

Earnings Inequality in Spain: New Evidence Using Tax Data*

Stéphane Bonhomme[†]

Laura Hospido[‡]

CEMFI

Bank of Spain and IZA

Revised version: November 2012

Abstract

We use tax files from 2004 to 2010 to document the recent evolution of earnings inequality in Spain. We find that inequality went in parallel with the evolution of the unemployment rate during the period. This evolution is consistent with the evidence from Social Security records recently documented in Bonhomme and Hospido (2012). Quantitatively, the 90/10 percentile ratio of daily earnings experienced a 10% increase between 2007 and 2010, which is partly but not fully explained by changes in labor force composition. We also use the tax data to study the evolution of the gender earnings gap, and find that it has decreased throughout the distribution during the period. Lastly, we tentatively exploit the panel dimension of the data to explore the permanent and temporary dimensions of Spanish inequality.

JEL classification: D31, J21, J31.

Keywords: Earnings Inequality, Tax files, Business cycle, Gender gap.

*We thank an anonymous referee and Olympia Bover for useful comments. Support from the European Research Council/ ERC grant agreement n^o263107 is gratefully acknowledged. All remaining errors are our own. The opinions and analyses are the responsibility of the authors and, therefore, do not necessarily coincide with those of the Bank of Spain or the Eurosystem.

[†]Corresponding author. Email: bonhomme@cemfi.es. Mail address: CEMFI, Casado del Alisal 5, 28014 Madrid, Spain.

[‡]Email: laura.hospido@bde.es. Mail address: Bank of Spain, Alcalá 48, 28014 Madrid, Spain.

1 Introduction

In this paper we use tax files from 2004 to 2010 to document the recent evolution of earnings inequality in Spain. The evidence that we present complements the results we obtained using Social Security data (see Bonhomme and Hospido, 2012).

While evidence for the US and other developed countries has long been available (see for example Autor *et al.*, 2008, or Dustmann *et al.*, 2009), the study of Spanish earnings inequality has recently been made possible by the availability of Social Security records from the Continuous Sample of Working Histories (MCVL in Spanish). Bonhomme and Hospido (2012) used this source of consistent earnings data to document the evolution of earnings inequality over a long period, from 1988 to 2010.¹ Here we use the tax files, which were matched to the MCVL by the Social Security administration starting in 2004, to provide additional evidence for the more recent period.

Tax files provide a complementary perspective on earnings distributions. Taxable labor income and social security contributions are closely related— though conceptually distinct— measures of workers’ earnings. For example, in a comparison exercise Bonhomme and Hospido (2012) found a correlation of about 90% between the two earnings measures for the recent period. However, social security data are subject to top and bottom-coding. As a consequence, making inference on the entire distributions of earnings requires extrapolation based on a parametric model. Bonhomme and Hospido (2012) used cell-specific tobit regressions for extrapolation, and exploited the tax data to validate this parametric extrapolation method. In contrast, this paper directly relies on the uncensored tax records.

The absence of censoring allows for a more complete description of earnings distributions. As an example, we are able to compute different indices of inequality, including earnings percentile ratios but also the standard deviation of log-earnings and the Gini index. In addition, using the tax files allows us to document the level and evolution of top earnings percentiles; specifically, the 90th, 95th, and 99th percentiles. The absence of censoring also enables the analysis of longitudinal dimensions of inequality, taking advantage of the fact that the tax data follow the same individuals over the 2004-2010 period.

The goal of this paper is twofold. First, we document the evolution of overall earn-

¹Previous evidence on this evolution was mostly based on survey data. See Pijoan-Mas and Sanchez-Marcos (2010), Carrasco *et al.* (2011), and Izquierdo and Lacuesta (2012) among other references.

ings inequality during the recent period. We find that earnings inequality, as measured by the 90/10 percentile ratio of daily earnings, evolved in parallel with the unemployment rate during the period. Between 2007 and 2010— a severe recession period in Spain— inequality increased by 10 percentage points. This increase was more salient for men than for women, and was concentrated in the bottom part of the earnings distribution. This provides further evidence on the strong cyclicity of Spanish earnings inequality, which we documented in our earlier paper.

The second goal of this paper is to explore several dimensions of earnings inequality using the tax data. The first aspect we study is the effect of changes in labor force composition on earnings inequality. Using a simple reweighting approach, we find that while changes in age and occupation composition account for a small part of the increase in inequality during the recent recession, allowing for changes in sectoral composition explains up to half of the increase in both male and female inequality. The sectoral dimension seems to have played an important role in the recent evolution of the Spanish labor market, in particular due to the weight of the construction sector in the economy.

Gender differences are another interesting dimension of earnings inequality that may be studied using the tax data. We document that the gender gap in daily earnings is highest at the bottom of the distribution, but remains substantial at the top— where males earn 20% to 30% more than females. We find that differences in age, occupation (broadly defined using 10 occupation groups from the social security), immigrant status, type of contract and sector of activity do not fully explain this gap. Moreover, our results suggest that the earnings difference between males and females has decreased between 2004 and 2010, by up to 5%, throughout the distribution. The evidence of a decrease in the gender gap is robust to considering full-time workers only.²

Lastly, we take advantage of the panel dimension of the tax data to tentatively explore the permanent or transitory nature of earnings inequality. For this, we start by documenting the evolution of intertemporal correlations in annual earnings during the period. We find that earnings mobility remained rather constant, at the same time as earnings inequality was rising. We then compute averages of annual earnings over time for every individual in our sample. We find that the distribution of individual intertemporal averages of annual earnings is substantially less dispersed than the cross-

²De la Rica *et al.* (2008) study the gender gap in hourly earnings in Spain using the European Community Household Panel for 1999. See also Gardeazábal and Ugidos (2005), and Guner *et al.* (2012) for further evidence on the Spanish gender gap.

sectional distributions, suggesting that part of the inequality that this paper documents is transitory.

The outline of the paper is as follows. In Section 2 we describe the tax data and our main earnings measures. In Section 3 we show the results on the evolution of earnings inequality. In Section 4 we assess the impact of changes in labor force composition in this evolution, study the gender gap in earnings, and explore the permanent and transitory dimensions of earnings inequality. Lastly, Section 5 concludes.

2 The data

This paper uses the tax files that have been recently matched to micro-level social security records from the Continuous Sample of Working Histories (*Muestra Continua de Vidas Laborales*, MCVL, in Spanish). The matched data are so far only available from 2004 to 2010, so we will limit our study to this period.

2.1 Sample selection

The MCVL is a representative sample of the population registered with the social security administration. From 2004 to 2010, the MCVL has a proper longitudinal design, that is, an individual who is present in a wave and subsequently remains registered with the social security administration stays as a sample member. In addition, the sample is refreshed with new sample members so it remains representative of the population in each wave.³

The population of reference of the MCVL consists of individuals registered with the social security administration at any time in the reference year, including pension earners, recipients of unemployment benefits, employees and self-employed workers, but excluding individuals registered only as medical care recipients, or those with a different social assistance system (part of the public sector, such as the armed forces or the judicial power). The raw data represent a 4 per cent non-stratified random sample of this reference population, and consist of nearly 1.1 million individuals each year.

