide summary statistics regarding the variables observed in the experiment described above.

Tables 4.2 and 4.3 show the results regarding the EMU estimates gauged from the 16th question of the RP and DA, respectively (i.e. risk aversion and inequality aversion, respectively). For the first the median and mode are 0.71-1.16 and >3.6, respectively. For inequality aversion, the median and mode are both >7.5. The risk aversion median is consistent with the risk aversion estimated in Harrison et al. (2007) (0.43-0.91) and Taylor (2013) (0.23-1.1), and the inequlity aversion median is consistent with the inequality aversion estimated by the second study of Pirttila and Uusitalo (2008) (>3). For the EIS the median midpoint is 7.34, which is lower than Barsky et al.'s (1997) midpoint estimate of 8.7.[214]

Table 4.2. Frequency distribution for risk aversion

Range	Frequency	Percent	Cumulative percent
<0	9	10	10
0-0.5	6	7	17
0.5-0.71	9	10	27
0.71-1.16	23	26	53
1.16-3.6	16	18	71
3.6>	24	28	100
Total	87	100	

Source: see text.

Table 4.3. Frequency distribution for inequality aversion

Range	Frequency	Percent	Cumulative percent
<0	8	10	10
0-1.5	3	4	14
1.5-2.11	5	6	20
2.11-3.75	7	8	28
3.75-7.75	11	13	41
7.75>	49	59	100
Total	83	100	

Source: see text.

Table 4.4 shows the summary statistics for all other variables observed in the experiment, and table 4.5 gives additional details regarding the CRT, the TC, gender and age.

Table 4.4. Summary statistics for CRT, MFMB, EQ, AQ, RME, TC, AGE

Var	Mean	Stdev	Median	Max	Min	NbObs
CRT	0.81	1.07	0	3	0	93
MFMB	0.04	0.22	0.03	0.6	-0.4	92[215]

[214] I do not include a table for the EIS because each choice combination estimates a specific EIS value range.

[215] The test was not registered for 1 participant.

EQ	45.37	12.51	46	72	16	93
AQ	17.55	5.89	17	34	5	93
RME	25.68	3.66	26	36	16	93
TC	0.12	0.33	0	1	0	83
AGE	21.34	3.9	20	33	18	86[216]

Source: see text

Table 4.5. Additional details for the CRT, TC, gender and age

No of right questions in the CRT test	Frequency	Percentage
0	52	56%
1	18	19%
2	12	13%
3	11	12%
Choosing time consistently	10	12%
Not choosing time consistently	73	88%
Female[217]	73	78%
Above 24	17	18%

Source: see text

4.3.4. Empirical strategy

In this section I give a brief explanation regarding interval regressions, the statistical procedure used throughout. I then analyse the context-sensitivity of the EMU estimates and its relation with socially desirable traits and lastly examine the influence of these traits on the EMU.

4.3.4.1. Interval regressions

Consider the following linear structural model[218]

$$y = \boldsymbol{x\beta} + u \tag{4.7}$$

$$E(u|\boldsymbol{x}) = 0, \tag{4.8}$$

where $\boldsymbol{x}$ is the vector of explanatory variables, and $\boldsymbol{\beta}$ is the vector of coefficients.

In the case of interval regressions, the dependent variable y is observed in intervals (which is the case for risk aversion, inequality aversion and the EIS in this study), so we actually observe a different variable w, which can be described as $w = 0$ if $y \leq r_1$, $w = 1$ if $r_1 < y \leq r_2$, …, $w = J$ if $y > r_J$, where $r_1 < r_2 \ldots < r_J$ are known interval limits.

If we assume

$$u_i|\boldsymbol{x}_i, r_i \sim Normal\ (0, \sigma^2), \tag{4.9}$$

[216] Age was not registered for 7 people.
[217] Gender was not registered for 3 people.
[218] The explanation is based on Wooldridge's (2010).

where the subscript i for u and $\boldsymbol{x}$ indicates we refer to the observations, not to the structural model, we can derive the conditional probabilities $P(w = j|\boldsymbol{x})$ for $j = 0, 1, \dots, J$ and thus obtain the maximum likelihood estimators $\widehat{\boldsymbol{\beta}}$ and $\widehat{\sigma^2}$ for $\boldsymbol{\beta}$ and σ^2, respectively.

For each random draw i the conditional log likelihood ($logP(w|\boldsymbol{x})$) used for obtaining the estimators is given by

$$l_i(\boldsymbol{\beta}, \sigma) = 1[w_i = 0]\log\{\Phi[(r_1 - \boldsymbol{x}_i\boldsymbol{\beta})/\sigma]\} + 1[w_i = 1]\log\{\Phi[(r_2 - \boldsymbol{x}_i\boldsymbol{\beta})/\sigma - \Phi[(r_1 - \boldsymbol{x}_i\boldsymbol{\beta})/\sigma]\} + \cdots + 1[w_i = J]\log\{1 - \Phi[(r_J - \boldsymbol{x}_i\boldsymbol{\beta})/\sigma]\}, \tag{4.10}$$

where $1[.]$ refers to a dummy variable equal to 1 when what is inside the brackets is true and equal to 0 otherwise, and $\Phi(.)$ represents the standard normal cumulative distribution function (cdf).[219]

The $\widehat{\boldsymbol{\beta}}$ and $\widehat{\sigma^2}$ are the parameters that solve

$$\max_{\boldsymbol{\beta},\sigma} E[l_i(\boldsymbol{\beta}, \sigma)], \tag{4.11}$$

whose sample analogue is

$$\max_{\boldsymbol{\beta},\sigma} N^{-1} \textstyle\sum_{i=1}^{N} logP(w_i|\boldsymbol{x}_i; \boldsymbol{\beta}, \sigma). \tag{4.12}$$

Finally, it must be noticed that if the interval limits change across i, which is the case for the analyses regarding the EIS, we have to make a further assumption, which is

$$D\big(y_i\big|\boldsymbol{x}_i, r_{i1}, \dots, r_{iJ}\big) = D(y_i|\boldsymbol{x}_i), \tag{4.13}$$

which basically says the interval limits must be independent of y_i conditional on $\boldsymbol{x}_i$ [$D(.)$ stands for distribution].

4.3.4.2. Context analysis

The main question regarding the context analysis is whether highly empathetic/reflective participants' EMU varies across contexts more than does lowly empathetic/reflective participants' EMU. I also look at the influence of age and time consistency.

To answer this question we stack the data[220] and then divide the participants into highly and lowly empathetic/reflective according to the different empathy and reflectiveness measures. Such a division was made in two ways. In the first we divided the sample in two after ordering it

[219] The cdf is given by $\Phi(x) = 1/2\pi \int_{-\infty}^{x} e^{-t^2/2}\, dt$.
[220] This is, we stacked the estimates of the EMU as risk aversion, inequality aversion and as the EIS and the corresponding trait measurements.

according to the empathy/reflectiveness measures, obtaining the 50% most and least empathetic/reflective. In the second just the 33% at the beginning and end of the distribution are considered, rendering an analysis relative to the third at the top and bottom of the ordered measures. Throughout this analysis I account for clustering at the level of the respondent given that for most participants the EMU is measured as risk/inequality aversion and EIS.[221]

We conduct the same analysis for those older than 24 years as opposed to those younger (or exactly at 24 years) and those choosing time consistently for others as opposed to those not choosing time consistently. The results obtained for each group referred above are also compared with the ones obtained considering all the participants.

To test whether the EMU (η) estimated as the EIS is equal to the EMU estimated as risk and inequality aversion inside the groups referred above we use interval regressions of the following form:

$$\eta = \beta_0 + \beta_1 D_{RP} + \beta_2 D_{DA}, \tag{4.14}$$

where D_{RP} is a dummy variable indicating the EMU intervals were estimated as risk aversion and D_{DA} is a dummy variable indicating the EMU intervals were estimated as inequality aversion. In such a framework, the EMU estimated as the EIS equals to β_0, the EMU estimated as risk aversion equals to $\beta_0 + \beta_1$ and the EMU estimated as inequality aversion equals to $\beta_0 + \beta_2$. Thus, to test whether the EMU as risk (inequality) aversion is equal to the EMU as the EIS we have to look at whether $\beta_1 = 0$ ($\beta_2 = 0$).

To test whether the EMU as risk aversion is equal to the EMU as inequality aversion we use the following equation:

$$\eta = \beta_0 + \beta_1 D_{RP} + \beta_2 D_{EIS}, \tag{4.15}$$

in which D_{EIS} is a dummy variable indicating the EMU intervals were estimated as the EIS. In this case, the EMU estimated as risk aversion equals to $\beta_0 + \beta_1$, as in the last equation. The EMU estimated as inequality aversion, on the other hand, equals to β_0. Thus, to test whether the EMU

[221] To be more precise that is the case for 52 participants out of 93. 31 participants answered just the RP and DA due to their not answering consistently the questionnaire eliciting the EMU as the EIS, thus having their EMU estimated only as risk aversion and inequality aversion. 5 participants answered just the RP and other five just answered the DA (having thus their EMU estimated as risk and inequality aversion only), due to being interviewed at the beginning of the research (they were the first 10 to participate) where the task relative to estimating the EMU as the EIS was not included and the participants were given just one of the two tasks regarding EMU estimation. Moreover, the 5 participants answering just the DA had not their EMU as inequality aversion considered, since after the pilot the method to estimate the EMU as inequality aversion changed.

as risk aversion is equal to the EMU as inequality aversion we examine whether β_1 is statistically different from 0.

4.3.4.3. EMU estimates analysis

For this analysis I do three kind of analysis (in all I use interval regressions). In the first I do not stack the data, but estimate an EMU parameter for each group[222] in each of the three contexts and then compare them. In the same way as for the context analysis, the estimations obtained for each group are also compared with the estimations obtained considering all the participants (within each context).

In the second I stack the data and run 6 pooled regressions (accounting for clustering at the level of the participant), each one with an index of empathy, an index of reflectiveness, a dummy indicating those choosing time consistently for others, another dummy indicating those with more than 24 years of age, and more two dummies indicating context as explanatory variables.[223] Then I run another regression containing just variables that were significant in at least one of the 6 previous regressions. From the latter I derive conclusions on how the traits affect the EMU for the sample.

The third is the same as the second, but without the context dummies. The goal is to observe how the EMU varies in response to the psychological traits only.

4.4. Results and discussion

In the first subsection of this section I present the results regarding the context analysis (in which I analyse the constancy of the EMU across the contexts). In the second I look at the results regarding the EMU estimates analysis (in which I analyse how the traits considered affect the EMU estimates). In the final subsection I discuss the results in light of the literature reviewed earlier (section 4.2).

4.4.1. Context analysis

The context analysis results for all participants show in general they treat the risk prospect (RP) context[224] differently from the distributional alternatives (DA) and EIS contexts,[225] but treat the

[222] The framework for the analysis is similar to the one regarding the previous analysis. The data is divided between the 50% and 30% most (least) reflective and empathetic participants, those with more than 24 years old as opposed to those with less than (or equal to) 24 and those choosing time consistently for others as opposed to those not choosing time consistently for others. So these are the groups referred.
[223] It must be 6 regressions in order to contemplate all the combinations regarding the 2 indexes of reflectiveness and the 3 indexes of empathy considered.
[224] The context in which the EMU as risk aversion is estimated.
[225] The contexts in which the EMU is estimated as inequality aversion and as the EIS, respectively.

DA context in the same way they treat the EIS.[226] The results for the 50% most (least) empathetic/reflective participants can be visualized on table 4.6. Notice highly empathetic people as measured by the EQ and the RME scale treat the DA context differently from the EIS, suggesting more empathetic people tend to treat all contexts differently, as opposed to low empathetic people whose EMU as inequality aversion is not significantly different from the EMU as EIS.

Table 4.6. Context analysis for the 50% most (least) empathetic/reflective participants

	H0: RP=EIS	H0: DA=EIS	H0: RP=DA
High MFMB	-4.206** (2.60e-05)	0.701 (4.83e-01)	-4.778** (1.77e-06)
Low MFMB	-4.59** (4.49e-06)	1.81° (7.08e-02)	-6.93** (4.20e-12)
High CRT	-5.091** (3.57e-07)	0.719 (4.72e-01)	-4.633** (3.60e-06)
Low CRT	-3.746** (1.80e-04)	0.948 (3.43e-01)	-6.312** (2.75e-10)
High EQ	-4.52** (6.10e-06)	2.53* (1.14e-02)	-6.42** (1.36e-10)
Low EQ	-4.245** (2.18e-05)	-0.246 (8.05e-01)	-4.595** (4.33e-06)
High AQ	-5.20** (2.01e-07)	1.21 (2.28e-01)	-6.05** (1.48e-09)
Low AQ	-3.87** (1.11e-04)	1.18 (2.38e-01)	-5.21** (1.91e-07)
High RME	-4.85** (1.26e-06)	2.56* (1.06e-02)	-6.69** (2.27e-11)
Low RME	-3.791** (1.50e-04)	-0.298 (7.66e-01)	-4.322** (1.55e-05)

*The table presents z-statistics. The numbers in parenthesis are p-values. ** indicates the relevant coefficient is significant at 1%, * at 5% and ° at 10%.*

The results regarding the 33% most (least) empathetic/reflective participants in table 4.7 also suggest more empathetic people tend to treat the DA context differently from the EIS, but in addition show low reflective people as measured by the MFMB scale tend to do the same. Notice the fact the effect of empathy on how participants treat the contexts was detected in the comparison between the 50% and 33% most (least) empathetic and the effect of reflectiveness appeared only when we compared the most (least) 33% reflective[227] suggests the empathy effect is stronger (more easily detectable).

[226] It basically means that only their EMU estimated as inequality aversion is equal to their EMU estimated as the EIS. Considering all participants, the coefficient testing RP=EIS (which is the null hypothesis) is -6.11 (the p-value being 9.97e-10), the coefficient testing RP=DA (H0) is -7.91 (the p-value being 2.48e-15) and finally the coefficient testing DA=EIS (H0) is 1.69 (the p-value being 9.10e-02). Notice we consider the 5% significance level as the criterion to regard equality between contexts.

[227] To be more precise, the same effect is shown in the 50% comparison (table 4.6) for the MFMB, but just if we consider a 10% significance level.

Considering that empathy is an intuitive trait (not the result of reflection or slow thinking), it does not come as a surprise that more empathetic tend to treat intra-temporal inequality (DA) aversion differently from inter-temporal inequality aversion (EIS). To regard these two contexts as equals one would have to think about one person in different points of time as different people, which requires the ability to abstract inequality itself from the two different kinds of inequality (intra and inter-temporal) implied in the contexts. Such exercise is arguably more likely to occur when intuitive feelings caused by empathy are not much salient, given human limited capacity to direct attention to multiple objects.

In other words, if someone is highly empathetic his attention will be in great part directed to the emotional charge released by the situation, which makes him less likely to realize (given attention and processing limitation) the similarity between redistributing money for one person or group over time (inter-temporal inequality) and across people (intra-temporal inequality).

