Journal of the Royal Statistical Society
Statistics in Society
Series A

J. R. Statist. Soc. A (2018)

Modelling earnings dynamics and inequality: foreign workers and inequality trends in Luxembourg, 1988–2009

Denisa M. Sologon

Luxembourg Institute of Socio-Economic Research, Esch-sur-Alzette, Luxembourg

and Philippe Van Kerm

Luxembourg Institute of Socio-Economic Research and University of Luxembourg, Esch-sur-Alzette, Luxembourg

[Received July 2015. Final revision May 2017]

Summary. The paper exploits large-scale administrative data to analyse trends in male earnings inequality in Luxembourg during 20 years of rapid economic growth, industrial redevelopment and massive inflow of foreign workers. A detailed error components model is estimated to identify persistent and transitory components of (the trends of) log-earnings variance and to disentangle the contributions to it of native, immigrant and cross-border workers. The model is flexible and allows for a high degree of individual, age, time and cohort heterogeneity. We observe a surprising stability in overall earnings inequality as a result of more complex underlying changes, with marked increases in persistent inequality (except among natives), a growing contribution of foreigners and a decrease in earnings instability (primarily for natives).

Keywords: Administrative data; Cross-border and immigrant workers; Earnings dynamics; Minimum distance estimation; Persistent inequality; Transitory inequality

1. Introduction

Following a seminal analysis for the USA by Gottschalk and Moffitt (1994), much of the empirical literature on earnings inequality trends has explored the extent to which long-term changes in earnings inequality reflect an increase in persistent wage differentials between workers or whether it reflects increased transitory earnings variations. The former is consistent with explanations that are related to increasing returns to skills and education—which are essentially permanent individual characteristics—whereas the latter are associated with increased labour market risks and volatility (see, for example, Haider (2001)). Globalization and skill-biased technological change have arguably amplified returns to skills and are typically seen as key forces behind increasing earnings inequality in the last three decades (see, for example, Freeman and Katz (1994), Organisation for Economic Co-operation and Development (2011), Jaumotte *et al.* (2013) and Autor (2014)). Additionally, a role for labour market institutions in curtailing increases in inequality has been suggested to account for the different trends that have been observed in the USA and continental Europe (Freeman and Katz, 1994; Acemoglu, 2002).

Address for correspondence: Denisa M. Sologon, Luxembourg Institute of Socio-Economic Research, 11 Porte des Sciences, L-4366 Esch-sur-Alzette, Luxembourg.
E-mail: denisa.sologon@liser.lu

 0964–1998/18/181000

Empirical strategies to decompose inequality trends into permanent and transitory components typically consist in exploiting dynamic error components models for individual earnings. Earnings dynamics processes incorporate both persistent terms (that affect earnings permanently) and transitory terms (that have short-lived effects), and model parameter estimates are then used to decompose the overall log-earnings variance into permanent and transitory factors whose relative contributions can be tracked over time (see Meghir and Pistaferri (2011) and Jäntti and Jenkins (2015) for reviews). Most recent studies based on panel data with a long time series dimension have found that permanent inequality increased in most industrialized countries between the 1970–1980s and the 1990–2000s, both in Europe and in North America. (See, among others, Haider (2001), Kopczuk *et al.* (2010), Moffitt and Gottschalk (2012) and DeBacker *et al.* (2013) on the USA, Baker and Solon (2003) on Canada, Dickens (2000) and Kalwij and Alessie (2007) on the UK, Cappellari (2004) and Cappellari and Leonardi (2013) on Italy, Bingley *et al.* (2013) on Denmark and Bönke *et al.* (2015) on Germany. As a matter of exception, Gustavsson (2007) observed a decrease in persistent inequality in Sweden until 1990 and an increase thereafter. Also, in a study of 15 European Union countries based on a relatively short panel, Sologon and O'Donoghue (2012) found that only Denmark stands out with the lowest and decreasing overall permanent variance in the 1990s–early 2000s.) Results on trends in transitory variance are somewhat more mixed. (Moffitt and Gottschalk (2012) found a dramatic increase in transitory variance in the USA in the 1980s, a levelling-off in the late 1980s, followed by a decrease in the 1990s and a further increase in the early 2000s. In Canada, most of the increase in earnings instability occurred during the early 1980s and early 1990s (Baker and Solon, 2003). Across Europe, a strong increase in transitory inequality was found by Kalwij and Alessie (2007) in the UK, by Cappellari and Leonardi (2013) in Italy, and by Bönke *et al.* (2015) in Germany. In the UK, whereas Kalwij and Alessie (2007) found that transitory inequality increased to a larger extent than permanent inequality, Dickens (2000) found similar increases in both components. The difference was attributed to the methodological advancements that were brought by Kalwij and Alessie (2007) which account for age, time and cohort effects in their model specification (see Section 5). Bingley *et al.* (2013) found an increase in earnings instability in Denmark starting with the mid-1990s, and this appears to be the trend across most other European countries, at least until the early 2000s (Sologon and O'Donoghue, 2012).) Building on this literature, the present paper exploits a large-scale administrative data set to estimate a rich model of earnings dynamics and to analyse trends in persistent and transitory earnings inequality among male workers in Luxembourg between 1988 and 2009, which was a period of rapid economic growth for this small open economy.

The originality of the study is threefold. Firstly, we take advantage of a large-scale administrative data set on earnings and employment which allows us to specify and estimate a flexible model of earnings dynamics. This paper is still one of the few studies to date based on a large administrative data set with complete coverage of the working-age population in the country—we analyse just under 370000 men contributing more than 3 million person-year observations (see Section 2). To the best of our knowledge, only Blundell *et al.* (2014) have exploited larger data for analyses of this type. (Other studies which have exploited administrative registers have generally analysed smaller extracts (see, for example, Baker and Solon (2003), Cappellari (2004), and DeBacker *et al.* (2013).) The data are derived from social security administration registers and provide annual information on earnings spanning 22 years about each person ever employed in Luxembourg at any point in time during this period. The size of the data set both in the cross-section and the time dimensions enables us to estimate a flexible and comprehensive earnings dynamics model that nests the specifications that have been proven by most recent studies as crucial in capturing accurately the dynamics in individual earnings (see Meghir and

Pistaferri (2011) for a review). Reliable inference on flexible models earnings dynamics requires access to data with both a high number of observations and a long time frame, as Doris *et al.* (2013) emphasized. We can allow the variance of both permanent and transitory shocks to vary flexibly with workers' age—an essential feature emphasized in Blundell *et al.* (2014)—the permanent component via a random-walk specification with age-specific innovation variances and the transitory component via an auto-regressive moving average ARMA(1,1) with age-specific heteroscedastic transitory variances. The non-stationary pattern of earnings is accommodated by time-specific loading factors on both earnings components. Cohort heterogeneity is accommodated by allowing both the permanent and the transitory component to vary by cohort—an essential feature emphasized in Kalwij and Alessie (2007). In addition, we introduce a correction for left censoring for each cohort in the first year observed, following Moffitt and Gottschalk (2012).

