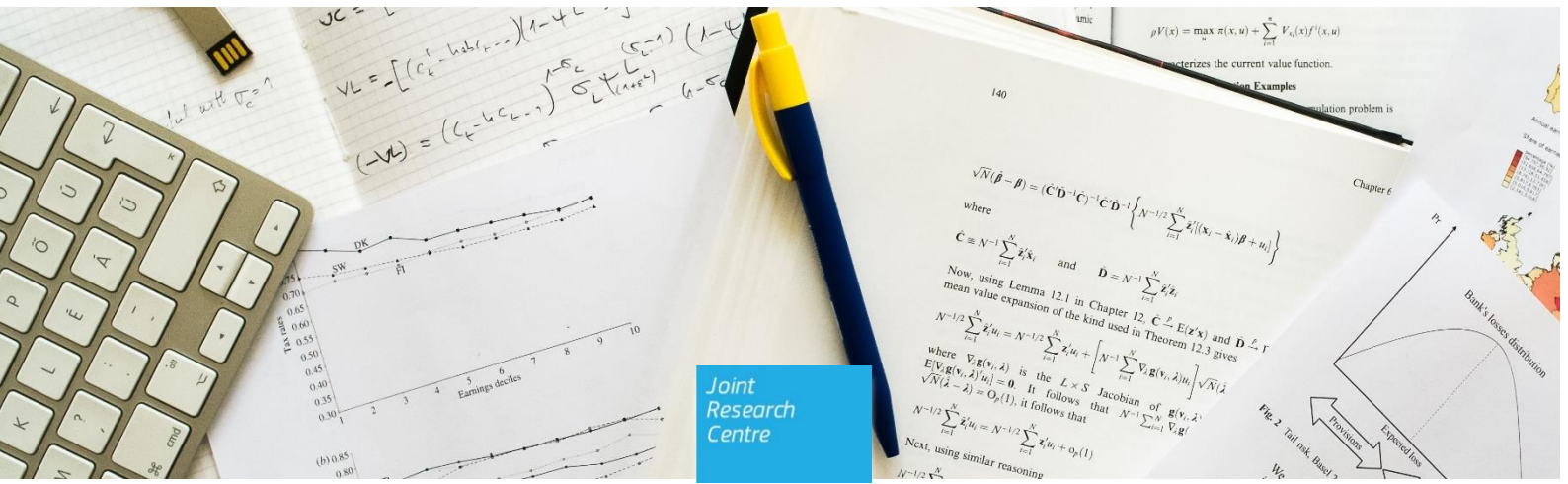
European Commission

JRC TECHNICAL REPORTS

EU-wide income inequality in the era of the Great Recession

Péter Benczúr
Zsombor Cseres-Gergely
Péter Harasztosi

JRC Working Papers in Economics and Finance, 2017/14



This publication is a Technical report by the Joint Research Centre (JRC), the European Commission's science and knowledge service. It aims to provide evidence-based scientific support to the European policymaking process. The scientific output expressed does not imply a policy position of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use that might be made of this publication.

Contact information

Name: Péter Benczúr
Address: European Commission, Joint Research Centre (JRC), Via E. Fermi 2749, 21027 Ispra, Italy.
E-mail: peter.benczur@ec.europa.eu
Tel.: +39 033278-5772

Name: Zsombor Cseres-Gergely
Address: European Commission, Joint Research Centre (JRC), Via E. Fermi 2749, 21027 Ispra, Italy.
E-mail: zsombor.cseres-gergely@ec.europa.eu
Tel.: +39 033278-9681

Name: Péter Harasztosi
Address: European Commission, Joint Research Centre (JRC), Via E. Fermi 2749, 21027 Ispra, Italy.
E-mail: peter.harasztosi@ec.europa.eu
Tel.: +39 0332 783837

JRC Science Hub
https://ec.europa.eu/jrc

JRC109805

PDF ISBN 978-92-79-77303-7 ISSN 2467-2203 doi:10.2760/244655

Luxembourg: Publications Office of the European Union, 2017

© European Union, 2017

The reuse of the document is authorised, provided the source is acknowledged and the original meaning or message of the texts are not distorted. The European Commission shall not be held liable for any consequences stemming from the reuse.

How to cite this report: Benczur P.; Cseres-Gergely Zs.; Harasztosi P. EU-wide income inequality in the era of the Great Recession, JRC Working Papers in Economics and Finance, 2017/14, doi:10.2760/244655

All images © European Union 2017

EU-wide income inequality in the era of the Great Recession

Péter Benczúr[1], Zsombor Cseres-Gergely[1] and Péter Harasztosi[1]

[1]*Joint-Research Centre, European Commission*

Abstract

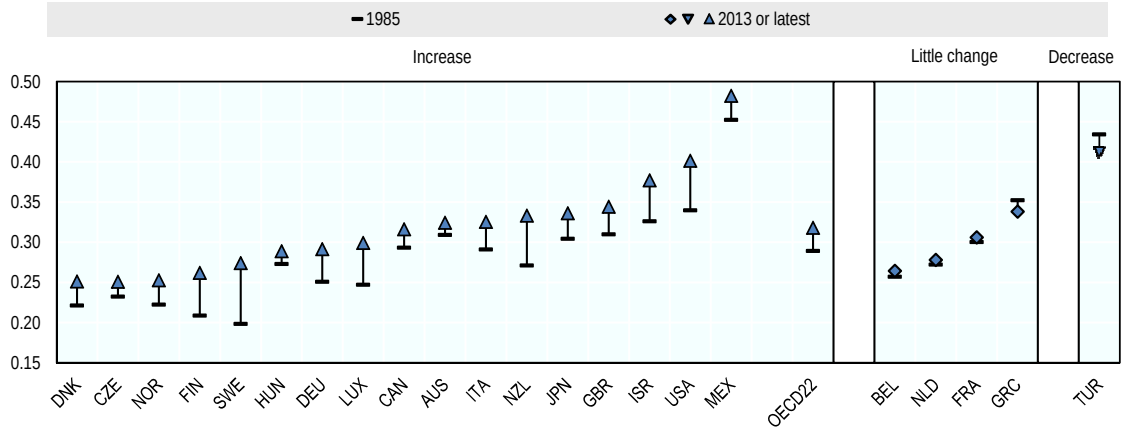
This paper uses European micro-data to look at stylised facts of EU-wide income inequality during the 2006-2013 period. Our contribution is to bring together four elements of the analysis that has appeared only in separation so far. Our analysis is EU-wide, but regionally detailed, looks at the longest possible term with harmonized survey data, uses inequality indicators sensitive to different parts of the income distribution and shows the contribution of different income components to income inequality. To our knowledge, this is the first attempt to produce stylised facts with these features at the same time, which, we believe is important in order to ask and answer questions related to the welfare of the people of the EU as a whole. We also take great care of rendering our work as homogeneous and transparent as possible. Because of this, we believe that the evidence we provide can be a starting point to a line of research on EU-wide income inequality.

We would like to thank the comments of György Tóth István made as a discussant at the 2016 conference of the Hungarian Society of Economics (MKE). We also thank seminar participants in Ispra, Brussels, at MKE 2016 and the comments of Stefano Filauro. We thank Virmantas Kvedaras for his ideas and significant contributions that helped us a lot in setting up a consistent dataset. We would also like to acknowledge the help of Nadja Kernchen at the initial phase of the work during the spring and summer of 2016 when she was an intern at the JRC. Naturally, all errors are ours only. *Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use which might be made of the following information.*

1 Introduction

The crisis of 2009 has resulted in an unprecedented loss of income in many countries, among others EU member states. This income loss was not only large in size, but also highly uneven in its distribution. Reports of the OECD have monitored several aspects of this process and have warned about its potential adverse effects – see OECD (2008, 2011, 2013, 2015). The message of these analyses is clearly formulated: inequality has increased during the crisis and it has undesirable effects on multiple layers of socio-economic outcomes. A number of documents produced by the European Commission share this concern. Recently, Maquet et al. (2015) has showed that the former convergence pattern of inequality across member states has stopped, and did so mostly driven by a change in market income inequality. Changes in and the importance of access to services creating equality of opportunity such as health and education are pointed out, along with the potentially detrimental effect of increased inequality on growth.

Figure 1: Income inequality trends in selected Member States and OECD countries



Source: OECD (2015), Figure 1.3. The figure displays the Gini coefficient of household disposable income. 'Little change' in inequality refers to changes of less than 1.5 percentage points. Data year for 2013 or latest year (2013 for FI, HU, NL and the United States, 2009 for Japan, and 2012 for the other countries).

This increase in inequality is not specific to the crisis period. There is clear evidence that income inequality has increased markedly in EU countries since the mid-1980s (see Figure 1 for selected Member States and other OECD countries). This development is mainly due to a broadening the gap between the top and the bottom earners. From the mid-1980s to 2008 (OECD & The World Bank 2012), the average annual real disposable

income of the richest 10% increased more than 2.5 times faster than that of the poorest 10%.

Analysis concentrating mostly on inequality developments within individual countries has a hard time in putting changes in the income distribution of a given country into the context of income levels and distribution of its peers. One cannot judge, for example, whether various income segments of New Member States in the former Eastern block have been closing their income gaps relative to corresponding income segments of EU countries in North-west Europe. Or in other words, changes in between-country income disparities will not be captured in any set of country-specific calculations. For example, the substantial work documented in Nolan et al. (2014) gives very interesting details about countries, but these are often specific to them only. Indeed, the synthesis of Salverda et al. (2014) cannot pool these formally, and needs to bridge the gaps verbally. Looking at the income distribution of the whole EU solves this problem, albeit at the cost of compromises to be discussed later. Nevertheless, we believe that the gains are substantial and that such a distribution is also an important object on its own right. Our first objective is thus to study the whole EU-wide income distribution and also its various subsets.

Such an EU-wide approach has major advantages relative to a country-based one, both for political-philosophical (at the moment mostly normative) reasons, and for that of economic theory too (mostly positive). On the one hand, the current leadership of the Union is considering measures and institutions that strengthen the cooperation between member states to an extent of providing EU-wide homogeneous minimum services. With one of the first of such institutions, the European Pillar of Social Rights already being approved, it is imperative to understand the pan-European social situation and within that, income inequality. On the other hand, the existing integration of Europe into world trade and the increasingly integrated internal markets both require that we think about Europe as a common market for goods, labour, and also income.

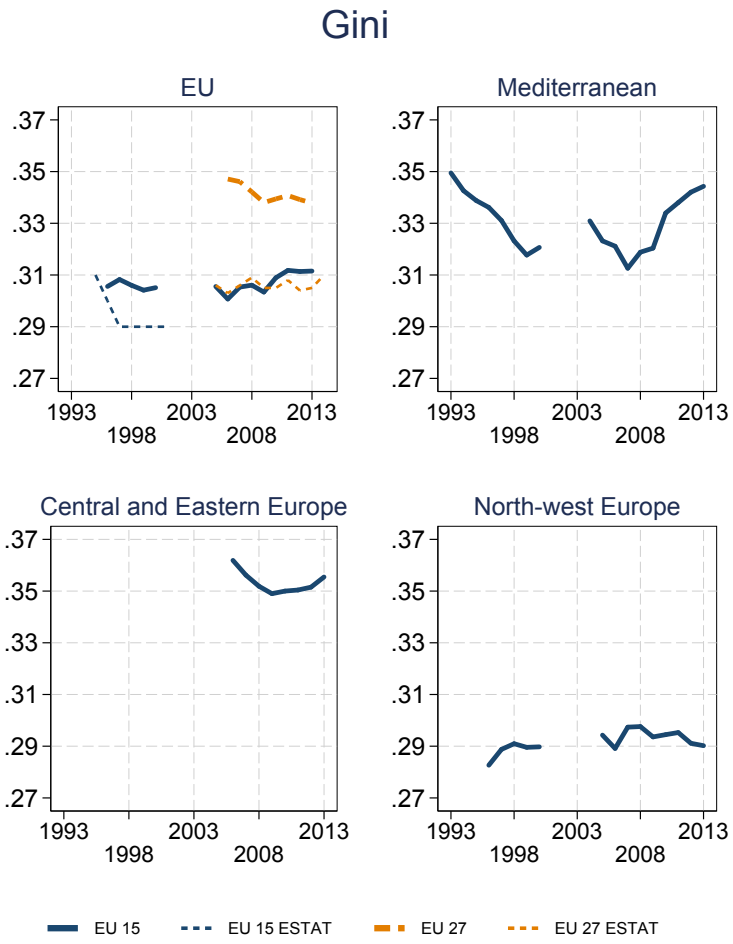
One of the difficulties of such an EU-wide approach is the availability of data. Though there are suitable micro-level datasets which are comparable across countries (most importantly, the European Community Household Panel and the EU Survey on Income and Living Conditions – the ECHP and the EU-SILC, respectively), it is not straightforward to merge the income data across different countries. As discussed and

demonstrated by Brandolini (2007), major issues include the conversion into common currencies, controlling for price (purchasing parity) differentials, and the choice of common or country-specific corrections for family size (equalisation). Moreover, consumption expenditures are not included in the EU-SILC, so the analysis has to stop at per capita equivalised net household income as the closest proxy of utility. Also, it is important to realise that harmonised data come most often from surveys, which themselves have limitations. One of these is the coverage of the very rich. We have to live with the fact that we cannot contribute much to the otherwise lively debate on the contribution of 'top inequality' to the overall one.

As opposed to country-studies, when looking at the income distribution of, and the development of inequality within the EU as a whole, a different pattern emerges. Figure 2 on the next page displays the evolution of a frequently used inequality measure, the Gini coefficient of household net income for the EU15 and two groups of countries, during the years 1993-2013. Until 2000, inequality within the EU was mildly decreasing, whereas during the 2000s, a similarly mild increase is to be observed. Considering the EU27 (existing after 2006 only), we observe that despite the similar effect of the crisis, the trend of inequality is downward-sloping at the EU27 level, while the average Gini has slightly increased after 2006. Observe that the official figures, based on the average of country-specific Gini indices, are almost always below the Gini calculated from a pooled population (for the EU15 before and for the EU27 after 2003).

The behaviour of the EU-wide income inequality is affected by within- and between-country income differences. As Milanovic & Lakner (2013) points out on a global scale, it is possible that while inequality increases within countries, global inequality decreases. By the same token, it is also possible that inequality increases within many member states of the EU, but EU-wide inequality decreases at the same time. As we shall see later, this happens when levels of income in member states converge as it happened during the 2000s. Filauro (2017), a paper similar to ours in its motivation and use of the EU-SILC, shows that between-country differences account for 20% of the inequality measured in the full EU income distribution.

Figure 2: Gini of per capita household income in the EU countries (net)

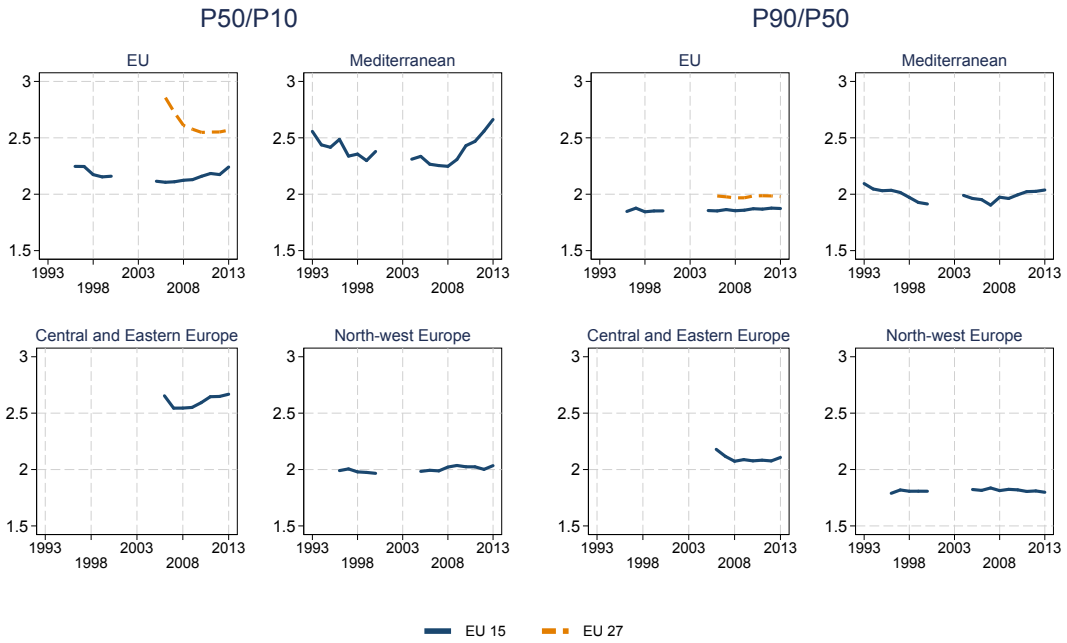


Source: Own calculations from the ECHP and the EU-SILC. Croatia is not covered, due to limited data availability. Series marked with 'ESTAT' are ilc_di12 Eurostat indicator values.

Looking at country-groups – which we define later and use for the rest of the paper –, Figure 2 shows substantial differences. Members of the EU15 had very different experiences during the two decades, as the initial decrease was a net effect of a sharp decrease in the Mediterranean countries and some increase in the North-west. Events of the 2000s formed a mirror image of that in the 1990s: inequality in the Mediterranean increased to 1993 levels by 2013, while that in the North-west decreased. Relative to the EU15, the full EU27 income inequality is much higher, due to the large income differences between Central and Eastern Europe and other parts of the EU. The economic crisis of 2009 had seemingly little effect on the EU-wide inequality, but it did have a rather important, trend-setting effect on the Mediterranean country group.