We use data from the 2004-2010 original samples. We keep prime-age employees

³As a complement, the MCVL also tries to reconstruct the labor market histories of the individuals in the sample back to 1967, earnings data being available since 1980. However, in this study we do not use this retrospective information.

(aged 25-54) enrolled in the general regime,⁴ with positive taxable income in the reference year. We obtain a panel of 696,223 individuals (374,260 men and 321,963 women) and more than 3.3 million annual observations for the 2004-2010 period. We present descriptive statistics on sample composition and demographics by gender in Appendix B.

2.2 Tax data and earnings measure

The tax data come from the “Annual summary of retentions and payments for the personal income tax on earnings, economic activities, awards and income imputations” (model 190), which is a declaration filled by the employer that includes the total amount of individual compensation paid to employees in the reference year.⁵ This information is available for all employees, whether or not they are exempted from paying the labor income tax. For each individual present in the MCVL and each year, we pool all records corresponding to the different employers she has had during the year.

In our analysis, we use *daily earnings* as our main earnings measure, computed as the ratio between annual taxable labor income and the number of days worked in a particular year. Taxable labor income is computed as the full amount of monetary compensation effectively received by employees.⁶ Earnings are deflated using the 2010 general price index. However, as the social security data do not record hours of work, we cannot compute a hourly wage measure.⁷ As a complement, in Subsection 4.3 we will also use annual taxable labor income in order to compare cross-sectional inequality and permanent inequality.

⁴In Spain, more than 80 per cent of workers are enrolled in the general scheme of the social security administration. Separate schemes exist for some civil servants, workers in fishing, mining and agricultural activities, and the self employed. This means that these categories are not considered in this study.

⁵See Appendix A for details.

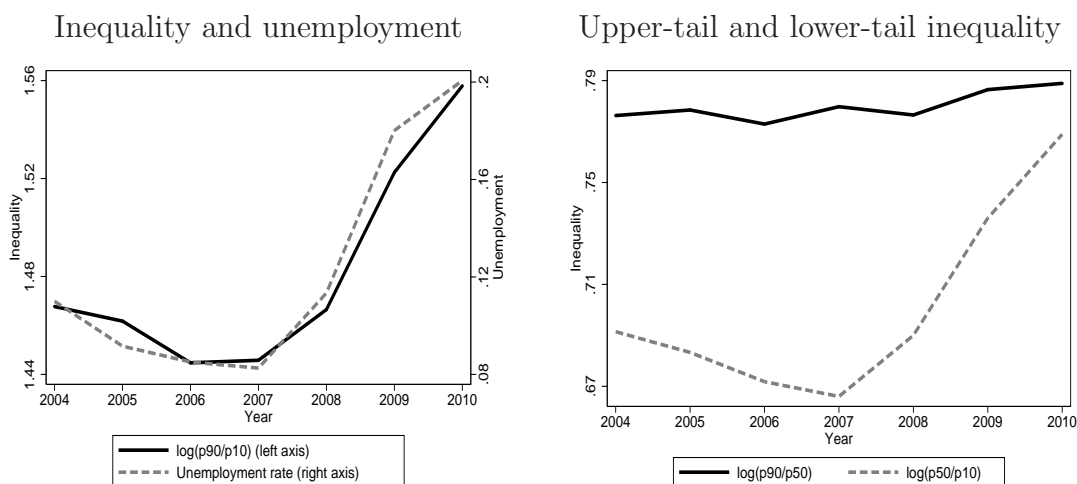
⁶This amount may also include travel allowances and other non-taxable expenses.

⁷The data contain measures of part-time and full-time work. In Bonhomme and Hospido (2012) we re-weighted daily earnings using these measures and found little difference for males, although it did somewhat affect the results for females, especially at the bottom of the earnings distribution. In Subsection 4.2, when comparing males and females, we will also report results using full-time workers only.

3 Overall evolution of earnings inequality

On the left panel of Figure 1 we start by showing the evolution of earnings inequality, as measured by the logarithm of the ratio between the 90th and 10th percentiles of daily earnings. The numbers include both males and females. We see that inequality decreased slightly up to 2007, before increasing sharply— by more than 10 log-points— between 2007 and 2010. On the same graph we report the evolution of the unemployment rate (dashed line, right axis). We see that the evolution of inequality and that of the unemployment rate went in parallel during the period. This evidence points to the fact that Spanish inequality is markedly cyclical, as documented in Bonhomme and Hospido (2012).

Figure 1: Earnings inequality

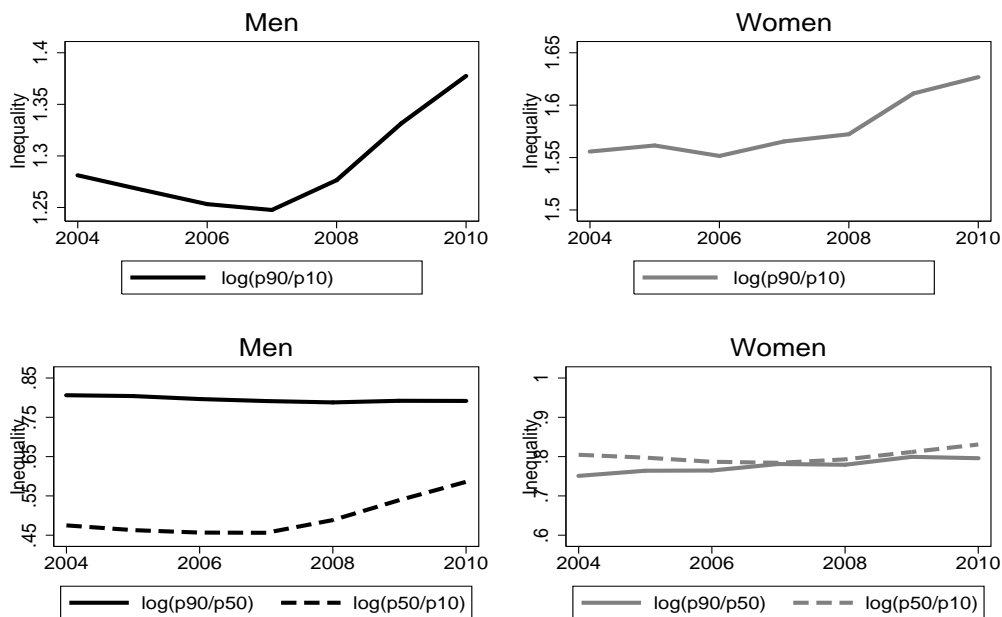


Notes: Source Tax data. Inequality is measured by the logarithm of the ratio of the 90th and 10th percentiles of daily earnings. Upper-tail inequality is measured by the logarithm of the ratio of the 90th and 50th percentiles of daily earnings. Lower-tail inequality is measured by the logarithm of the ratio of the 50th and 10th percentiles of daily earnings.

In addition, the right panel of Figure 1 shows the evolution of the logarithms of the 90th and 50th earnings percentile (upper-tail inequality) and of the 50th and 10th percentiles (lower-tail inequality). According to the tax data, most of the inequality increase during the recent recession occurred in the lower half of the earnings distribu-

tion.