Use of administrative data brings further advantages compared with survey data such as very low reporting or recollection error and the absence of selective attrition (other than through migration or death). However, information on earnings is affected by top coding. To address this issue, we implement a multiple-imputation procedure as proposed by Jenkins *et al.* (2011) and incorporate this in the process of estimating the parameters of our earnings dynamics model.

Secondly, the Luxembourg case-study is yet unexplored and is of interest *per se*. We look at a period during which this small economy experienced sustained economic growth and an industrial redevelopment from an industry-driven economy to an economy dominated by the tertiary sector, the financial sector in particular (Annaert, 2004; Allegrezza *et al.*, 2004; Fusco *et al.*, 2014). The transition from the steel industry towards the specialization in financial and banking sectors recorded a strong upswing of gross domestic product (GDP) growth from the mid-1990s. Sustained economic growth increased labour demand to levels that could not be matched by the resident population alone (especially for high skilled workers) and soaring labour demand led to a massive inflow of foreign workers—both of immigrants and of cross-border workers residing in Belgium, France and Germany (Amétépé and Hartmann-Hirsch, 2011). According to our calculations, the share of cross-border workers among male workers aged 20–57 years recorded an increase from over 20% in the late 1980s to close to 45% in the late 2000s. By 2009, foreign workers represented 75% of workers in this group. We conjecture that rising demand for high skill labour (in the financial sector in particular) and the limited supply of domestic workers put strong upward pressure on earnings inequality. However, this may have been mitigated by a growth-induced general increase in the demand for labour across the overall skill distribution, the abundant supply of foreign labour from neighbouring countries and relatively strong labour market institutions—in particular, influential collective bargaining institutions, a high statutory minimum wage and relatively strict employment protection regulation. The trends in earnings inequality in Luxembourg can therefore provide some empirical indication about whether strong labour market regulation and large foreign labour supply can counterbalance otherwise inequality increasing pressures.

Thirdly, owing to the scale of our data set, we can examine the contribution of foreign workers in detail by estimating models separately for native, immigrant and cross-border workers. We then use the separate model parameters to estimate the contributions of each of the subgroups to the overall trends in inequality (and to its permanent and transitory components), disentangling trends in within-group inequality, in between-group differentials and in the relative share of each group in total employment. As far as we are aware, no previous study has distinguished these trends for native and foreign workers, and identified their relative contributions to the overall long-term earnings inequality trends. This is a distinctive feature of our analysis which is particularly relevant here given the magnitude of changes in the employment composition throughout the period, the different skill composition of these three groups of workers, the limited access to public sector jobs by foreign workers and the variations in social insurance and fiscal policy to which non-resident workers are exposed (Choe and Van Kerm, 2014; Fusco *et al.*, 2014).

To preview our results, we find evidence of only a relatively modest increase in earnings inequality. However, this surprising stability in light of the drastic labour market changes in the period that is analysed is the net result of somewhat more complex underlying changes, with marked increases in persistent inequality among cross-border workers and among immigrants, a growing contribution of foreign workers, divergence in persistent differentials between subgroups and a decrease in earnings instability (but primarily for native workers). Native workers appear to have experienced particularly favourable trends.

The paper is structured as follows. Section 2 describes the data that are used in the analysis, our sample selection and the strategy implemented to address top coding of earnings. Section 3 sets the scene by documenting the trends in mean earnings and in inequality observed in the data and Section 4 describes the general autocovariance structure of earnings. Our model of earnings dynamics is detailed in Section 5. Section 6 exploits model estimates to disentangle persistent and transitory components in the variance of log-earnings and reveals the long-run increase in persistent inequality and the contribution of foreign workers to these trends. In Section 7, we examine the correlation between the trends revealed and macroeconomic and institutional factors. Our main results are finally contrasted with comparable estimates from other countries in Section 8. Section 9 concludes.

The programs that were used to analyse the data can be obtained from

`http://wileyonlinelibrary.com/journal/rss-datasets`

2. Data

2.1. Data frame and sample description

Each person with a paid occupation in Luxembourg is registered to the social security administration (*Inspection Générale de la Sécurité Sociale*) from the date of their first job in the country. Information is subsequently recorded on various aspects of individual employment histories for calculating public pension entitlements. Our analysis exploits a large-scale anonymized scientific use extract from these registers. Our data set covers the period 1988–2009 and contains information about all people ever working for an employer based in Luxembourg in this 22-years period. We can observe individual level data on gross annual labour income during each worker's career, number of months of employment each year, occupational status, nationality and country of residence. (Note that, because of the purpose of these registers, they contain no information on potentially relevant variables such as educational achievements, non-labour incomes and household level contextual and demographic information.)

Our analysis focuses on men aged between 20 and 57 years to avoid issues that are related to labour market participation at the end of the career; see the discussion of monthly wage calculation below. We consider individuals who were born in 41 birth cohorts between 1940 and 1980 who have been recorded working in Luxembourg at least in one year between 1988 and 2009. The 41 cohorts are observed at least 10 years over the timespan of the data. (See Baker and Solon (2003) for the rationale of such a cohort selection rule in the context of error components model estimation.) Individuals who experienced at least five years of inactivity gaps between 1950 and 2009 because of disability or who retired before the age of 57 years with a disability benefit are disregarded because they have irregular earnings profiles. Individuals may exit and (re-)enter the data set at any year because of death or migration. The resulting data set (after additional selection based on earnings described below) contains data on 369288 men providing an unbalanced panel of 3 265 927 person-year observations with positive annual earnings. (This is a large population in comparison with the sample sizes of 3115, 2988, 76079 and 169 877 individuals that were used in similar studies in the USA by Haider (2001) and Moffitt and Gottschalk (2002), in Sweden by Gustavsson (2008) and in the UK by Dickens (2000). Samples from administrative sources of smaller sizes were used by Cappellari (2004) for a study in Italy (67 768 individuals) and by Baker and Solon (2003) for a study in Canada (31 105 individuals). Blundell *et al.* (2014) in contrast analysed a data set of 1 004 294 Norwegian men.)

Table 1 shows, by year, the size of the population, the age range, the share of observations whose income is top coded (see below) and the distribution across native, immigrant and cross-border workers. (Table A.1 in the on-line appendix A details the population composition in persons and person-years, years observed and age range for each of the 41 cohorts.) Whereas the share of immigrant workers in our data set remained stable throughout the period, the share of cross-border workers increased sharply over time; Fig. 1. As a result, the share of native workers in employment fell from about 51% in 1988 to only 25% in 2009.