Adding the EU-wide perspective in Figure 2 yields interesting insights already, but it

Figure 3: Quantile ratios of per capita household income in EU countries (net)



Source: Own calculations from the ECHP and the EU-SILC. Croatia is not covered, due to limited data availability.

remains in the realm of usual inequality analysis in many ways. It uses a single statistics, the Gini that is an overall measure of income inequality. However, there is evidence that major events and shocks such as the one in 2009 have heterogeneous effects on different parts of the income distribution. An increase in the inequality as measured by the Gini coefficient can come from a worsening of bottom or an improvement of top incomes. Moreover, since a drop in income has more severe consequences for the poor, an income drop with decreasing inequality is compatible with a rise of poverty risk, as suggested by Honohan (2005). There is also evidence that the burden of debt service was much more severe in the bottom of the income distribution – see ECB (2013). Our second aim is thus to use indicators beyond the Gini to provide a richer characterisation of inequality retaining nonetheless the unified framework we are working in.

Working with microdata allows not only creating economically meaningful groups of many kind, but also calculating further statistics. Figure 3 shows the P50/P10 and the P90/P50 quantile ratios in the same spirit as we have seen the Gini before. By assigning separate inequality indicators to the lower- and upper part of the distribution, it is possible to see that much of the development we have observed comes from the

lower and only to little or no extent from the upper part of the distribution. Again, this process was driven mostly by changes for the Mediterranean, so powerful that it could induce a decrease and subsequently an increase in the overall EU15 inequality, despite the relatively small share of this country group. After the enlargement in the beginning of the 2000s, the accession of Central and Eastern European countries increased the EU-wide (now EU27-wide) inequality to previously unprecedented levels. This has subsequently decreased after the times characterised by the outbreak of the crisis. CEE and Mediterranean income levels have converged, but inequality in the Mediterranean countries has increased strongly.[1]

Yet another characteristic of most existing inequality research is that it looks primarily at the inequality of household net income, just as we have done before. When one needs to be concise and can only use a single chart to show a welfare effect, household equivalised net income is arguably the most relevant income concept to base discussions. It includes all available financial resources at the household level, and incorporates the effect of the size of the household, including economies of scale. Not having consumption data, it is the closest proxy to utility, especially if it is credibly corrected for differences in prices. However, a focus on a single income measure precludes the separation to factors contributing to final observed income inequality. These include different sources of income such as labour- and asset income, automatic stabilisers, such as unemployment benefits, but some on-off, crisis-related measures themselves might have compressed the income distributions. Domnisoru (2014) for example documents a larger than average drop in inequality traced back to cuts in relatively high public sector wages on the higher, the kick-in of automatic stabilisers and the introduction of a minimum guaranteed pension on the lower end. Our third aim is to take into account of such income elements to identify how the observed, final income inequality is built up.

Our approach is based on the individual, and traces the 'production chain' of utility within the household. In this framework, each income element is attached to the individual and yields an income stream whose level and volatility is equally important. It is built on the income-consumption insurance mechanism of Blundell & Preston (1998) and Blundell et al. (2008), summarized in Blundell (2011). In order to operationalise

[1]Note that the convergence could have been largely affected by the crisis, but also by other forces. One of these is the effect of the new member states' start of using EU structural funds and its spillover-effects.

this concept, we adapt the approach used in the special issue (13 of 2010) of the Review of Economic Dynamics, presenting analysis of the evolution of inequality in a unified framework (Krueger et al. (2010)). Although the issue includes articles on EU member states such as Great Britain, Germany, Italy, Spain and Sweden, our analysis is more akin to that of Heathcote et al. (2010) of the US, in the sense that it discusses a heterogeneous political entity. Our approach is not able to explicitly incorporate dynamics into the analysis, but by considering annual income, it subsumes the effect of moving out of and into different labour market statuses.[2]

Beyond producing stylized facts, we believe that the construction of a database of ten years (ten plus ten in case of household income) of the full EU income distribution is also an important achievement. In the forthcoming JRC Report on Fairness, and its preview in the JRC 2016 Annual Science Lecture,[3] we have already used this dataset to produce the EU version of the well-known 'elephant chart' of Milanovic & Lakner (2013). We hope to see an increasing number of studies following a similar approach, potentially utilizing our dataset.

From our results, there are a few specific observations worth emphasizing, most of them concerning different geographic patterns within Europe. We find that the relative average stability of the ultimate utility measure, net household equivalised income masks important heterogeneity at the country-group level. The increase of inequality in the second half of the 2010s was very strong for the Mediterranean, but rather weak or non-existent for the North-west and the Central and Eastern European country groups. This decrease of inequality in the Mediterranean before 2000 has steered downwards also the EU15 inequality. Later this pattern was reversed, inequality started to increase both in the Mediterranean and in the EU15, while it stayed rather flat in North-west Europe EU. EU level inequality increased moderately with the enlargement (i.e., moving from EU15 to the EU27 in our data), due to large income disparities between new and old member states. Though inequality in the EU27 decreased afterwards, its 2013 level

[2]One of the few papers looking explicitly at the dynamics of inequality is Raitano (2016), using longitudinal EU-SILC microdata pooled for multiple countries. He analyses the change of the Gini index for four country groups in the EU between 2008 and 2011, that is during the recovery from the crisis. His interest centres on the contribution of different income types and redistribution to the overall change. He finds that the decline of inequality after the 2009 crisis was largely due to government intervention. Working with a four-year longitudinal sample enables him to look at and incorporate zero incomes due to job loss, yielding higher levels of measured inequality.

[3]https://ec.europa.eu/jrc/sites/jrcsh/files/20161010-annual-lecture-2016_en.pdf

in the whole EU27 was higher than it had ever been before in the EU15. The strength of this finding varies across income concepts, but it is rather robust, and importantly, it is local to the lower part of the distribution, not shown by completely aggregate indicators.

The paper is structured as follows. Section 2 discusses our data sources, and lists all the cleaning and adjustment steps we implemented. Section 3 presents the stylised facts and their analysis. Section 4 concludes, while the Appendix offers further details about the data and some additional results.

2 Data and methods

Our detailed, individual-based approach requires the use of microdata to calculate various statistics of inequality for various subsamples and various income concepts. To encompass the whole of Europe, we need to obtain, adjust and restrict the microdata as needed for comparability.

2.1 Data source: the EU-SILC

Our data source is the EU-SILC,[4] the primer micro-data source for income with a complete European coverage.[5] The EU-SILC is a household survey supplying additional information on the lower end of the income distribution and asks questions designed to detect difficulties of the less well to do. However, it is representative for the whole population and includes sufficient amount of information on demography and income for our analysis. The current analysis uses the July 2016 distribution of the EU-SILC with data extending to 2013.[6] Croatia had to be excluded from the analysis, due to limited data availability.

[4]For detailed information on the EU-SILC, see its official website at http://ec.europa.eu/eurostat/web/microdata/european-union-statistics-on-income-and-living-conditions. The base document for the EU-SILC is EUROSTAT (2013). For in-depth technical information on the EU-SILC, see 'EU-SILC' under 'Eurostat' under 'European Commission' under 'Browse categories' at https://circabc.europa.eu/. Quality reports on the data are to be found on this site.

[5]Note that for the long-term graph in the Introduction, we also used the predecessor of EU-SILC, the ECHP (European Community Household Panel), covering ultimately all of the EU15 member states. Initially, our ambition was to use the ECHP for the complete analysis, but the fact that only household net income could be followed throughout the two decades in both surveys in a comparable way prevented us from doing so. The main reason for this is that not having access to components of total income is not compatible with our individual approach. The base document for the ECHP is EUROSTAT (2003 december), while EUROSTAT (2005) provides description of the continuity issues during the transition from the EHCP to the EU-SILC.

[6]The 2016 July distribution includes only part of the data collected in 2015, referring to the year 2014. For this reason, the last reference year the current analysis can consider is 2013.

2.2 The unit of observations and income concepts

Our analysis is individual-based, focusing on persons rather than on households. Inequality research often considers households as the unit of analysis and this happens because of both theoretical and pragmatic, data-related reasons. Yet, we have many reasons to look at the individual level instead. First, households can potentially reconfigure. Second, income is closely linked to the the generation of utility at the individual level. And finally, many components of income (labour income and certain taxes and transfers) are at the individual level, so it is important to study their distribution at the same level. In order to do so in a meaningful way, we have to select appropriate populations and income concepts for our analysis.

To avoid the complexities associated with the beginning and the end of the life-cycle, we consider only the 25-60 year old adult population. We group different sources of income together according to the potential role they can play in the ultimate distribution, level and the variability of income. We base our grouping on how much control individuals' own actions have on the influence of these income elements on final disposable income. Overall, we define the following categories:

1. Personal labour income (decomposable to the wage rate and hours worked over the year)

2. Personal income: labour income plus person-related benefits (transfers provided by the state)

3. Shared partner income: the share from total personal income of partners (solely personal and with benefits)

4. Shared household income: equivalised per capita household income (including every household-level income, such as asset income and household-related transfers)[7]

5. Shared household net income: share from net (after tax) overall household income (equivalised)[8]

[7] We have decided against including imputed rent as an income component as we consider it as a type of pre-committed consumption. To be consistent, we do not include any consumption of this type.

[8] In fact, the EU-SILC does not use the concept of net income, but that of 'Total household disposable income', variable, HY020. Because net income is defined as gross income minus taxes and contributions

This categorisation provides a clear path from strictly personal income to total overall household income, but data availability forces us to make compromises:

- Our preferred income concept would be *'supergross'* income that includes contributions paid by employers too.[9] Unfortunately, the quality of the data at hand does not permit using this concept in the current study as contributions paid by the employer are not available for Germany and Romania at all, missing in 2006 for the UK and for two additional years in Lithuania.

- There are many income types that SILC reports at the household level and thus we shall consider in household overall income, but are in fact personal. These include asset income from financial assets and renting real estate. In case some individuals with this kind of income do not have labour income at all, we shall register them as having zero personal income.

- Taxes and many transfer items are only available at the household level. For this and the above reason, we cannot use the otherwise meaningful and popular concept of 'market income' and 'post-government' income, as neither can be defined at the individual level. The former would miss asset income, the latter the deduction of taxes.

- We have no knowledge about sharing rules among couples and within the household and thus we have to assume them. The shift of such sharing rules can be a source of adjustment that we cannot take into account.

- In the step from concept 3 to 4, we in fact increase the volume of overall household

paid by the household, disposable income can be different from that only due to expenses that are counted as pre-committed or unavoidable the same was as taxes are. There is only one such item in the HY020 variable, regular inter-household cash transfers paid. Although the question whether such transfers are more or less avoidable than taxes is a real and substantial one, we adhere to the clear definition of net income and subtract HY020 from this to get Total net household income as the result. Somewhat more confusingly, imputed rent (variables HY030G) is categorised as part of the gross income component (a negative one). Fortunately, the description of Total household gross income (variable HY010) makes it clear that even though the collection of such components is mandatory, they are not included in the calculation of aggregates.

[9]The rationale behind this choice would be threefold. Firstly and most importantly, supergross constitutes the full labour cost to employers. Given that contributions constitute a large share of the gross wage (sometimes up to around 30%, but sometimes none), and their importance varies across countries and even over time, its distribution can evolve very differently from that of gross income. Secondly, supergross would be an income concept truly free from redistribution, a perfect departure for separating different factors of inequality. Thirdly, it usually feeds directly into social security provision very much the same way as contributions paid by the employee and can be interpreted as pre-committed consumption.

income as we use equivalent units instead of per capita, but keep the number of observations constant.

We start our individual-based analysis of the evolution of inequality by looking at annual labour income, the main source of income for most, including income from self-employment.[10] For the subsequent analysis, we also look at the development of individual's hourly wage rates. Since income is reported only on an annual basis, we generate (hourly) wages by dividing annual labour income by annual working hours. We construct a measure for annual working hours by using information on the number of hours that individuals usually work per week in their main job and the number of months that they worked.[11]

Within the overall sample, we would like to differentiate three characteristically different groups of countries, which we define as follows:

1. North-west Europe, abbreviated by NW (includes AT, BE, DE, DK, FI, FR, IE, LU, NL, SE and UK)

2. Mediterranean, abbreviated by MED (includes CY, EL, ES, IT, MT and PT)

3. Central and Eastern Europe, abbreviated by CEE (includes BG, CZ, EE, HU, LT, LV, PL, RO, SI and SK)

This grouping can be motivated in various ways, ours being a two-step procedure. First we calculate per capita GDP in 2006, which separates the NW countries from the others. Then within the latter, we separate the CEE from the MED by geographic location. Interestingly, this definition (exogenous to inequality developments) gives the same grouping to which Heidenreich (2016) arrives inductively from a larger set of groups.

2.3 The working sample and data adjustments

Our working dataset contains 25-60 year-old individuals with reasonable income values, after taking successive cleaning steps.

[10] For comparability with other analyses and data, we replace negative values with zero income.

[11] An alternative to this procedure would be to rely on the actual monthly income figure for the time of the interview, as described by Engel & Schaffner (2012). We do not pursue this here due to its detachment from the income reference period where other income types are measured exclusively.

2.3.1 Creating the working samples

The full 2006-13 EU-SILC database covers 4,149,713 individuals, of which 2,447,357 are in the 25-60 age group. From this sample, we create consecutive sub-samples, as shown in Table 1. First, we eliminate observations with implausibly low income. In particular, we look at the personal income variables and mark cases with values that are non-zero but smaller than 1/12 of the first percentile of the (purchasing-power adjusted) distribution for that particular variable in the country the person is resident in. This essentially assumes that anyone having income, but less than one month worth of a very small income is not credible. A person is marked if any of the personal income variables is lower than the threshold defined this way or if the household's recorded disposable income is zero (a very rare event). Through this procedure, we have marked 13,934 individuals (0.57 percent). In a further step, every household is marked if at least one member is marked this way. Finally, we mark a very small number of observations with zero registered household income. We drop all members of such households, increasing the loss to 21,690 observations (0.89 percent). This leaves us with a sample of 2,425,667 individuals (sample C).

Table 1: Number of observations in the working samples

	SILC	Working age	'Base'	Subsample	Subsample	'Working'	Subsample
			C	C1	C2	D	D2
Restriction	None	25-60	25-60, clean, positive household income	Positive shared (partner) income	Positive personal income	Positive personal labour income	Hours info present
N in 2006 - 2013	4,149,713	2,447,357	2,425,667	1,714,356	2,080,229	1,856,489	1,638,328

Because we want to focus on individuals who are most likely to be active, we restrict age to 25-60 in our 'base' sample (C). Doing so, we potentially drop household members, but because all information (also relating to couples and level at the household level) is assigned to the individual, this does not create any problem. Within the base sample (C), we define two subsamples, C1 and C2, including only individuals with non-zero shared partner- and personal income, respectively (both including benefits over and

above labour income). For inclusion in C1, we also require a partner to be present.

Our 'working' sample (D) is a subset of the 'base' sample, including only those with nonzero personal work-related income (1,856,489 individuals). Keeping only those having reported nonzero working hours, we define a subsample (D2) of them.

Note that only the base sample is defined using exogenous characteristics, the others depend on decisions and random factors. As a result, actual individuals move in- and out of the samples over time, which is important to keep in mind when interpreting the time series of indicators.

The number of observations in each of our samples is shown in Table 1 on the preceding page, for the entire 2006-2013 period. Note that there are two important ways one can read the above table. From left to right, we describe the technical process of restricting the sample more and more. From right to left (except for sample D2), we go through expanding the budget constraint by adding elements to it, fundamental to our approach.

Figure A.1 on page 38 in the Appendix gives an overview of the time trend of the proportion of the subsamples. It shows among others that employment of MED men has declined most and that the share of women with positive earning is highest in the NW.

2.3.2 Cross-country and across-time comparability: PPP and CPI adjustments

In order to carry out an EU-wide analysis, we have to apply adjustments to the data to make them comparable across time, space and to each other.

The effect of changing relative prices needs to be adjusted across countries and time periods. Notwithstanding the known difficulties of PPP adjustment,[12] we believe that when working with standard survey data, using adjusted as opposed to unadjusted values in an EU-wide comparison makes more sense. This way, the consumption value of one EUR in Belgium in 1993 can be compared to one EUR in Bulgaria in 2013.

[12]The main simplification is of course that this treatment excludes the possibility of adjustment through shifting consumption patterns. See the exposition of the problem and a potential solution based on attitudes and special data in Deaton (2010) and in Deaton & Dupriez (2011). Our ambition here is to build on a widely available survey and to consider the whole distribution, not just poverty, therefore we cannot implement the suggestions embedded in these works. The problem is further exacerbated by cross-border spatial mobility and migration, because local prices are relevant for the less mobile only. Given that we do not observe the really rich and mobile, we do not think that PPP-related problems have a great influence on our conclusions.

To make the adjustment exercise feasible, we separate adjustment in the cross-section and over time. Data are already given in or transformed into EUR.[13] First, we adjust income given in EUR so that the overall price level of the EU27 is 1 and then adjust this with the EU28-wide consumer price index to 2015 levels.[14] A comparison of some results based on PPP adjusted and -unadjusted data are presented in Section A.3 on page 40 in the Appendix.