Figure 2: Earnings inequality by gender



Notes: See notes to Figure 1.

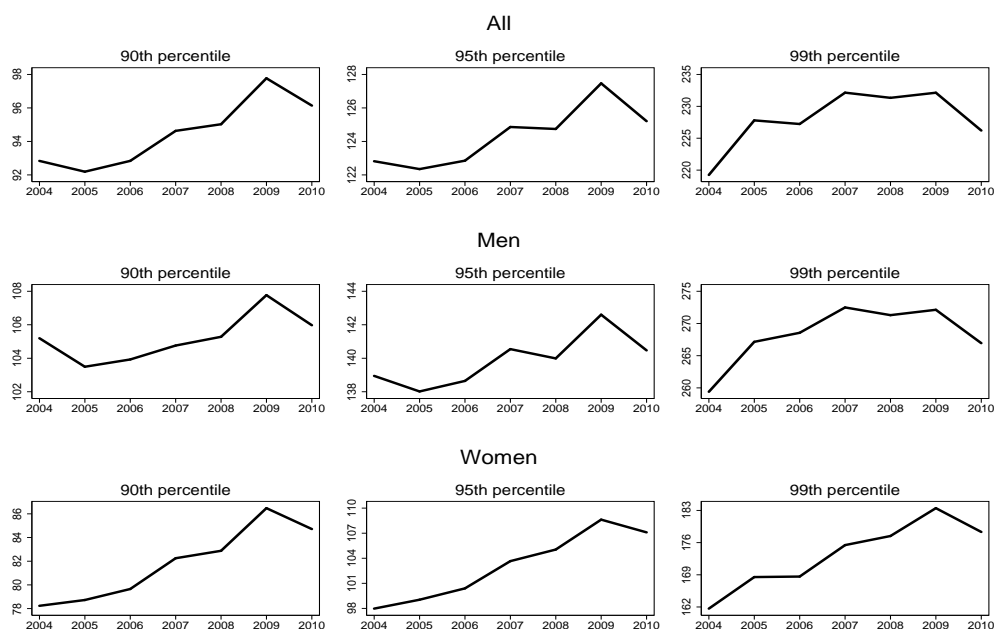
Figure 2 shows the evolution of the 90/10 percentile ratio (upper panel) and of the 90/50 and 50/10 ratios (lower panel), by gender. Although female inequality also shows an increase during the period, its evolution appears to less clearly follow the business cycle than male inequality. In addition, while the 90/50 ratio remains rather stable for males, it shows some increase for females.

The evolution of upper- and lower-tail inequality differs somewhat from the results in Bonhomme and Hospido (2012), which are based on social security records. Indeed, while according to the tax data upper-tail inequality remained rather constant for males, according to the social security data the 90/10 percentile ratio increased by 3.8 percentage points between 2007 and 2010. One possible reason for this discrepancy is that wage bonuses, which are not contained in social security earnings but are part of taxable income, may have decreased during the recent recession.⁸

⁸Consistently with this interpretation, the percentage difference between tax earnings and uncensored social security earnings of high-skilled workers (occupation groups 1-3) decreased from 3.3% in 2007 to 2.3% in 2010, while it remained stable at 2% for middle- and low-skilled workers (occupation

As a complement, Table C.1 in Appendix C shows the evolution of several indices of inequality: the 90/10 percentile ratio, the 90/50 and 50/10 ratio, the standard deviation of log-earnings, and the Gini index. Results are reported for the full sample, and for males and females separately. All indices show an increase in the second part of the period, which coincides with the recession. In particular, we see that inequality has increased for both men and women. The increase is most pronounced when using the 90/10 percentile ratio, and least pronounced when using the Gini index.

Figure 3: Top earnings percentiles



Notes: Source Tax data.

Lastly, tax data allow to document the right tail of the distribution of labor earnings; see for example Piketty and Saez (2006), and Alvaredo and Saez (2009) for evidence for Spain until 2005. In Figure 3 we show the evolution of the 90th, 95th, and 99th percentiles of daily earnings. We see that the three percentiles increased until 2009, and fell at the end of the period. For males the increase until 2009 amounts to 3%, 3%, and 5% at the 90th, 95th, and 99th percentiles, respectively. This almost parallel evolution for males is consistent with the fact that the recent increase of male inequality was essentially due to an increase in the lower part of the distribution, and not at the groups 4-10).

top. For females, the 90th, 95th, and 99th percentiles increased by 10%, 11% and 13%, respectively, between 2004 and 2009. In Subsection 4.2 we will more precisely document gender differences in earnings.⁹

4 Three aspects of earnings inequality

In this section we document three dimensions of earnings inequality: the impact of changes in labor force composition, earnings differences between men and women, and the permanent and transitory nature of inequality.

4.1 Changes in labor force composition

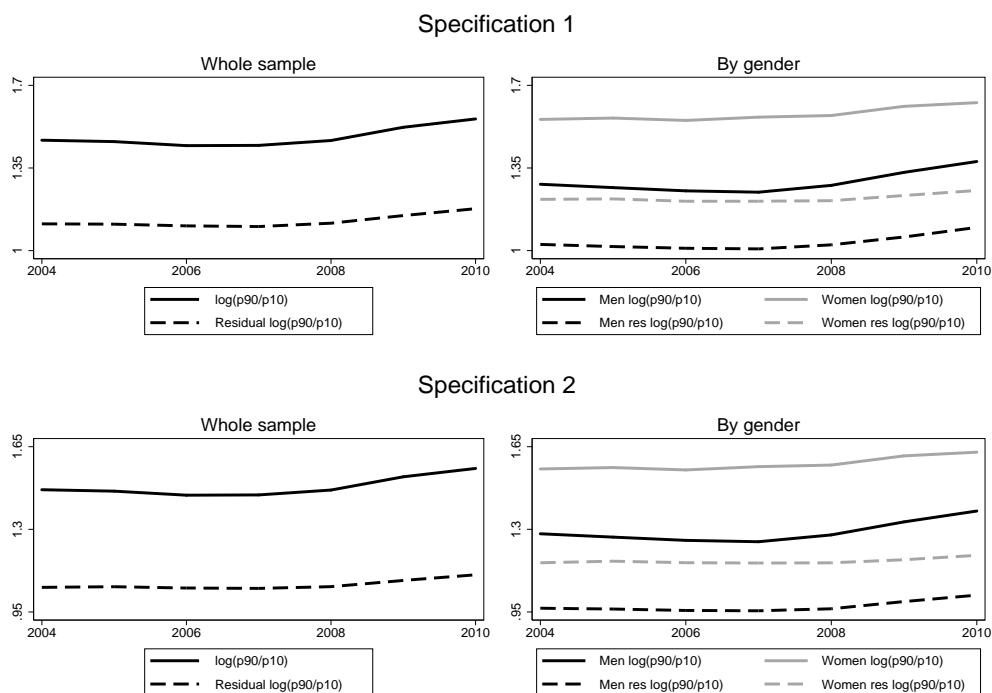
Residual inequality. Here we aim at assessing the impact of changes in labor force composition on the evolution of earnings inequality. We start by documenting the evolution of *residual inequality*. The latter is computed using least-squares residuals from a pooled OLS regression of log-earnings on a number of individual characteristics. We compare two specifications: one where the controls include gender, age, year, occupation and immigrant dummies (Specification 1), and another specification that includes in addition sectoral dummies and a dummy for the type of contract (permanent versus fixed-term).¹⁰