This same account helps to explain why highly reflective participants are more likely to treat the two contexts referred equally. This is, as more reflective they are capable to think more considerately about the situation, which increases the probability to realize the two contexts are not so different as it may look at first.[228]

Table 4.7. Context analysis for the 33% most (least) empathetic/reflective participants

	H0: RP=EIS	H0: DA=EIS	H0: RP=DA
High MFMB[229]	-3.122** (1.80e-03)	-0.159 (8.73e-01)	-2.969** (2.99e-03)
Low MFMB[230]	-3.04** (2.34e-03)	2.34* (1.95e-02)	-5.28** (1.28e-07)
High CRT[231]	-5.083** (3.72e-07)	0.813 (4.16e-01)	-3.891** (9.97e-05)
Low CRT[232]	-3.64** (2.68e-04)	1.11 (2.67e-01)	-5.88** (4.04e-09)
High EQ[233]	-3.10** (1.92e-03)	2.45* (1.41e-02)	-4.40** (1.09e-05)
Low EQ[234]	-3.928** (8.56e-05)	-0.743 (4.58e-01)	-3.864** (1.11e-04)
High AQ[235]	-3.996** (6.43e-05)	0.889 (3.74e-01)	-5.817** (6.01e-09)
Low AQ[236]	-2.78** (5.45e-03)	1.93° (5.38e-02)	-4.25** (2.16e-05)

[228] Moreover, it is interesting to note reflectivity as measured by the CRT and empathy as measured by the AQ have no effect on the EMU context-sensitivity.
[229] Observations for DA=27; EIS=18; RP=28.
[230] Observations for DA=29; EIS=14; RP=30.
[231] Observations for DA=26; EIS=27; RP=20.
[232] Observations for DA=24; EIS=24; RP=24.
[233] Observations for DA=26; EIS=20; RP=27.
[234] Observations for DA=28; EIS=13; RP=32.
[235] Observations for DA=29; EIS=15; RP=29.
[236] Observations for DA=27; EIS=16; RP=30.

High RME[237]	-3.19** (1.42e-03)	2.45* (1.42e-02)	-4.79** (1.63e-06)
Low RME[238]	-2.378* (1.74e-02)	0.577 (5.64e-01)	-3.620** (2.94e-04)

*The table presents z-statistics. The numbers in parenthesis are p-values. ** indicates the relevant coefficient is significant at 1%, * at 5% and ° at 10%.*

The results regarding those choosing more time consistently for others and with more than 24 years are shown in table 4.8 and table 4.9 respectively. The first shows participants choosing time consistently for others tend to treat the DA and EIS contexts differently, just like the most empathetic ones.[239] The second shows participants with more than 24 years treat just the RP and the DA contexts differently, being thus those getting closer to act as predicted by the SWEM. Those with less than 24 treat all contexts differently, as the most empathetic and those choosing time consistently for others.[240]

As long as maturity is concerned, the results are consistent with immature people being more impulsive.[241] The basic idea is that given reflectiveness is associated with being able to refrain from giving impulsive responses, it is expected that younger people act more similarly to those with low reflectivity, which seems to be indeed the case. This shows the dangers of choosing convenience samples comprising students for estimating the EMU, given the SWEM assumptions do not fit their behaviour.[242]

Table 4.8. Context analysis for participants (not) choosing time consistently for others

	H0: RP=EIS	**H0: DA=EIS**	**H0: RP=DA**
TC	-3.27** (1.09e-03)	3.94** (8.12e-05)	-4.59** (4.49e-06)
Not TC	-5.46** (4.69e-08)	1.11 (2.66e-01)	-7.13** (9.92e-13)

*The table presents z-statistics. The numbers in parenthesis are p-values. ** indicates the relevant coefficient is significant at 1%, * at 5% and ° at 10%.*

Table 4.9. Context analysis for participants with more (less) than 24 years old[243]

	H0: RP=EIS	**H0: DA=EIS**	**H0: RP=DA**
More than 24[244]	-1.327 (1.85e-01)	-0.295 (7.68e-01)	-2.556* (1.06e-02)

[237] Observations for DA=25; EIS=23; RP=25.
[238] Observations for DA=23; EIS=12; RP=38.
[239] It is important to note there are just 10 participants who chose time consistently for others. All of them answered the RP and the DA, but only six answered the RP, the DA and the EIS.
[240] It is important to notice there are only 17 people with more than 24 years old in the sample. From those we have 6 observations of the EMU as the EIS, 11 observations of the EMU as inequality aversion and 14 observations of the EMU as risk aversion.
[241] See the introduction for more details on works investigating the relation between age and impulsivity.
[242] It is important to notice these results were not known before applying the experiment, but as a result of the analysis, such that our choice to use a convenience sample could be not influenced by this finding.
[243] If we consider those below 24 (L_age) and above or equal to 24 (H_age) we have that $RP/DA = EIS$ and $RP \neq DA$ for H_age and $RP \neq DA/EIS$ and $DA = EIS$ for L_age, which does not change the general conclusion regarding being older than 24.
[244] DA=11, EIS=6, RP=14 observations.

Less or equal to 24	-7.44** (1.01e-13)	2.17* (3.02e-02)	-5.51** (3.63e-08)

*The table presents z-statistics. The numbers in parenthesis are p-values. ** indicates the relevant coefficient is significant at 1%, * at 5% and ° at 10%.*

It is interesting to note at this point that different to Atkinson et al. (2009), Groom and Maddison's (2013) results (which are mostly based on revealed preferences, therefore not experimentally) suggest people in general do not treat the three contexts differently. Since Atkinson et al.'s (2009) sample is probably formed by younger people than the ones used by Groom and Maddison (2013),[245] based on the results shown on table 4.9 it is reasonable to speculate that differences in age between the samples is an important factor generating the different conclusions. It must be noticed however, that there are many other differences between Atkinson et al. (2009) and Groom and Maddison (2013) that may cause the results to differ, especially variations regarding the methodologies used.

Therefore, overall the results regarding the context analysis suggest more empathetic people tend to treat contexts more differently than less empathetic. More specifically, they treat the DA and EIS contexts differently, whereas less empathetic people treat them equally. The same effect is observed in participants who choose time consistently for other people, in participants with less or equal to 24 years and in less reflective participants, although the latter effect is probably weaker than the empathy effect. On the other hand, people with more than 24 years treated two of the three comparisons equally, being those acting closer to predicted by SWEM in the study. Such result may help to understand why Atkinson et al. (2009) and Groom and Maddison (2013) got results suggesting different conclusions with regards the constancy of the EMU across the contexts.

4.4.2. EMU estimates analysis

The EMU point estimates (calculated per context by interval regression) considering all participants can be seen on table 4.10. The EMU estimated as risk aversion is about 2, the EMU as inequality aversion is about 9 and the EMU estimated as EIS is about 5.5.

The point estimate for the EMU as risk aversion is compatible with the median values obtained by works like Carlsson et al. (2005) (with median EMU in the interval 2-3) and Hryshko et al. (2011) (with EMU median in the interval 2 to 3.76), while the median interval obtained in the study (0.71-1.16) is compatible with values obtained by Taylor (2013) (0.237 - 1.1) and Harrison et al. (2007) (0.43 - 0.91).

[245] Atkinson et al.'s (2009) sample overrepresents young people, which is not the case for Groom and Maddison (2013), who in most cases use samples obtained from the general public to derive their results.

The point estimate for the EMU as inequality aversion is far above the median values obtained by the works gathered in the literature review (with median varying from 0 to 3); that is also the case for the median interval in the study (above 7.5). However, the modal value is the same as for Atkinson et al. (2009) (above 7.5), which indicates the point estimate is probably compatible with the value they would have obtained had they estimated a point estimate using interval regression.

Finally, the point estimate for the EMU as EIS is below the interval midpoint of the median individual in Atkinson et al. (2009) (8.8) and the modal response in Barsky et al. (1997) (midpoint of 8.7). However, the mode and the median midpoint obtained in the study (7.34) are reasonably compatible with the values obtained in the studies mentioned.

Table 4.10. EMU estimation per context considering all participants

	EMU
RP (risk aversion)	1.950 (0.237)
DA (inequality aversion)	9.34 (1.198)
EIS	5.565 (0.463)

Source: see text. Standard errors are shown in parenthesis.

In the following I present the results regarding the first kind of analysis referred in section 4.3.4.3, in which I compare EMU estimates for low and highly reflective participants, low and highly empathetic participants, mature and imature participants and finally time consistent and not time consistent participants.

The estimates regarding the 50% most (least) empathetic/reflective are presented on table 4.11. Notice the estimates for the highly reflective (as measured by both the CRT and the MFMB scales) are lower than the ones referring to low reflective in the RP and DA contexts. The differences are larger than 0.5[246] in all cases except for the MFMB scale in the RP context. It suggests the EMU estimated is smaller for highly reflective people. Also notice all estimates for the highly empathetic in the RP and DA contexts are smaller than the ones obtained considering all participants. Figure 4.3 makes it easier to visualize the results.

Table 4.11. EMU estimation per context regarding the 50% most/least empathetic/reflective

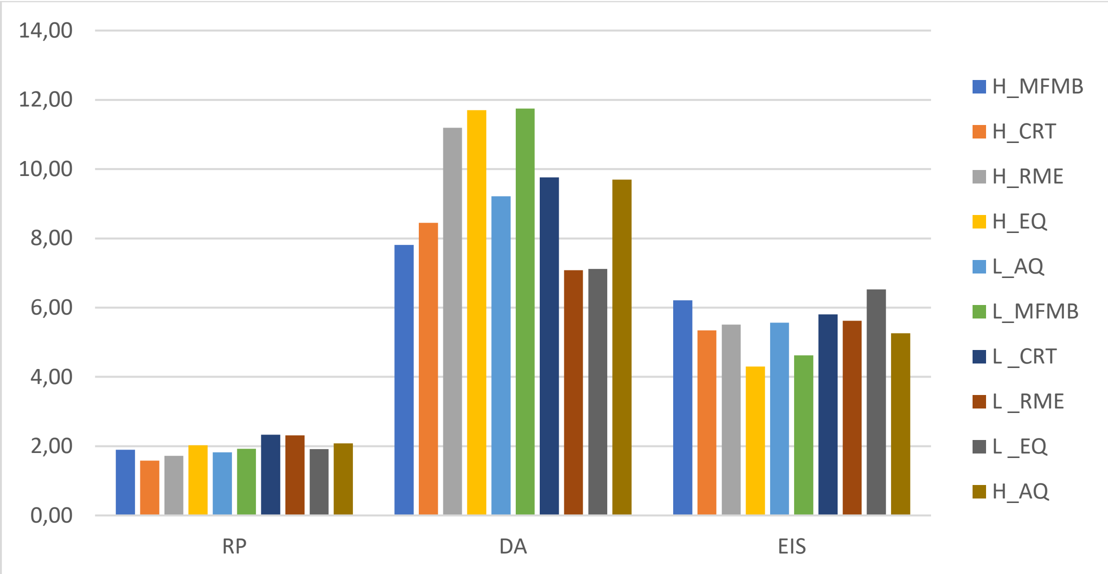
HIGHLY EMP/REF	RP	DA	EIS
H_MFMB	1.893 (0.320)	7.81 (1.104)	6.21 (0.739)
H_CRT	1.586 (0.305)	8.45 (1.60)	5.344 (0.611)

[246] Which is significant in the SDR context. Nonehteless, Appendix 4.D shows the EMU estimates 95% confidence interval for this analysis.

H_RME	1.721 (0.257)	11.19 (1.956)	5.511 (0.652)
H_EQ	2.030 (0.331)	11.7 (2.275)	4.303 (0.389)
L_AQ	1.821 (0.331)	9.22 (1.549)	5.57 (0.737)
LOW EMP/REF	-	-	-
L_MFMB	1.921 (0.342)	11.75 (2.791)	4.624 (0.490)
L _CRT	2.334 (0.368)	9.76 (1.597)	5.81 (0.701)
L _RME	2.31 (0.468)	7.08 (1.338)	5.619 (0.659)
L _EQ	1.915 (0.347)	7.12 (1.233)	6.53 (0.799)
H_AQ	2.079 (0.349)	9.70 (1.898)	5.258 (0.552)

Source: see text. Standard errors are shown in parenthesis.

Figure 4.3. EMU estimation per context regarding the 50% most/least empathetic/reflective



Source: see text. Notes: (1) L_ stands for the 50% least empathetic/reflective and H_ stands for the 50% most empathetic/reflective.

The estimates for the 30% most (least) empathetic/reflective people are shown on table 4.12. Notice the estimates for the highly empathetic in the RP and DA contexts are higher than the ones regarding the low empathetic, and it is the opposite in the EIS context. It suggests EMU estimates tend to be higher for more empathetic people in the RP and DA contexts and tend to be lower in the EIS context.

However, some results from table 4.11 contradict those shown on table 4.12, which brings into question the conclusions regarding the effects of reflectiveness and empathy on the EMU estimates referred. Figure 4.4 makes it easier to visualize the results.

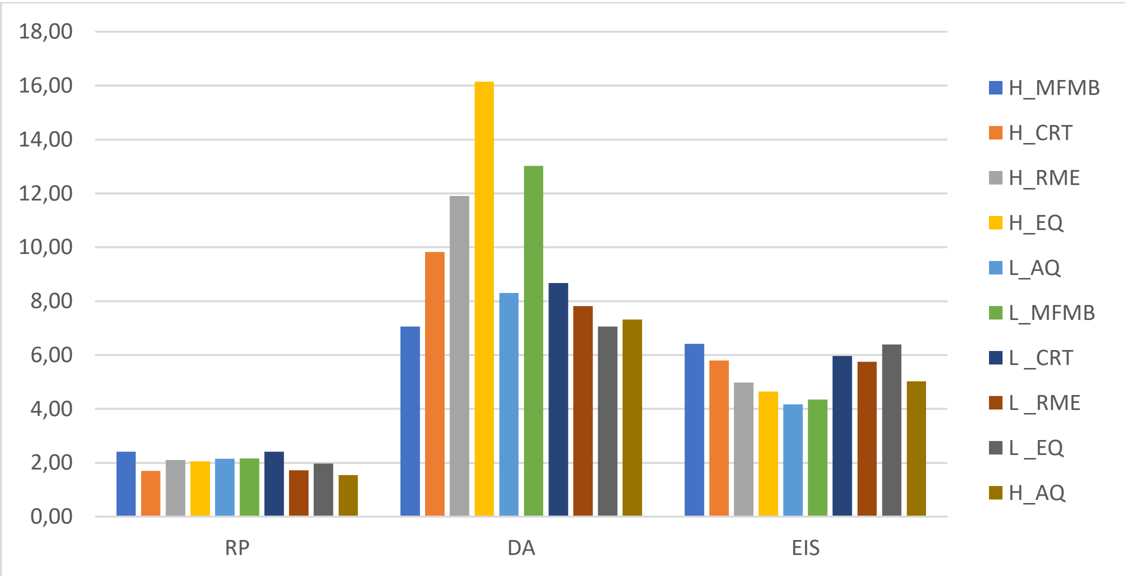
Table 4.12. EMU estimation per context regarding the 30% most/least empathetic/reflective

HIGHLY EMP/REF	**RP**	**DA**	**EIS**
H_MFMB	2.405	7.06	6.42

	(0.550)	(1.431)	(1.059)
H_CRT	1.703 (0.331)	9.82 (2.71)	5.80 (0.88)
H_RME	2.099 (0.399)	11.9 (2.859)	4.976 (0.580)
H_EQ	2.052 (0.480)	16.15 (5.383)	4.639 (0.524)
L_AQ	2.145 (0.449)	8.30 (1.552)	4.174 (0.486)
LOW EMP/REF	-	-	-
L_MFMB	2.156 (0.398)	13.02 (4.286)	4.351 (0.469)
L _CRT	2.41 (0.564)	8.67 (1.588)	5.961 (0.837)
L _RME	1.725 (0.469)	7.82 (1.908)	5.75 (0.966)
L _EQ	1.970 (0.485)	7.05 (1.631)	6.39 (1.035)
H_AQ	1.537 (0.361)	7.32 (1.820)	5.02 (0.712)

Source: see text. Standard errors are shown in parenthesis.