Note that cross-border workers—and immigrants to a lesser extent—pose a specific problem since their earnings are recorded only for the years worked in Luxembourg. Although they are properly followed on re-entry into the data frame, no information is available in the years worked abroad. Similarly, migrant workers who leave the country are not tracked until they return in Luxembourg; nor are they observed before they enter the country. Immigrants, however, exhibit relatively rich longitudinal profiles. Tables B.2–B.4 in the on-line appendix B display detailed

Table 1. Sample size by year

Year	*Number of people*	*Age range (years)*	*% top coded*	*Nationals*	*Immigrants*	*Cross-border workers*
1988	74785	20–48	10.85	38675	18543	17567
1989	81609	20–49	10.58	40036	20999	20574
1990	89621	20–50	11.26	41363	24084	24174
1991	97504	20–51	10.31	42587	26641	28276
1992	104417	20–52	5.50	43698	28615	32104
1993	109890	20–53	4.92	44667	30090	35133
1994	116849	20–54	5.14	45647	32043	39159
1995	125868	20–55	4.54	49392	34115	42361
1996	133124	20–56	4.93	50563	36234	46327
1997	141196	20–57	4.47	51945	38357	50894
1998	149607	20–57	4.91	52360	40474	56773
1999	165208	20–57	5.21	59255	43119	62834
2000	174490	20–57	5.14	59428	45698	69364
2001	181030	21–57	5.43	58779	47229	75022
2002	183103	22–57	5.67	57840	48104	77159
2003	185291	23–57	5.08	56977	48901	79413
2004	187474	24–57	5.34	55749	49792	81933
2005	189317	25–57	5.38	54379	50625	84313
2006	192061	26–57	6.13	52946	51842	87273
2007	194849	27–57	5.97	51530	52626	90693
2008	196625	28–57	6.44	50088	54064	92473
2009	192009	29–57	5.96	48477	53805	89727
Total	3265927			1106381	876000	1283546

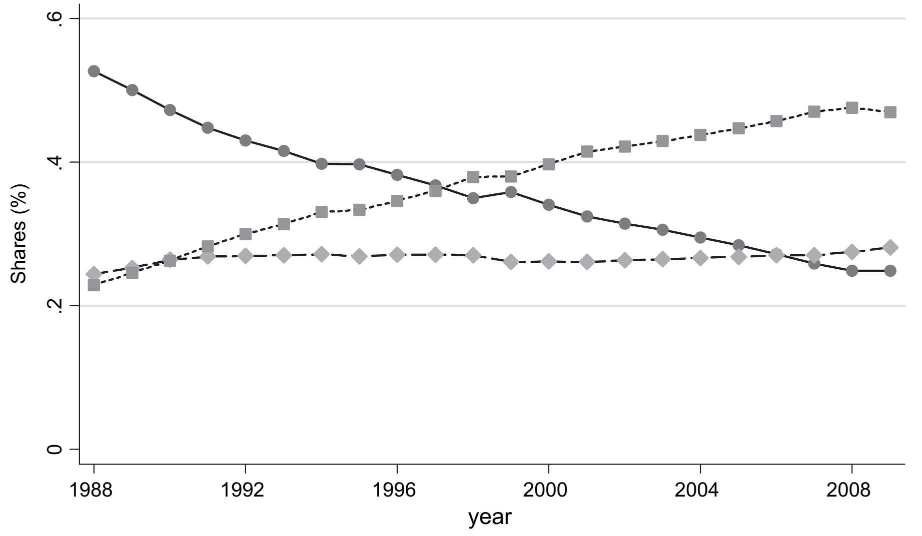


Fig. 1. Share of nationals, immigrants and cross-border workers (men, aged 20–57 years, born between 1940 and 1980, with positive earnings): ●, share nationals; ◆, share immigrants; ■, share cross-border workers

population composition information by worker type. For example, the share of immigrants who were active for all possible years in the timespan of analysis in each cohort ranges from close to 71% for those in the oldest cohort born in 1940, to 33% for the cohort born in 1960 and the youngest cohort born in 1980. The corresponding percentages for Luxembourg nationals are 73%, 66% and 46%. Cross-border workers have more incomplete profiles with corresponding figures of 54%, 22% and 14%. Like first-generation immigrants, many such workers have held jobs abroad before being employed in Luxembourg. They also tend to have higher rates of tertiary education attainment with correspondingly more frequent entry in employment after the age of 20 years. But perhaps more importantly some cross-border workers alternate spells of employment in Luxembourg and in neighbouring countries. (Voluntary mobility towards jobs outside Luxembourg is, however, mitigated by the relatively large wage differences across countries.) This limitation of our data does not prevent estimation of their contribution to the trends in persistent and transitory earnings inequality *in* Luxembourg, which is presented below. However, since we only partially observe job-to-job transitions for cross-border workers, we possibly underestimate the overall variability of their earnings; we return to this issue in Section 5.5.

2.2. Monthly earnings calculation and adjustments for top coding

Our analysis focuses on individuals' average real gross monthly wage, which we refer to as 'earnings'. The average monthly wage is calculated as total gross annual earnings reported to the social security administration divided by the number of months during which a person has been employed in Luxembourg and paid social security contribution during the year. Examination of average monthly wage is preferable to annual earnings for workers with incomplete employment in Luxembourg during a given year. Incomplete annual employment is common for cross-border workers and immigrants in their first job in Luxembourg. For other male workers, the difference between total annual earnings and annualized monthly wages is unlikely to be large since rates of unemployment were low among men in Luxembourg in the period that is covered by the analysis.

(Figs C.1 and C.2 in the on-line appendix show that the mean and the variance of log-annual-earnings evolve parallel with the monthly figures for Luxembourgish workers. The similarity in trends between the annual and monthly figures weakens somewhat when all workers—foreign and nationals—are considered.) (Note that actual hours worked are not available in the data.) All earnings are inflated to 2009 prices by using the consumer price index to obtain our final measure of real average gross monthly wage.

Our annual earnings data are affected by top coding. The monthly reports by employers to the social security administration are top coded at four times the monthly minimum wage until 1991 and five times thereafter. This top coding in the monthly employer reports translates into truncated annual earnings in our data set. The fourth column of Table 1 gives the share of observations with top-coded earnings. (An observation is considered to be top coded if the annual earnings recorded in the data are equal to the monthly top coding threshold times the number of contributory months. Not all employers appear to apply the top code, so we actually observe a mix of complete and top-coded data.) The change in the legislation for reporting wages after 1991 is reflected in the share of top-coded observations which drops by almost a half afterwards.