Figure 4 shows the evolution of total earnings inequality and residual inequality, overall and by gender. Table C.2 in Appendix C shows the percentage changes during the subperiod of expansion (2004-2007) and during the subperiod of recession (2008-2010). The results suggest that changes in labor force composition have partly contributed to the evolution of earnings inequality during the two subperiods. For example, for males the 90/10 ratio of earnings residuals decreased by 1% during the expansion and increased by 6% during the recession according to Specification 2. In comparison, the decrease and subsequent increase in total inequality were 3% and 8%, respectively. Interestingly, changes in composition seem to have contributed to half of the increase in the 50/10 ratio for males (i.e., 9.9% versus 19.9%).

⁹Note that our earnings measure is daily and only includes labor income. Income from other sources, assets and wealth are available for Spain for the years 2002, 2005 and 2008 from the *Encuesta Financiera de las Familias* (EFF).

¹⁰The full set of covariates is indicated in the notes to Figure 4.

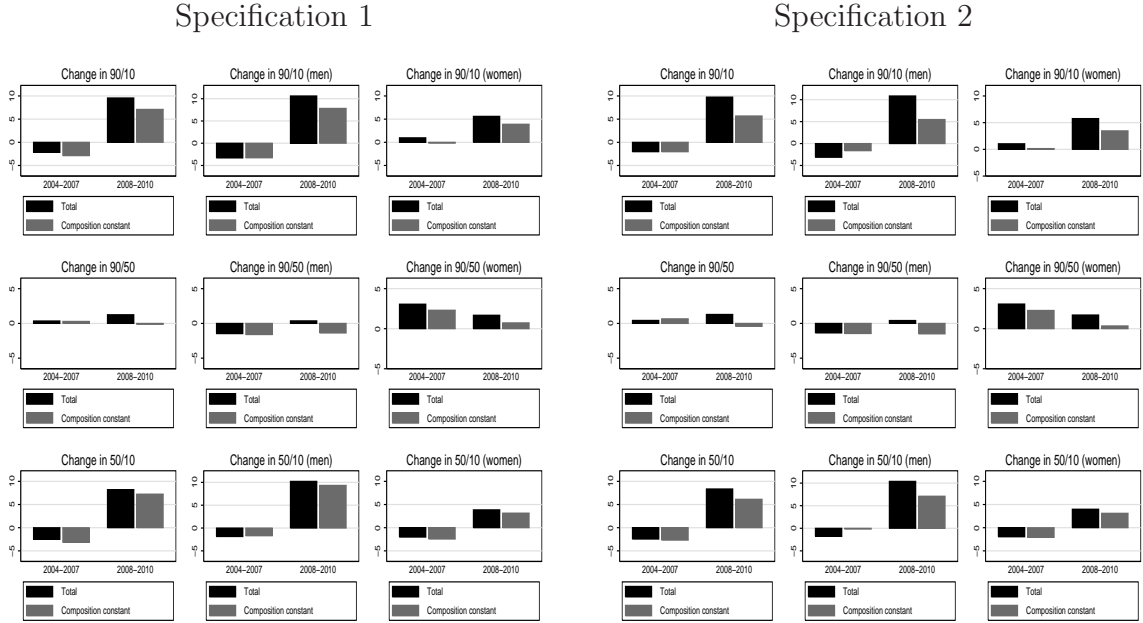
Figure 4: Overall and residual inequality



Notes: Source Tax data. Specification 1 includes a gender dummy (only for the whole sample), age dummies (25-54), year dummies (2004-2010), occupation group (1-10), and an immigrant/native dummy. Specification 2 includes in addition dummies of sector of economic activity (10 sectors, see Table B.2 in Appendix B), and a dummy of type of contract (permanent/fixed-term).

It is also interesting to document between-group differences during the period. As an example, in Figure C.1 in Appendix C we show the evolution of the earnings gap between workers with a permanent or a fixed-term contract. The duality of the Spanish labor market created by these two types of labor contracts has received attention in the academic literature (e.g., Dolado *et al.*, 2011). We see that, for male workers, the gap decreased at the beginning of the period, and that it increased slightly during the recent recession. This evolution is consistent with the evidence reported in Bonhomme and Hospido (2012). For females, the evolution shows a continued decrease in the difference between permanent and fixed-term workers.

Figure 5: Inequality decomposition



Notes: Source Tax data. The grey bars (“composition constant”) report the evolution of inequality that would have been observed if labor force composition had been fixed to the end-of-subsample composition (i.e., either 2007 or 2010). See notes to Figure 4 for the definition of covariates specifications.

A decomposition exercise. In the last part of this subsection we show the results of a formal decomposition exercise. The methodology that we use is based on a simple reweighting approach, and complements the above evidence on residual inequality using a nonparametric technique. To illustrate our approach, let us suppose that one wants to compute the evolution of earnings inequality between 2008 and 2010 net of composition changes. For this, we group individual covariates into discrete cells, and reweigh the 2008 observations using cell percentages of 2010. The distribution of 2008 earnings in this reweighted sample is then used to compute a counterfactual measure of inequality net of composition changes. We proceed similarly to compare 2004 and 2007, taking 2007 as the reference year. Consistently with an extensive literature on wage decompositions (e.g., Autor *et al.* 2008), we refer to changes in inequality net of composition changes as “price effects”.

The dark bars in Figures 5 show the percentage changes in total (i.e., observed) inequality, while the grey bars show the percentage changes net of composition effects. We see that composition changes in terms of age and occupation (Specification 1) explain a relatively small part of the evolution. Figure C.2 in Appendix C shows that also including immigrant status and type of contract does not increase the contribution of composition changes. However, when accounting for changes in sectoral composition (right panel in Figure 5), composition changes explain up to half of the increase in inequality during the recent recession for males, slightly less so for females. This evidence is consistent with the evolution of residual inequality reported in Figure 4, and suggests that composition and price effects have both had a substantial impact on the evolution of Spanish inequality. The empirical relevance of sectoral changes has been put forward by Bonhomme and Hospido (2012) as one of the factors behind the recent evolution of Spanish earnings inequality, particularly due to the weight of the construction sector in the economy.