Figure 4.4. EMU estimation per context regarding the 30% most/least empathetic/reflective



Source: see text.

The estimates for participants with more (less) than 24 years old and choosing (not choosing) time consistently for others are shown in tables 4.13 and 4.14, respectively. Participants with more than 24 have higher estimates than participants with less than 24 for the RP and EIS contexts; it is the opposite for the DA context.[247] On the other hand, in this analysis it is not possible to say whether there are significant differences in EMU estimates for participants choosing consistently and not consistently for others in the DA context, considering that all time

[247] Considering those above or equal 24 (H_age) and below 24 (L_age) we have for the RP: EMU=2.439 (H_age) and EMU=1.777 (L_age); for the DA: EMU=10.91 (H_age) and EMU=9.03 (L_age); for the EIS: EMU=6.63 (H_age) and EMU=4.871 (L_age). So the general analysis would change just for the DA context.

consistent chose the highest interval (above 7.5). In the other 2 contexts there are no significant differences.

Table 4.13. EMU estimation per context for participants with more (less) than 24 years old

H_AGE (more than 24 years old)	EMU
RP	2.357 (0.679)
DA	8.66 (4.214)
EIS	7.26 (1.683)
L_AGE (less or equal than 24)	-
RP	1.808 (0.256)
DA	9.28 (1.224)
EIS	4.794 (0.442)

Source: see text.

Table 4.14. EMU estimation per context for participants choosing (not choosing) time consistently for others

	EMU
TC (choosing time consistently for others)_RP	1.997 (0.887)
nTC (not choosing time consistently)_RP	1.910 (0.251)
TC_DA	NA[248]
nTC_DA	8.18 (1.083)
TC_EIS	5.639 (0.696)
nTC_EIS	5.451 (0.496)

Source: see text.

Thus the general conclusion is that more empathetic people tend to have different values for the EMU than low empathetic people, but the way they differ varies across contexts. More empathetic have higher values for the RP and DA contexts and lower values for the EIS contexts. Similarly, those with more than 24 years differ from those with 24 years or more in EMU values, but the direction of the difference varies across contexts. They have higher values for the RP and EIS contexts and lower values for the DA context.[249]

In Appendix 4.D we show the 95% confidence interval for all EMU parameters estimated in this analysis. It shows the estimates vary significantly across contexts, but no significant variation

[248] All participants in this category have EMU above 7.75.

[249] Another interesting result worthy to mention is that we found a significant correlation between left handed people and choosing time consistently for others.

across traits is found. The lowest EMU was obtained in the risk aversion context and the highest in the inequality aversion context.

Now I present the results for the second analysis referred to in section 4.3.4.3, for which I run interval regressions in order to determine what traits and contexts significantly affect the EMU and to measure their influence on the parameter. For both the second and third analysis presented in this section the final regression and the 6 regressions indicating the significant variables to be included in it are shown in Appendix 4.E.

In all six regressions run to identify traits and contexts significantly affecting the EMU the dummy indicating risk aversion context was significant (meaning the RP context was treated differently while the DA and EIS context were treated equally);[250] in the two regressions where the RME index for empathy was used it was significant. Also, the dummy indicating those choosing time consistently for others was significant in all cases. Both the empathy measure and the time consistency dummy are positively related to the EMU estimate, suggesting those possessing such characteristics tend to have higher estimates than the general population.

For those being maximally empathetic according to the RME[251] measure and choosing time consistently for others the EMU as risk aversion is equal to 3.63; as inequality aversion and EIS the EMU is 8.27. For the least empathetic[252] and not choosing time consistently for others the EMU as risk aversion is -4.64; as inequality aversion and EIS the EMU is 0.

Finally, the results regarding the third analysis referred in section 4.3.4.3 (which is equal to the previous analysis without the context dummies)[253] are as follows. Both the RME and the time consistency regarding others variables were significant at least once in the first 6 regressions,[254] being thus included in the final regression. Considering the latter we have an EMU of 9 for the most empathetic and time consistent people, an EMU not significantly different from 0 for the least empathetic and not time consistent people and an EMU of 4.75 for the "average person" (not choosing time consistently for others and at the mean RME score).

4.4.3. Discussion

In this section I relate the results from the EMU estimates analysis with the literature review comments made in section 4.2.6. Notice I do not address the context analysis in this section

[250] Which is in accordance with the context analysis considering all participants in all contexts.
[251] This is, scoring 36 points in the RME test.
[252] Scoring 0 in the RME test.
[253] It allows us to concentrate on the traits.
[254] The time consistency variable was significant in all 6 and the RME was significant in the two regressions it is included in.

because no other work was found in the literature review dealing with the same issues (i.e. with how psychological traits affect the EMU context-sensitivity).

The findings shown in analyses 2 and 3 that empathy is positively related with the EMU is consistent with our expectations, which are based on the relationship between empathy and political views found in Johansson-Stenman et al. (2002) and Carlsson et al. (2005). However, in the present study the relation is found for the EMU in general (considered equally across the contexts), not for inequality aversion, which is the factor investigated by the articles.

The first study's results (in which the EMU estimates are shown for each context and measure) regarding the effect of empathy on the EMU are also consistent with the papers cited early (and therefore with our expectations). In almost all cases (there is one exception)[255] the EMU as inequality aversion is higher for highly empathetic people.

Also in the first analysis, almost all the results[256] regarding the relationship between the EMU as risk aversion and reflectiveness show a negative correlation between the variables. This is consistent with the Taylor (2013) results.

The negative correlation between maturity and the EMU as inequality aversion (also found in the first analysis) is consistent with Harrison et a. (2007), but differs from results found in Hryshko et al. (2011) and works in psychology (like Steinberg, 2008 and Steinberg and Scott, 2003) which show a positive correlation between age and risk-taking behaviour. This result is fragile though, for If we regard those with 24 years as mature the correlation becomes positive.

The findings regarding the relation between choosing time consistently for others (in analyses 2 and 3) and the EMU is (to knowledge) unprecedented. However, it again refers to the EMU in general.

4.5. Conclusion

In this work, I estimate the EMU and measure psychological traits which are arguably perceived as socially desirable for a sample of roughly 93 people (the majority are students).[257] The parameter is estimated experimentally in three different ways, each of which interpret it differently as risk aversion, intra-temporal inequality aversion (or simply inequality aversion)

[255] The exception being for the AQ when the comparison is done for 50% most and least empathetic.
[256] Once again there is one exception, which is for the MFMB measure when the comparison is done between the 30% most and least reflective.
[257] It is important to notice again that the influence of age (and therefore of using a sample consisting mostly of students) on the EMU was noticed as a result of the experiment and therefore could not be taken into consideration for the design of the experiment. Nonetheless, it would be certainly ideal to have a sample representative of the general population.

and inter-temporal inequality aversion (EIS). The traits referred are empathy, reflectiveness, maturity and time consistency when choosing for others.

The two main analyses carried out with the data aim to obtain EMU estimates with more normative significance than works estimating it using samples drawn from the general population or only students (i.e. which do not 'sample-frame') and to investigate whether the traits influence the constancy of the parameter across the contexts.

The analysis drew mainly from Drupp et al. (2015) and Atkinson et al. (2009). The first survey among prestigious economists familiar with the subject what EMU they think is ideal to be considered in the (long-term) SDR determination. The fact that scientists have been shown to be highly autistic and therefore less empathetic (Baron-Cohen et al., 2001) suggests another sample selection useful to deriving an ideal EMU can be obtained based on psychological traits perceived as socially desirable, such as empathy. The second tests the constancy of the EMU across the different contexts considered herein, but they do not select their sample in any way to reflect ideal social planners,[258] such that their study does not test the SWEM normative validity.

In the first analysis the parameter estimates regarding those possessing a high level of the named traits are compared to those possessing a low level of them in order to identify the direction the parameter goes when the traits vary (EMU estimates analysis). In the second the two groups are again compared in terms of which has their EMU varying more across contexts (context analysis).

Regarding the context analysis the results suggest more empathetic (as measured by the RME test and the EQ questionnaire) people's EMU estimates vary more across contexts than low empathetic people EMU estimates, and for people in general (i.e. when the whole sample is considered). The same happens for those choosing time consistently compared to others. On the other hand, more reflective people, as measured by the MFMB test, have their estimates varying less than those less reflective, but not less than people in general.[259] For those possessing more than 24 years we find they treat contexts more equally than those with less than or equal 24 and also more equally than the general analysis (whole sample).

No group treated all three contexts equally as predicted by the SWEM, which suggests besides not describing the behaviour of the general population, the model also does not describe the

[258] In this study it is done by comparing those possessing a high (low) level of psychological traits perceived as socially desirable.

[259] Less reflective people (MFMB), however, treat the context more differently than people in general.

behaviour of ideal social planners (in terms of socially desirable traits), not serving thus as a good normative model. Such results reinforce the need to develop a theoretical framework capable of disentangling risk, inequality and intertemporal inequality,[260] but also calls attention to the fact some specific groups may be better described by the model than others.

Moreover, the results relative to how maturity influences the constancy of the parameter across contexts may help to explain why results in Atkinson et al. (2009) and Groom and Maddison (2013) lead to different conclusions regarding the issue. It is also possible to conjecture that the findings relative to the other traits also contribute in this respect if one assumes their samples differ in the degree of empathy, reflectiveness, etc. (which is possible given how different they are).

Regarding the first analysis the results are for the most part mixed. More empathetic participants had lower estimates in the RP context (risk aversion) and DA context (inequality aversion) and higher for the EIS context. Participants with more than 24 had higher estimates for the RP and EIS contexts and lower for the DA context. The results for more reflective participants were not consistent and there is no significant difference between those choosing and not choosing time consistently for others in the RP and EIS contexts, while in the DA context it is not possible to determine whether it is different or not.

For the second and third analysis the results show more empathetic participants and participants choosing time consistently for others tend to have higher estimates. Such results suggest the ideal social planners as defined in the study tend to have higher EMU estimates than the general population.

A direct policy implication of the results presented above is that the ideal EMU is higher than the one obtained from the general population. It must be underlined that such recommendation relies on the assumption that the ideal EMU does not simply reflect the preferences of the majority of the population, but reflects to a larger extent the preferences of the individuals possessing socially desirable psychological traits.

Notice, however, that the EMU values obtained in the EMU estimates analysis are the result of a first attempt at estimating how much larger (assuming empathy is a desirable trait) the ideal EMU value is relative to the EMU value of the overall population. Therefore, it is important that more works investigate the issue – in special works trying to integrate ‘sample framing’ with

[260] As noticed in chapter 1 Epstein and Zin (1989) give an important contribution in that sense by disentangling the EIS and risk aversion.

revealed-preferences techniques of estimating the EMU – before additional and more specific policy recommendations are given.

References

Amiel, Y., Creedy, J., & Hurn, S. (1999). Measuring Attitudes Towards Inequality. Scand. J. of Economics, 101(1), 83-96.

Andersson, O., Holm, H. J., Tyran, J., & Wengstrom, E. (2016). Journal of the European Economic Association, 14(5), 1129–1154.

Atkinson, G., Dietz, S., Helgeson, J., Hepburn, C., & Sælen, H. (2009). Siblings, Not Triplets: Social Preferences for Risk, Inequality and Time in Discounting Climate Change. Economics: The Open-Access, Open-Assessment E-Journal, 3, 2009-26.

Attanasio, O. and Weber, G. (1989). Intertemporal Substitution, Risk Aversion and the Euler Equation for Consumption. The Economic Journal, 99, 59-73.

Baron-Cohen, S., Wheelwright, S., Hill, J. J., & Raste, Y. (2001). The "Reading the Mind in the Eyes" Test Revised Version: A Study with Normal Adults, and Adults with Asperger Syndrome or High-functioning Autism. Journal of Child Psychology and Psychiatry, 42, 241-252.

Baron-Cohen, S., & Wheelwright, S. (2004). The Empathy Quotient: An Investigation of Adults with Asperger Syndrome or High Functioning Autism, and Normal Sex Differences. Journal of Autism and Developmental Disorders, 34(2), 163-175.

Baron-Cohen, S. (2011). The Science of Evil. New York: Basic Books.

Barsky, R. B., Juster F. T., Kimball M. S., & Shapiro M. D. (1997). Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study. Quarterly Journal of Economics 112(2), 537–579.

Bloom, P. (2016). Against Empathy: the Case for Rational Compassion. London: Bodley Head.

Blundell, R., Browning, M., & Meghir, C. (1994). Consumer Demand and the Life-Cycle Allocation of Household Expenditures. Review of Economic Studies, 61, 57-80.

Carlsson, F., Daruvala, D., & Johansson-Stenman, O. (2005). Are People Inequality-Averse, or Just Risk-Averse? Economica, 72, 375–396.

Carlsson, F., Gupta, & Johansson-Stennman, O. (2003). Choosing from behind a veil of ignorance in India. Applied Economics Letters, 10, 825–827.

Chark, R., Chew, S. H., & Zhong, S. (2015). Extended Present Bias: a Direct Experimental Test. Theory and Decision, 79, 151-165.

Cowell, F A., & Gardiner, K (1999) Welfare Weights. STICERD, London School of Economics.

Crone, E. A., & Dahl, R. E. (2012). Understanding Adolescence as a Period of Social–Affective Engagement and Goal Flexibility. Cognitive Control, 13, 636-650.

Dodd, M. D., Hibbing, J. R., & Smith, K. B. (2011). The Politics of Attention: Gaze-Cuing Effects are Moderated by Political Temperament. Attention, Perception, & Psychophysics, 73, 24-29.

Doll, B. B., Simon, D. A., & Daw, N. D. (2012). The Ubiquity of Model-Based Reinforcement Learning. Current Opinion in Neurobiology, 22, 1075–1081.

Drupp, M., Freeman, M., Groom, B., & Nesje, F. (2015). Discounting Disentangled. Centre for Climate Change Economics and Policy Working Paper No. 195 and Grantham Research Institute on Climate Change and the Environment Working Paper No. 172.

Eckel, C. C., & Grossman, P. J. (2002). Sex Differences and Statistical Stereotyping in Attitudes Toward Financial Risk. Evolution and Human Behavior, 23, 281–295.

Eppinger, B., Walter, M., Heekeren, H. R., & Li, S. (2013). Of Goals and Habits: Age-Related and Individual Differences in Goal-Directed Decision-Making. Frontiers in Neuroscience, 7, article 253.

Epstein, L. G., & Zin S. E. (1989). Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: a Theoretical Framework. Econometrica 57(4): 937– 968.