We address this issue by treating top-coding as a missing data problem and (multiply) imputing simulated values for top-coded earnings. We follow Jenkins *et al.* (2011) and first conduct (censored) maximum likelihood estimation of a parametrically specified distribution for top incomes and then multiply impute each top-coded earnings observation with m independent draws from the estimated top income distribution. Multiple imputation allows us to account for the variability that is introduced by the stochastic nature of the imputation. As is now common (see, for example, Atkinson and Piketty (2010), Kopczuk *et al.* (2010), Atkinson *et al.* (2011) and Alfons *et al.* (2013)), we assume that the upper tail of the annual earnings distribution for each year is described by a Pareto distribution with cumulative distribution function

$$F_\theta(y) = 1 - \left(\frac{y}{y_0}\right)^{-\theta}, \qquad y \geqslant y_0, \tag{1}$$

where $y_0 > 0$ is a threshold beyond which data are assumed Pareto distributed and $\theta > 0$ is a parameter to be estimated.

We estimate the θ-parameter independently for each year in 1988–2009 by fitting a Pareto distribution to observations with earnings above or equal to y_0 set at 0.7 of the top coding threshold. Estimation is conducted by maximum likelihood where, crucially, the likelihood function accounts for the top coding of observed earnings: the log-likelihood contribution of observation i is 0 if observed earnings y_i is below y_0 and is otherwise

$$\ln(L_i) = c_i \ln\{1 - F(y_i)\} + (1 - c_i) \ln\{f(y_i)\} \tag{2}$$

where $c_i = 1$ if i's earnings have been top coded and $c_i = 0$ otherwise and f is the Pareto density function.

Parameter estimates $\hat{\theta}$ are then used to draw imputed values for top-coded earnings for each year by using the inverse transform sampling method based on the standard formula for truncated distributions (Jenkins *et al.*, 2011). To account for the imputation variance, we draw $m = 20$ imputed values for each top-coded observation and thereby generate 20 partially synthetic data sets composed of reported non-top-coded data and an imputed value for all top-coded earnings. We finally retain in each of the synthetic data sets all observations with positive earnings and, following common practice (see, for example, Moffitt and Gottschalk (2012)), we drop the highest and lowest 1% of monthly earnings to prevent outlying observations from driving our model and inequality estimates.

All calculations and estimations conducted in our analysis were subsequently replicated on each of the 20 synthetic data sets and the estimates that are reported in the paper were obtained by using the combination formula proposed in Reiter (2003) as recommended in Jenkins *et al.* (2011):

$$\bar{q}_m = \sum_{j=1}^{m} \frac{q_j}{m} \tag{3}$$

where q_j is an estimate from data replication j ($j=1,\ldots,m=20$). (The sampling variance of $\bar{q}_m$ can be estimated as

$$T_p = \frac{1}{m} \sum_{j=1}^{m} \frac{(q_i - \bar{q}_m)^2}{m-1} + \sum_{j=1}^{m} \frac{v_j}{m} \tag{4}$$

where v_j is an estimate of the sampling variance of q_j (Reiter, 2003).) This procedure ensures that we properly account for the variability that is introduced by the stochastic nature of the imputation process.

Note that our imputation procedure randomly imputes earnings independently on each occurrence at which it is top coded: we do not attempt to take into account any potential within-person correlation in earnings beyond the top coding threshold at the imputation stage. We therefore expect that this procedure will overestimate the variability of top earnings. As we show below, we do not, however, find any evidence of a sharp change in the overall autocovariance of earnings after 1991 when the top coding threshold was increased to affect only about 5% of observations (from about 10% from 1988 to 1991), nor any break in our estimates of the transitory and persistent components of log-earnings variance. We therefore did not attempt to introduce refinements to incorporate varying degrees of dynamic dependence in earnings at the imputation stage.

3. Trends in the mean and variance of earnings

Before proceeding to the error components model and to the main part of our analysis, we describe the broad empirical patterns that are observed in our data. Fig. 2 shows the evolution of the variance and mean log-monthly-earnings in our population of men aged between 20 and 57 years and born between 1940 and 1980.

Throughout the period, there is an overall increase in both earnings inequality and mean earnings. Mean earnings are relatively stable between 1988 and 2000 (barring a jump between 1998 and 1999 which is due to an increase of 8% in the gross wage of civil servants). It then increases continuously from 2000 to 2008—a period during which Luxembourg's GDP grew by 4.3% annually on average. Mean earnings finally drop sharply in the recession year of 2009. The variance of log-earnings appears to evolve less smoothly. It increased most sharply between 1993 and 1999 (when mean earnings were stable), declined until 2004 (when mean earnings were growing) and increased again until 2009. Although both variables trended upwards throughout the 22 years, the patterns of change do not exhibit any systematic association. The long-run relative increase in the variance of log-earnings is somewhat smaller than observed during the same period in the USA (Moffitt and Gottschalk, 2012), the UK (Kalwij and Alessie, 2007), Italy (Cappellari and Leonardi, 2013) and Germany (Bönke *et al.*, 2015), but higher than in Sweden (Gustavsson, 2007, 2008).

Fig. 2 distinguishes the trends for native, immigrant and cross-border workers. Patterns of