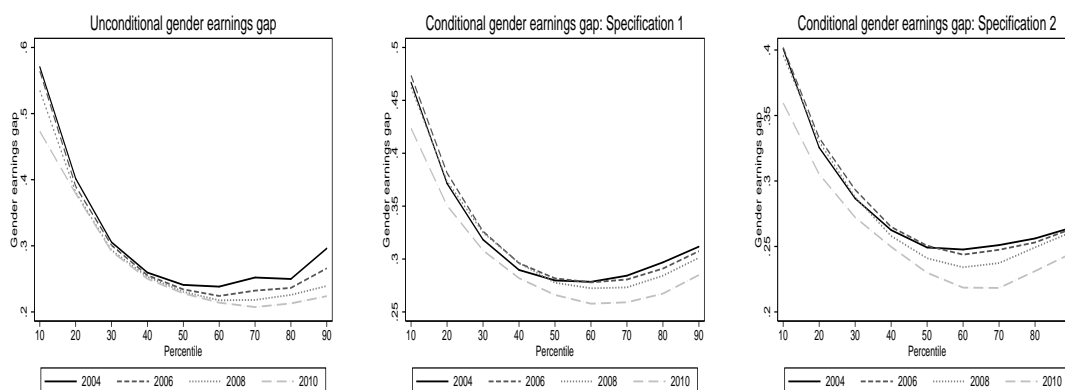
4.2 Gender differences in earnings

In this subsection we document the level and evolution of the gender gap in daily earnings during the recent period. When interpreting the evidence, it is important to keep in mind that our data do not measure hours of work, so the gaps in daily earnings that we document are not hourly wage gaps.

The left graph in Figure 6 shows the difference between the τ -unconditional percentiles of log-earnings of males and females, where τ percentiles are shown on the x -axis. Table C.3 in Appendix C shows the values for $\tau = .10, .50, .90$ and all years. We see that the gender gap is highest at the bottom of the distribution, that it decreases to about 25% at the median and increases slightly at the top of the distribution. Moreover, the earnings gap decreased monotonically during the period, especially at the tails: at the 10th percentile it went from 57% in 2004 to 47% in 2010, while at the 90th percentile the earnings gap decreased from 30% in 2004 to 22% in 2010.

In the central and right graphs of Figure 6 we attempt to control for individual characteristics when comparing male and female earnings. To do so we run quantile regressions of log-earnings on gender and additional characteristics (specifically, those of covariates Specifications 1 and 2 above) and report the gender coefficients— for various values of the percentile index τ — on the graphs. We see that accounting for differences in characteristics decreases the gender gap at the bottom of the distribution: at the

Figure 6: Gender gap in daily earnings



Notes: Source Tax data. The left graph shows the difference between the τ -unconditional percentiles of log-earnings of males and females. The central and right graphs show the gender coefficient in a τ -quantile regression of log-earnings where covariates follow Specification 1 and 2, respectively (see the notes to Figure 4). τ is shown on the x -axis. Various years– 2004, 2006, 2008 and 2010– are shown on the graphs.

10th percentile the gap is now smaller than 40% according to Specification 2. At the same time, accounting for characteristics tends to increase the gender gap in the middle of the distribution.¹¹ Interestingly, the results that account for differences in individual characteristics suggest that the gender gap in daily earnings has decreased between 2004 and 2010 throughout the distribution, with the decrease ranging between 2% and 5% according to Specification 2.

One limitation of the above results is that they do not account for differences in hours worked. In order to partly capture these differences, we next document the gender earnings gap for full-time workers only. Table C.4 in Appendix C shows that the proportion of full-time workers for both males and females has decreased during the period. Moreover, the proportion of full-time working females (about 70%) is substantially lower than the proportion of full-time working males (about 90%). Table C.5 in Appendix C shows that, once differences between full-time and part-time workers are taken into account, the gender gap at the bottom of the earnings distribution is

¹¹Though surprising at first sight, this evidence is consistent with the descriptive statistics reported in Table B.1 in Appendix B: in Spain women are more educated than men, and work in higher-skilled occupations.

greatly reduced. Nevertheless, the evidence of a decrease in the gender earnings gap between 2004 and 2010— throughout the distribution— is robust to considering full-time workers only.

4.3 Permanent and transitory inequality

The literature has emphasized that the welfare consequences of earnings inequality depend on its degree of persistence. A large amount of work has aimed at separating permanent and transitory components of inequality, see among others Hall and Mishkin (1982), and Gottschalk and Moffitt (1994).¹²

The goal of this subsection is to provide some preliminary evidence on the longitudinal information in the tax data. In Table 1 we start by computing first-order correlations in annual earnings, i.e. correlations between year- $(t - 1)$ earnings and year- t earnings. We also report second-order correlations. These measures may be interpreted as indices of earnings mobility, which are indicative of the persistence of inequality. To compute the correlations we have considered only individuals with at least three years of non-zero annual earnings.¹³ The table shows that, though rather volatile, the evolution of earnings mobility does not exhibit a clear trend over the period.

Table 1: t-1/t correlations and t-2/t correlations, 2004-2010 (%).

		2005	2006	2007	2008	2009	2010
All	t-1/t	92.9	84.2	82.3	90.0	88.1	86.8
	t-2/t		81.2	85.3	81.9	82.8	87.4
Men	t-1/t	93.1	82.6	80.0	89.6	87.2	85.6
	t-2/t		79.6	84.5	80.7	81.8	86.9
Women	t-1/t	91.4	89.1	90.2	90.7	90.3	90.1
	t-2/t		85.9	87.1	85.0	85.5	88.1

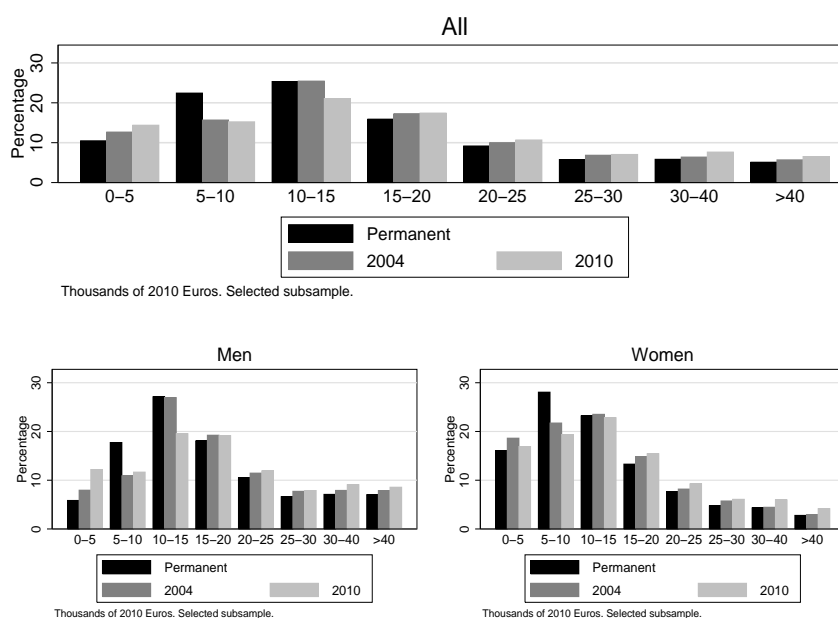
¹²For Spain, Pijoan-Mas and Sanchez-Marcos (2010) have used the European Community Household Panel (ECHP) and the *Encuesta Continua de Presupuestos Familiares* (ECPF) to estimate a stochastic model of earnings dynamics.