Ernst, M. & Fudge, J. L. (2009). A Developmental Neurobiological Model of Motivated Behavior: Anatomy, Connectivity and Ontogeny of the Triadic Nodes. Neuroscience & Biobehavioral Reviews, 33, 367-382.

Evans D. J. (2004). The Elevated Status of the Elasticity of Marginal Utility. Applied Economics Letters, 11, 443-447.

Evans, D. J. (2005). The Elasticity of Marginal Utility of Consumption: Estimates for 20 OECD Countries. Fiscal Studies, 26(2), 197-224.

Evans, D. J., Kula, E., & Sezer, H. (2005). Regional welfare weights for the UK: England, Scotland, Wales and Northern Ireland. Regional Studies, 39(7), 923-937.

Frederick, S. (2005). Cognitive Reflection and Decision Making. Journal of Economic Perspectives, 19(4), 25-42.

Frisch, R. (1959). A complete system for computing all direct and cross demand elasticities in a model with many sectors. Econometrica, 27, 177-96.

Groom, B., & Maddison, D. (2013). Non-identical Quadruplets: Four new estimates. Centre for Climate Change Economics and Policy, Working Paper n. 141 and Grantham Research Institute on Climate Change and the Environment, Working Paper n. 121.

Halicioglu, F., & Karatas, C. (2013). A social Discount Rate for Turkey. Quality & Quantity, 47(2), 1085-1091.

Hanoch, Y., Johnson, J.G., & Wilke, A. (2006). Domain specificity in experimental measures and participant recruitment. Psychological Science, 17 (4), 300-304.

Harden, K. P. & Tucker-Drob, E. M. (2011). Individual differences in the development of sensation seeking and impulsivity during adolescence: further evidence for a dual systems model. Developmental Psychology, 47, 739-746.

Harrison, G. W., & List, J. A. (2004). Field Experiments. Journal of Economic Literature, 42 (4), 1009-1055.

Harrison, G. W., Lau, M. I., & Rutstrom, E. E. (2007). Estimating Risk Attitudes in Denmark: a Field Experiment. The Scandinavian Journal of Economics, 109(2), 341-368.

Holt, Charles A. and Susan K. Laury. (2002). Risk Aversion and Incentive Effects. American Economic Review, 92(5), 1644-1655.

Hoppe, E. I., & Kusterer, D. J. (2011). Behavioral Biases and Cognitive Reflection. Economics Letters, 110, 97-100.

Hryshko, D., Luengo-Prado, M. J., & Sørensen, B. E. (2011). Childhood Determinants of Risk Aversion: the Long Shadow of Compulsory Education. Quantitative Economics, 2, 37-72.

Johansson-Stenman, O., Carlsson, F., & Daruvala, D. (2002). Measuring Future Grandparents' Preferences for Equality and Relative Standing. The Economic Journal, 112, 362-383.

Michaelson, Z. (2015). Biases in choices about fairness: Psychology and economic inequality. Judgment and Decision Making, 10(2), 198-203.

Mondak, J. J., Hibbing, M. V., Canache, D., Seligson, M. A., & Anderson, M. R. (2010). Personality and Civic Engagement: An Integrative Framework for the Study of Trait Effects on Political Behavior. American Political Science Review, 104(1), 85-110.

Office of National Statistics (2016). Population Estimates for UK, England and Wales, Scotland and Northern Ireland: mid-2016.

Pfeifer, J. H. & Allen, N. B. (2012). Arrested development? Reconsidering dual-systems models of brain function in adolescence and disorders. Trends in Cognitive Science, 16, 322-329.

Pirttila, J., & Uusitalo, R. (2008). A 'Leaky Bucket' in the Real World: Estimating Inequality Aversion using Survey Data. Economica, 77, 60-76.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a 'theory of mind'? Behaviour and Brain Sciences, 4, 515-526. In: Baron-Cohen, S., Wheelwright, S., Hill, J. J., & Raste, Y. (2001). The "Reading the Mind in the Eyes" Test Revised Version: A Study with Normal Adults, and Adults with Asperger Syndrome or High-functioning Autism. Journal of Child Psychology and Psychiatry, 42, 241-252.

Rifkin, J. (2009). The Empathic Civilization: The Race to Global Consciousness in a World in Crisis. London: Penguin.

Somerville, L. H., Jones, R. M. & Casey, B. J. (2010). A time of change: behavioral and neural correlates of adolescent sensitivity to appetitive and aversive environmental cues. Brain and Cognition, 72, 124-133.

Stanovich, K. (2011). Rationality and the Reflective Mind (First ed.). New York: Oxford University.

Steinberg, L. & Scott, E. S. (2003). Less guilty by reason of adolescence: developmental immaturity, diminished responsibility, and the juvenile death penalty. American Psychologist, 58, 1009-1018.

Steinberg, L. (2008). A social neuroscience perspective on adolescent risk-taking. Developmental Review, 28, 78-106.

Steinberg, L. (2013). The Influence of Neuroscience on US Supreme Court Decisions About Adolescents' Criminal Culpability. Perspectives, 14, 513-518.

Stern, N. (1977) Welfare weights and the elasticity of the marginal valuation of income. In: Artis, M. and Nobay, R. Proceedings of the Aute Edinburgh Meeting of April 1976. Basil Blackwell.

Szpiro, G. G. (1986). Measuring Risk Aversion: An Alternative Approach. The Review of Economics and Statistics, 68(1), 156-159.

Taylor, M. P. (2013). Bias and Brains: Risk Aversion and Cognitive Ability Across Real and Hypothetical Settings. Journal of Risk and Uncertainty, 46, 299-320.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a Predictor of Performance on Heuristics-and-Biases Tasks. Memory & Cognition, 39, 1275-1289.

Tversky, A., & Kahneman, D. (1979). Prospect Theory: an Analysis of Decision under Risk. Econometrica, 47(2), 263-291.

Weber, E.U., Blais, A.R., & Betz, N.E. (2002). A domain-specific risk-attitude scale: measuring risk perceptions and risk behaviors. Journal of Behavioral Decision Making, 15 (4), 263-290.

Weitzman, M. L. (2001). Gamma Discounting. American Economic Review, 91(1), 260-271.

Welsh , M. B., Burns, N. R., & Delfabbro, P. H. (n. d.). The Cognitive Reflection Test: how much more than Numerical Ability? University of Adelaide, Australia.

Wooldridge, J. M. (2010). Econometric Analysis of Cross Section and Panel Data. Cambridge, Mass.: MIT Press.

Appendix 4.A – Tasks on economic decision making

INSTRUCTIONS FOR THE RP:

Imagine you are actually faced with the choice described in the problems below and indicate the decision you would make in each case. There is no correct answer to such problems.

Type 'a' to choose option a, 'b' to choose option b, 'c' to choose option c and so on.

After reading these instructions, if you have no questions, type spacebar to go to the questionnaire.

QUESTIONS 1 TO 15 OF THE RP:

Questions	optionA_1	optionB_1
1) In addition to what you own, you have been given 1,000 pounds. You are now asked to choose between	A) 50% chance of a further 1,000 pounds	B) A further 500 pounds for sure
2) In addition to what you own, you have been given 2,000 pounds. You are now asked to choose between	A) Lose 500 pounds for sure	B) 50% chance to lose 1,000 pounds
3) Choose between	A) 33% chance of 2,500 pounds; 66% chance of 2,400 pounds; 1% chance of 0 pounds	B) 2,400 pounds for sure
Choose between	A) 34% chance of 2,400 pounds	B) 33% chance of 2,500 pounds
5) Choose between	A) 3,000 pounds for sure	B) 80% chance of 4,000 pounds
Choose between	A) 25% chance of 3,000 pounds	B) 20% chance of 4,000 pounds
7) Choose between	A) 90% chance of 3,000 pounds	B) 45% chance of 6,000 pounds
Choose between	A) 2% chance of 3,000 pounds	B) 1% chance of 6,000 pounds
9) Choose between	A) 7% chance of 10 pounds	B) 0.7 pounds for sure
Choose between	A) 7% chance of 1,000 pounds	B) 70 pounds for sure
11) Choose between	A) 80% chance to lose 4,000 pounds	B) Lose 3,000 pounds for sure
Choose between	A) 20% chance to lose 4,000 pounds	B) 25% chance to lose 3,000 pounds
Choose between	A) 45% chance to lose 6,000 pounds	B) 90% chance to lose 3,000 pounds
Choose between	A) 1% chance to lose 6,000 pounds	B) 2% chance to lose 3,000 pounds
15) Consider the following two-stage game. In the first stage, there is a probability of	A) 3,000 pounds for sure	B) 80% chance of 4,000 pounds

75% to end the game without winning anything, and a probability of 25% to move into the second stage. If you reach the second stage you have a choice between		

QUESTION 16 OF THE RP (MEASURING THE EMU AS INEQUALITY AVERSION):

Indicate among the six following gambles the one you would most like to participate in:

A) 28 pound for sure

B) A 50% chance of 24 pounds and a 50% chance of 36 pounds

C) A 50% chance of 20 pounds and a 50% chance of 44 pounds

D) A 50% chance of 16 pounds and a 50% chance of 52 pounds

E) A 50% chance of 12 pounds and a 50% chance of 60 pounds

F) A 50% chance of 2 pounds and a 50% chance of 70 pounds

INSTRUCTIONS FOR THE DA:

You are serving as the administrator of a GAME OF CHANCE THAT IS TO BE PLAYED BY A GROUP OF PEOPLE. Being a game of chance, OUTCOMES ARE ASSIGNED COMPLETELY AT RANDOM, e.g., by a computer generating random numbers. There are 100 persons in this group. All 100 members are entered in the game. REMEMBER YOU ARE ADMINISTRATING AND NOT PLAYING THE GAME OF CHANCE YOURSELF.

In the following questions you will be asked to make choices about what outcomes will be assigned to the people playing. You do not know any of these persons, and none of them know you or know that you are the person selecting the outcomes that they will be assigned. The game is conducted publicly in an assembly; the people will be aware of what they receive and what others receive. You will not be there while the game is conducted, nor get reports back afterwards.

In the following questions you will be given a choice between different packages of outcomes. Each question is completely independent. That is, when you are offered the choice between the outcome packages, they are in no way an alternative to previous choices and no one is aware of the choices you made when given previous offerings or that such alternatives existed.

You may use a calculator or any other tool, if you would like. Your job as administrator is not an indirect attempt to test your mental math abilities.

Type 'a' to choose option a, 'b' to choose option b, 'c' to choose option c and so on.

After reading these instructions, if you have no questions, type spacebar to go to the questionnaire.

<u>QUESTIONS 1 TO 15 OF THE DA:</u>

Questions	optionA_1	optionB_1
1) In addition to what all people in the group own, they have been given 1,000 pounds. You are now asked to choose between	A) 50 people receive a further 1,000 pounds	B) All people receive a further 500 pounds
2) In addition to what all people in the group own, they have been given 2,000 pounds. You are now asked to choose between	A) 50 people lose 1,000 pounds	B) All people lose 500 pounds
Choose between	A) 33 people receive 2,500 pounds; 66 people receive 2,400 pounds; 1 person receives 0 pounds	B) All people receive 2,400 pounds
Choose between	A) 34 people receive 2,400 pounds	B) 33 people receive 2,500 pounds
Choose between	A) 80 people receive 4,000 pounds	B) All people receive 3,000 pounds
Choose between	A) 25 people receive 3,000 pounds	B) 20 people receive 4,000 pounds
Choose between	A) 90 people receive 3,000 pounds	B) 45 people receive 6,000 pounds
Choose between	A) 1 person receives 6,000 pounds	B) 2 people receive 3,000 pounds
Choose between	A) All people receive 0.7 pounds	B) 7 people receive 10 pounds
Choose between	A) 7 people receive 1,000 pounds	B) All people receive 70 pounds
Choose between	A) 80 people lose 4,000 pounds	B) All people lose 3,000 pounds
Choose between	A) 20 people lose 4,000 pounds	B) 25 people lose 3,000 pounds
Choose between	A) 45 people lose 6,000 pounds	B) 90 people lose 3,000 pounds
Choose between	A) 1 people lose 6,000 pounds	B) 2 people lose 3,000 pounds
This game will take place in two rounds and both stages are public. In Round 1, 75 randomly selected people in the group are eliminated and receive nothing. 25 people move on to Stage 2 of the game. In stage 2, you may award one of following packages	A) All 25 people receive 3,000	B) 20 people receive 4,000 pounds

QUESTION 16 OF THE DA (MEASURING INEQUALITY AVERSION THE ETA/EMU):

Imagine now that MANY OTHER PEOPLE WERE INCLUDED IN THE GROUP and that they will be UNIFORMLY DISTRIBUTED BETWEEN A MAXIMUM AND A MINIMUM INCOME AMOUNT YOU HAVE TO CHOOSE. So, for example, if you choose a maximum of 100 and a minimum of 90 pounds, 10% of the people will receive 91 pounds, other 10% will receive 92, and so on up to 100 pounds. If you choose a maximum of 110 and a minimum of 85 pounds, 4% of the people will get 86 pounds, other 4% will get 87 pounds, and so on up to 110 pounds. Indicate among the 6 following distributions the one you would like the people in the group to participate in:

A) People will be uniformly distributed between 308 and 319 pounds

B) People will be uniformly distributed between 264 and 396 pounds

C) People will be uniformly distributed between 220 and 484 pounds

D) People will be uniformly distributed between 176 and 572 pounds

E) People will be uniformly distributed between 132 and 660 pounds

F) People will be uniformly distributed between 22 and 770 pounds

MEASURING TIME CONSISTENCY:

This task involves your choosing on behalf of someone you have never and will never meet between receiving a sum of money on a specific day and another sum of money on another specific day. There is no risk that this person will not receive the money. There are 40 choices to make. The first ten pairs of choices are about receiving £100 today versus receiving a larger amount 7 days from now; the second ten pairs of choices are about receiving £100 in 2 days from now versus receiving a larger amount of money in 9 days from now; the third ten pairs of choices are about receiving £100 in 31 days from now versus receiving a larger amount of money in 38 days from now; the fourth ten pairs of choices are about receiving £100 in 31 days from now versus receiving a larger amount of money in 38 days from now.

REMEMBER THE MONEY WILL BE RECEIVED BY SOMEONE ABOUT WHOM YOU KNOW NOTHING AND THERE IS NO RISK SHE WILL NOT RECEIVE THE MONEY.

Starting with the today versus 7 days from now questions please indicate your most preferred choice across the questions with a single tick (√).

Example: question 1 is asking whether you prefer the referred person to receive £100 today or £101 in 7 days from now. Tick the option you prefer (for all the questions).