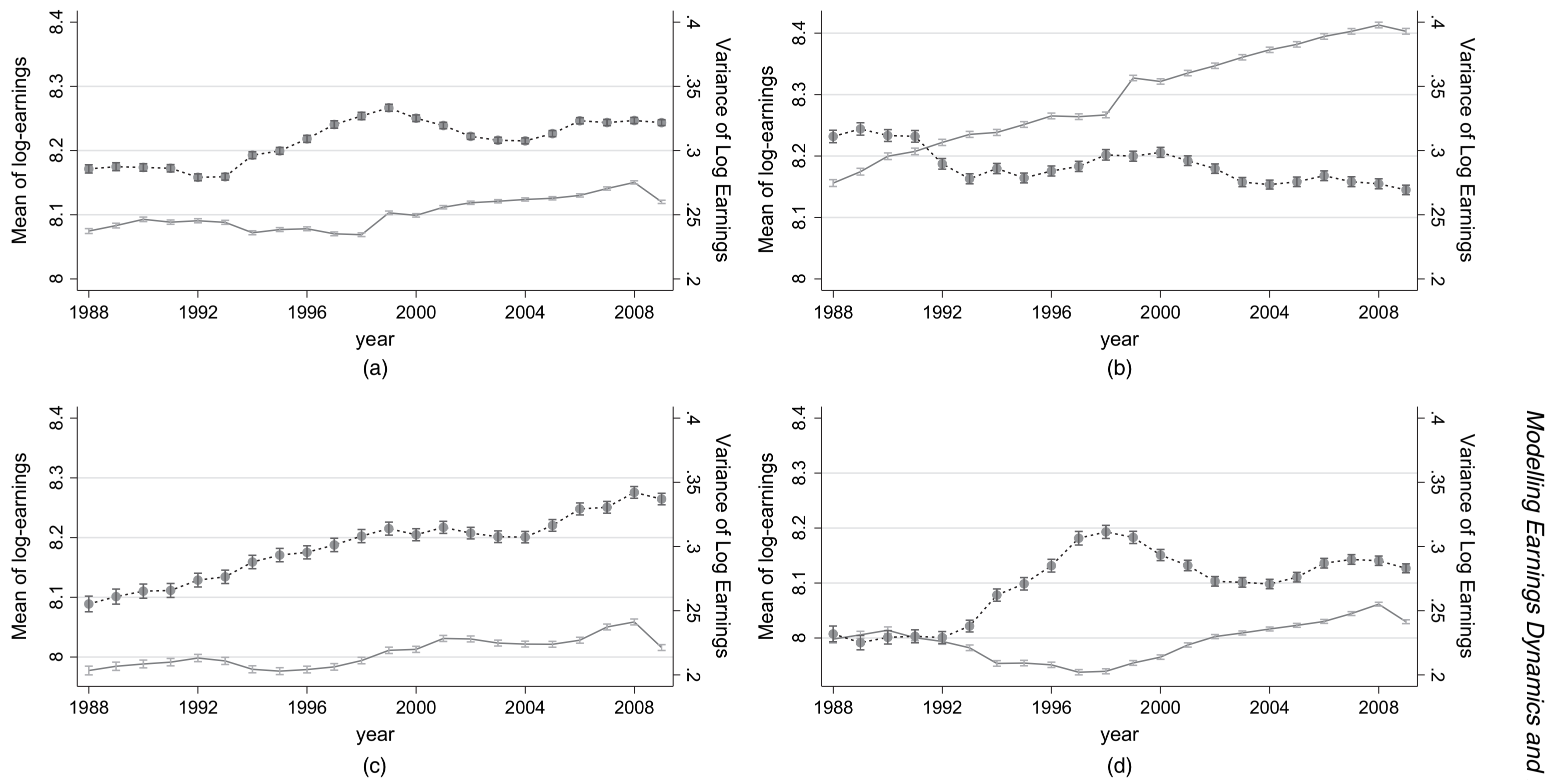


Fig. 2. Variance and mean of log-monthly-earnings, 1988–2009 (——, mean log-monthly-earnings; ⌶, confidence interval; ● variance of log-monthly-earnings; ⌶, confidence interval): (a) all men; (b) nationals; (c) immigrants; (d) cross-border workers

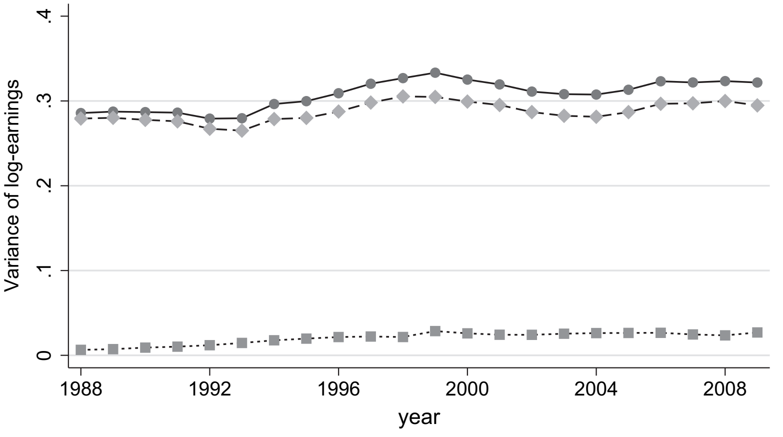


Fig. 3. Decomposition of the variance of log-earnings by population subgroups native, immigrant and cross-border workers, 1988–2009: ●, variance of log-monthly-earnings; ◆, within-groups variance of log-monthly-earnings; ■, between-groups variance of log-monthly-earnings

change differ between the three groups. Mean earnings grew faster for nationals than for immigrants or cross-border workers (and fell less in the recession year of 2009). For cross-border workers, mean earnings decreased between the early 1990s and the late 1990s and increased fast thereafter. Inequality overall decreased among nationals, whereas it increased among immigrants and cross-border workers. Cross-border workers' earnings exhibit still less inequality than residents' but have had a steep rate of increase over a period during which their share of total employment increased significantly (see Fig. 1).

Fig. 3 finally shows a decomposition of the trends in overall log-earnings variance into trends in within-group variances (defined as the population-weighted average of within-group variances shown in Fig. 2) and in between-group variances (defined as the residual difference between total variance and within-group variance). The increase in overall inequality was driven by an increase in both within- and between-group components, most of the increase occurring between the late 1980s and the late 1990s. Within-group inequality was the dominant component throughout the period, following the pattern that was observed for the overall inequality. The increase in the within-group inequality was mostly driven by increasing inequality between cross-border workers and immigrants. Between-groups differentials gradually increased from 1988 to 1999 but then remained stable afterwards.

4. The autocovariance structure of earnings

Taking advantage of our large-scale longitudinal data on individual earnings profiles, we seek to ascertain whether the trends in the variance of log-earnings primarily reflect an increase in short-run earnings variability or an increase in persistent, long-run earnings differences between workers. Answers to such a question are to be found in the autocovariance structure of earnings and its development over time. A selection of the long-run autocovariance structure of monthly earnings for all workers is shown in Fig. 4. The autocovariance structure of earnings is estimated for each cohort separately (adding up to 7513 population moments).

The autocovariances display different patterns across cohorts. The variance of log-monthly-earnings increases gradually over time for most cohorts—except the youngest and oldest (but those are observed for only a limited time window). Autocovariances also increased over time