¹³This selection results in a sample of 544,088 individuals, out of the 696,223 individuals in the full sample. Figure C.3 in Appendix C compares the level and evolution of annual earnings inequality in the full sample and in the selected subsample.

Notes: Source Tax data. The sample contains individuals with at least three years of non-zero annual earnings.

Next we compute individual averages of earnings over time, as a proxy for “permanent” earnings. As before we focus on individuals with at least three years of non-zero annual earnings. In Figure 7 we show the histogram of annual earnings in 2004 and 2010 (light bars), as well as the histogram of 2004-2010 individual averages of annual earnings (dark bars).

Figure 7: Current annual earnings (2004 and 2010) and earnings averages: histograms

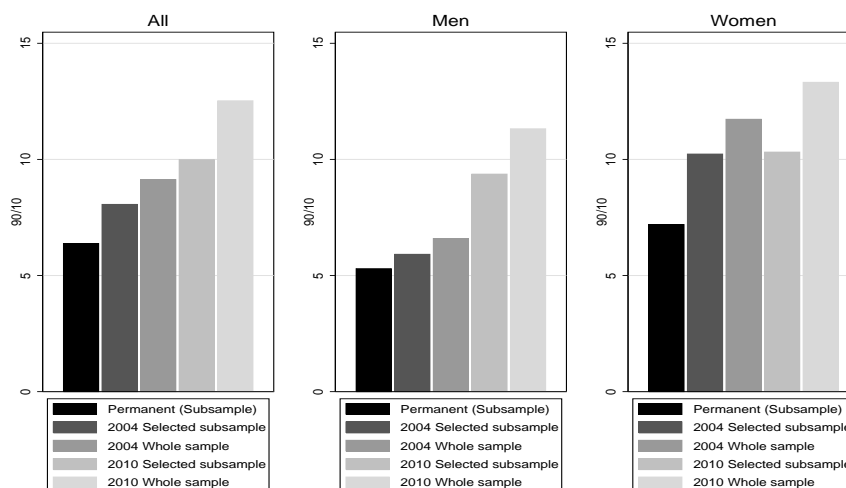


Notes: Source Tax data. “Permanent” earnings are individual averages of annual earnings over all available years. The sample contains individuals with at least three years of non-zero annual earnings.

Figure 7 shows that the left and right tails of the distribution of average earnings are thinner than those of the distribution of cross-sectional annual earnings. As a result, we expect the level of inequality to be lower when using intertemporal averages. This is what Figure 8 shows: the level of “permanent” earnings inequality, as proxied by the inequality of intertemporal earnings averages, is roughly 70% of the average of cross-sectional inequality during the 2004-2010 period. This finding is consistent with part of earnings inequality being transitory and vanishing out over time.

However, obtaining a precise decomposition of Spanish inequality into a temporary

Figure 8: Current annual earnings (2004 and 2010) and earnings averages: inequality



Notes: See notes to Figure 7. The “selected subsample” refers to individuals with at least three years of non-zero annual earnings.

part and a permanent part is a challenging exercise. Indeed, as a consequence of the level of Spanish unemployment, the number of observations with zero annual earnings is very high in our sample: 12.8% out of the nearly 4 million annual observations. This suggests that, in order to correctly assess the longitudinal dimension of Spanish inequality, a joint analysis of earnings and employment trajectories would be required. We leave this important issue for future work.

5 Conclusion

The Social Security data and matched tax files that have recently become available to researchers provide an opportunity to document the level and evolution of Spanish earnings inequality. In this paper we have provided some preliminary evidence, which complements the analysis of Bonhomme and Hospido (2012) using the tax data for the 2004-2010 period.

The analysis has provided further evidence on the cyclicity of Spanish earnings inequality. As the recession started and the unemployment rate increased, the level of inequality increased substantially— by 10 percentage points according to the 90/10 percentiles ratio. Changes in labor force composition, and in particular changes in

sectoral composition, partly explain this evolution.

Gender differences seem worth studying further. We have provided evidence that the gender earnings gap has decreased throughout the distribution between 2004 and 2010. Moreover, although one important shortcoming of our data is that hours worked are not recorded, we have shown that this conclusion is robust to considering full-time workers only. The recent recession, which particularly affected male earnings in the lower-middle part of the distribution, may be one factor explaining the decrease in the gender gap. It would be interesting to study if this decrease continues when the economic situation improves.

A second important avenue for research is to exploit the panel dimension of the data to analyze individual trajectories of employment and earnings over time, in order to document the longitudinal dimension of earnings inequality. In particular, assessing how the cyclical nature of Spanish earnings inequality interacts with individual earnings and employment dynamics seems an interesting question. We hope that this paper will stimulate future work on these and related themes.

References

- [1] Alvaredo, F., and E. Saez (2009), “Income and Wealth Concentration in Spain from a Historical and Fiscal Perspective”, *Journal of the European Economic Association*, 7, 1140-1167.
- [2] Autor, D. H., L. F. Katz, and M. S. Kearney (2008), “Trends in U.S. Wage Inequality: Re-assessing the Revisionists”, *Review of Economics and Statistics*, 90, 300-323.
- [3] Bonhomme, S. and L. Hospido (2012), “The Cycle of Earnings Inequality: Evidence from Spanish Social Security Data”, IZA Discussion Paper 6669.
- [4] Carrasco, R., J. F. Jimeno, and A. C. Ortega (2011), “Accounting for changes in the Spanish Wage Distribution: The Role of Employment Composition Effects”, Banco de España Working Paper 1120.
- [5] De la Rica, S., J. J. Dolado, and V. Llorens (2008), “Ceilings or floors? Gender wage gaps by education in Spain”, *Journal of Population Economics*, 21, 751-776.

- [6] Dolado, J. J., S. Ortigueira, and R. Stucchi (2011), “Does dual employment protection affect TFP? Evidence from Spanish manufacturing firms”, mimeo.
- [7] Dustmann, C., J. Ludsteck, and U. Schonberg (2009), “Revisiting the German Wage Structure”, *Quarterly Journal of Economics*, 124, 843-881.
- [8] Gardeazábal, J. and A. Ugidos (2005), “Gender wage discrimination at quantiles”, *Journal of Population Economics*, 18, 165-179.
- [9] Gottschalk, P. and R. Moffitt (1994), “The Growth of Earnings Instability in the U.S. Labor Market”, *Brookings Papers on Economic Activity*, 25, 217-272.
- [10] Guner, N., E. Kaya, and V. Sánchez-Marcos (2012), “Gender Gaps in Spain: Policies and Outcomes over the Last Three Decades”, IZA Discussion Paper 6812.
- [11] R. E. Hall, and F. S. Mishkin (1982), “The Sensitivity of Consumption to Transitory Income: Estimates from Panel Data on Households”, *Econometrica*, 50, 461-481.
- [12] Izquierdo, M. and A. Lacuesta (2012), “The contribution of changes in employment composition and relative returns to the evolution of wage inequality: the case of Spain”, *Journal of Population Economics*, 25, 511–543.
- [13] Pijoan-Mas, J., and V. Sanchez-Marcos (2010), “Spain is Different: Falling Trends of Inequality”, *Review of Economic Dynamics*, 13, 154-178.
- [14] Piketty, T. and Saez, E. (2006), “The Evolution of Top Incomes: A Historical and International Perspective,” *American Economic Review, Papers and Proceedings*, 96, 200-205.