	A	B
	Today	7 days from now
Question 1)	£100	£101
Question 2)	£100	£104
Question 3)	£100	£107
Question 4)	£100	£110
Question 5)	£100	£113
Question 6)	£100	£116
Question 7)	£100	£119
Question 8)	£100	£122
Question 9)	£100	£125
Question 10)	£100	£128

	2 days from now	9 days from now
Question 11)	£100	£101
Question 12)	£100	£104
Question 13)	£100	£107
Question 14)	£100	£110
Question 15)	£100	£113
Question 16)	£100	£116
Question 17)	£100	£119
Question 18)	£100	£122
Question 19)	£100	£125
Question 20)	£100	£128

	31 days from now	38 days from now
Question 21)	£100	£101
Question 22)	£100	£104
Question 23)	£100	£107
Question 24)	£100	£110
Question 25)	£100	£113

Question 26) £100	£116
Question 27) £100	£119
Question 28) £100	£122
Question 29) £100	£125
Question 30) £100	£128

301 days from now	308 days from now
Question 31) £100	£101
Question 32) £100	£104
Question 33) £100	£107
Question 34) £100	£110
Question 35) £100	£113
Question 36) £100	£116
Question 37) £100	£119
Question 38) £100	£122
Question 39) £100	£125
Question 40) £100	£128

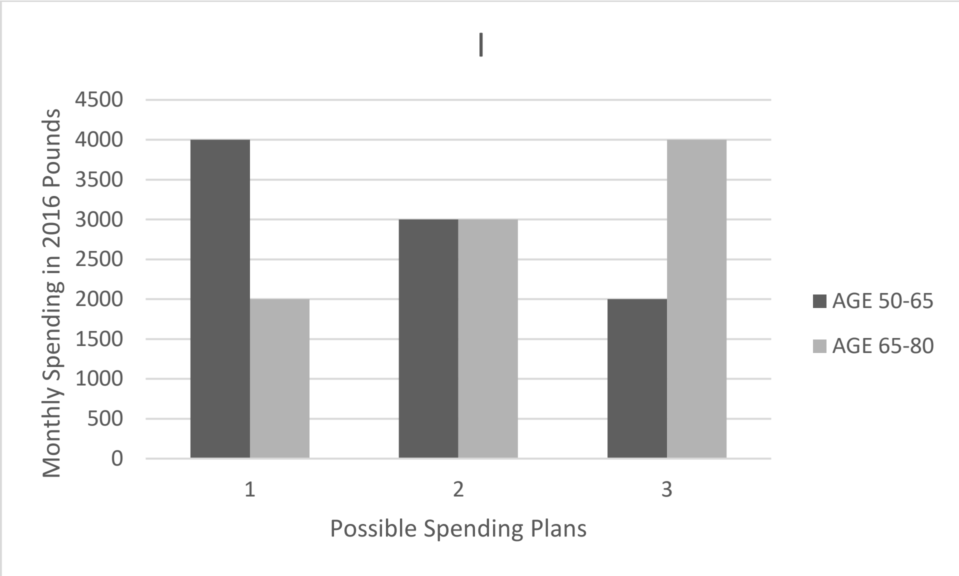
MEASURING THE EMU AS INTERTEMPORAL INEQUALITY AVERSION:[261]

Now I have a few questions about your preferences for OTHER PEOPLE'S spending and saving as they get older. To make the questions comparable for all respondents in the survey, let's suppose that you will have to choose for a group of people who are now 50 years old, that they [and their partners] will live to be 80. Furthermore, suppose that there will be no inflation, and their income after taxes is guaranteed to be £3000 each month from age 50 to age 80.

Card I contains several possible patterns of monthly spending before retirement, the darker bars, and after retirement, the clearer bars. By saving part of their income before retirement, they can have more to spend after retirement, as in choice 3. Or they could borrow and spend more before retirement, spending less and repaying the loan after retirement, as in choice 1.

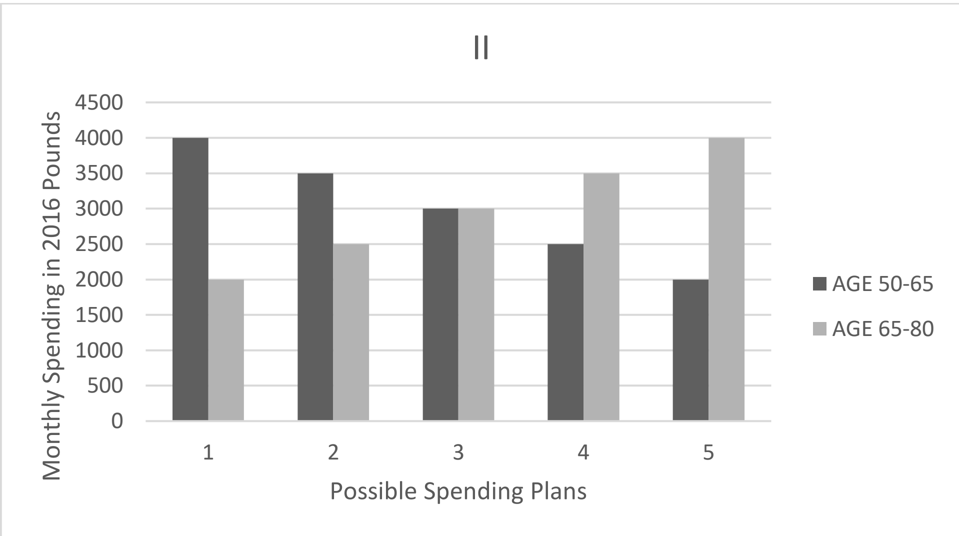
What would you choose for them?

[261] From participants 11 to 32 the questionnaire was slightly different. They were presented with 4 cards (I, II, VI and V), where II had just 4 options. The only difference is that for the first participants some estimated intervals were slightly larger.



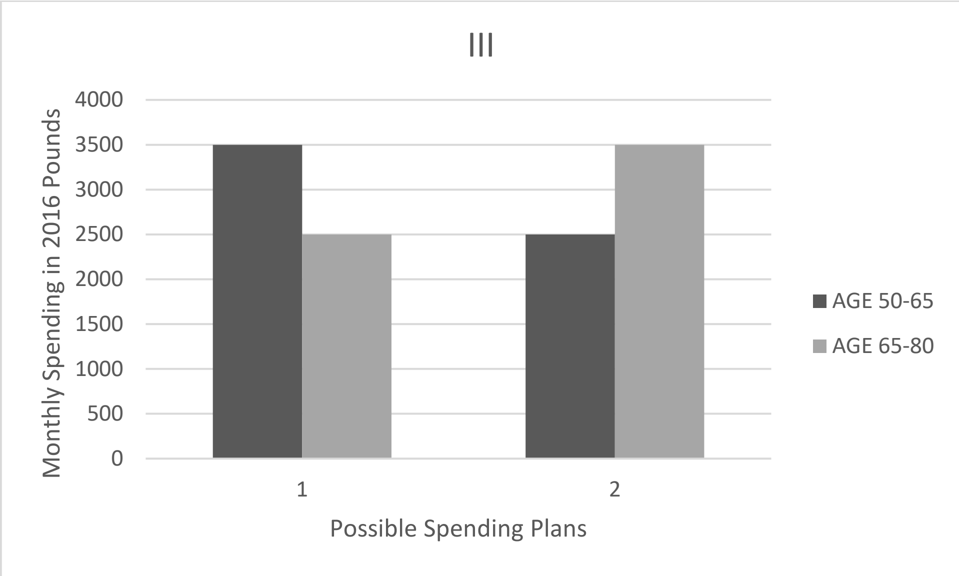
Here is another card (II) with 5 spending patterns for before and after retirement. As before, by saving part of their income before retirement, they can have more to spend after retirement.

What would you choose for them?



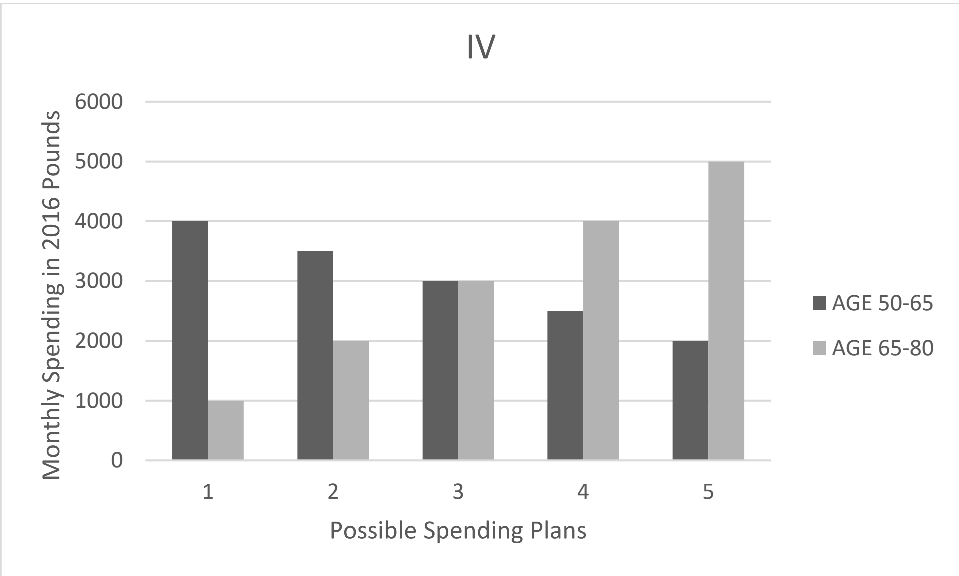
Here is another card (III) with 2 spending patterns for before and after retirement. As before, by saving part of their income before retirement, they can have more to spend after retirement.

What would you choose for them?



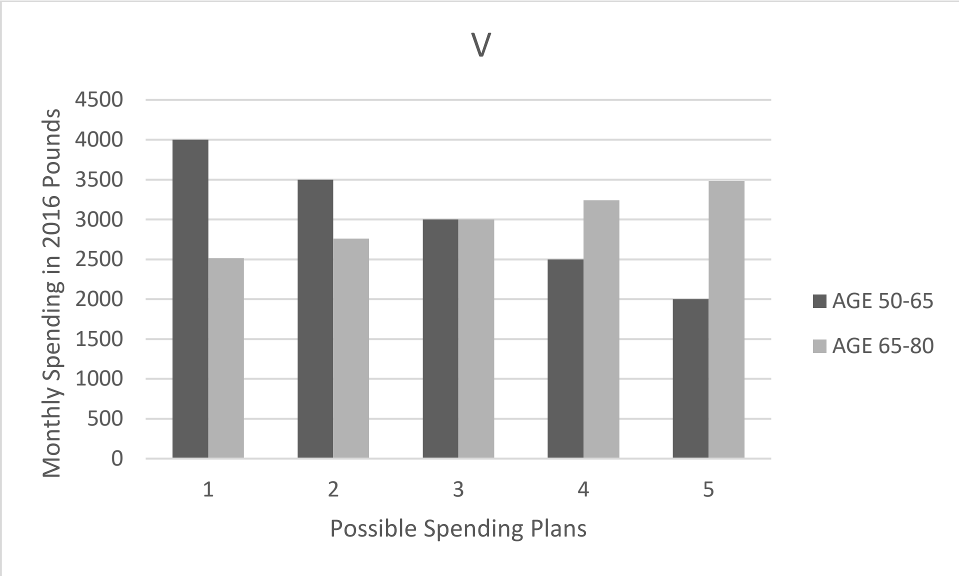
Here is another card (IV) with 5 spending patterns for before and after retirement. As before, by saving part of their income before retirement, they can have more to spend after retirement.

What would you choose for them?



Here is another card (V) with 5 spending patterns for before and after retirement. As before, by saving part of their income before retirement, they can have more to spend after retirement.

What would you choose for them?


V
Monthly Spending in 2016 Pounds
4500
4000
3500
3000
2500
2000
1500
1000
500
0
1
2
3
4
5
Possible Spending Plans
AGE 50-65
AGE 65-80

Appendix 4.B – Tasks on psychological traits assessment

INSTRUCTIONS FOR THE MODEL FREE MODEL BASED TASK:

Welcome to SPACE VOYAGER! You will be playing a game where your goal is to find as many gems as possible. On every turn, you will make a choice between two spaceships, which will bring you to a planet, which may or may not yield a gem. You must read all of the instructions carefully. The experimenter will ask you some questions about the way the task works before you begin.

Choose a spaceship:

or

Get taken to a planet:

or

Collect your yield:

 or 

There are two things for you to learn in order to do well.

1. Each spaceship has a preferred planet. At the start of each turn, you will be able to choose between two spaceships, which appear on the screen. One of these spaceships usually brings you to one of the planets, and the other spaceship usually brings you to the other planet. For example, one spaceship that you choose might bring you to one of the planets on 7 out of 10 turns. But that means that on 3 out of 10 turns, it will take you to the other planet, by mistake. These chances are fixed and you will have a sheet that tells you which spaceship leads to which planet. Learning which spaceships are more (or less) likely to bring you to each planet is very important to playing the game well. If you can do this, you will be able to make good choices that will bring you to the planets that are currently best.

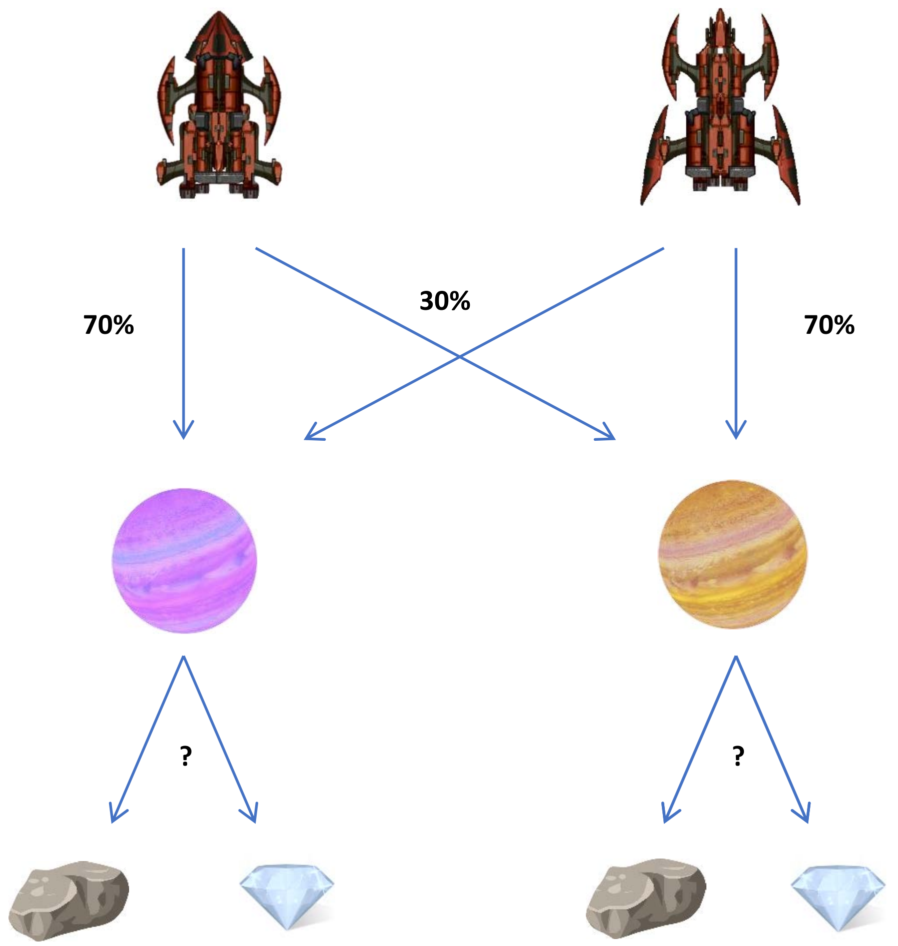
2. You need to keep track of which planets have the highest chance of yielding a gem. If a planet yields a gem, it will appear after the planet. If not, you will see a rock instead. Every time you go to a planet, the computer will decide whether or not that planet will yield a gem based on a 'chance', which has been assigned to that planet. The two planets have different chances of yielding a gem. Importantly, the chances of each planet yielding a gem will change slowly, and independently, over time. It is your job to keep track of which planet is currently best and to try and get to this planet. There are no strange patterns to this game, such as a planet yielding gems on every other choice. The computer is not trying to play tricks on you; it strictly works on the chance assigned to each planet, which will change slowly over time.