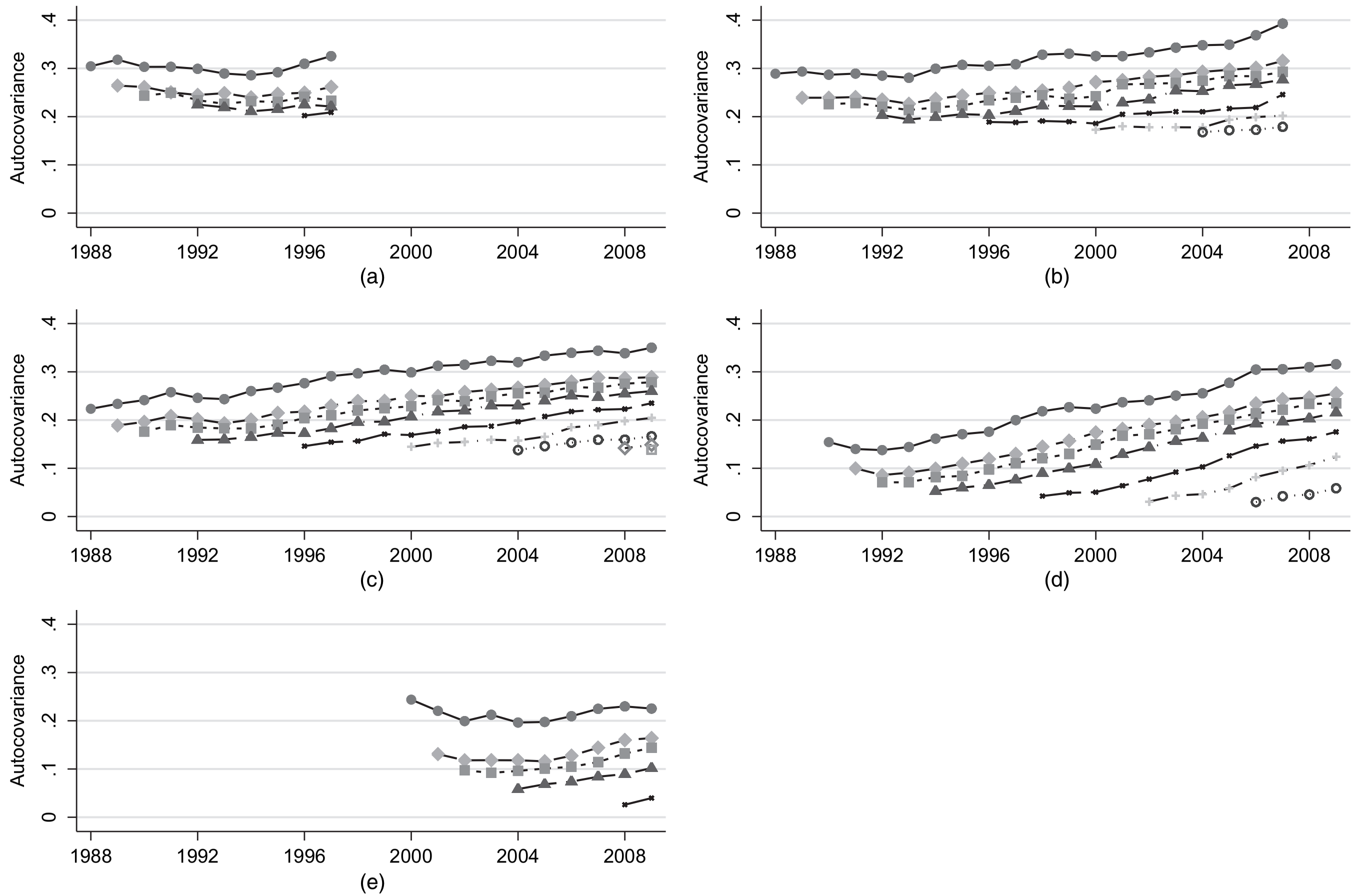


Fig. 4. Autocovariance structure of log-earnings for selected cohorts ('t' stands for the year displayed on the horizontal axis) (●, variance(t); ◆, covariance($t, t-1$); ■, covariance($t, t-2$); ▲, covariance($t, t-4$); ✖, covariance($t, t-8$); +, covariance($t, t-12$); ○, covariance($t, t-16$); ◇, covariance($t, t-20$); □, covariance($t, t-21$)): (a) cohort born 1940; (b) cohort born 1950; (c) cohort born 1960; (d) cohort born 1970; (e) cohort born 1980

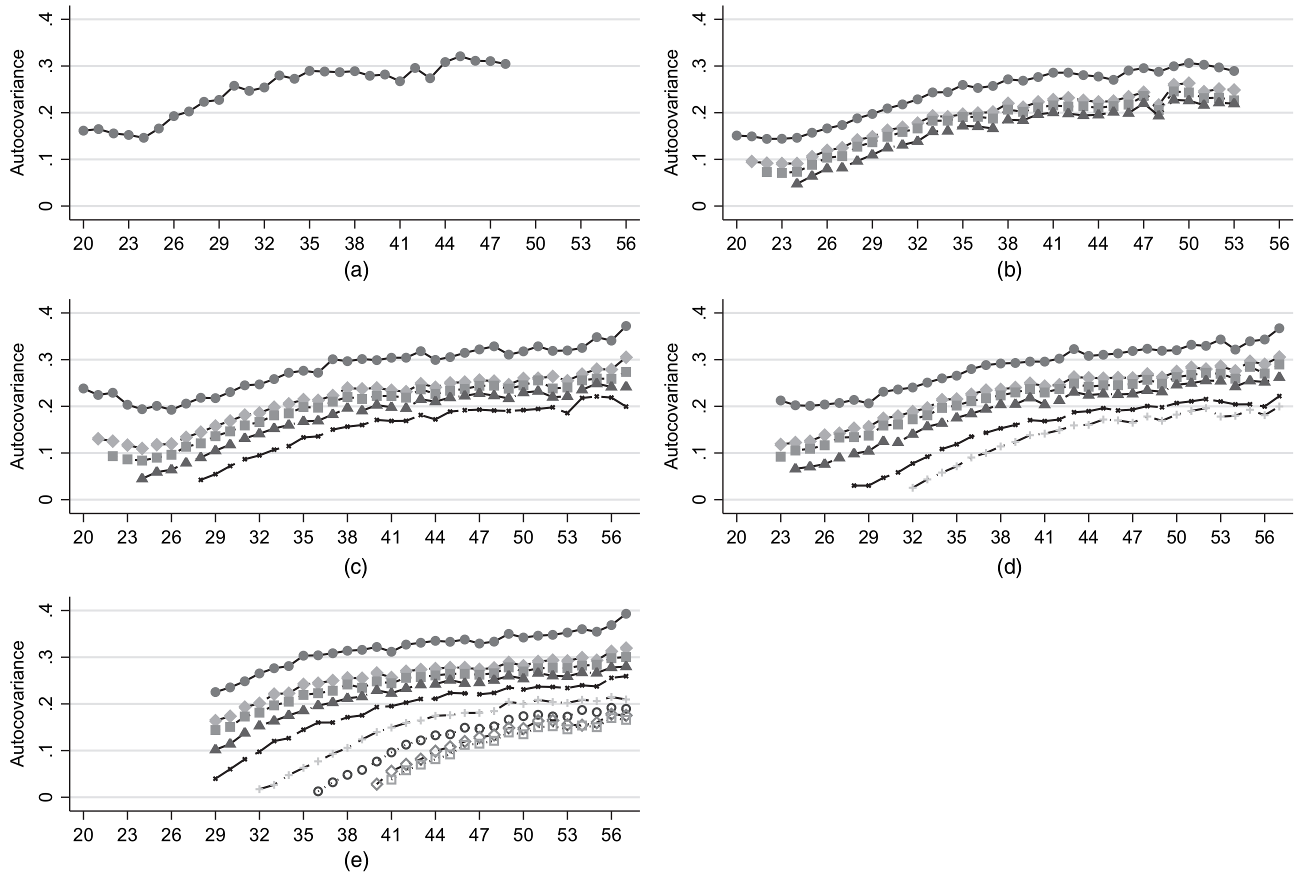


Fig. 5. Life cycle autocovariances of log-earnings for selected years (‘t’ stands for the age displayed on the horizontal axis) (●, variance(t); ◆, covariance($t, t-1$); ■, covariance($t, t-2$); ▲, covariance($t, t-4$); ✖, covariance($t, t-8$); +, covariance($t, t-12$); ○, covariance($t, t-16$); ◇, covariance($t, t-20$); □, covariance($t, t-21$)): (a) 1988; (b) 1993; (c) 1998; (d) 2003; (e) 2009