The EU-SILC comes with individual cross-sectional weights for each country and we use these for weighting the pooled EU sample (variable PB040). Country weights sum up to the size of the population in each country, thus using them for the whole of Europe results in proper weighting at that level too.[15]

2.4 Characterising complex distributions

Our analysis aims at disentangling and describing components that make up income distributions. We would like to characterise well the temporal evolution of a distribution of a complex income measure for a complex population. This means that we have to part with the commonly used Gini index, as it is neither capable of showing inequality in different parts of the distribution nor is decomposable.

In order to characterise the income distribution well, we need more than one index relating to important parts of the distribution. Taking the usual practice as the example, this means that we want to learn how the whole distribution of household income per capita has changed over time in the whole EU, which means that we want information about those with lower and higher income. To this end we use the P90/P50 as well as the P50/P10 ratio of income levels right at the respective percentiles, making the different

[13]The exception is the UK in 2004, where we have imported data from the BHPS and the ECHP, which we used to create a few graphs only. We used the annual average spot GBP/EUR rate of the ECB for the former and the going exchange rates distributed along with the ECHP in the country file.

[14]The cross-sectional adjustment is done with the Price Level Index of actual individual consumption (SDMX code is prc_ppp_ind/A.PLI_EU28.A01 for the EUROSTAT download) indicator in the Eurostat database. There are other possibilities, in particular the GDP-based deflator, but we believe that given our aims and measure of household welfare, the chosen one is the most appropriate one. The over-time adjustment uses the CPI for the EU, which is the EU27-EU28 for most of the time considered (makes no practical difference). The same procedure is applied to the ECHP data, once they have been converted to EUR, using the EU15 price index before 2006 (SDMX code is prc_hicp_aind/A.INX_A_AVG.CP00.EU for the EUROSTAT download).

[15]In case of the ECHP and the 2004 BHPS, weights correct only for the sample distortions. To make them compatible with the EU-SILC, we have used the 16+ population weights and the January 1 population figures scaled by 79%, respectively (SDMX code is demo_pjan/A.NR.TOTAL.T.UK for the EUROSTAT download).

behaviour of upper and lower part of the distribution explicit. We shall refer to the inequality detected by these indices also as 'upper'- and 'lower' inequality, respectively.

In order to show how different income sources and different parts of the population have contributed to the overall change, we need a decomposable measure. This is important, because the income measure is built up from different income components and studying them separately gives an idea of their contribution to overall inequality. Also, even if the ultimate object of interest is often the overall EU-wide income distribution, it is important to see which country groups changes come from.

For this exercise, we use the variance of the logarithm of the income measure, referred to as the 'log variance' of income. It allows for additive decompositions that make not only statistical, but also economic sense. This can be useful for decomposition by country groups and by certain income types. Section A.2 on page 39 in the Appendix shows that the overall variance is a population-weighted sum of the within-region variances and the contributions to the between-region variances – these are the first and the second term respectively in the following formula:

$$\widehat{Var(y)} = \underbrace{\sum_{r=1}^{R} \frac{N_r}{N} \widehat{Var_r(y)}}_{within-group} + \underbrace{\sum_{r=1}^{R} \frac{N_r}{N} (\bar{y} - \bar{y_r})^2}_{betweeen-group}, \tag{1}$$

where N_r is the number of observations in group r, $N = \sum_{r=1}^{R} N_r$ is the total number of observations, $\bar{y} = \frac{1}{N} \sum_{r=1}^{R} \sum_{i=1}^{N_r} y_{ri}$ is the overall sample average. In group r, $\bar{y_r} = \frac{1}{N_r} \sum_{i=1}^{N_r} y_{ri}$ is the average and $\widehat{Var_r(y)} = \frac{1}{N_r} \sum_{i=1}^{N_r} (y_{ir} - \bar{y_r})^2$ is the variance of the measure of interest.

Observing these two variance components tells us how the variance within a group contributes to the overall variance. To facilitate visual comparison of the components, we scale the results, thus one has to bear in mind the population shares as weights. These are fairly stable over time, which helps the analysis: North-west Europe, Mediterranean and Central and Eastern Europe have a 53%, 26% and 21% share of the total population during just about the whole period.

Because our main tools are graphs, we take care that they are informative for the comparison we intend to facilitate. All graphs are scaled so that series can be compared across regions and with the EU, as well as across income concepts. In the case of variance decompositions, scaling enables 'visual adding up' of the trend lines.

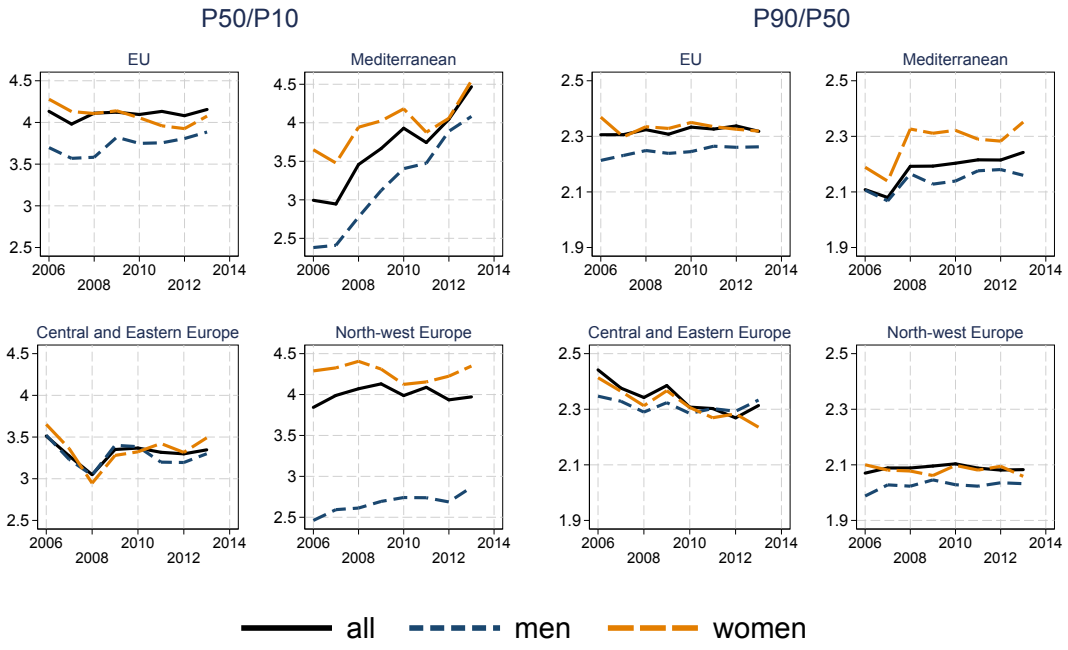
3 Analysis

In what follows, we use pooled EU-wide microdata to create the stylised facts we are after. We look at those who work first, decomposing correlation of hours and hourly wages, finally proceeding to every individual. In both cases, we consider only resources that an individual possesses. In a next step, we include resources from increasingly larger coalitions of individuals: a partner, the household and finally the society.

3.1 Personal income inequality of those who work

In order to build up income elements and the effect of insurance institutions gradually, we turn to the most individualistic income source, personal labour-income. Our starting point is the sample of individuals who are 'working', that is included in sample D (see Subsection 2.3.1 on page 12 for a precise definition).

Figure 4: Quantile ratios for annual labour income of those working (gross)



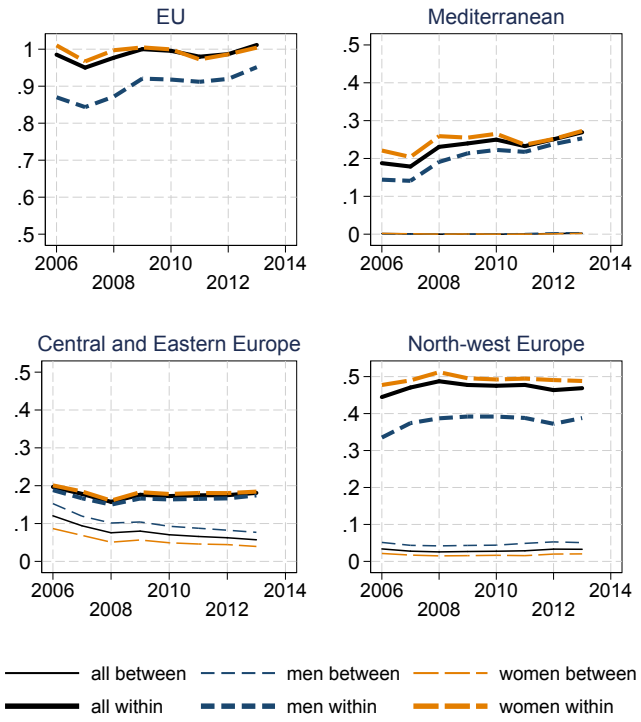
Source: Own calculations from the EU-SILC.

Both quantile ratios are somewhat flat for the whole of the EU in the post-2005 period, but there are remarkable differences in levels, trends and across genders among country groups – see Figure 4. The overall difference in the initial level of the P50/P10 ratio is not big in 2006, but it diverges substantially after 7 years, increasing by 50% in

the Mediterranean and only very little in the CEEs. There are larger differences between genders, those in North-west Europe being the largest and those in the CEEs being rather small. The gender gap closes to the greatest extent in the Mediterranean, where the overall increase is the largest. In general, the increases over time for men are larger than those for women at the lower part of the income distribution.

Inequality at the upper part tells a different story. Levels in the P90/P50 ratio in the NW and in the CEE are higher and lower, respectively, and the gender gap is smaller almost everywhere (beware of the different scale however). The time-trends are almost flat, except for a jump in the MED in 2008, and in the CEE, where it is declining slightly, but steadily.

Figure 5: Additive decomposition of the log variance of annual earnings of those who work by country-groups' contributions (gross)



Source: Own calculations from the EU-SILC.

Although we show the Gini and the log-variance in the Appendix only, they provide a way to compare our results to international evidence. The log variance (for both men and women, see Figure B.1 on page 46) in the EU and the CEE group is higher than in

the US.[16] The NW values are similar to the US, while in the MED, they have increased from below- to above that.

The additive decomposition of the total log variance into the population-weighted within- and between components can make it explicit why overall EU-wide inequality can have both different level and time-trend than its components – see Figure 5 on the preceding page. Driven by its population size, the 'NW-within' is the largest component, making stability a decisive factor for the whole of the EU. Scaled by its weight, the strong trend of MED is also much smoother. It is important to note that in case of the CEE's, the importance of the between-group contribution is comparable to that of the within-group contribution, and also that it declines, implying relatively strong convergence towards the EU-average.

3.1.1 Decomposing annual earnings to wage rates and hours

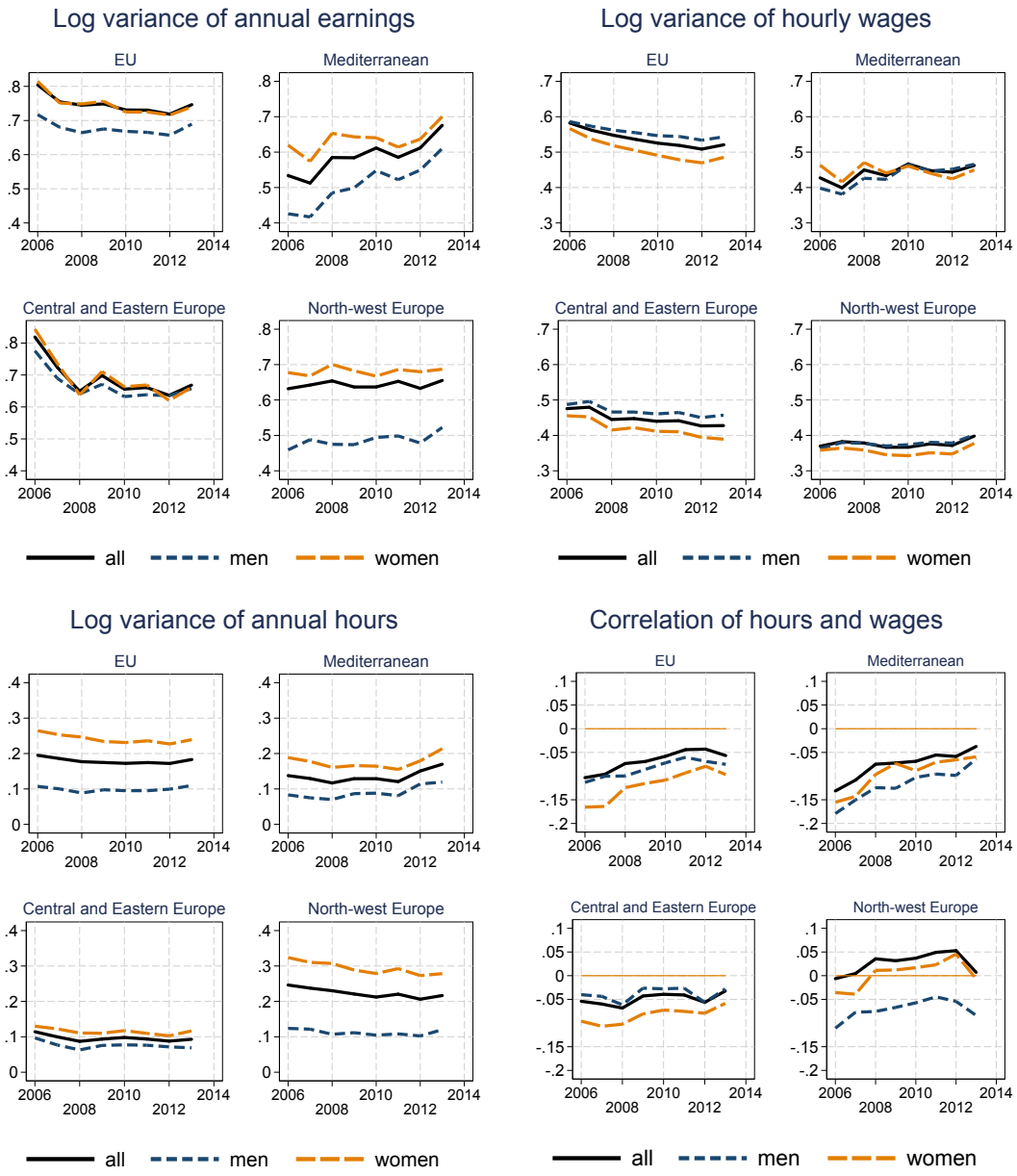
To see the role quantities and prices play in annual labour income inequality, we decompose its log variance to the contribution of the log of annual hours worked and that of hourly wage rates, as well as their correlation for the sample of working individuals with data on hours (sample D2).[17] If this latter component is large, it can increase overall variance if it is positive, but can also decrease it if negative. Those who earn more per hour also work more hours in the first case, while in the second case those who earn more per hour work less hours.

The log variance of annual labour income of those with hours data is on the top left panel of Figure 6 on the following page, while the bottom two panels show the two components of this variance, log hourly wage rates and log annual hours, and the top right panel shows their correlation. We already knew that the time-path of log annual earnings is fairly flat on the EU-level. Here we also see however that the negative correlation increases and the variance of log hours is rather flat, thus fluctuations in the contribution of wage rates dominate.

[16] We use Figure 6 of Heathcote et al. (2010) for comparison. It reports the log variance for men and women, but not for the total population, and no other inequality measures. Note that because of the differences in PPP-adjustments, the PPP-unadjusted figures in Section B on page 46 in the Appendix can also be relevant for international comparison.

[17] Although inequality is somewhat smaller in this population in general, both the trends and cross-region patterns are similar to what we have seen on sample D (see Figure B.2 on page 47 in the Appendix using sample D2 for a comparison).

Figure 6: Variance decomposition of log annual earnings to the contribution of log annual working hours and log wage rates (gross)



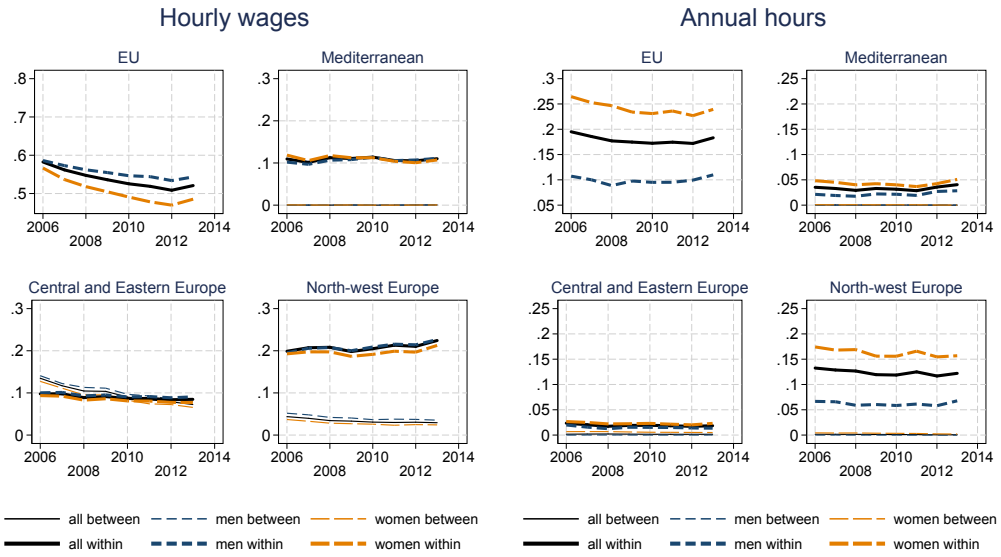
Source: Own calculations from the EU-SILC.