APPENDIX

A The tax data: model 190

The tax data come from the information in model 190, i.e., the “Annual summary of retentions and payments for the personal income tax on earnings, economic activities, awards and income imputations”. The need to fill-in this information is not subject to any kind of exemption. Thus, all entities that pay wages, pensions or benefits for unemployment are required to provide this information, regardless of their personality, activity, size or their public or private nature. The information is thus comprehensive and very detailed.

The total amount of compensation paid by the employer are given in model 190 for each year. Each employer should report a summary and a list of beneficiaries with individual information on both compensation and retentions. The lists of beneficiaries include all people who receive payments, even those with wages below the legal minimum of exemption for the obligation to declare their personal income tax.

The population is composed of wage earners, pensioners and beneficiaries of unemployment benefits that belong to the common tax regime, which includes the whole country except the Basque country and Navarra.

B Additional information on the data

Table B.1: Descriptive statistics by gender

	Total			Men			Women		
	Number	%		Number	%		Number	%	
Individuals	696,223	100.00		374,260	53.76		321,963	46.24	
Observations	3,322,129	100.00		1,817,712	54.72		1,504,417	45.28	
	2004	2008	2010	2004	2008	2010	2004	2008	2010
Age	37.56 (8.45)	38.07 (8.34)	38.64 (8.32)	38.04 (8.52)	38.36 (8.39)	38.92 (8.32)	36.94 (8.31)	37.73 (8.28)	38.33 (8.30)
Labor Income - 2010 €	17,227 (17,340)	17,274 (18,808)	16,932 (19,275)	19,940 (20,177)	19,682 (22,299)	19,183 (23221)	13,778 (12,006)	14,390 (12,905)	14,397 (13,058)
% days worked in a year	84.65 (26.81)	82.81 (27.76)	81.12 (29.90)	87.12 (24.62)	83.93 (26.84)	81.44 (29.82)	81.51 (29.07)	81.47 (28.77)	80.76 (29.99)
Daily Earnings - 2010 €	56.32 (208.41)	57.87 (186.93)	57.30 (158.33)	64.32 (230.78)	65.88 (216.46)	65.06 (184.39)	46.15 (175.38)	48.26 (143.21)	48.56 (121.91)
College	15.97	16.45	17.35	12.53	12.85	13.68	20.29	20.69	21.43
High-skilled	19.14	18.67	19.61	18.63	17.82	18.72	19.78	19.68	20.62
Medium-skilled	34.02	33.28	34.32	25.06	23.57	24.94	45.40	44.91	44.89
Low-skilled	46.84	48.05	46.06	56.30	58.60	56.34	34.82	35.41	34.48
Fixed-term	34.58	34.51	34.97	32.92	34.04	34.33	36.70	35.08	35.69
Immigrants	2.71	15.57	14.97	2.54	17.19	16.01	2.92	13.62	13.80

Notes: Source Tax data. Mean values and standard deviations of non-binary variables in parentheses. High-skilled, Medium-skilled and Low-skilled refer to occupation groups 1-3, 4-7, and 8-10, respectively.

Table B.2: Description of sectors

Industry:	Agriculture, mining, food and tobacco industry, clothing and footwear industry, metal industry, paper industry, timber industry, plastics industry, chemical industry, machinery and car industry, furniture industry and manufacturing.
Construction:	All general building works, installation systems and extensions (electrical system, painting, plumbing and tiling, carpentry, flooring, plastering), civil engineering works, renting of the building equipment.
Transportation:	Storing, transport, telecommunications and energy.
Sales:	Sales.
Hotels:	Hotels.
Finance:	Financial services.
Corporate services:	Corporate services, personal services, social activities.
Administration:	Public administration.
Education:	Education.
Health:	Health.

C Additional results

Table C.1: Earnings inequality, 2004-2010.

		2004	2005	2006	2007	2008	2009	2010
Whole sample	p90/p10	4.34	4.31	4.24	4.25	4.33	4.58	4.75
	p90/p50	2.17	2.18	2.17	2.18	2.17	2.20	2.20
	p50/p10	2.00	1.98	1.96	1.95	1.99	2.09	2.16
	SD of logs	.660	.662	.653	.660	.668	.683	.690
	Gini	.372	.373	.386	.376	.377	.379	.381
Men	p90/p10	3.60	3.55	3.50	3.48	3.58	3.79	3.97
	p90/p50	2.24	2.24	2.22	2.21	2.20	2.21	2.21
	p50/p10	1.61	1.59	1.58	1.58	1.63	1.72	1.80
	SD of logs	.610	.612	.607	.612	.623	.642	.652
	Gini	.364	.368	.387	.372	.371	.374	.375
Women	p90/p10	4.74	4.77	4.72	4.79	4.82	5.01	5.09
	p90/p50	2.12	2.15	2.15	2.18	2.18	2.22	2.22
	p50/p10	2.24	2.22	2.20	2.19	2.21	2.25	2.30
	SD of logs	.673	.677	.665	.676	.681	.695	.699
	Gini	.362	.360	.363	.363	.367	.371	.372

Notes: Source Tax data. p90 (resp., p50, p10) is the 90th (resp., median, 10th) percentile of daily earnings. “SD of logs” is the standard deviation of log-daily earnings.

Table C.2: Changes in earnings inequality, 2004-2007 and 2008-2010
(100×change in log percentile ratios)

		Change 2004-2007			Change 2008-2010		
		Overall	Residual		Overall	Residual	
			Spec. 1	Spec. 2		Spec. 1	Spec. 2
Whole sample	p90/p10	-1.5	-1.0	-.4	6.2	5.5	4.8
	p90/p50	.5	-1.1	.0	1.6	3.2	2.0
	p50/p10	-3.7	-1.0	-.8	11.1	7.5	7.0
Men	p90/p10	-2.6	-1.9	-1.1	7.9	7.2	6.0
	p90/p50	-1.8	-2.5	-1.0	.5	2.9	2.1
	p50/p10	-4.0	-1.2	-1.2	19.9	11.5	9.9
Women	p90/p10	.6	-.7	-.1	3.5	3.6	2.7
	p90/p50	4.0	-.2	-.0	2.1	3.2	1.9
	p50/p10	-2.5	-1.0	-.2	4.8	3.8	3.3

Notes: Source Tax data. See notes to Figure 4.

Table C.3: Gender earnings gap by percentile

	Percentile	2004	2005	2006	2007	2008	2009	2010
Unconditional	10th	.57	.57	.56	.56	.54	.50	.47
	50th	.24	.23	.23	.23	.23	.23	.23
	90th	.30	.27	.27	.24	.24	.22	.22
Conditional: Specification 1	10th	.47	.47	.47	.48	.46	.43	.42
	50th	.28	.28	.28	.28	.28	.27	.27
	90th	.31	.31	.31	.31	.30	.28	.28
Conditional: Specification 2	10th	.40	.40	.40	.41	.40	.37	.36
	50th	.25	.25	.25	.25	.24	.23	.23
	90th	.26	.26	.26	.27	.26	.25	.24

Notes: Source Tax data. Difference in the τ th ($\tau = .10, .50, .90$) quantiles of log-earnings between males and females (top panel), and gender coefficient in the τ th quantile regression ($\tau = .10, .50, .90$) of log-earnings that controls for the covariates in Specifications 1 and 2 (bottom panels). See notes to Figure 4 for the definition of covariates specifications.