except for the oldest cohort. The rate of those increases differs across cohorts. Similarly to the results of Dickens (2000) for the UK, the younger the cohort the faster the rise in the autocovariances over time. The absolute magnitude of the autocovariance structure has a hump-shaped pattern. The youngest cohort shown (1980) exhibits low lag autocovariances. Lag autocovariances are much higher for the middle cohorts (1950 and 1960 especially), at long lags in particular. The oldest cohort then exhibits somewhat lower autocovariances.

The distance between autocovariances at consecutive lags falls at a decreasing rate. The biggest fall is registered by the lag 1 autocovariance, after which the covariances appear to converge gradually at a positive level. As variances reflect both the permanent and the transitory components of earnings, and higher order covariances reflect the permanent component of earnings, the evolution of covariances at all orders suggests the presence of a permanent individual component of wages and a transitory component which is serially correlated.

Fig. 5 presents the variance–covariance structure by age for the selected years. All lag autocovariances of log-earnings show a similar pattern to that of the variance. They are positive and evolve parallel to the variance, yet at different rates over the life cycle. They rise sharply until the late 30s and early 40s, after which the rate of increase slows down. The diminishing rate of increase of all lag autocovariances observed from age 20 years until the late 50s is consistent with the presence of a permanent component of earnings that rises with age at a decreasing rate. Across years, the life cycle profile of the autocovariances become somewhat steeper. If the slope of the life cycle profile is interpreted as the permanent increase in earnings, steeper slopes in later years imply increasing returns to the permanent component of earnings over time.

The autocovariances structure and the life cycle profiles for national, immigrant and cross-border workers differ somewhat in levels but the general patterns of lag auto-correlation and life cycle variances all follow broadly similar patterns. Note that, for cross-border workers, the slowdown in the rate of increase in life cycle variances after the late 30s is stronger compared with the other labour market groups. (The full long-run autocovariance structure of monthly earnings for all workers, as well as the patterns for natives, immigrants and cross-border workers, is reported in the on-line appendix Figs D.3, D.4, D.5 and D.6 for autocovariances and appendix Figs D.7, D.8 and D.9 for life cycle profiles.)

5. Persistent and transitory inequality in a model of earnings dynamics

We propose a flexible error components model to fit the autocovariance structure just described. To separate out life cycle dynamics from secular changes in earnings inequality, earnings trajectories are analysed within each of the 41 birth cohorts. Models are estimated separately for natives, immigrants and cross-border workers to compare and account for their different earnings dynamics and variances. Combining model parameter estimates then allows us to disentangle permanent and transitory components in the level and trends of earnings inequality and the contribution of the different worker types to these trends.

5.1. Model specification

We first detrend earnings and model earnings as zero-mean deviations from yearly cohort means:

$$r_{it} = Y_{it} - \bar{Y}_{c(i)m(i)t} \tag{5}$$

where Y_{it} is the natural logarithm of real average monthly earnings of individual i in year t and $\bar{Y}_{c(i)m(i)t}$ is the average in year t of Y_{it} over all workers of the same cohort and of the same type as individual i—whether native, immigrant or non-resident. (Demeaning is a standard procedure in administrative data (Bingley *et al.*, 2013; Baker and Solon, 2003). Survey data studies more frequently rely on regression adjustment (e.g. Moffitt and Gottschalk (2012)). Worker type is treated as a time invariant status. Individuals are classified on the basis of their most frequent status. Individuals with multiple status are primarily cross-border workers who later migrate to Luxembourg. Only 11.07% of workers whom we classify as immigrants have multiple status over time. This causes 2.6% of person-year observations for immigrants to be effectively periods spent as cross-border workers. Similarly 2.68% of workers whom we classify as cross-border workers have multiple status over time. This causes 1.2% of person-year observations for this group to be effectively periods spent as resident.) Individual-specific deviations from year-cohort means r_{it} are then assumed to be independently distributed across individuals, but auto-correlated over time. So, the structure of earnings differentials within each cohort and worker type is fully characterized by modelling the covariance structure of individual (demeaned) earnings: $E(r_{it}r_{it-s})$ for $t=t^0_{c(i)},\ldots,(t^0_{c(i)}+T_{c(i)})$ and $s=0,\ldots,t-t^0_{c(i)}$. ($t^0_{c(i)}$ is the first year at which the cohort of individual i is observed in the data (e.g. 1988 for the 1940 cohort) and $T_{c(i)}$ represents the total number of years that the cohort is observed.)

For clarity of exposition, we ignore indices for worker type throughout this section. All model parameters will be estimated separately for each of the three groups of workers. For the sake of exposition, we also denote by c instead of $c(i)$ the cohort of individual i.

As in much of the literature, our model is an extension of the canonical model of earnings dynamics of Lillard and Willis (1978) in which r_{it} is assumed to be the sum of two orthogonal terms:

$$r_{it}=\mu_i+\upsilon_{it}, \qquad \mu_i\sim \mathrm{IID}(0,\sigma^2_\mu), \quad \upsilon_{it}\sim \mathrm{IID}(0,\sigma^2_\upsilon). \tag{6}$$

This canonical model decomposes earnings into a permanent, time invariant individual-specific component μ_i (reflecting labour market returns to innate ability and pre-labour-market human capital accumulation) and a transitory component (reflecting any yearly deviation from the permanent component) υ_{it}. Both components in this model are independent both across individuals and over time. The implied covariance structure of earnings then takes the form

$$\mathrm{cov}(r_{it},r_{is})=\begin{cases}\sigma^2_\mu+\sigma^2_\upsilon, & t=s,\\ \sigma^2_\mu, & t\neq s,\end{cases} \tag{7}$$

where σ^2_μ is the persistent dispersion of earnings (permanent earnings inequality) and σ^2_υ is the variance of transitory deviations. The variance of earnings at a given year t is given by $\sigma^2_r=\sigma^2_\mu+\sigma^2_\upsilon$ and deviates from the persistent dispersion by the variance of the transitory shocks. This canonical model obviously imposes severe restrictions on the covariance structure of earnings. More sophisticated specifications are now routinely estimated (see Meghir and Pistaferri (2011) for a comprehensive review).