Now that you understand these two parts, we will remind you how they fit together in the game you are about to play. On each turn, you have a choice between two spaceships, you will choose a spaceship by clicking on it with the mouse. It will take you to a planet and you will see if that planet yields a gem. After you find out whether or not your planet yielded a gem, you will go back to the start and make another choice and try to earn another gem and so on.

While some of the planets may become very good at times (that is, they often yield a gem), these same planets may become bad later in the game. You need to stay on top of which planets are best. You must use this information to make good choices that are likely to bring you to the planets that are currently good.

INSTRUCTIONS KEPT BY THE PARTICIPANT WHILST GOING THROUGH MODEL FREE MODEL BASED TASK (IN TREATMENT 1):

Spaceship preferences (1):



INSTRUCTIONS FOR THE RME TASK:

Task instructions

I will show you 36 different set of eyes. For each set of eyes, 4 different words will be shown in each corner of the screen.

Choose which word best describes what the person in the picture is thinking or feeling by pressing:

'Q' for the upper left corner

'P' for the upper right corner

'A' for the lower left corner

'L' for the lower right corner

Your answer will be highlighted in blue.

You may feel that more than one word is applicable but please choose just one word, the word which you consider to be most suitable. Before making a choice, make sure that you have read all the 4 words.

You should try to do the task as quickly as possible but you will not be timed. If you really do not know what a word means you can look it up in the definition handout.

Press the spacebar to view an example.

EXAMPLE OF A TRIAL IN THE RME:

SERIOUS ALARMED

ASHAMED BEWILDERED

INSTRUCTIONS FOR THE DIGIT SPAN (PART 1 AND PART 2):

PART 1:

You will see digits appearing on the screen. Read each digit aloud when presented.

Your task is to remember the list of digits in the same order as presented.

After you are showed the complete list of digits, please type your response on the keyboard.

PLEASE DO NOT USE THE NUMBER KEYBOARD THAT CAN BE LOCKED BY THE 'NUM LOCK' KEY.

Press the spacebar to proceed

The blue multiplication sign indicates the beginning of a trial. You will hear a short beep when each digit is presented.

We will start with a serie of 2 digits. After three trials we will add an extra digit to the serie of digits.

Remember that the order of digits is important!

If you do not have any questions, press the spacebar to start.

PART 2

Welcome to the second part!

You will again see digits appearing on the screen.

Now your task is to remember them in the REVERSE order as presented.

In other words, you have to start with the last number I presented.

For example I show you 3 and 7, then you enter 7 and then 3.

Press the spacebar to start!

QUESTIONS AND EXPLAINED ANSWERS FOR THE CRT:

(1) A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost? _____ cents

Say the ball costs X. Then the bat costs $1 more, so it is X + 1. So we have bat + ball = X + (X + 1) = 1.1 because together they cost $1.10. This means 2X + 1 = 1.1, then 2X = 0.1, so X = 0.05. This means the ball costs 5 cents and the bat costs $1.05

(2) If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? _____ minutes

If it takes 5 machines 5 minutes to make 5 widgets, then it takes 1 machine 5 minutes to make 1 widget (each machine is making a widget in 5 minutes). If we have 100 machines working together, then each can make a widget in 5 minutes. So there will be 100 widgets in 5 minutes.

(3) In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? _____ days

Every day FORWARD the patch doubles in size. So every day BACKWARDS means the patch halves in size. So on day 47 the lake is half full.

THE EMPATHY QUOTIENT (EQ) QUESTIONNAIRE:

	Definitely agree	Slightly agree	Slightly disagree	Definitely disagree
1. I can easily tell if someone else wants to enter a conversation.	○	○	○	○
2. I prefer animals to humans.	○	○	○	○
3. I try to keep up with the current trends and fashions.	○	○	○	○
4. I find it difficult to explain to others things that I understand easily, when they don't understand it first time.	○	○	○	○
5. I dream most nights.	○	○	○	○
6. I really enjoy caring for other people.	○	○	○	○

7. I try to solve my own problems rather than discussing them with others.	○	○	○	○
8. I find it hard to know what to do in a social situation.	○	○	○	○
9. I am at my best first thing in the morning.	○	○	○	○
10. People often tell me that I went too far in driving my point home in a discussion.	○	○	○	○
11. It doesn't bother me too much if I am late meeting a friend.	○	○	○	○
12. Friendships and relationships are just too difficult, so I tend not to bother with them.	○	○	○	○
13. I would never break a law, no matter how minor.	○	○	○	○
14. I often find it difficult to judge if something is rude or polite.	○	○	○	○
15. In a conversation, I tend to focus on my own thoughts rather than on what my listener might be thinking.	○	○	○	○
16. I prefer practical jokes to verbal humour.	○	○	○	○
17. I live life for today rather than the future.	○	○	○	○
18. When I was a child, I enjoyed cutting up worms to see what would happen.	○	○	○	○
19. I can pick up quickly if someone says one thing but means another.	○	○	○	○
20. I tend to have very strong opinions about morality.	○	○	○	○
21. It is hard for me to see why some things upset people so much.	○	○	○	○
22. I find it easy to put myself in somebody else's shoes.	○	○	○	○
23. I think that good manners are the most important thing a parent can teach their child.	○	○	○	○
24. I like to do things on the spur of the moment.	○	○	○	○
25. I am good at predicting how someone will feel.	○	○	○	○

26. I am quick to spot when someone in a group is feeling awkward or uncomfortable.	○	○	○	○
27. If I say something that someone else is offended by, I think that that's their problem, not mine.	○	○	○	○
28. If anyone asked me if I liked their haircut, I would reply truthfully, even if I didn't like it.	○	○	○	○
29. I can't always see why someone should have felt offended by a remark.	○	○	○	○
30. People often tell me that I am very unpredictable.	○	○	○	○
31. I enjoy being the centre of attention at any social gathering.	○	○	○	○
32. Seeing people cry doesn't really upset me.	○	○	○	○
33. I enjoy having discussions about politics.	○	○	○	○
34. I am very blunt, which some people take to be rudeness, even though this is unintentional.	○	○	○	○
35. I don't tend to find social situations confusing.	○	○	○	○
36. Other people tell me I am good at understanding how they are feeling and what they are thinking.	○	○	○	○
37. When I talk to people, I tend to talk about their experiences rather than my own.	○	○	○	○
38. It upsets me to see an animal in pain.	○	○	○	○
39. I am able to make decisions without being influenced by people's feelings.	○	○	○	○
40. I can't relax until I have done everything I had planned to do that day.	○	○	○	○
41. I can easily tell if someone else is interested or bored with what I am saying.	○	○	○	○
42. I get upset if I see people suffering on news programmes.	○	○	○	○
43. Friends usually talk to me about their problems as they say that I am very understanding.	○	○	○	○

44. I can sense if I am intruding, even if the other person doesn't tell me.	○	○	○	○
45. I often start new hobbies but quickly become bored with them and move on to something else.	○	○	○	○
46. People sometimes tell me that I have gone too far with teasing.	○	○	○	○
47. I would be too nervous to go on a big roller coaster.	○	○	○	○
48. Other people often say that I am insensitive, though I don't always see why.	○	○	○	○
49. If I see a stranger in a group, I think that it is up to them to make an effort to join in.	○	○	○	○
50. I usually stay emotionally detached when watching a film.	○	○	○	○
51. I like to be very organized in day-to-day life and often make lists of the chores I have to do.	○	○	○	○
52. I can tune in to how someone else feels rapidly and intuitively.	○	○	○	○
53. I don't like to take risks.	○	○	○	○
54. I can easily work out what another person might want to talk about.	○	○	○	○
55. I can tell if someone is masking their true emotion.	○	○	○	○
56. Before making a decision I always weigh up the pros and cons.	○	○	○	○
57. I don't consciously work out the rules of social situations.	○	○	○	○
58. I am good at predicting what someone will do.	○	○	○	○
59. I tend to get emotionally involved with a friend's problems.	○	○	○	○
60. I can usually appreciate the other person's viewpoint, even if I don't agree with it.	○	○	○	○

EQ SCORING:

"Definitely agree" responses scored 2 points and "slightly agree" responses scored 1 point on the following items: 1, 6, 19, 22, 25, 26, 35, 36, 37, 38, 41, 42, 43, 44, 52, 54, 55, 57, 58, 59, 60.

“Definitely disagree” responses scored 2 points and “slightly disagree” responses scored 1 point on the following items: 4, 8, 10, 11, 12, 14, 15, 18, 21, 27, 28, 29, 32, 34, 39, 46, 48, 49, 50.

Items 2, 3, 5, 7, 9, 13, 16, 17, 20, 23, 24, 30, 31, 33, 40, 45, 47, 51, 53, and 56 are just fillers.

THE AUTISM QUOTIENT (AQ) QUESTIONNAIRE:

1. I prefer to do things with others rather than on my own.	definitely agree	slightly agree	slightly disagree	definitely disagree
2. I prefer to do things the same way over and over again.	definitely agree	slightly agree	slightly disagree	definitely disagree
3. If I try to imagine something, I find it very easy to create a picture in my mind.	definitely agree	slightly agree	slightly disagree	definitely disagree
4. I frequently get so strongly absorbed in one thing that I lose sight of other things.	definitely agree	slightly agree	slightly disagree	definitely disagree
5. I often notice small sounds when others do not.	definitely agree	slightly agree	slightly disagree	definitely disagree
6. I usually notice car number plates or similar strings of information.	definitely agree	slightly agree	slightly disagree	definitely disagree
7. Other people frequently tell me that what I’ve said is impolite, even though I think it is polite.	definitely agree	slightly agree	slightly disagree	definitely disagree
8. When I’m reading a story, I can easily imagine what the characters might look like.	definitely agree	slightly agree	slightly disagree	definitely disagree
9. I am fascinated by dates.	definitely agree	slightly agree	slightly disagree	definitely disagree
10. In a social group, I can easily keep track of several different people’s conversations.	definitely agree	slightly agree	slightly disagree	definitely disagree

11. I find social situations easy.	definitely agree	slightly agree	slightly disagree	definitely disagree
12. I tend to notice details that others do not.	definitely agree	slightly agree	slightly disagree	definitely disagree
13. I would rather go to a library than a party.	definitely agree	slightly agree	slightly disagree	definitely disagree
14. I find making up stories easy.	definitely agree	slightly agree	slightly disagree	definitely disagree
15. I find myself drawn more strongly to people than to things.	definitely agree	slightly agree	slightly disagree	definitely disagree
16. I tend to have very strong interests which I get upset about if I can’t pursue.	definitely agree	slightly agree	slightly disagree	definitely disagree
17. I enjoy social chit-chat.	definitely agree	slightly agree	slightly disagree	definitely disagree
18. When I talk, it isn’t always easy for others to get a word in edgeways.	definitely agree	slightly agree	slightly disagree	definitely disagree
19. I am fascinated by numbers.	definitely agree	slightly agree	slightly disagree	definitely disagree
20. When I’m reading a story, I find it difficult to work out the characters’ intentions.	definitely agree	slightly agree	slightly disagree	definitely disagree
21. I don’t particularly enjoy reading fiction.	definitely agree	slightly agree	slightly disagree	definitely disagree

22. I find it hard to make new friends.	definitely agree / slightly agree / slightly disagree / definitely disagree
23. I notice patterns in things all the time.	definitely agree / slightly agree / slightly disagree / definitely disagree
24. I would rather go to the theatre than a museum.	definitely agree / slightly agree / slightly disagree / definitely disagree
25. It does not upset me if my daily routine is disturbed.	definitely agree / slightly agree / slightly disagree / definitely disagree
26. I frequently find that I don't know how to keep a conversation going.	definitely agree / slightly agree / slightly disagree / definitely disagree
27. I find it easy to "read between the lines" when someone is talking to me.	definitely agree / slightly agree / slightly disagree / definitely disagree
28. I usually concentrate more on the whole picture, rather than the small details.	definitely agree / slightly agree / slightly disagree / definitely disagree
29. I am not very good at remembering phone numbers.	definitely agree / slightly agree / slightly disagree / definitely disagree
30. I don't usually notice small changes in a situation, or a person's appearance.	definitely agree / slightly agree / slightly disagree / definitely disagree
31. I know how to tell if someone listening to me is getting bored.	definitely agree / slightly agree / slightly disagree / definitely disagree
32. I find it easy to do more than one thing at once.	definitely agree / slightly agree / slightly disagree / definitely disagree
33. When I talk on the phone, I'm not sure when it's my turn to speak.	definitely agree / slightly agree / slightly disagree / definitely disagree

34. I enjoy doing things spontaneously.	definitely agree	slightly agree	slightly disagree	definitely disagree
35. I am often the last to understand the point of a joke.	definitely agree	slightly agree	slightly disagree	definitely disagree
36. I find it easy to work out what someone is thinking or feeling just by looking at their face.	definitely agree	slightly agree	slightly disagree	definitely disagree
37. If there is an interruption, I can switch back to what I was doing very quickly.	definitely agree	slightly agree	slightly disagree	definitely disagree
38. I am good at social chit-chat.	definitely agree	slightly agree	slightly disagree	definitely disagree
39. People often tell me that I keep going on and on about the same thing.	definitely agree	slightly agree	slightly disagree	definitely disagree
40. When I was young, I used to enjoy playing games involving pretending with other children.	definitely agree	slightly agree	slightly disagree	definitely disagree
41. I like to collect information about categories of things (e.g. types of car, types of bird, types of train, types of plant, etc.).	definitely agree	slightly agree	slightly disagree	definitely disagree
42. I find it difficult to imagine what it would be like to be someone else.	definitely agree	slightly agree	slightly disagree	definitely disagree
43. I like to plan any activities I participate in carefully.	definitely agree	slightly agree	slightly disagree	definitely disagree
44. I enjoy social occasions.	definitely agree	slightly agree	slightly disagree	definitely disagree

45. I find it difficult to work out people's intentions.	definitely agree	slightly agree	slightly disagree	definitely disagree
46. New situations make me anxious.	definitely agree	slightly agree	slightly disagree	definitely disagree
47. I enjoy meeting new people.	definitely agree	slightly agree	slightly disagree	definitely disagree
48. I am a good diplomat.	definitely agree	slightly agree	slightly disagree	definitely disagree
49. I am not very good at remembering people's date of birth.	definitely agree	slightly agree	slightly disagree	definitely disagree
50. I find it very easy to play games with children that involve pretending.	definitely agree	slightly agree	slightly disagree	definitely disagree

<u>SCORING THE AQ:</u>

"Definitely agree" or "slightly agree" responses scored 1 point, on the following items: 1, 2, 4, 5, 6, 7, 9, 12, 13, 16, 18, 19, 20, 21, 22, 23, 26, 33, 35, 39, 41, 42, 43, 45, 46. "Definitely disagree" or "slightly disagree" responses scored 1 point, on the following items: 3, 8, 10, 11, 14, 15, 17, 24, 25, 27, 28, 29, 30, 31, 32, 34, 36, 37, 38, 40, 44, 47, 48, 49, 50.