Inequality is strongly increasing in the MED, slightly in NW and almost not at all in the CEEs. The three country groups differ not only in the time trend, but also in the level and main drivers of the inequality in annual labour income. The strong increase in the MED is a result of growing variance of hourly wages in the beginning and that of hours later, further amplified by the increasing correlation (which is nevertheless still

strongly negative). The slight increase in the NW is primarily a result of two cancelling effects. On the one hand the correlation is strongly increasing for both sexes, of which that of women turns and stays positive as of 2007.[18] On the other hand, the variance of log hours declines. Finally, as both the correlation and the variance of log hours are rather stable, the time pattern in the CEE is influenced mostly by the log wage rate. The development of the correlation in the US was similar, its level starting at -0.14 for men and 0.01 for women in 1967 and having increased to around 0.1 by 2005.[19]

Due to its relatively small population-share, the regional decomposition of the variance of log hours worked show only a reduced gender-difference in the contribution of the MED – see Figure 7. There is more to see in case of wage rates. The within-contribution of CEE wage rates is rather substantial, but declining over time. Variance of wage rates in the NW also play an important role. The large gender difference in the NW makes a distinctive mark on the whole of Europe.

Figure 7: Additive decomposition of the log variance of hourly wages (left) and annual hours (right) of those who work by country groups' contributions (gross)



Source: Own calculations from the EU-SILC.

[18]See the detailes analysis of Checchi et al. (2016), analysing the role of hours and wage rates as well as their correlation in shaping inequality. They document the same increasing pattern for Germany and France on a longer time horizon, but find a more stable, already higher correlation in the UK.

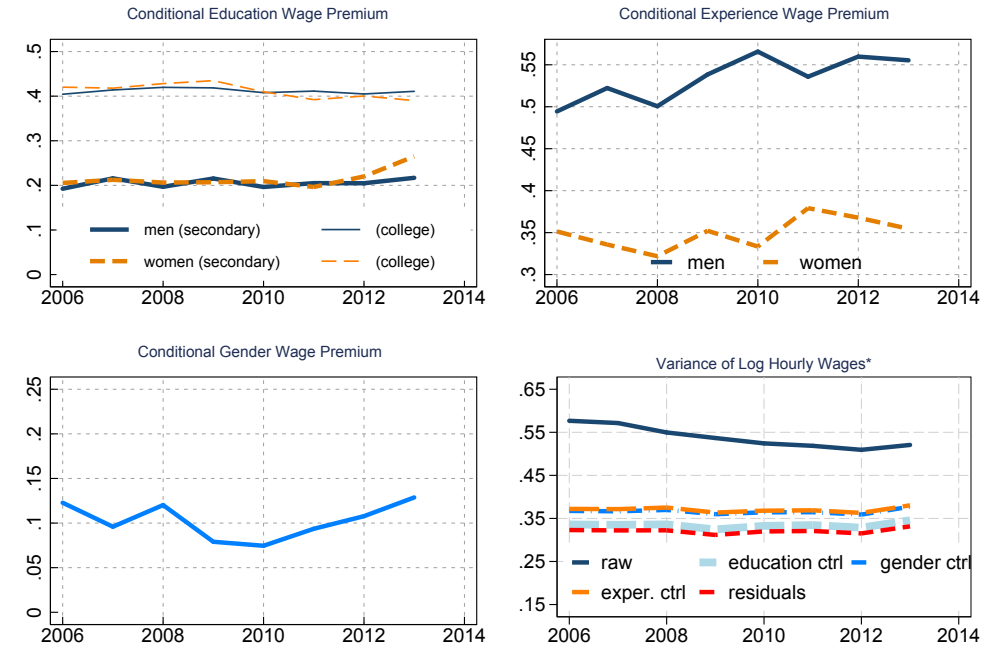
[19]See Figure 6 of Heathcote et al. (2010)

3.1.2 Hourly wages: observable and unobservable influences

In order to separate the main observable characteristics usually associated with wage rates from unobservable influences, we run standard Mincerian wage regressions to estimate the time-path of wage premia of secondary- and college education, of experience and of gender, controlled for geographic location (see details in Section B.1.1 on page 48 of the Appendix).

The observable factors explain somewhat less than half of the variance in wage rates, the share of both these and that of the residual factor being constant throughout the period, (see the bottom left panel of Figure 8).

Figure 8: Wage premia estimated through a Mincerian wage regression

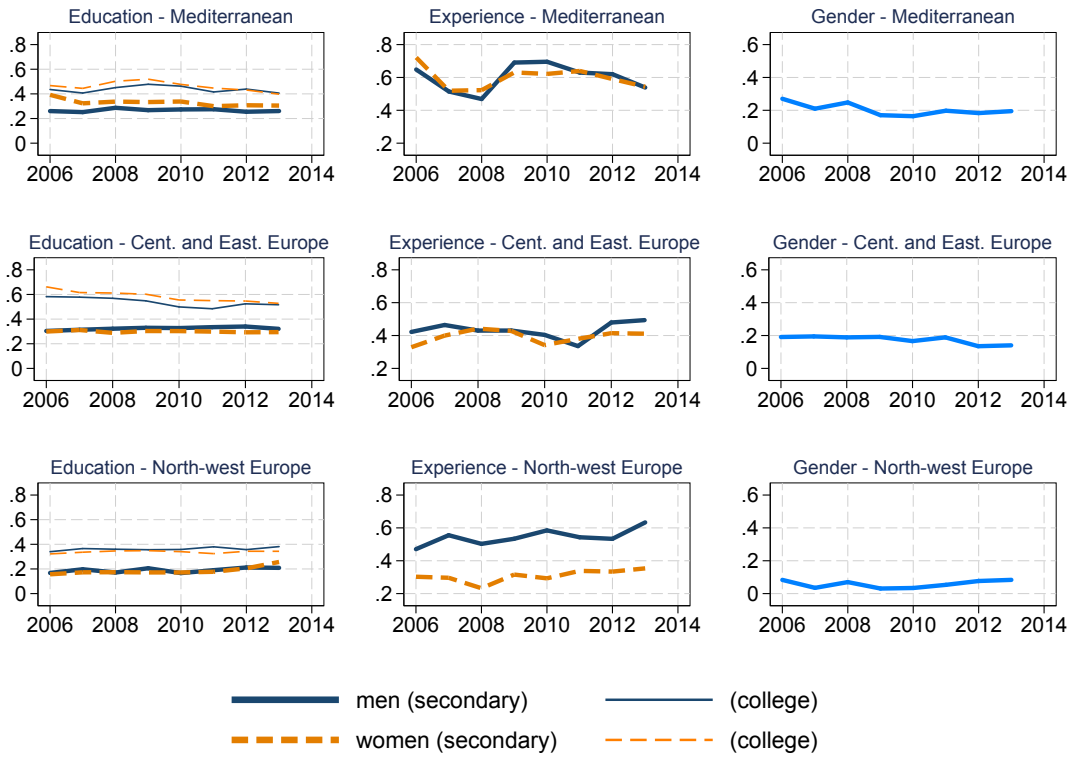


*Country controls, no HH distiction

Source: own calculations using EU-SILC microdata
Remarks: including controls for secondary- and college education, experience, and gender. Details are in Section B.1.1 on page 48.

The college wage-premium has increased until the outbreak of the crisis, but has decreased afterwards. It did so more for women, while the premium for secondary school has increased at the same time. The conditional experience wage-premium for women was slightly fluctuating over a stable level, while that of men has increased until 2010, when it has levelled out after a correction. The conditional gender-wage

Figure 9: Conditional wage premia of key characteristics by country group



Source: own calculations using EU-SILC microdata
Remarks: see Figure 8 on the preceding page

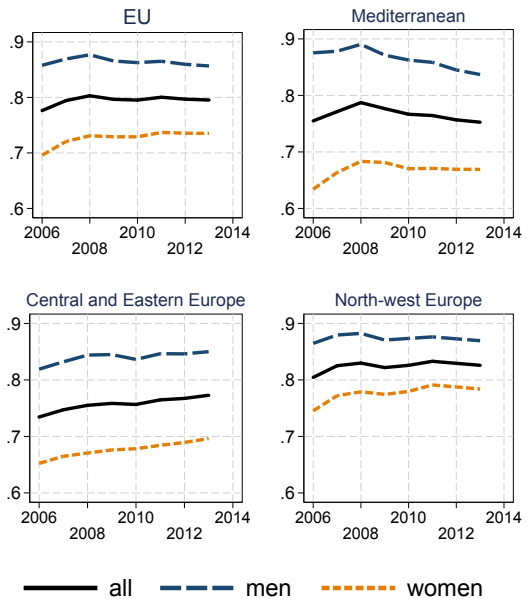
premium (that is the advantage of men) was slightly decreasing until the crisis, and rising thereafter.

Just as the overall variance, returns on the three key characteristics show different levels and time trends for the three country groups we are looking at – see Figure 9. There was not much action in the NW (except for the 2010 dip in the gender-gap, see the rightmost panel) and also the CEE has witnessed mostly some slight decline in the returns to college. In the MED, there was some increase in the returns to college in the first crisis years, followed by a slight decrease. After a substantial decline before the crisis, the experience-premium has increased strongly for both sexes during the crisis, returning to its former lower levels by 2013.

3.2 Personal income inequality, with non-work income included

The share of those working[20] in the whole of the EU is relatively high at 80% and is also relatively stable over time – see Figure 10. We find a lower level, a larger post-crisis drop and a wider gender-employment gap, together with a steady increase in the CEEs. Employment has not only recovered after the crisis, but also slightly increased in the NW in a gender-balanced way. In the MED however, employment decreased constantly after the crisis and did so most remarkably for men, narrowing somewhat the still large gender gap.

Figure 10: Share of the working within the base sample

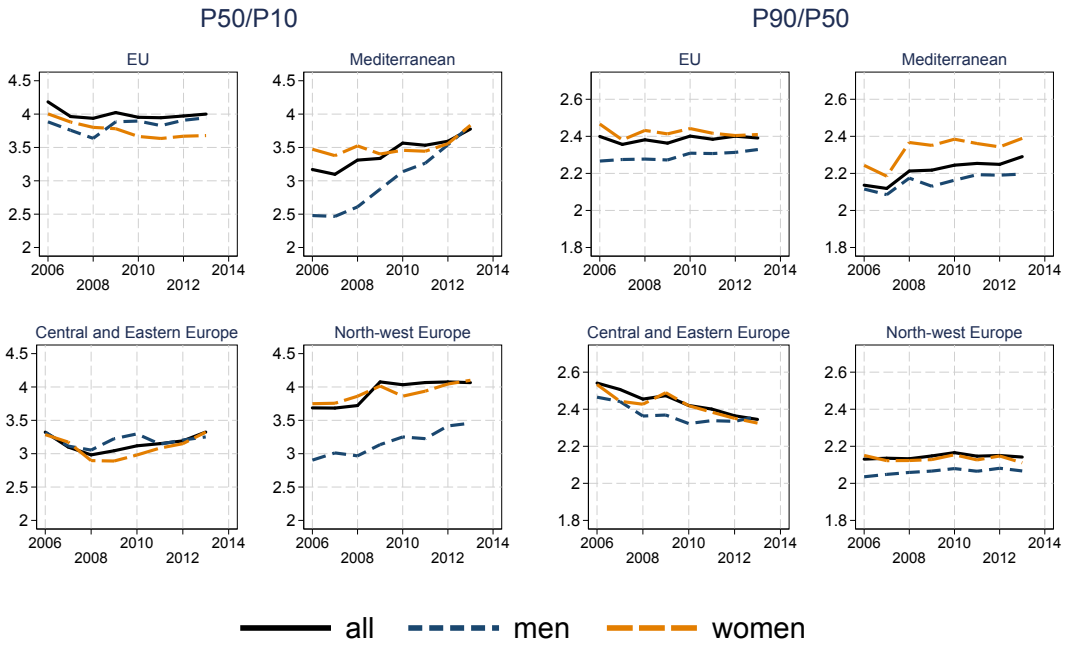


Source: own calculations using EU-SILC microdata.

A substantial part of the 25-60 year old population does not work and thus its personal income comes from other sources, such as unemployment benefits, social assistance and pensions received, which we now include in the analysis. Figure 11 on the following page shows quantile ratios of non-zero personal income (as before, Gini and log variance are given in Figure B.3 on page 49 in the Appendix).

[20] Recall that we mark a person as working if the associated annual labour income is larger than zero. Note that this measure can be more forgiving than the official employment statistics coming from Eurostat, based on the Labour Force Survey and the ILO definition of employment counting anyone who has worked at least an hour for money during the preceding week as employed.

Figure 11: Quantile ratios of annual personal income (gross)

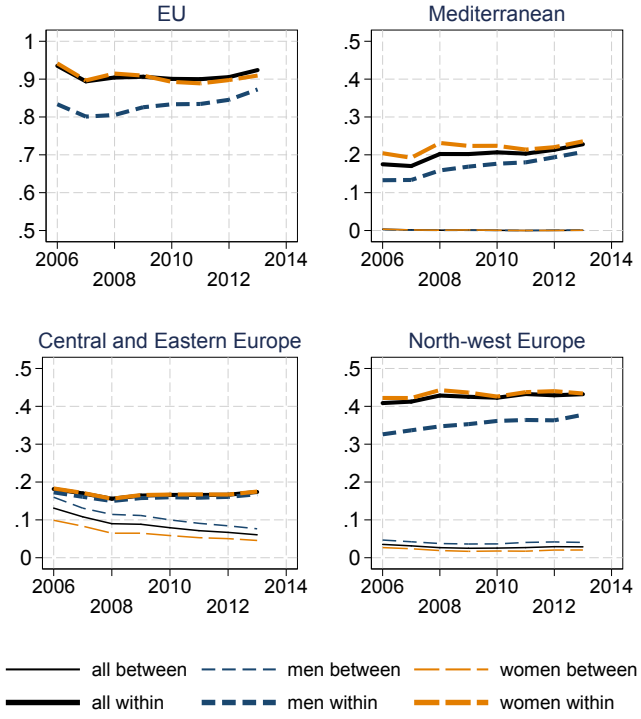


Source: own calculations using the EU-SILC.

Compared with Figure 4 on page 16, the similar graph for work income, inequality appears to be *higher* in almost all cases (with the strong exception of the gender-gap in the NW), and exhibits an even sharper than before increase of inequality in case of the MED. All this is somewhat counter-intuitive, as we have also included benefits among the sources of income, whose availability is supposed to help *reducing* variability of income for an individual (this is the idea of automatic stabilisers, such as unemployment benefit). At the same time however, we add people to our sample who might not work full time or during the year and thus receive much less income than the labour income of the steadily working, thus widen the distribution of income we account for. The increase in inequality is the net of these two effects. Comparing Figure 4 on page 16 to Figure B.4 on page 50 in the Appendix confirms this. Using the new, broader income concept, but the previous sample of the working, the latter shows a decrease of inequality across the board, smaller gender gaps and less pronounced trends too, just as we expect.

In the same comparison (with respect to Figure 5 on page 17), the regional decomposition shows smaller levels of inequality within the NW, dominating the overall EU picture – see Figure 12 on the following page. Neither the dynamics, nor the between country-group contributions are affected.

Figure 12: Additive decomposition of the log variance of annual earnings of those who work by country-groups' contributions (gross)



Source: own calculations using the EU-SILC.

3.3 Inequality beyond individual resources

So far, we have looked at income that is registered at the individual level, being available even if a person lives alone.[21] This also means however that in case of a labour market shock such as a job loss, only welfare transfers are available for smoothing. Living in a community can make an individual more resilient to shocks in various ways. Partners living together can decide jointly about their behaviour in general, and in case of a labour market shock in particular. These include labour market participation, hours worked and the level of earnings they aim at. Households (with children and more than two adults) can have income sources that individuals might not, both in the form of benefits and in terms of joint ownership. Household size also matters because when income is spent on consumption, economies of scale can reduce the per capita resource needs. Finally, tax systems can also express preference towards larger households (e.g.

[21]Note that due to data limitations, we cannot take into account individual asset holding of any kind. These can be accounted for only among household resources, introduction a distortion into our estimates at the upper part of the income distribution.

those with children), which can further change their net income position.