Table C.4: Proportion of full-time workers (%).

	2004	2005	2006	2007	2008	2009	2010
All	87.2	86.5	86.4	86.1	85.2	83.3	81.4
Men	94.7	94.5	94.6	94.6	94.0	92.5	90.8
Women	77.6	76.5	76.2	75.8	74.7	72.8	70.7

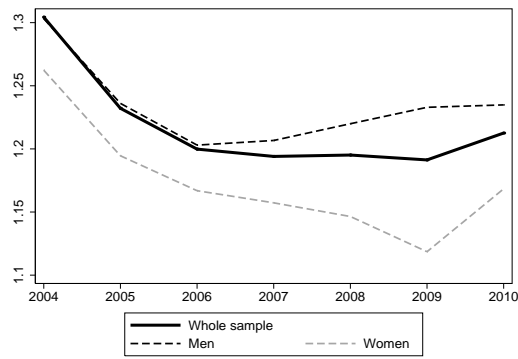
Notes: Source Tax data.

Table C.5: Gender earnings gap by percentile, full-time workers

	Percentile	2004	2005	2006	2007	2008	2009	2010
Unconditional	10th	.24	.23	.23	.22	.20	.17	.16
	50th	.13	.12	.12	.12	.11	.10	.10
	90th	.24	.21	.20	.17	.17	.15	.15
Conditional: Specification 1	10th	.22	.22	.22	.22	.21	.18	.17
	50th	.20	.20	.20	.20	.19	.18	.17
	90th	.27	.26	.26	.26	.26	.23	.23
Conditional: Specification 2	10th	.22	.21	.21	.20	.19	.17	.16
	50th	.19	.18	.18	.18	.17	.16	.15
	90th	.23	.22	.22	.22	.22	.20	.20

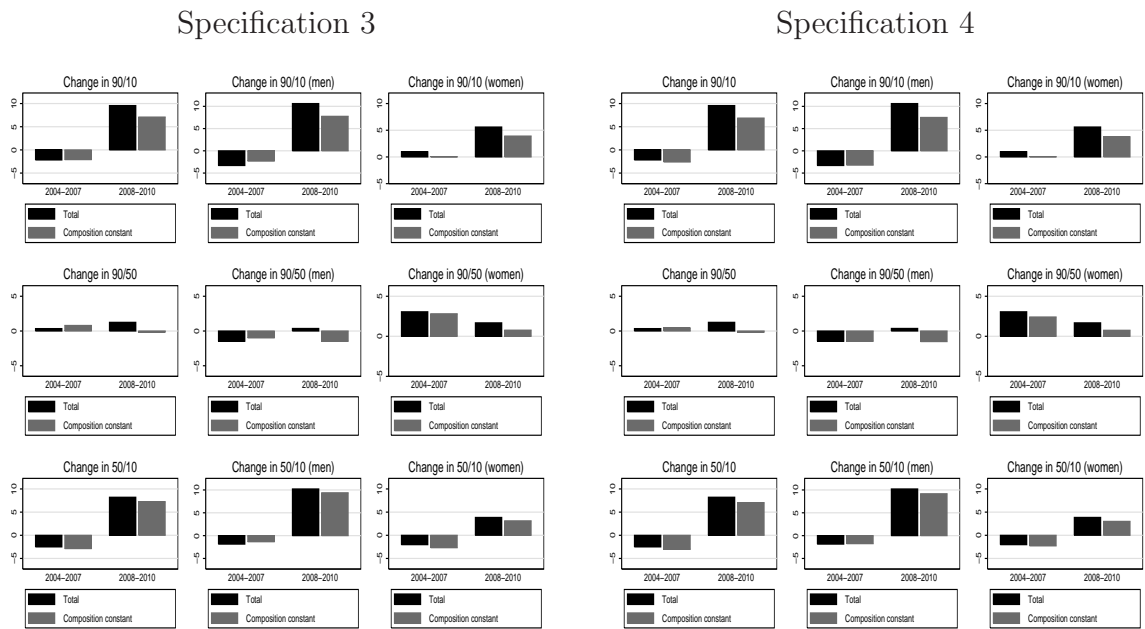
Notes: See notes to Table C.3. Full-time workers only.

Figure C.1: Permanent/fixed-term earnings gap



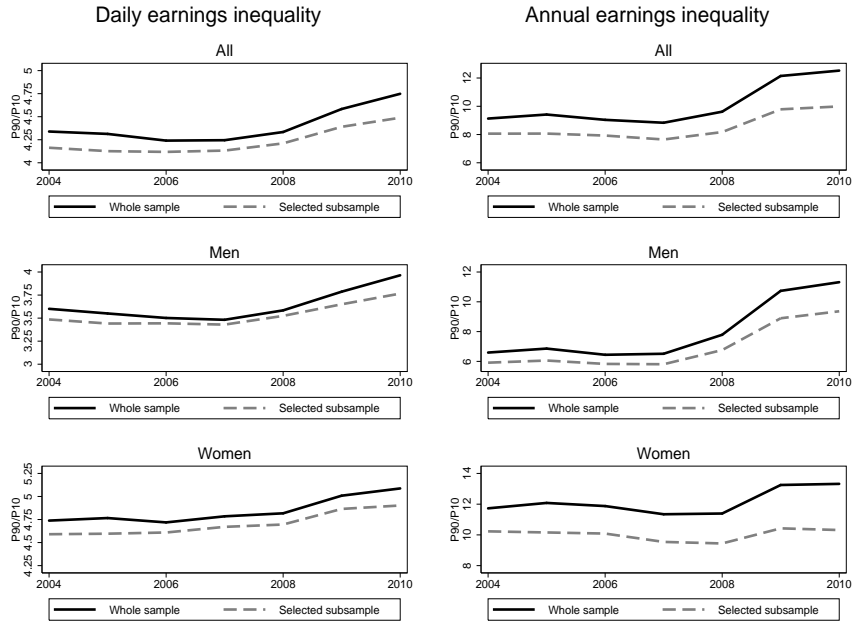
Notes: Source Tax data.

Figure C.2: Inequality decomposition (alternative specifications)



Notes: See notes to Figure 5. Specification 3 includes age and occupation dummies only. Specification 4 includes, in addition to Specification 1, an immigrant/native dummy and a dummy for the type of contract (permanent/fixed-term).

Figure C.3: Inequality in the full sample, and in a restricted subsample



Notes: Source Tax data. Solid lines show the results for the full sample (696,223 individuals), while dashed lines show the result in a subsample of individuals with at least three years of non-zero annual earnings (544,088 individuals).