Appendix 4.C – Drupp et al. (2015) survey test

Imagine that you are asked for your advice by an international governmental organization that needs to determine the appropriate social discount rate for calculating the present value of risk-free cash flows of public projects with intergenerational consequences. For its calculations, the organization needs single values for the components of the social discount rate. While this does not capture all of the important complexities of social discounting, including time horizon-dependent individual discount rates, it does reflect most existing policy guidance on the matter. Your answers will therefore help to improve the current state of decision-making for public investments. Specifically, you are asked to provide your recommendations on the single number, global average and long-term (>100 years) values of the following determinants of the social discount rate:

1. Growth rate of real per-capita consumption [X percent per year].

2. Rate of societal pure time preference (or utility discount rate) [X percent].

3. Elasticity of the marginal utility of consumption [X].

4. Real risk-free interest rate [X percent per year]. Remember that this should be a global average and long-term forecast.

5. What relative weight (summing up to 100 percent) should the governmental body place on the following rationales for determining the social discount rate:

   (a) Normative issues, involving justice towards future generations [X percent], and

   (b) Descriptive issues, involving forecasted average future returns to financial assets [X percent]?

6. What is your recommended real social discount rate for evaluating the certainty-equivalent cash flows of a global public project with intergenerational consequences [X percent per year]?

7. What minimum and maximum real social discount rate would you be comfortable with recommending [X percent to X percent per year]? 8. Do you have any additional comments [X]?

Appendix 4.D – EMU estimates analysis: intervals

Table 4D.1. EMU estimation per context considering all participants

	EMU
RP (risk aversion)	1.485509-2.413786
DA (inequality aversion)	6.989844-11.68653
EIS	4.65625-6.472807

Source: see text. Standard errors are shown in parenthesis.

Table 4D.2. EMU estimation per context regarding the 50% most/least empathetic/reflective

HIGHLY EMP/REF	**RP**	**DA**	**EIS**
H_MFMB	1.264709-2.5207	5.64163-9.96907	4.765167-7.6616
H_CRT	0.9879005-2.183	5.325162-11.578	4.146398-6.5415
H_RME	1.216407-2.2254	7.360069-15.027	4.233577-6.7880
H_EQ	1.38184-2.67818	7.29103-16.2071	3.539261-5.0659
L_AQ	1.171377-2.4701	6.18446-12.2567	4.122461-7.0119
LOW EMP/REF	-	-	-
L_MFMB	1.250395-2.5917	6.278338-17.219	3.662341-5.5848
L _CRT	1.611494-3.0558	6.633779-12.892	4.430905-7.1803
L _RME	1.391348-3.2260	4.453534-9.7002	4.326367-6.9110
L _EQ	1.234697-2.5960	4.704884-9.5382	4.965778-8.0986
H_AQ	1.39441-2.76323	5.980205-13.421	4.175888-6.3410

Source: see text. Standard errors are shown in parenthesis.

Table 4D.3. EMU estimation per context regarding the 30% most/least empathetic/reflective

HIGHLY EMP/REF	**RP**	**DA**	**EIS**
H_MFMB	1.325643-3.4835	4.251437-9.8618	4.341467-8.4936
H_CRT	1.054058-2.3524	4.507457-15.139	4.071101-7.5221
H_RME	1.3165-2.882179	6.300845-17.509	3.84038-6.11204
H_EQ	1.111342-2.9934	5.59933-26.6995	3.612555-5.6658
L_AQ	1.264803-3.0247	5.257327-11.340	3.22143-5.12713
LOW EMP/REF	-	-	-
L_MFMB	1.374933-2.9367	4.621913-21.422	3.43105-5.27117
L _CRT	1.301647-3.5126	5.554348-11.779	4.320541-7.6004
L _RME	0.8054184-2.645	4.077658-11.555	3.861011-7.6465
L _EQ	1.019741-2.9206	3.855784-10.250	4.358516-8.4165
H_AQ	0.8291521-2.244	3.755888-10.890	3.626937-6.4168

Source: see text. Standard errors are shown in parenthesis.

Table 4D.4. EMU estimation per context for participants with more (less) than 24 years old

H_AGE (more than 24 years old)	**EMU**
RP	1.026883-3.687311
DA	0.3972343-16.9174
EIS	3.964503-10.56241
L_AGE (less or equal than 24)	**-**
RP	1.306147-2.309499
DA	6.883885-11.68161
EIS	3.927621-5.660364

Source: see text.

Table 4D.5. EMU estimation per context for participants choosing (not choosing) time consistently for others

	EMU
TC (choosing time consistently for others)_RP	0.2579241-3.735429
nTC (not choosing time consistently)_RP	1.418534-2.400881
TC_DA	NA[262]
nTC_DA	6.061456-10.30831
TC_EIS	4.275277-7.002719
nTC_EIS	4.479242-6.422172

Source: see text.

[262] All participants in this category have EMU above 7.75.

Appendix 4.E – Second and third analyses regressions (EMU estimates)

<u>6 initial regressions for the second analysis:</u>

1) EMU ~ MFMB + RME + DUMrp[263] + DUMda[264] + AGE24 + TIME.CONS

```
                Value Std. Err (Naive SE)      z        p
(Intercept)     0.356   2.4384     2.2865  0.146 8.84e-01
data$MFMB      -0.778   1.5128     1.3176 -0.515 6.07e-01
data$RME        0.198   0.0873     0.0825  2.262 2.37e-02
data$DUMrp     -3.654   0.6332     0.7284 -5.771 7.90e-09
data$DUMda      1.524   0.7982     0.7531  1.910 5.61e-02
data$AGE24      0.303   0.9398     0.7985  0.322 7.47e-01
data$TIME.CONS  2.098   0.7919     0.9167  2.649 8.07e-03
Log(scale)      1.271   0.1045     0.0762 12.165 4.80e-34

Gaussian distribution
Loglik(model)= -331.7   Loglik(intercept only)= -365
        Chisq= 66.51 on 6 degrees of freedom, p= 2.1e-12
(Loglikelihood assumes independent observations)
Number of Newton-Raphson Iterations: 5
n=205 (17 observations deleted due to missingness)
```

```
2) EMU ~ MFMB + EQ + DUMrp + DUMda + AGE24 + TIME

                 Value Std. Err (Naive SE)      z        p
(Intercept)     4.7107   1.4726     1.3843  3.199 1.38e-03
data$MFMB      -0.2602   1.4767     1.3385 -0.176 8.60e-01
data$EQ         0.0186   0.0247     0.0244  0.753 4.51e-01
data$DUMrp     -3.7398   0.6500     0.7399 -5.753 8.75e-09
data$DUMda      1.4137   0.7964     0.7630  1.775 7.59e-02
data$AGE24      0.1326   0.9663     0.8290  0.137 8.91e-01
data$TIME.CONS  2.3174   0.7337     0.9307  3.158 1.59e-03
Log(scale)      1.2889   0.1051     0.0764 12.267 1.37e-34

Gaussian distribution
Loglik(model)= -334.3   Loglik(intercept only)= -365
        Chisq= 61.39 on 6 degrees of freedom, p= 2.4e-11
(Loglikelihood assumes independent observations)
Number of Newton-Raphson Iterations: 5
n=205 (17 observations deleted due to missingness)

3) EMU ~ MFMB + AQ + DUMrp + DUMda + AGE24 + TIME.CONS

                 Value Std. Err (Naive SE)      z        p
(Intercept)     5.2467   1.0522     1.0249  4.986 6.15e-07
data$MFMB      -0.2960   1.4483     1.3428 -0.204 8.38e-01
data$AQ         0.0258   0.0596     0.0510  0.432 6.65e-01
data$DUMrp     -3.8018   0.6356     0.7425 -5.981 2.22e-09
data$DUMda      1.3374   0.7887     0.7632  1.696 9.00e-02
data$AGE24     -0.1235   0.9383     0.8205 -0.132 8.95e-01
data$TIME.CONS  2.4042   0.6997     0.9494  3.436 5.90e-04
Log(scale)      1.2915   0.1060     0.0762 12.179 4.02e-34

Gaussian distribution
Loglik(model)= -334.4   Loglik(intercept only)= -365
```

[263] It represents a dummy indicating EMUs measured as risk aversion.
[264] It represents a dummy indicating EMUs measured as inequality aversion.

```
        Chisq= 61.06 on 6 degrees of freedom, p= 2.7e-11
(Loglikelihood assumes independent observations)
Number of Newton-Raphson Iterations: 5
n=205 (17 observations deleted due to missingness)

4) EMU ~ CRT + RME + DUMrp + DUMda + AGE24 + TIME.CONS

                 Value Std. Err (Naive SE)      z        p
(Intercept)      0.616   2.4729     2.2898  0.249 8.03e-01
data$CRT        -0.162   0.2115     0.2639 -0.766 4.44e-01
data$RME         0.192   0.0859     0.0819  2.230 2.58e-02
data$DUMrp      -3.679   0.6337     0.7300 -5.806 6.42e-09
data$DUMda       1.490   0.7972     0.7549  1.869 6.16e-02
data$AGE24       0.337   0.9025     0.8033  0.374 7.08e-01
data$TIME.CONS   2.140   0.7299     0.9137  2.932 3.37e-03
Log(scale)       1.271   0.1045     0.0763 12.161 5.03e-34

Gaussian distribution
Loglik(model)= -331.7   Loglik(intercept only)= -365
        Chisq= 66.54 on 6 degrees of freedom, p= 2.1e-12
(Loglikelihood assumes independent observations)
Number of Newton-Raphson Iterations: 5
n=205 (17 observations deleted due to missingness)

5) EMU ~ CRT + EQ + DUMrp + DUMda + AGE24 + TIME.CONS

                  Value Std. Err (Naive SE)      z        p
(Intercept)      4.8033   1.4803     1.3830  3.245 1.18e-03
data$CRT        -0.1643   0.2160     0.2678 -0.761 4.47e-01
data$EQ          0.0194   0.0242     0.0242  0.804 4.22e-01
data$DUMrp      -3.7666   0.6474     0.7404 -5.818 5.95e-09
data$DUMda       1.3802   0.7956     0.7637  1.735 8.28e-02
data$AGE24       0.1955   0.9386     0.8347  0.208 8.35e-01
data$TIME.CONS   2.3280   0.6972     0.9269  3.339 8.40e-04
Log(scale)       1.2875   0.1056     0.0764 12.188 3.59e-34

Gaussian distribution
Loglik(model)= -334.1   Loglik(intercept only)= -365
        Chisq= 61.72 on 6 degrees of freedom, p= 2e-11
(Loglikelihood assumes independent observations)
Number of Newton-Raphson Iterations: 5
n=205 (17 observations deleted due to missingness)

6) EMU ~ CRT + AQ + DUMrp + DUMda + AGE24 + TIME.CONS

                  Value Std. Err (Naive SE)       z        p
(Intercept)      5.4056   1.0683     1.0804  5.0600 4.19e-07
data$CRT        -0.1404   0.2183     0.2726 -0.6432 5.20e-01
data$AQ          0.0226   0.0595     0.0513  0.3800 7.04e-01
data$DUMrp      -3.8222   0.6339     0.7431 -6.0296 1.64e-09
data$DUMda       1.3108   0.7874     0.7642  1.6648 9.60e-02
data$AGE24      -0.0669   0.9140     0.8289 -0.0732 9.42e-01
data$TIME.CONS   2.4109   0.6512     0.9446  3.7022 2.14e-04
Log(scale)       1.2905   0.1062     0.0763 12.1471 5.95e-34

Scale= 3.63

Gaussian distribution
Loglik(model)= -334.3   Loglik(intercept only)= -365
        Chisq= 61.28 on 6 degrees of freedom, p= 2.5e-11
(Loglikelihood assumes independent observations)
Number of Newton-Raphson Iterations: 5
```

```
n=205 (17 observations deleted due to missingness)
```

<u>Final regression for second analysis:</u>

```
EMU ~ RME + DUMrp + TIME.CONS

               Value Std. Err (Naive SE)      z        p
(Intercept)     2.10   2.2601     2.1149  0.927 3.54e-01
data$RME        0.17   0.0821     0.0804  2.074 3.81e-02
data$DUMrp     -4.64   0.4925     0.5612 -9.421 4.45e-21
data$TIME.CONS  2.15   0.7312     0.9210  2.942 3.26e-03
Log(scale)      1.29   0.0992     0.0738 13.002 1.20e-38

Gaussian distribution
Loglik(model)= -349.2   Loglik(intercept only)= -381.7
        Chisq= 64.88 on 3 degrees of freedom, p= 5.3e-14
(Loglikelihood assumes independent observations)
Number of Newton-Raphson Iterations: 4
n=217 (5 observations deleted due to missingness)
```