3.3.1 Couples and inequality

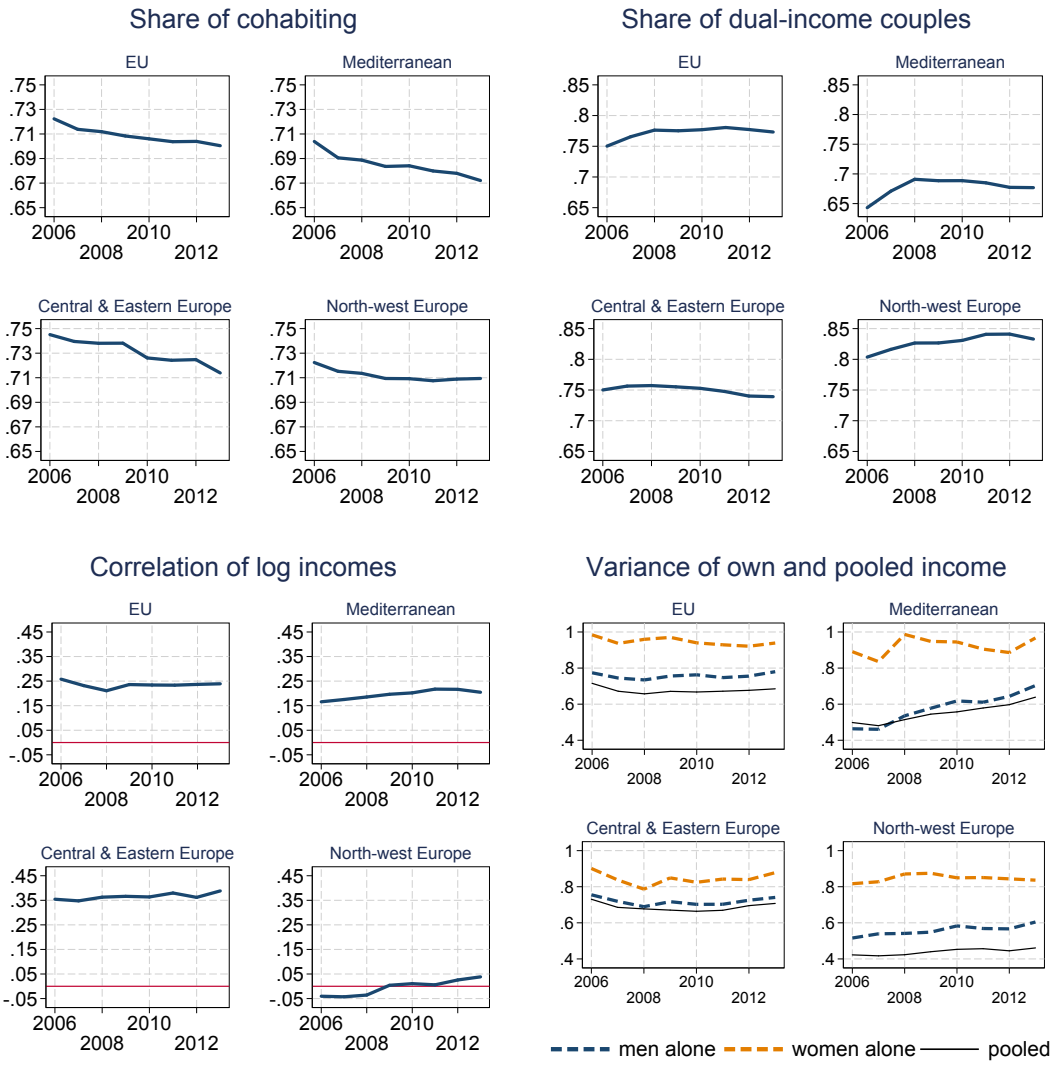
On average 71% of the 25-60 year old working-age population lived together with a partner in the period we are looking at – see the top left panel of Figure 13 on the next page. Although there is variation in this share across country groups already in 2006, a subsequent decrease in the CEE produces a 71% share by 2013, similar to the NW. MED also witnesses a substantial decline, albeit from an already low level. For these country groups, the extent of a potential insurance coming from a partner is monotonically decreasing over time.

However, even if a spouse is present, such an insurance is realised only if both parties do actually work. The top-right panel shows the fraction of dual-earners among those with a partner, which is non-decreasing for all groups before the crisis. While the increase was followed by a decrease in the MED and the CEEs, it continued slowly but steadily in the NW, which had already the highest level. Much of this was realised through increased employment rates of women: they had relatively low employment rates and were able to enter the labour force. At the same time, male employment often went down. (The time-path of our 'working' sample D in Figure A.1 on page 38 in the Appendix shows these changes.)

Income streams of members of a couple act as revenue streams in the portfolio of an investor, having two effects: increasing available resources on the one hand, and changing the variance of total resources on the other. Both effects depend on the correlation between partners' incomes. A nonnegative contemporaneous correlation in the cross-section implies that individuals earning more live together and pool income with other high earners, while low earners pool with low earners. This results in a potentially higher inequality than considering personal income only. However, if the correlation is negative, high earners pool with low earners and thus pooled income is less unequal than personal income, the dampening effect being dependent on the absolute size of the correlation.[22]

[22] All this can be shown by the same logic we used for the decomposition of hourly wages and hours worked, using levels rather than the log of income.

Figure 13: Income pooling and inequality



Source: Own calculations from the EU-SILC.

The actual correlation (see the bottom-left panel) is positive, moderately large and small for the CEE and MED, respectively, and is slightly increasing over time. It was somewhat below zero in NW in 2006, but has substantially increased thereafter, to about 5%. This is a rather low level, but a substantial increase, showing a move towards stronger matching within the couples. Pooling income benefits mostly women (see the lower right panel) and this is in line with their average lower employment rate and earnings when employed (see Figure A.1 on page 38 and Figure A.2 on page 39 in the Appendix, respectively). It is important to note that in NW, men also gain from pooling, in line with the high share of dual-income couples combined with small correlation of incomes.[23]

[23]The weight of the individual income streams will depend on the sharing rule between partners, which

3.3.2 Inequality and risk sharing in coalitions of persons

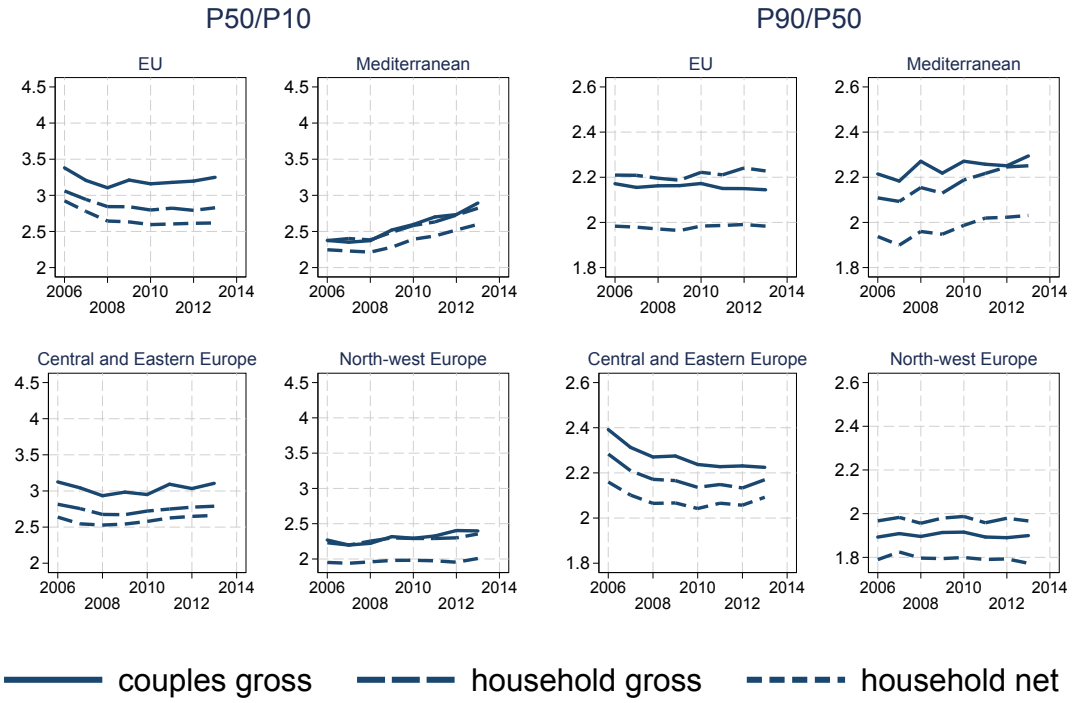
People share income with each other in different coalitions and different income types, of which living with another adult and sharing labour income is just one option. Household-level income that would not be present without the existence of the household can also be shared with children and parents if present. Income net of taxes represents another, social level of income and risk sharing. In order to look at these, we also consider per capita gross- and net household income. Gross household income includes income from other household members outside our age group of interest, the balance of household-level transfers from other households and benefits from the state. Any type of asset held by members of the household is accounted for at the household level in EU-SILC, so assets as such appear here too. Besides taking into account these new elements, we project total income onto equivalent consumption units to take into account economies of scale in household production and thus get closer to actual welfare. Following EUROSTAT, we use the modified OECD scale supplied with EU-SILC for this purpose. Figure 14 on the next page shows lower and upper inequality as we move from couples' gross- to households' gross- and net income.

Comparing income inequality in the case when income is pooled between partners (solid line) with the case when relying only on individual resources only (see Figure 11 on page 24) shows that inequality levels drop to some extent in all cases. This drop is almost zero in the CEE countries, but rather large in the other two country groups, where women's employment rate is much smaller – note that the current graph does not differentiate women and men due to the assumed equal sharing. As Figure B.5 on page 51 in the Appendix shows with the same series calculated on sample C2 – requiring that the partners have at least some personal income –, the drop is even smaller if we are looking at the same individuals, therefore likely to be due mostly to income pooling.[24] These broad patterns hold no matter if we are looking at lower or upper inequality.

we have assumed to be half-half. We are aware of the literature on intra-household bargaining, such as the seminal work of Browning & Chiappori (1998) and the more recent contribution of Knowles (2013), for example. Nevertheless, lacking information on sharing, we cannot do anything but assume a fixed share.

[24]Note that this particular comparison is imperfect by its nature, since it also requires to have a partner, thus the effective sample is the intersection of sample C1 and sample C2. However, it is still informative.

Figure 14: Quantile ratios of average of annual income in different coalitions (gross)

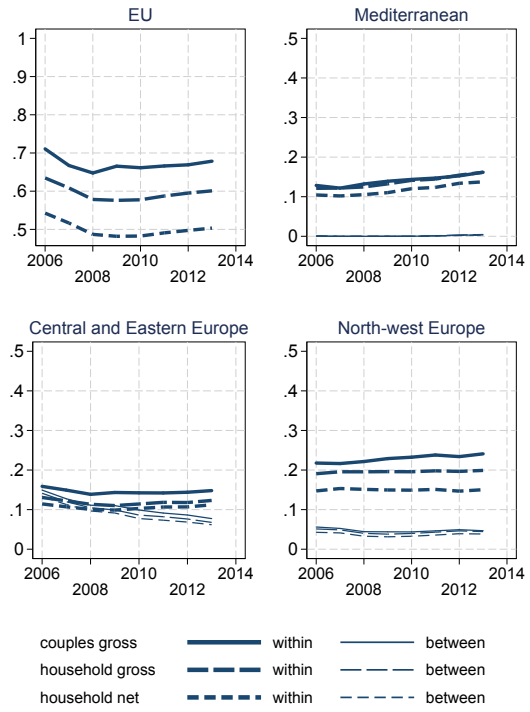


Source: own calculations using the EU-SILC

Sharing income with a larger coalition lowers inequality in many cases – compare the solid and long-dashed line on Figure 14. The change is largest in the CEE countries, but nonexistent in the two other groups for lower-, and in fact reversed for the North-west in case of upper inequality. This seemingly strange situation is only partially due to a selection effect, because it is very small in all cases (the picture is almost similar when estimated on sample C1 – see Figure B.7 on page 53 in the Appendix). It is likely to be due also to the inclusion of more volatile elements in the income of the household, including income from members outside our target population, even though our stylised analysis is unable to show this. Indeed, more volatile market income has a much larger-, and the less volatile transfers a smaller additional share in the NW than elsewhere, most importantly among women (as shown on Figure A.1 on page 38 and Figure A.2 on page 39 in the Appendix).[25] Interestingly, the extent of the change in inequality is almost constant for all country groups, except for upper inequality in the MED.

[25]Note that these individuals are still not included in the analysis, but contribute only through their headcount and shared income.

Figure 15: Additive decomposition of the log variance of annual income in different coalitions (gross)



Source: own calculations using the EU-SILC

Moving to the level of the society and considering net- instead of gross income has the expected effect of decreasing inequality across the board. The fact that the drop is larger for upper inequality shows the overall progressive nature of the tax system.

A similar picture emerges if one looks at single indicators of inequality instead of one targeted at a specific part of the distribution (see Figure B.6 on page 52 in the Appendix). Both the Gini and the log variance show the decrease in inequality as we move towards larger coalitions, but the latter does not register the peculiarities of the NW. Nevertheless, the Gini provides a useful point of reference: compared to the value of around 0.43 for gross individual annual earnings, the same measure is 0.38 for couples' and household' per capita gross income. This change is larger than both cross-country-group and over-time variation within the EU over the period we are looking at, including the crisis.

Being based on the log variance measure, the regional decomposition does not cause surprises either – see Figure 15. Within-group inequality levels are decreasing as coalitions get larger and the weight of the between-component is decreasing as the average income level is rising in the CEE countries.

3.4 Personal income inequality over the life-cycle and cohorts

An important dimension of inequality is the life-cycle. Income shocks accumulate over the lifetime of an individual, leaving a mark on their financial position, but there is a subjective aspect too. Individuals might care for inequality around themselves (that is: for inequality in the cross-section), but their own experience might relate closer to that of their peers, who they observe the best over time. Life-cycle profiles for wages are well-known objects to study the constraints under which individuals plan on the long term. These profiles of levels have an inverted U-shape on average due to the lifetime time-constraint, but profiles of variances are different. If earnings shocks hitting individuals follow a random walk and there are no insurance mechanisms available, lifetime income profiles diverge and inequality is expected to increase over time. The more of such insurance instruments exist, the more these shocks are insurable and the less inequality grows. In turn, cross-section inequality becomes smaller too.[26]

Here we chose to concentrate on inequality over the life-cycle, characterising inequality among individuals of different ages. Because we do not have individual panel data, we construct a synthetic panel of cohorts to accomplish this following Deaton & Paxson (1994).

Figure 16 on the following page shows the observed time-path of the variance of log income of men and women born within 5-year intervals from 1945 to 1980. These pieces of the cohort-profiles together trace the whole underlying profile, but also indicate changes over time. Should both the society and the economy be stationary, they would trace out the underlying single life-cycle profile without variation. An upward-sloping profile indicates that shocks are stronger than insurance, while a decrease indicates the opposite. If the profile does not degenerate to a single line, the spread of the profile depends on how stationary the income process is over time.

The two panels of the figure show the richest case for individual and household-level insurance, total personal income on the left, equivalised household income on the right. 'Individual' profiles show a strong trend, dispersion and gender-gap, while 'household' profiles are flatter, less dispersed and show less gender-differences.

[26]Inequality as measured in the cross-section can also be represented as an outcome of the over-time evolution of inequality among cohorts. Similarly to the additive decomposition by regions, we can decompose each cross-section to the variance within a cohort and the deviation of the cohort's income from the prevailing mean income level.

Figure 16: Log variance-profiles for different cohorts using household (left panel) and personal (right panel) income concepts (gross)

Source: own calculations using the EU-SILC

With little insurance at work, the personal profiles tell a rather rich story. In line with what we expect based on a simple story of wage-evolution, they are increasing for males in all cases. This means that the weight of insurance is less than that of shocks in shaping inequality as individuals age. This does not, however necessarily leads to an increase of cross-sectional inequality, if, as in the NW, new cohorts experience similar levels of inequality as they age as older ones. On the other hand, if their experience is similar to that of the Mediterranean, experiencing higher inequality already at an earlier age, cross-sectional inequality increases too, just as we have seen. The exception from the above is men over 35-40 in the CEE, where insurance (such as some form of unemployment benefit or pension) or significantly slower wage-growth with seniority seems to effectively counteract the dispersion of income.

The case of women is different. Firstly, inequality among them is higher than among men, which is likely to be associated with less labour-market participation and perhaps with less stable jobs on average. Secondly, women's profiles exhibit a marked jump at the age of 30, most likely associated with career breaks related to family formation, the profile sloping downward slightly thereafter (as modelled in the seminal work of Weiss & Gronau (1981)). Interestingly, there is no such tendency to be observed in MED.

Women's profiles rise similarly to men's, even though it indicates less non-stationarity. The low share of dual-earner couples make it more likely that what we see here is only the active part of the otherwise polarised female labour market.

As we have seen earlier, equivalized household income (the most widely used measure) includes all insurance mechanisms at work and as such, its profile is supposed to be as flat as it is. Its level represents the already seen long-term differences across regions. Its spread relates to the non-stationarity within the region, apparent to the greatest extent in the Mediterranean, but also present in CEE too.

4 Conclusions

In this paper, we set out to produce stylised facts on inequality for the whole of the EU, for a turbulent period including and following the economic crisis of 2009. Despite the breadth and depth of the contributions available so far, we are the first to provide a picture that is geographically comprehensive for Europe and looks at the contribution of different income components. In order to do so, we have employed an empirical framework based on a set of income concepts that allows following the evolution of individual inequality as institutions of risk-sharing take effect. We also used indicators that grasp changes at both the top and the bottom of the income distribution. Doing so, we have demonstrated that calculating inequality statistics in a detailed yet harmonised way can yield insights that are difficult to obtain with an analysis restricted to a short period or to few geographic units.

Our main result is a rich collection of stylised facts, but there are a few observations worth emphasizing, most of them concerning the centre-periphery dynamics within Europe. We have seen that the relative average stability of the ultimate utility measure, net household equivalised income masks important heterogeneity at the country-group level. EU inequality increased somewhat with the enlargement (i.e., moving from EU15 to the EU27 in our case), due to large income disparities between new and old member states. Though inequality in the EU27 decreased afterwards, its 2013 level in the whole EU27 is higher than it was ever before in the EU15. Overall, the increase of inequality in the second half of the 2010s was very strong for the Mediterranean, but rather weak or non-existent for the North-west Europe the Central and Eastern European countries.

This effect varies across income concepts, but is rather robust. Importantly, it comes from the lower part of the distribution, not shown by completely aggregate indicators. Our results suggest that in terms of personal income inequality, the Mediterranean has surpassed the CEE group during the crisis years, and this reverses only as we consider the effect of income pooling within the household. This is particularly true if we focus on the lower half of the distribution, much of which is occupied by citizens of these two groups.

There is much we can do to go further down the road to exploit our framework of a pan-European perspective for inequality research, and to follow the effects we have identified. Firstly, we observe that there is much inequality in the cross section during the 2000s, which calls for more detailed analysis at a finer regional level, more connected to substantial economic differences than political borders. Secondly, we confirm the intuition that despite the importance of governmental intervention, much of the insurance against shocks takes place at the household level.

References

Blundell, R. (2011), 'From Income to Consumption: Understanding the Transmission', *Focus* **28**(1), 23–30.

Blundell, R., Pistaferri, L. & Preston, I. (2008), 'Consumption Inequality and Partial Insurance', *American Economic Review* **98**(5), 1887–1921.

Blundell, R. & Preston, I. (1998), 'Consumption Inequality and Income Uncertainty', *The Quarterly Journal of Economics* **113**(2), 603–640. ArticleType: research-article / Full publication date: May, 1998 / Copyright © 1998 Oxford University Press.

Brandolini, A. (2007), Measurement of Income Distribution in Supranational Entities: The Case of the European Union, SSRN Scholarly Paper ID 988025, Social Science Research Network, Rochester, NY.

Browning, M. & Chiappori, P. A. (1998), 'Efficient Intra-Household Allocations: A General Characterization and Empirical Tests', *Econometrica* **66**(6), 1241–1278.

Checchi, D., García-Peñalosa, C. & Vivian, L. (2016), 'Are changes in the dispersion of

hours worked a cause of increased earnings inequality?', *IZA Journal of European Labor Studies* **5**(1).

Deaton, A. (2010), 'Price Indexes, Inequality, and the Measurement of World Poverty', *American Economic Review* **100**(1), 5–34.

Deaton, A. & Dupriez, O. (2011), Spatial price differences within large countries, Technical report.

Deaton, A. & Paxson, C. (1994), 'Intertemporal Choice and Inequality', *Journal of Political Economy* **102**(3), 437–467.

Domeij, D. & Flodén, M. (2010), 'Inequality trends in Sweden 1978–2004', *Review of Economic Dynamics* **13**(1), 179–208.

Domnisoru, C. (2014), The largest drop in income inequality in the European Union during the Great Recession: Romania's puzzling case, Ilo working paper, International Labour Organization.

ECB (2013), The Eurosystem Household Finance and Consumption Survey Results from the first wave, Technical report, European Central Bank.

Engel, M. & Schaffner, S. (2012), How to Use the EU-SILC Panel to Analyse Monthly and Hourly Wages, Technical Report 390, RWI - Leibniz-Institut für Wirtschaftsforschung, Ruhr-University Bochum, TU Dortmund University, University of Duisburg-Essen.

EUROSTAT (2003 december), ECHP UDB Description of variables, Technical Report DOC. PAN 166 / 2003-12, EUROSTAT.

EUROSTAT (2005), *The Continuity of Indicators during the Transition between ECHP and EU-SILC*, Office for Official Publications of the European Communities, Luxembourg. OCLC: 315964614.

EUROSTAT (2013), Methodological Guidelines and Description of EU-SILC Target Variables, Technical report.

Filauro, S. (2017), European incomes, national advantages: EU-wide inequality and its decomposition by country and region, Research Paper 05/2017, Economics and Econometrics Research Institute.

Heathcote, J., Perri, F. & Violante, G. L. (2010), 'Unequal we stand: An empirical analysis of economic inequality in the United States, 1967–2006', *Review of Economic Dynamics* **13**(1), 15–51.

Heidenreich, M. (2016), Introduction: The double dualization of inequality in Europe, *in* 'Exploring Inequality in Europe: Diverging Income and Employment Opportunities in the Crisis', Edward Elgar Pub, Northampton, MA.

Honohan, P. (2005), Measuring microfinance access : Building on existing cross-country data, Technical Report WPS3606, The World Bank.

Knowles, J. A. (2013), 'Why are Married Men Working So Much? An Aggregate Analysis of Intra-Household Bargaining and Labour Supply', *The Review of Economic Studies* **80**(3), 1055–1085.

Krueger, D., Perri, F., Pistaferri, L. & Violante, G. L. (2010), 'Cross-sectional facts for macroeconomists', *Review of Economic Dynamics* **13**(1), 1–14.

Maquet, I., Meyermans, E. & Duiella, M. (2015), High and rising inequalities; what can be done about it (at EU level)?, Technical Report 6/2015, European Commission.

Milanovic, B. & Lakner, C. (2013), 'Global income distribution - From the fall of the Berlin wall to the great recession'.

Mincer, J. (1958), 'Investment in Human Capital and Personal Income Distribution', *Journal of Political Economy* **66**(4), 281–302.

Nolan, B., Salverda, W., Checchi, D., Marx, I., McKnight, A., Tóth, I. G. & van de Werfhorst, H. G. (2014), *Changing Inequalities and Societal Impacts in Rich Countries: Thirty Countries' Experiences*, 1 edn, OUP Oxford.

OECD (2008), Growing Unequal? Income Distribution and Poverty in OECD Countries, Technical report, OECD, Paris.

OECD (2011), Divided We Stand, Technical report, OECD, Paris.

OECD (2013), Crisis squeezes income and puts pressure on inequality and poverty, Technical report.

OECD (2015), *In It Together: Why Less Inequality Benefits All*, OECD, Paris.

OECD & The World Bank (2012), *Promoting Inclusive Growth*, Organisation for Economic Co-operation and Development, Paris.

Raitano, M. (2016), 'Income Inequality in Europe Since the Crisis', *Intereconomics* **51**(2), 67–72.

Salverda, W., Nolan, B., Checchi, D., McKnight, A., Tóth, I. G. & van de Werfhorst, H. G. (2014), *Changing Inequalities in Rich Countries Analytical and Comparative Perspectives*, Oxford University Press.

Weiss, Y. & Gronau, R. (1981), 'Expected Interruptions in Labour Force Participation and Sex-Related Differences in Earnings Growth', *The Review of Economic Studies* **48**(4), 607–619.

Appendices

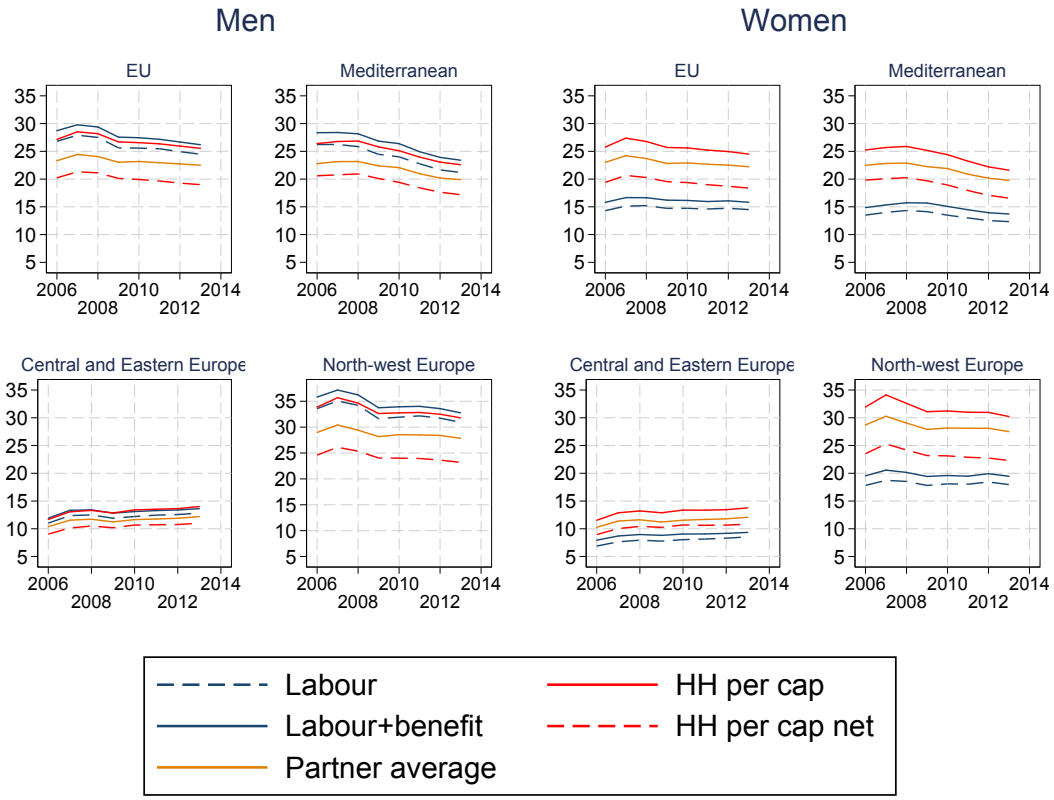
A Data

A.1 Additional tables and figures

Figure A.1: Share of subsamples within the whole sample

Source: own calculations based on the EU-SILC (top row) and ECHP (bottom row) microdata

Figure A.2: Means of different income types over time by gender



Source: Own calculations from EU-SILC microdata

A.2 Variance decomposition

If we want to look at how stable supra-individual entities such as regions contribute to the variance of an income measure (log or levels), we simply use the fact that the estimate of the variance is a sum. We start from a simple elaboration of the overall sample variance estimator

$$\widehat{Var(y)} = \frac{1}{N}\sum_{r=1}^{R}\sum_{i=1}^{N_r}(y_{ri} - \bar{y})^2, \tag{2}$$

where N_r is the number of observations in unit r, $N = \sum_{r=1}^{R} N_r$ is the total number of observations, $\bar{y} = \frac{1}{N}\sum_{r=1}^{R}\sum_{i=1}^{N_r} y_{ri}$ the sample average. We can and subtract the unit averages $\bar{y}_r = \sum_{i=1}^{N_r} y_{ri}$ without changing the value of the measure to get

$$\widehat{Var(y)} = \frac{1}{N}\sum_{r=1}^{R}\sum_{i=1}^{N_r}[(y_{ri} - \bar{y}_r) - (\bar{y} - \bar{y}_r)]^2. \tag{3}$$

Summing over the units first, then take a weighted average of the results as

$$\widehat{Var(y)} = \sum_{r=1}^{R} \frac{N_r}{N} \frac{1}{N_r} \sum_{i=1}^{N_r} ((y_{ir} - \bar{y_r}) - (\bar{y_r} - \bar{y}))^2. \tag{4}$$

Working out the square gives

$$\widehat{Var(y)} = \sum_{r=1}^{R} \frac{N_r}{N} \frac{1}{N_r} \sum_{i=1}^{N_r} (y_{ir} - \bar{y_r})^2 + (\bar{y_r} - \bar{y})^2 - 2(y_{ir} - \bar{y_r})(\bar{y} - \bar{y_r}), \tag{5}$$

which after collecting terms is

$$\widehat{Var(y)} = \sum_{r=1}^{R} \frac{N_r}{N} \left[\frac{1}{N_r} \sum_{i=1}^{N_r} (y_{ir} - \bar{y_r})^2 + \frac{1}{N_r} \sum_{i=1}^{N_r} (\bar{y_r} - \bar{y})^2 - 2(\bar{y} - \bar{y_r}) \frac{1}{N_r} \sum_{i=1}^{N_r} (y_{ir} - \bar{y_r}) \right]. \tag{6}$$

Note that the third term is zero (covariance with a constant) and the second term simplifies to $(\bar{y} - \bar{y_r})^2$ and thus we can write overall variance in two parts as

$$\widehat{Var(y)} = \sum_{(r=1)}^{R} \frac{N_r}{N} \left[\frac{1}{N_r} \sum_{i=1}^{N_r} (\bar{y_i} r - \bar{y_r})^2 + (\bar{y} - \bar{y_r})^2 \right]. \tag{7}$$

The overall variance is thus a population-weighted average of each unit's within-variance and its contribution to between-unit variance. To make this more visible, we can denote the variance within unit r as $\widehat{Var_r(y)} = \frac{1}{N_r} \sum_{i=1}^{N_r} (y_{ir} - \bar{y_r})^2$ and rearrange to get

$$\widehat{Var(y)} = \underbrace{\sum_{r=1}^{R} \frac{N_r}{N} \widehat{Var_r(y)}}_{Within-unit} + \underbrace{\sum_{r=1}^{R} \frac{N_r}{N} (\bar{y_r} - \bar{y})^2}_{Betweeen-unit}. \tag{8}$$

This shows that we can express the overall variance as the sum of two measures: the contribution of units with their within-unit variance and their contribution to the between-unit variance, which is essentially the squared distance of their mean from the overall mean. We introduce this measure in Section 2.4 on page 14.

A.3 The effect of Purchasing Power Parity adjustment

As already indicated in Subsection 2.3.2 on page 13, we are using a purchasing power parity (PPP) corrected measure of income throughout our analysis. Application of such

a correction is not innocuous and this is exactly why we find it important. Even though the correction works mostly only through level- and only to some extent through over-time variation in differences across countries, the data (shown in Table A1) indicates that these variations are considerable.

Table A1: Price level indices for ESA 2010 aggregates [prc_ppp_ind]

Country	2006	2007	2008	2009	2010	2011	2012	2013
EU28	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Belgium	106.1	107.8	110.9	113.4	111.8	112.3	111.4	110.8
Bulgaria	39.0	40.9	44.7	47.0	45.7	45.4	44.9	44.1
Czech Republic	57.9	58.7	67.6	64.4	66.5	68.8	66.8	63.2
Denmark	139.3	135.0	138.2	141.4	139.8	139.7	139.8	139.1
Germany	101.0	99.8	102.0	104.4	102.9	102.1	101.0	102.4
Estonia	61.1	64.7	68.6	67.4	66.0	67.2	67.5	69.8
Ireland	122.3	124.7	130.4	129.6	122.5	122.4	120.4	123.0
Greece	85.4	88.0	90.3	93.5	95.3	94.8	91.8	87.7
Spain	89.1	91.5	94.5	97.3	98.6	98.4	95.7	94.9
France	105.8	105.5	109.1	111.0	109.7	109.5	108.5	106.8
Italy	102.0	100.7	101.1	103.4	100.3	101.4	100.9	102.1
Cyprus	90.0	86.0	89.0	91.8	94.4	95.7	95.9	95.0
Latvia	57.3	63.5	70.2	68.0	62.2	65.4	65.8	65.9
Lithuania	51.2	53.9	59.5	60.2	57.8	58.3	57.9	57.7
Luxembourg	117.9	120.7	123.7	129.7	136.4	135.8	135.7	134.9
Hungary	56.7	61.9	64.0	58.0	57.3	56.3	54.8	54.3
Malta	71.0	69.8	73.7	75.6	76.5	77.9	77.9	80.7
Netherlands	102.2	101.3	104.0	108.3	112.2	112.5	111.1	111.7
Austria	102.3	103.8	106.9	109.7	109.1	109.9	108.7	109.0
Poland	55.7	55.5	62.5	53.0	55.1	53.6	52.0	52.4
Portugal	79.5	80.8	82.9	85.0	84.9	85.1	82.6	81.0
Romania	48.5	54.1	53.6	49.0	46.9	47.6	45.6	47.7
Slovenia	74.4	76.3	80.0	85.1	84.2	83.6	81.7	82.0
Slovakia	52.2	57.3	63.2	66.4	62.5	63.4	63.3	63.3
Finland	119.1	114.8	117.3	120.8	120.2	121.4	121.6	123.8
Sweden	119.7	117.4	115.5	109.3	124.4	131.6	134.1	140.2
United Kingdom	125.4	127.1	114.0	107.3	111.2	111.2	118.5	116.9