<u>6 initial regressions for the third analysis:</u>

```
1) EMU ~ MFMB + RME + AGE24 + TIME.CONS

                Value Std. Err (Naive SE)      z        p
(Intercept)    -0.328   2.7378     2.7637 -0.120 9.05e-01
data$MFMB      -0.675   1.6806     1.6732 -0.402 6.88e-01
data$RME        0.200   0.1014     0.1045  1.977 4.80e-02
data$AGE24      0.314   1.0620     1.0199  0.296 7.67e-01
data$TIME.CONS  2.146   0.9567     1.1361  2.243 2.49e-02
Log(scale)      1.520   0.0795     0.0749 19.124 1.59e-81

Gaussian distribution
Loglik(model)= -360.9   Loglik(intercept only)= -365
        Chisq= 8.19 on 4 degrees of freedom, p= 0.085
(Loglikelihood assumes independent observations)
Number of Newton-Raphson Iterations: 3
n=205 (17 observations deleted due to missingness)

2) EMU ~ MFMB + EQ + AGE24 + TIME.CONS

                 Value Std. Err (Naive SE)       z        p
(Intercept)     4.2004   1.5076     1.5605  2.7861 5.33e-03
data$MFMB      -0.1459   1.6323     1.6880 -0.0894 9.29e-01
data$EQ         0.0151   0.0282     0.0307  0.5352 5.92e-01
data$AGE24      0.1034   1.0927     1.0524  0.0946 9.25e-01
data$TIME.CONS  2.3738   0.8997     1.1467  2.6385 8.33e-03
Log(scale)      1.5318   0.0806     0.0751 19.0007 1.68e-80

Gaussian distribution
Loglik(model)= -362.6   Loglik(intercept only)= -365
        Chisq= 4.77 on 4 degrees of freedom, p= 0.31
(Loglikelihood assumes independent observations)
Number of Newton-Raphson Iterations: 3
n=205 (17 observations deleted due to missingness)

3) EMU ~ MFMB + AQ + AGE24 + TIME.CONS

                 Value Std. Err (Naive SE)      z        p
(Intercept)     4.5338   1.0899     1.1464  4.160 3.18e-05
```

```
data$MFMB      -0.1670   1.6108     1.6910 -0.104 9.17e-01
data$AQ         0.0243   0.0665     0.0642  0.366 7.14e-01
data$AGE24     -0.1122   1.0533     1.0391 -0.106 9.15e-01
data$TIME.CONS  2.4525   0.8715     1.1669  2.814 4.89e-03
Log(scale)      1.5333   0.0813     0.0750 18.851 2.88e-79

Gaussian distribution
Loglik(model)= -362.6   Loglik(intercept only)= -365
        Chisq= 4.67 on 4 degrees of freedom, p= 0.32
(Loglikelihood assumes independent observations)
Number of Newton-Raphson Iterations: 3
n=205 (17 observations deleted due to missingness)

4) EMU ~ CRT + RME + AGE24 + TIME.CONS

                Value Std. Err (Naive SE)       z        p
(Intercept)    -0.112   2.7520     2.7572 -0.0407 9.68e-01
data$CRT       -0.171   0.2299     0.3325 -0.7433 4.57e-01
data$RME        0.196   0.0989     0.1036  1.9796 4.77e-02
data$AGE24      0.360   1.0288     1.0275  0.3496 7.27e-01
data$TIME.CONS  2.179   0.9030     1.1322  2.4130 1.58e-02
Log(scale)      1.520   0.0800     0.0749 18.9856 2.24e-80

Gaussian distribution
Loglik(model)= -360.8   Loglik(intercept only)= -365
        Chisq= 8.29 on 4 degrees of freedom, p= 0.082
(Loglikelihood assumes independent observations)
Number of Newton-Raphson Iterations: 3
n=205 (17 observations deleted due to missingness)

5) EMU ~ CRT + EQ + AGE24 + TIME.CONS

                 Value Std. Err (Naive SE)      z        p
(Intercept)     4.2825   1.5008     1.5550  2.853 4.32e-03
data$CRT       -0.1642   0.2374     0.3359 -0.692 4.89e-01
data$EQ         0.0158   0.0276     0.0306  0.572 5.67e-01
data$AGE24      0.1707   1.0697     1.0615  0.160 8.73e-01
data$TIME.CONS  2.3765   0.8721     1.1427  2.725 6.43e-03
Log(scale)      1.5311   0.0814     0.0751 18.820 5.22e-79

Gaussian distribution
Loglik(model)= -362.5   Loglik(intercept only)= -365
        Chisq= 5 on 4 degrees of freedom, p= 0.29
(Loglikelihood assumes independent observations)
Number of Newton-Raphson Iterations: 3
n=205 (17 observations deleted due to missingness)

6) EMU ~ CRT + AQ + AGE24 + TIME.CONS

                 Value Std. Err (Naive SE)       z        p
(Intercept)     4.6959   1.1011     1.2123  4.2648 2.00e-05
data$CRT       -0.1413   0.2381     0.3415 -0.5937 5.53e-01
data$AQ         0.0204   0.0662     0.0648  0.3083 7.58e-01
data$AGE24     -0.0479   1.0342     1.0513 -0.0463 9.63e-01
data$TIME.CONS  2.4477   0.8321     1.1617  2.9415 3.27e-03
Log(scale)      1.5329   0.0818     0.0750 18.7429 2.21e-78

Gaussian distribution
Loglik(model)= -362.6   Loglik(intercept only)= -365
        Chisq= 4.83 on 4 degrees of freedom, p= 0.3
(Loglikelihood assumes independent observations)
Number of Newton-Raphson Iterations: 3
```

```
n=205 (17 observations deleted due to missingness)
```

<u>Final regression for third analysis:</u>

```
EMU ~ RME + TIME.CONS

                Value Std. Err (Naive SE)       z        p
(Intercept)    0.0271   2.5906     2.6728  0.0105 9.92e-01
data$RME       0.1889   0.0956     0.1020  1.9752 4.83e-02
data$TIME.CONS 2.1977   0.8911     1.1391  2.4662 1.37e-02
Log(scale)     1.5325   0.0786     0.0734 19.4897 1.34e-84

Gaussian distribution
Loglik(model)= -377.7   Loglik(intercept only)= -381.7
        Chisq= 7.86 on 2 degrees of freedom, p= 0.02
(Loglikelihood assumes independent observations)
Number of Newton-Raphson Iterations: 3
n=217 (5 observations deleted due to missingness)
```

Overall Conclusion

In chapter 1 I estimated the EMU having in view its use in the Ramsey rule (which has been increasingly used to determine the SDR) by the EASA. In doing so I compared different estimation methods inside the EASA framework and tested the constancy of the parameter across income level, countries, time, tax units, and in comparison with another two methodologies. This is, I systematically approached works estimating the EMU via EASA by testing the extent to which variants within the methodology produce different results, examined some of the EASA assumptions (constancy over time and income) and investigated the robustness of the EMU estimates across methodologies and countries in a way that was hitherto lacking.

After reviewing works estimating the EMU by the EASA I described the dataset on income tax obtained from the 9 countries considered (UK, Sweden, Norway, Italy, Ireland, Germany, France, Denmark and Belgium). After presenting results concerning the constancy of the parameter cross income level, countries, time and tax unit, and suggesting a value for the EMU based on the estimates obtained via EASA, I compared the EASA estimates with EMU estimates obtained with two other methodologies: one using data on consumption growth rate and other using data on insurance.

The value suggested for the parameter was 1.5. It was the estimate obtained utilizing all the data and the best (according to the study) methodology within the EASA (considering NICs, not considering SI and using single people as tax units). The constancy tests show the parameter varies with time and tax unit, and that although it also varies significantly across countries and income, the variation is very small. It suggests that not considering changes of the EMU through time may be an issue, but that it is reasonable to assume it is constant over income level. Moreover, although it can be regarded as robust across the countries considered, but not across tax units. The results regarding the comparison of different approaches in the EASA show the method is more suitable for single people as tax units.

The estimates obtained with the EASA are consistent with estimates obtained with the two other methodologies mentioned above. It reinforces the value of 1.5 suggested by the EASA. It also must be said that the methodologies compared are similar in that they are based on revealed preferences of the overall population, but different not just in terms of the estimation procedure, but also in that the EMU is interpreted differently in each approach.

The main shortcoming of the EMU estimate obtained with the EASA in chapter 1 is that the method assumes Governments set income tax schedules based solely on the EASP; something which can be disputed. Moreover, other ways to obtain the SI can be devised, such that the ones estimated in the study can be challenged.[265]

For future work it would be interesting to include other countries in the analysis, given all 9 nations considered are similar in many aspects. It would enrich the study by providing more material to test EMU estimates' robustness across countries. The reason more nations were not included is that to organize income tax information in order to obtain a large amount of observations on the ATR and MTR for a wide range of income levels in a dataset as the one used to derive the results would be significant research in itself. It would be, nonetheless a useful contribution. The necessary data to start to fill this gap might be found by simulating taxes according to the legislation of the different countries (as is the case for the dataset utilized herein). Yet another idea would be to expand the research in the sense of estimating the EMU using the general tax burden instead of just income taxes. This is, to derive the ATR and MTR e.g. including sales taxes.

In chapter 2 I investigated the determinants of the EMU parameter as estimated by the EASA. The variables hypothesized to affect the EMU were selected among those expected to affect tax progressivity, given the parameter can also be regarded as an index of tax progressivity. A fixed effects model with time trend was applied as the empirical strategy of investigation.

Notice the parameter concerned is seen as a constant given that the EASA assumes the iso-elastic utility function. Nonetheless, it is reasonable to think there are factors influencing people's (and Governments') inequality aversion as shown by the degree of progression of income tax schedules and thus that there are factors causing the parameter to vary. Another assumption underlying the EASA is that Government's inequality aversion represents peoples' inequality aversion because of the democratic process (obviously it only applies to democracies), which has not been clearly challenged so far.

In chapter 2 I allow for the possibility that the EMU, as measured by the EASA, varies according to political, economic and demographic variables, and allow instead for the possibility that Governments do not necessarily represent peoples' preferences, but are influenced by other factors in the way they set the progressivity level of income tax schedules. In doing so I advance the debate on the legitimacy of the EASA technique.

[265] As an example, one could think of the minimum wage as an alternative way to obtain a reasonable SI value instead of individual's the tax-free income used to estimate the SI herein.

After introducing the chapter and establishing that the EMU measures income tax progressivity (as well as inequality aversion), I reviewed works investigating the determinants of tax progressivity in order to get insight into possible factors determining the EMU. After, I give more details on the data used and on the empirical strategy implemented. Next, I present and discuss the results regarding the analysis and conclude.

The results show basically that the dependency ratio,[266] Government stability, openness, population size and governmental expenditure with education are all factors affecting the EMU. From these variables only the population size strongly affects the EMU, which in my opinion weakens the reliability of the EASA, but not much, given the evidence favouring the technique found in chapter 1 and the fact all other factors considered fail to cause strong variation on EMU estimated by the EASA.

In future work I think it would be important to revisit the explanations regarding the mechanism through which the explanatory variables considered affect the EMU – be it seen as reflecting people's or Governments' inequality aversion – and test whether some of them should be given preference in view of statistical evidence. For example, I argued population size may affect the EMU by affecting societal inequality aversion via the 'sense of community' (McMillan and Chavis, 1986) or by affecting governmental decision-taking by the remoteness of the central Government from the people. It would be interesting to gather statistics better measuring these two aspects and test which explanation is more appropriate. The research proposed would shed light on the extent to which the setting of income tax schedules by Governments depends on the general population's will.

In chapter 3 I investigate causality between income inequality and income tax progressivity. The study builds on Scheve and Stasavage (2016), which dedicate part of their study to the same issue. The main difference is that besides using other progressivity and inequality indexes, we apply new causality tests to approach the issue.

In the literature review I identify SS as the only work formally testing (i.e. by using statistical causality tests) causality between progressivity and inequality.[267] Other works are found addressing the same issue, but not formally. They suggest that if the relation between the variables existed before it does not anymore.

[266] Percentage of those with more than 65 and less than 14 in the population.
[267] It must be noticed their study was published when the present one was ongoing.

While SS used top marginal income tax rates as index of progressivity, we used two standard progressivity measures, the Liability Progression (LP) and Residual Progression (RP), both first presented by Musgrave and Thin (1948). We argued these are better measures of progressivity mainly due to the fact that, differently from top rates, they stem directly from the general definition of tax progression (that the average tax rate grows with income), and are evaluated at a representative income level (APW). Moreover, we used GINI coefficients to measure income inequality instead of top income shares as SS. The advantage of the GINI is that they take into consideration the whole income distribution, while the top income shares are restricted to top incomes.

Besides using standard Granger non-causality tests (similar to SS) we also used Dumitrescu and Hurlin's (2012) Granger non-causality in heterogeneous panels tests. The main difference between the two methodologies is that the latter allows for parameters to vary across panels, which thus produces tests that reflect better the variability across the countries considered. Also, I tested the variables stationarity before using them in the causality analysis – which is not done by SS.

After introducing the chapter, I reviewed the literature concerning the matter under consideration and described the methodology regarding the measurement of income inequality and income tax progressivity. Then I gave more details about the data used to perform the analysis. Next, I presented the results concerning the causality and stationarity tests and concluded the chapter.

The results showed causality in both directions, i.e. from progressivity to inequality and vice-versa. It suggests Governments take into consideration the level of income inequality to set income tax schedules, which challenges SS's conclusions. In terms of policy making it means administrations responsible for setting income tax schedule are sensitive to arguments regarding income inequality. It confirms the widespread assumption that Governments do use income tax to address distributional issues – even if some think that this link is now broken and that Governments have given up trying.

It must be noted that the RP and LP indexes are local progressivity measures calculated at the Average Production Wage (APW) (which is the most representative income level), and therefore do not consider progressivity for the whole income distribution. Also, it is important to notice the GINI coefficients are derived with slightly different income definitions and that I made use of interpolations to complete the GINI series.

In the fourth chapter I investigated how psychological traits perceived as socially desirable (empathy, reflectiveness, time consistency when choosing for others and maturity) affect the value of the EMU parameter. Such relationship sheds light on the normative significance of the EMU estimates. I also examined how such traits affect the EMU context-sensitivity, which gives insight into how well the Standard Welfare Economic Model (SWEM) describes normative behaviour. The EMU was estimated in three different contexts: as risk aversion, inequality aversion and the Elasticity of Intertemporal Substitution (EIS).

After introducing the chapter I review the literature on works relating psychological traits to the EMU and to the EMU context-sensitivity. There are works relating some of the traits considered with the value of the EMU, but none of them are primarily concerned with the EMU normative significance. This is, they do not use desirable traits and their potential relation with normative behaviour in order to derive more relevant EMU estimates. Moreover, there are no works relating psychological traits with EMU context-sensitivity.

In the methodology section I described the experiment undertaken in order to gauge the EMU estimates in the different contexts and to gauge the traits. I also described the data collected and the empirical strategy adopted to verify how the traits affect the parameter and its context-sensitivity. In both analyses I use different specifications of interval regressions.

Regarding the EMU context-sensitivity analysis the results showed basically that more empathetic people treat contexts more differently. More specifically, the EMU as inequality aversion is not significantly different from the EMU as the EIS for the low empathetic. The same qualitative results are found for those choosing time consistently for others. The more mature and reflective participants, on the other hand, treated the different contexts more equally. It suggests the SWEM describes better people who are highly reflective, mature, low empathetic and do not choose time consistently for others.

Regarding the relationship between psychological traits and the EMU value, the results show basically that the EMU tends to be higher for highly empathetic people, which suggests the ideal EMU is higher than indicated by EMU estimates derived with samples stemming from the general population (assuming empathy is a desirable trait and that psychological experiments are a better way of determining the EMU than the revealed preference estimates derived in earlier chapters).

It is important to note the tasks designed to gauge the EMU are based on the SWEM, such that the value estimated is limited by the assumptions made in this framework (which imply the parameter is constant across the three contexts mentioned before). Due to time and resources constraint it was not possible to obtain a sample as large and representative of the general population as I had desired. It caused me to have fewer observations on specific groups, such as those above 24 years old and non-students; also it would be ideal to have at least about 300 participants.[268] Finally, most of the tasks were not incentivized, which can give room for questions concerning the results.

A potential future work stemming from chapter four would be to repeat the experiment using a more representative and larger sample. Another suggestion regarding future works would be related to using experiments based on models relaxing some of the assumptions of the SWEM (as for example using a utility function different from the iso-elastic) in order to obtain more accurate values for the EMU.

Another outstanding challenge is to explain the apparently large difference between experimental and revealed-preference based estimates of the EMU. The main difficulty regarding this kind of investigation would be to obtain revealed preference data on the decisions made by individuals who are known to have particular psychological traits. The currently available microeconomic datasets do not give information which is relevant to determine respondent's key psychological traits such as the degree of empathy and reflectiveness, making it difficult to specify the agents behind the choices presented in the data.

[268] A sample of about 300 participants is what most works reviewed in the relevant chapter have.

References

Dumitrescu, E., & Hurlin, C. (2012). Testing for Granger Non-causality in Heterogeneous Panels. Economic Modelling, 29, 1450-1460.

McMillan, D.W. & Chavis, D.M. (1986). Sense of community: A definition and theory. Journal of Community Psychology, 14(1), 6-23.

Musgrave, R. A., & Thin, T. (1948). Income Tax Progression, 1929-48. Journal of Political Economy, 56, 498-514.

Scheve, K., & Stasavage, D. (2016). Taxing the Rich: a History of Fiscal Fairness in the United States and Europe. Princeton: Princeton University Press.