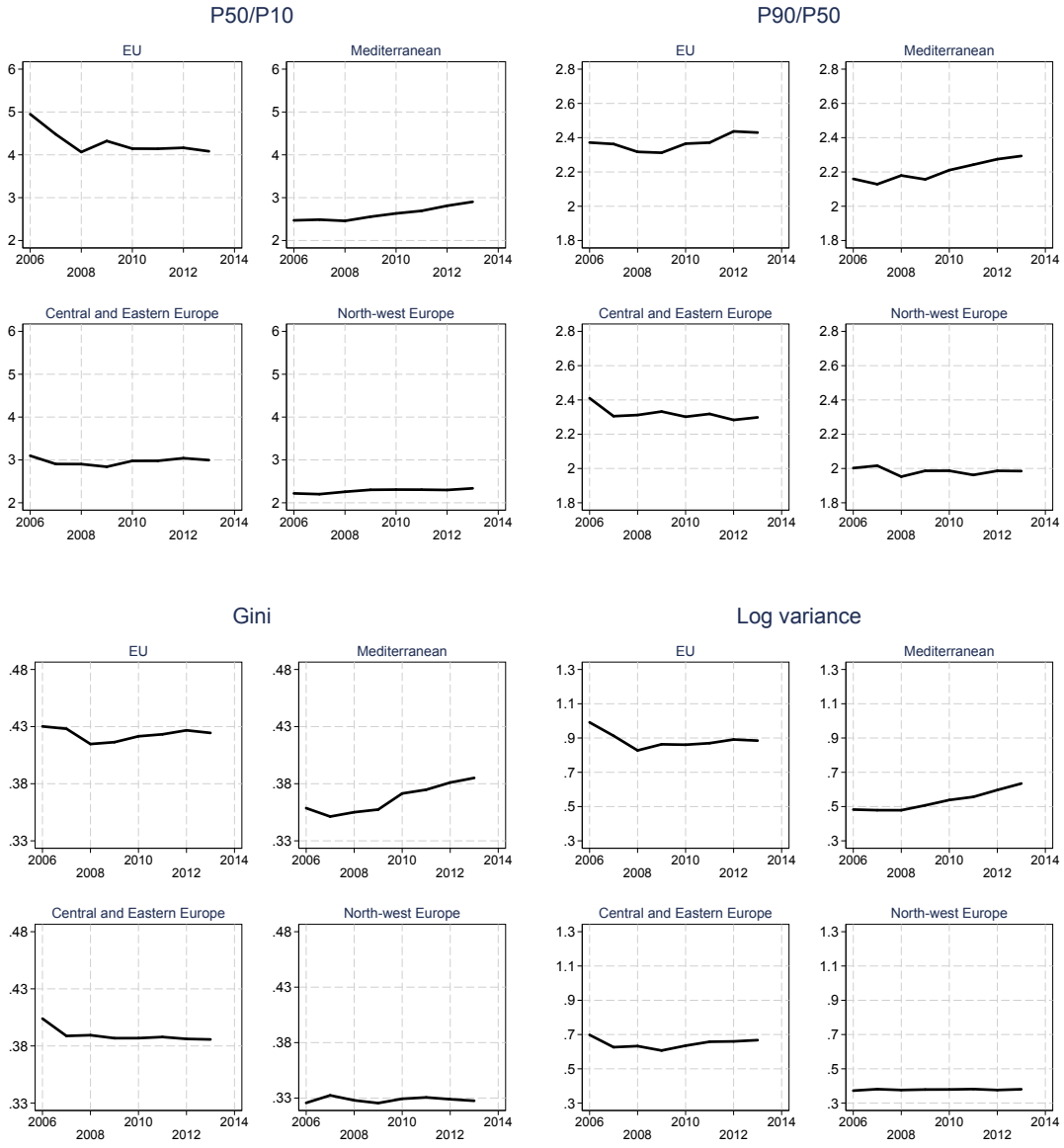
Source: extract from the EUROSTAT on-line database using 'Actual individual consumption'

Italy and Germany are the countries with price levels at about the EU28 average, but the overall range of the price variation extends from 39 to 141% points. This more than three-fold difference is hard to neglect. In general, CEE price levels are the lowest, not much larger than half of the EU28 average, whereas those on the NW are the largest, the MED being somewhat below-average. The relative position of the countries is fairly stable over time, but there are exceptions: Luxembourg and Sweden became relatively more expensive over time, while the UK relatively cheaper to a greater-, Poland and Hungary to a smaller extent. Trends that seemed to prevail formerly have changed in the UK, Ireland, Sweden, Hungary and Poland.

Figure A.3: Quantile ratios (top panel), Gini and log variance (bottom panel) of annual labour income of the working, without ppp adjustment (gross, cpi adjusted)

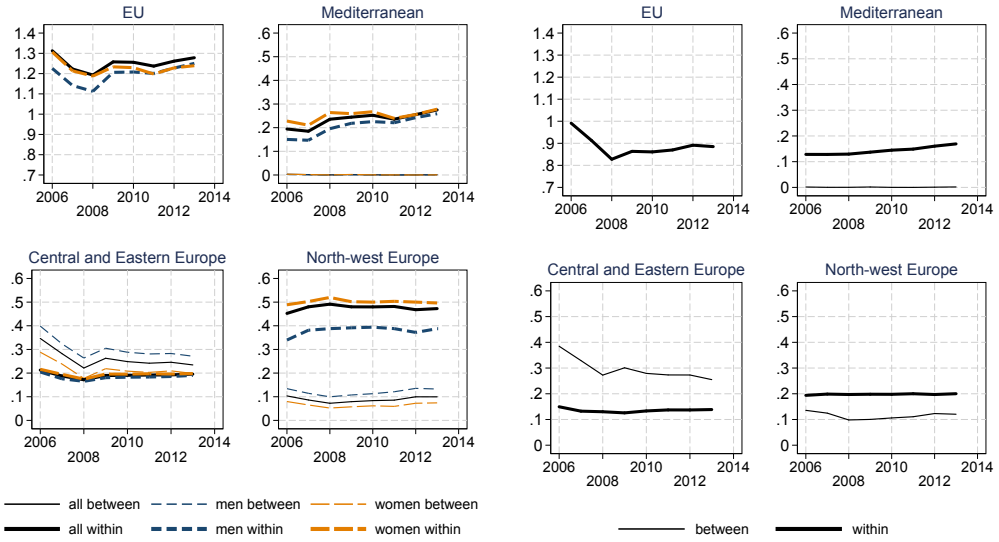
Source: Own calculations from EU-SILC microdata

Figure A.4: Quantile ratios (top panel), Gini and log variance (bottom panel) of household equivalised income, without ppp adjustment (gross, cpi adjusted)



Source: Own calculations from EU-SILC microdata

Figure A.5: Additive decomposition of the log variance of annual earnings of those who work by country-groups' contributions (left panel) and that of the log variance of household equivalised income (right panel), without ppp adjustment (gross, cpi adjusted)



Source: Own calculations from EU-SILC microdata

In order to show the effect of the PPP correction, we reproduce the key figures based on individual- (Figure 4 on page 16 and Figure B.1 on page 46) as well as based on household equivalised income concepts (Figure 14 on page 29 and Figure B.6 on page 52), but this time without PPP correction – see Figure A.3 on page 42 and Figure A.4 on the previous page, respectively. The percentile ratios for individual income suggest that the uncorrected inequality levels are always higher, in particular for the whole of the EU and the lower part of the distribution. (Note differences in scaling!) The difference at the lower end (about 1.5% point) is substantial, larger than the over time or the cross-country group differences observed. The one on the upper end is not significant at only 0.1% point. There are minor differences in trends and changes in gender gaps of different directions, due to the particular interaction of differences of employment of genders and relative price levels (NW at the bottom decreasing, EU on the top increasing).

Note that besides the inequality levels, over-time changes too are substantially influenced by the PPP correction. Both ratios are much smoother when corrected in case of the EU27. For the CEE on the other hand, the correction introduces an increase in inequality in the crisis years, visible both in case of the ratios, as well in case of the log

variance and the Gini – the effect being rather dramatic for the latter. More visible in the case of household data is the rather sharp drop in inequality from a high to lower levels for the lower part of the distribution.

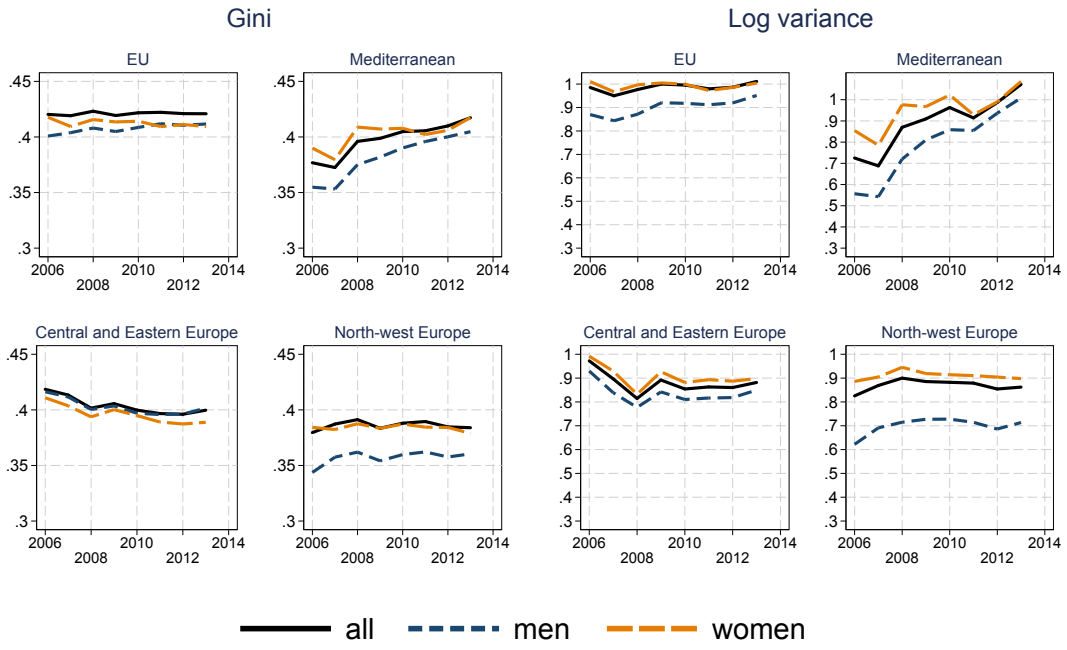
Really substantial differences are visible on the decomposition by country groups, especially in the case of the CEE (see Figure A.5 on the preceding page). Within-region log variances seem to be relatively stable or increasing just a tiny bit in all cases. The between-region differences however have changed quite substantially: higher prices in the NW and lower prices in the CEE have doubled the contribution. This was rather small for the former, but large in the latter case. It is in fact larger in the CEE than the contribution of the within-inequality. Note also the kink for the CEEs in 2008.

The question emerges naturally: which version of inequality measures should we use? The answer depends on what we think about the way utility is generated and the role of savings. If all income was spent at the time and in the country of earning it, then (a good) PPP correction was completely legitimate. Unfortunately we do not know a lot about the savings behaviour or households, but average savings rates seem to be much lower in the CEE than in the NW, for example and than in the case of richer households. Also, some households are more likely to spend their money abroad, partly because of their consumption structure, partly beause they physically are outside the country they make the money in. Both effects are likely to be correlated with income, making the ligitimacy of PPP correction dependent on income levels. Our position is thus that there are pros and cons to using PPP correction, but as we aim at an EU-wide investigation, we can not do without it. Still, this remains an interesting area of further research as we do not have a satisfactory solution for an important problem.

B Additional graphs for the Analysis

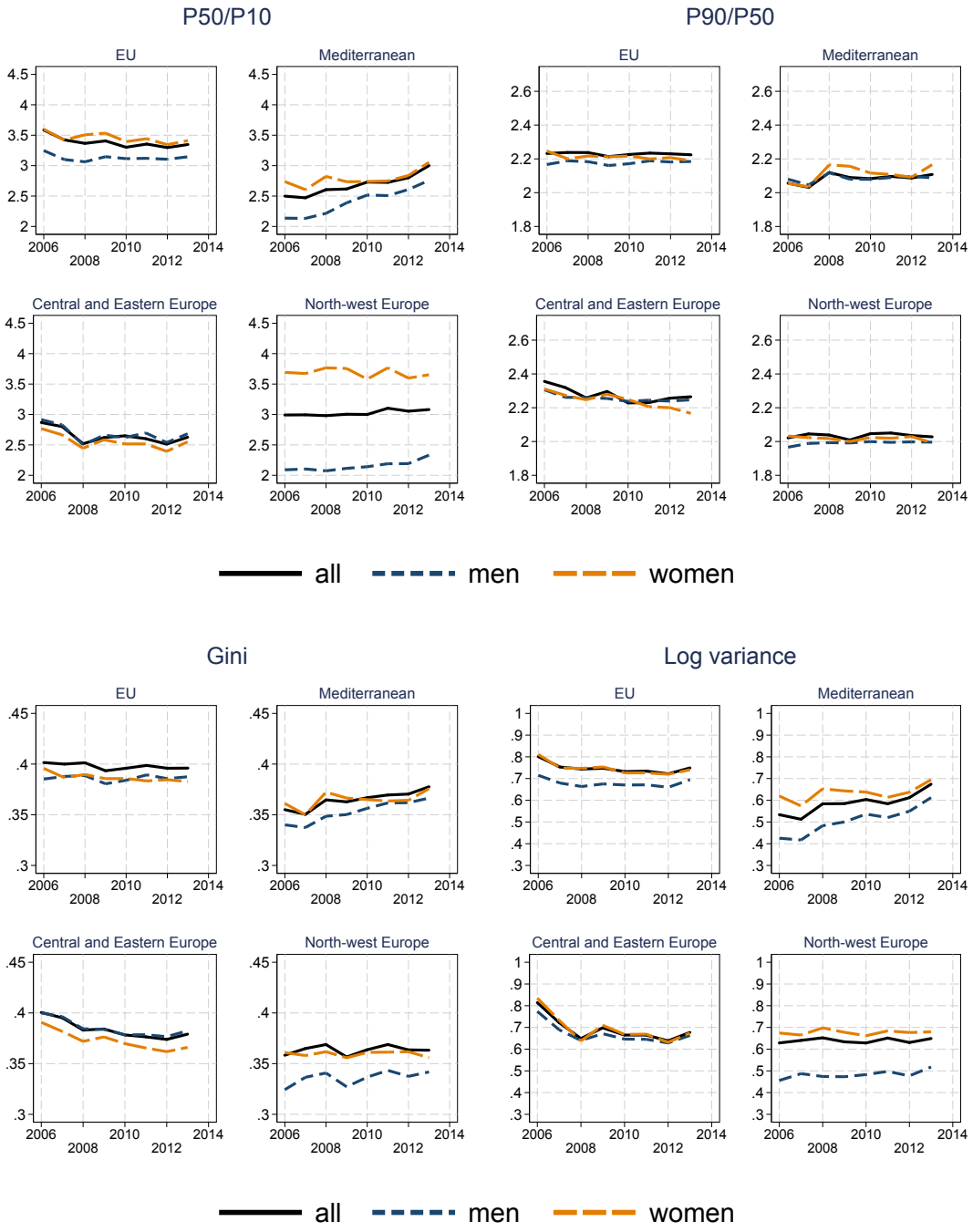
B.1 Personal income inequality of those who work

Figure B.1: Gini and log variance of annual earnings of those who work (gross)



Source: Own calculations from EU-SILC microdata

Figure B.2: Quantile ratios, Gini and log variance of annual income of the working, but for those with hours information as well (sample D2)



Source: Own calculations from EU-SILC microdata

B.1.1 Hourly wages: observable and unobservable influences

Wage rates can be decomposed into observable determinants and unobservable random shocks, estimated in a cross-sectional wage regression following the classic idea of Mincer (1958). This decomposition is relevant here, as it is important to isolate the effect of what is given to the individual, and what he or she receives as a shock as time goes by. We use a cross-sectional regression to trace the path of the parameters of key determinants over time, and separate the stochastic component to be analysed later. We use a simple specification as follows:

$$y = \alpha + \beta_1 s_1 + \beta_2 s_2 + \gamma_1 e + \gamma_2 e^2 + c + \varepsilon, \tag{9}$$

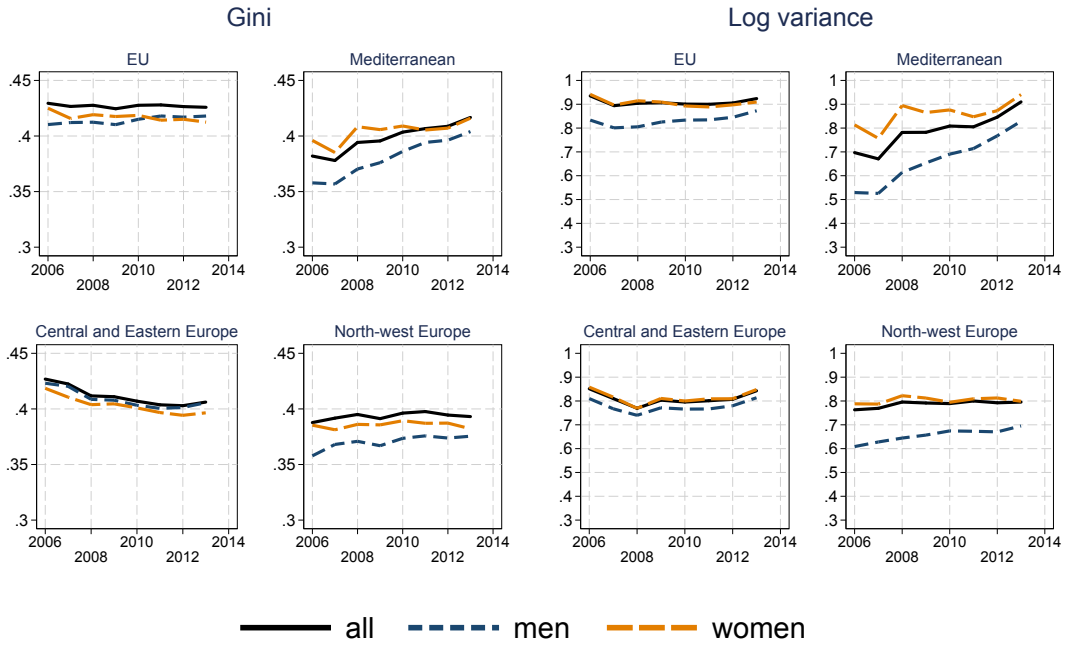
where y is the outcome variable, such as the log of annual income. On the right-hand side s_1 is an indicator for secondary- and s_2 for college education, e stands for experience, and c is a full set of country fixed-effects. All other unobserved stochastic influences are subsumed in ε, assumed to be orthogonal to everything already included. Mincerian wage regressions became a workhorse of labour economics over the years and so has estimating them become an industry. By today's micro-econometric standards, we have no chance to estimate a 'true' parameter such as the returns to education using the EU-SILC: we do not have a true hourly wage rate, have less than desirable observables and thus we are unable to either control for a number of actually observable characteristics or to undo selection to those working.

Keeping in mind the above limitations, we estimate the equation by years to look at the time path of the three key variables of elementary human capital research, schooling, experience and gender. We use the above specification and sample D2 (with valid hours and thus wage rate data). Estimating separately by gender, we are both allowing all parameters and the stochastic structure of ε to vary. [27]

[27] One could also make a next step and use the residuals to estimate the share of permanent and transitory shocks, using the panel EU-SILC files (similarly to Heathcote et al. (2010)) and Domeij & Flodén (2010), who themselves do not report encouraging results). We tried and failed modelling this dynamic behaviour. Unfortunately, the limitations of the 4 year rotating structure of the EU-SILC files and the simple nature of the model imply that the estimated time path of the components is very sensitive to the estimation method – both cited papers document that the choice of using levels or differences for the estimation has a substantial effect on the values one obtains.

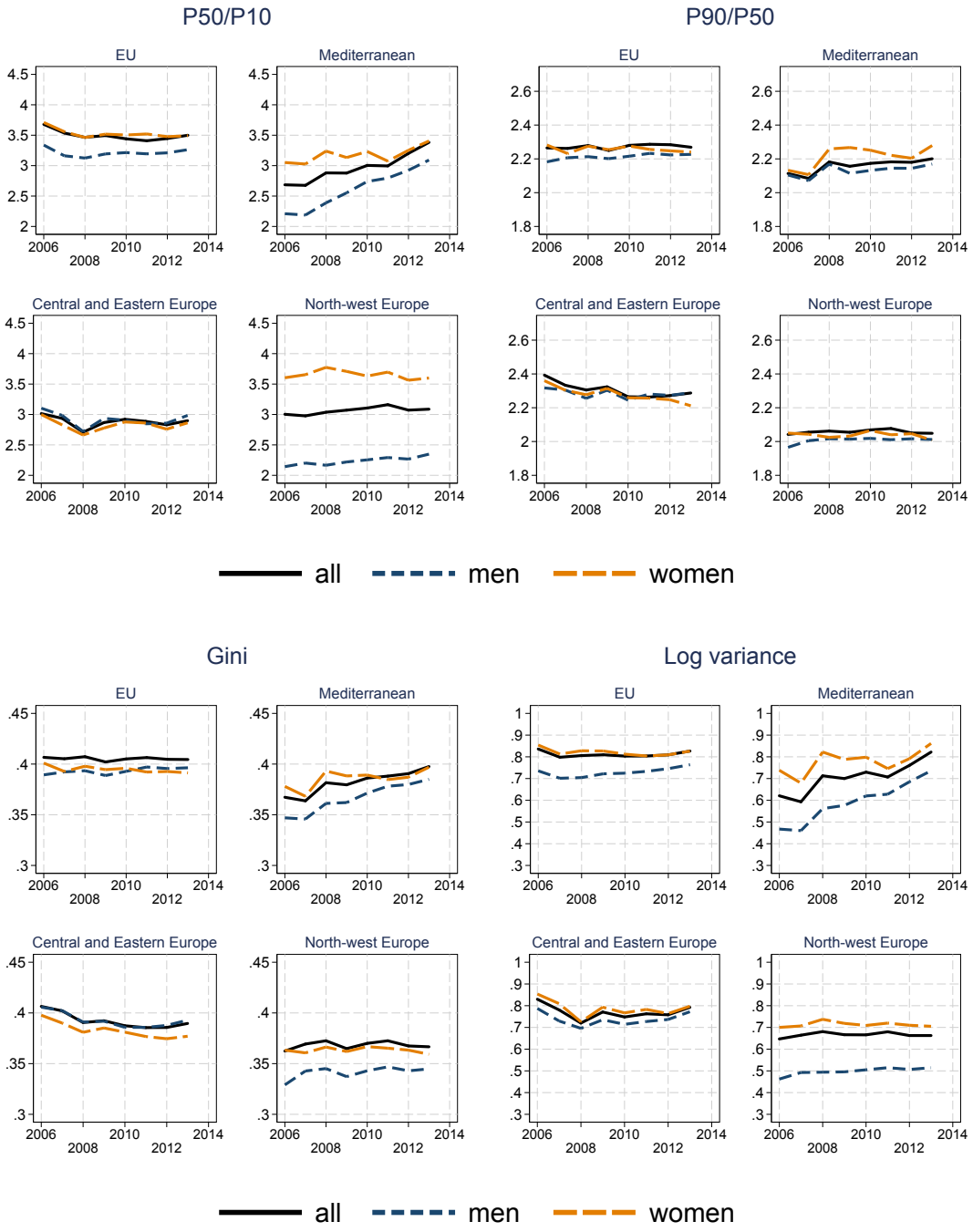
B.1.2 Personal income inequality non-work income included

Figure B.3: Gini and log variance of annual personal income (sample C2, gross)



Source: Own calculations from EU-SILC microdata

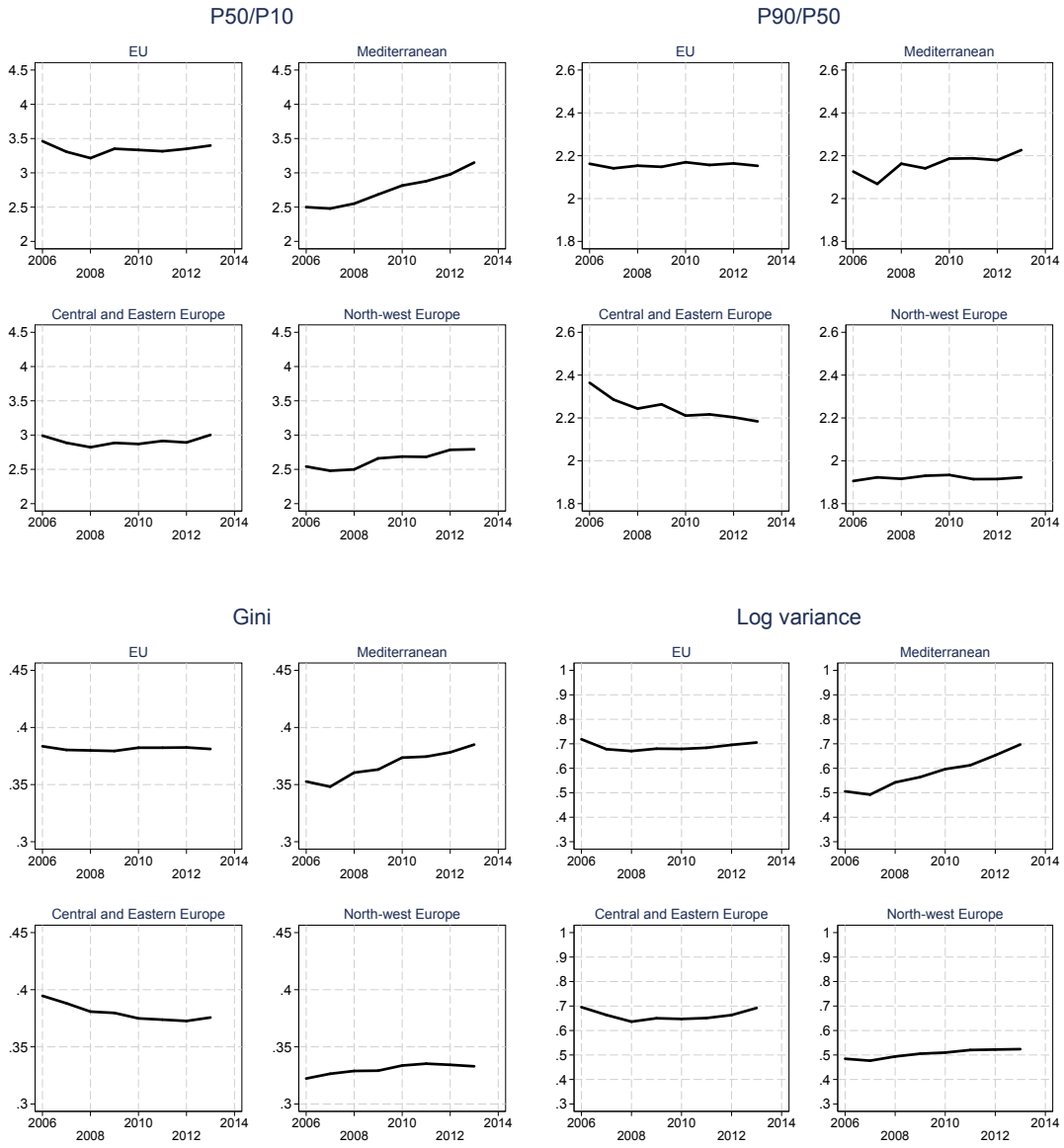
Figure B.4: Quantile ratios (top panel), Gini and log variance (bottom panel) of annual personal income for those who work, that is sample D (nonzero only, gross)



Source: Own calculations from EU-SILC microdata

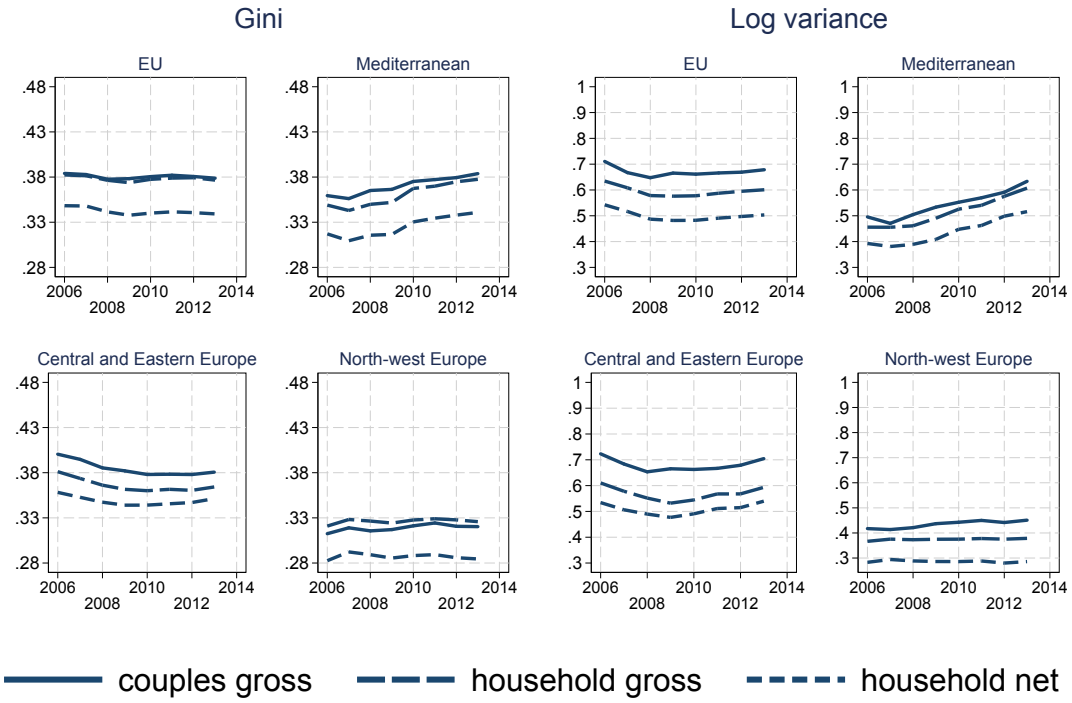
B.2 Inequality and risk sharing in coalitions of persons

Figure B.5: Quantile ratios (top panel), Gini and log variance (bottom panel) of annual personal income shared equally between partners (if any) among those having own in-come (sample C2)



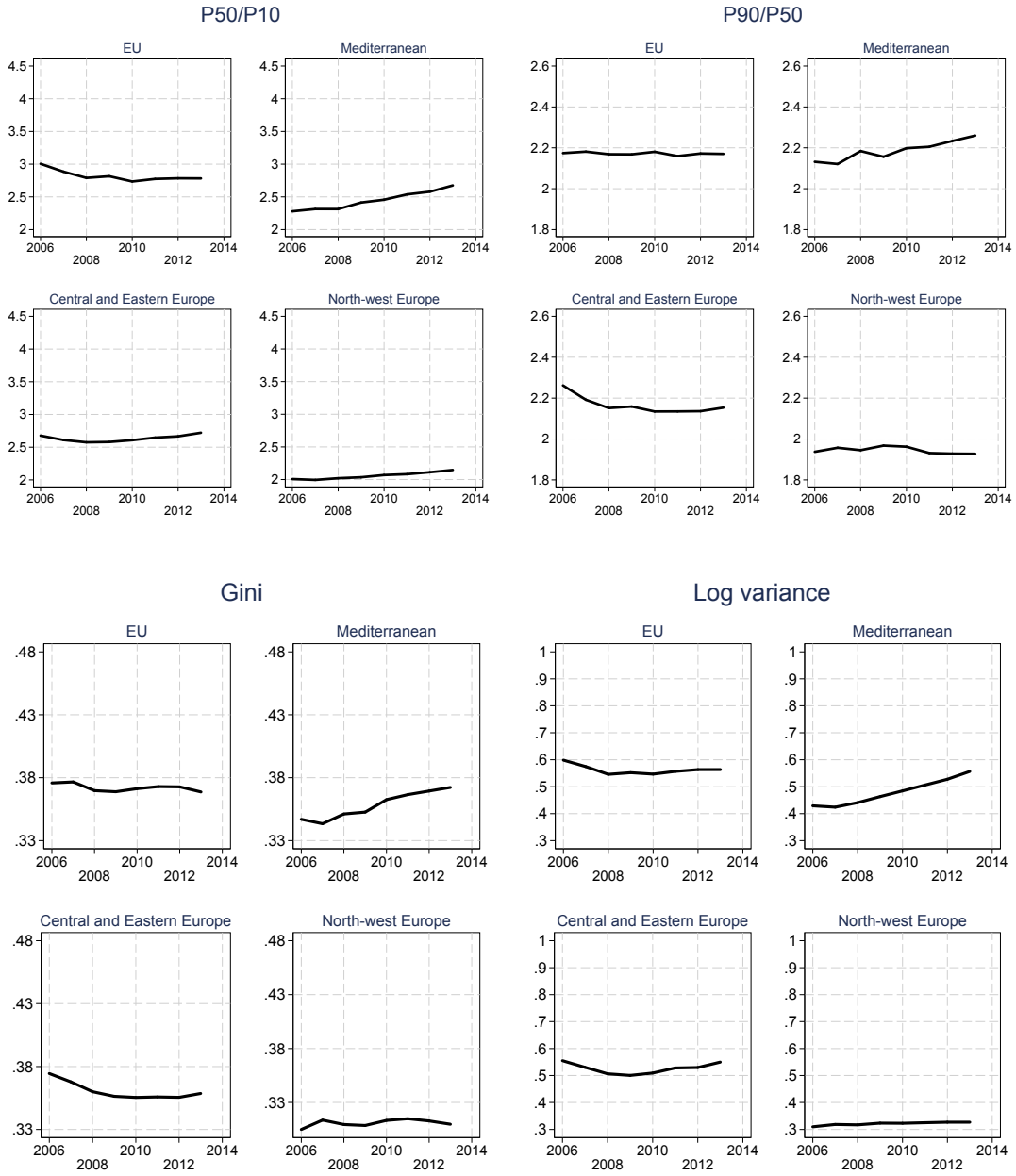
Source: Own calculations from EU-SILC microdata

Figure B.6: Gini and log variance of annual income of different coalitions (gross)



Source: Own calculations from EU-SILC microdata

Figure B.7: Quantile ratios of household equivalised income (gross) only for those with nonzero income shared with partner, that is sample C1



Source: Own calculations from EU-SILC microdata

Europe Direct is a service to help you find answers to your questions about the European Union.

Freephone number (*)**:**

00 800 6 7 8 9 10 11

(*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

More information on the European Union is available on the internet (http://europa.eu).

HOW TO OBTAIN EU PUBLICATIONS

Free publications:

- one copy:
  via EU Bookshop (http://bookshop.europa.eu);
- more than one copy or posters/maps:
  from the European Union's representations (http://ec.europa.eu/represent_en.htm);
  from the delegations in non-EU countries (http://eeas.europa.eu/delegations/index_en.htm);
  by contacting the Europe Direct service (http://europa.eu/europedirect/index_en.htm) or
  calling 00 800 6 7 8 9 10 11 (freephone number from anywhere in the EU) (*).

  (*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

Priced publications:

- via EU Bookshop (http://bookshop.europa.eu).

KJ-AE-17-014-EN-N

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub
ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

Joint Research Centre

EU Science Hub

Publications Office

doi:10.2760/244655

ISBN 978-92-79-77303-7