We specify and estimate a model which accommodates fine details of the autocovariance structure of earnings. We maintain the basic assumption that r_{it} is the sum of two orthogonal components, one persistent and one transitory, but we allow the relative weight of each of the two terms to vary over time and by cohort to examine changes in the relative weight of persistent and transitory components:

$$r_{it}=\gamma_{1c}\lambda_{1t}\mu_{it}+\gamma_{2c}\lambda_{2t}\upsilon_{it}. \tag{8}$$

We allow the permanent term μ_{it} to have a unit root and to evolve as a random walk with age,

$$\mu_{it}=\mu_{i(c+20)}\sim \mathrm{IID}(0,\sigma^2_{\mu_{c+20}}) \qquad \text{if } t=c+20, \tag{9}$$

$$\mu_{it}=\mu_{i,t-1}+\pi_{it} \qquad \text{if } t>c+20, \tag{10}$$

$$\pi_{it} \sim \text{IID}(0, \sigma^2_{\pi_{t-c}}),$$
$$E(\mu_{i,t-1}, \pi_{it}) = 0$$

and let the transitory term υ_{it} follow an ARMA(1,1) process:

$$\upsilon_{it} = \rho\upsilon_{i,t-1} + \epsilon_{it} + \theta\epsilon_{i,t-1}, \tag{11}$$
$$\epsilon_{it} \sim (0, \sigma^2_{\epsilon_{ct}}),$$
$$\upsilon_{i0} \sim (0, \sigma^2_{c0}).$$

The covariance structure of earnings is allowed to vary over time by incorporating time-specific shifters on the two main components, λ_{kt}, $k=1,2$, that allow for the relative contributions of the permanent and transitory components to change over time (see, for example, Dickens (2000), Haider (2001), Moffitt and Gottschalk (2002), Baker and Solon (2003), Ramos (2003), Cappellari (2004), Biewen (2005), Kalwij and Alessie (2007), Gustavsson (2007, 2008) and Sologon and O'Donoghue (2012)). λ_{kt}, $k=1,2$, is normalized to 1 in the first year (1988) for identification.

Allowing the relative contributions of the permanent and transitory components to vary also by cohort by incorporating cohort-specific loading factors γ_{kc}, $k=1,2$, is as in Kalwij and Alessie (2007), Gustavsson (2008) or Sologon and O'Donoghue (2012). γ_{kc}, $k=1,2$, is normalized to 1 for the cohort born in 1945.

Specification of a random walk in age for the permanent component of earnings follows MaCurdy (1982), Abowd and Card (1989), Dickens (2000), Baker and Solon (2003), Ramos (2003), Kalwij and Alessie (2007), Gustavsson (2008), Moffitt and Gottschalk (2012) and Sologon and O'Donoghue (2012). This specification captures earnings shocks with permanent effects. Whereas most studies restrict the innovation variance $\sigma^2_{\pi_{t-c}}$ to be constant, we estimate age-specific innovation variances (age is $a=t-c$) in a way similar to Dickens (2000), Gustavsson (2008) and Kalwij and Alessie (2007). (In the application, age-specific innovation variances are estimated from age 21 to 49 years, after which innovation variances are allowed to vary every 2 years at age 50–51, 52–53,…, 56–57. For cross-border workers the innovation variances vary only twice after age 39 years, namely for age 40–49 and 50–57 years.) The importance of allowing for age-specific variances was emphasized in Blundell *et al.* (2014). This specification accommodates the highly persistent increase in earnings variance with age, as observed in Fig. 4.

The ARMA(1,1) specification for the transitory component of earnings is as in MaCurdy (1982). The serial correlation parameter ρ captures the decreasing rate of decay of the covariances with the lag, the moving average parameter θ captures the sharp drop of the lag 1 autocovariance compared with the other autocovariances and ϵ_{ict} are white noise mean reverting transitory shocks. The cohort-specific variance σ^2_{c0} measures the volatility of shocks at the start of the observation period and the cohort-specific $\sigma^2_{\epsilon_{ct}}$ the volatility of shocks in subsequent years.

According to MaCurdy (1982), initial cohort transitory variances could be treated as additional parameters to be estimated. However, Ostrovsky (2010) and Moffitt and Gottschalk (2012) argued that treating the initial transitory variances of each cohort as unrestricted parameters is problematic because it affects the time trend for left-censored observations. They proposed instead to introduce a parameter α which allows cohort-specific transitory variances in the first wave to deviate from what they would be if $\lambda_{2t}=1$ for the years before the first wave, so

$$\sigma^2_{c0_{\text{left censored}}} = \{1+\alpha(a_{c0}-20)\}\sigma^2_0, \qquad c = 1940, \ldots, 1980, \tag{12}$$

where $a_{c0} = t^0_c - c$ is the age of the cohort in the first wave.

Finally, as recent studies found that the variance of the transitory component tends to be a U-shaped function of age or experience (Baker and Solon, 2003; Gustavsson, 2008), we also allow for age-related heteroscedasticity in the transitory shocks by letting a cohort-specific variance of ϵ_{it} vary as a polynomial in age:

$$\sigma^2_{\epsilon_{ct}} = \beta_0 + \beta_1(a_{ct}-20) + \beta_2(a_{ct}-20)^2 + \beta_3(a_{ct}-20)^3 + \beta_4(a_{ct}-20)^4 \tag{13}$$

where $a_{ct} = t - c$ is the age of cohort c at time t.

This model specification allows for a wide range of dynamics: a high degree of individual heterogeneity by allowing for individual and age-specific characteristics in the permanent component via a random-walk specification with age-specific innovation variances, a transitory component which evolves as an ARMA(1,1) process, with a correction for left censoring for each cohort in the first year observed, and with age-specific heteroscedastic transitory variances. The non-stationary pattern of earnings is accommodated by time-specific loading factors on both earnings components. Cohort heterogeneity is accommodated by allowing both the permanent and the transitory component to vary by cohort. The model is similar to that of Kalwij and Alessie (2007), with added features from Baker and Solon (2003) (age-specific heteroscedastic transitory variances), and Ostrovsky (2010) and Moffitt and Gottschalk (2012) for the correction for left censoring for each cohort in the first year observed.

5.2. Alternative models

Our model (henceforth called the base model) applies a random walk that uses many parameters to capture the age-specific characteristics in the permanent component. This specification is data demanding. It is amenable to estimation here thanks to the large size of our data but it may be difficult to estimate in smaller data sets. We therefore test three restricted models and assess their fit compared with the flexible base model.

In the first model, we replace the age-specific innovation variances in the random walk of our base model with the standard random-walk specification (which assumes an age invariant innovation variance) complemented by a random-growth component (e.g. Baker and Solon (2003)). Formally we allow the permanent term μ_{it} to evolve as a random walk with age u_{it} plus a random-growth factor (β_i):

$$\begin{aligned}
\mu_{it} &= \alpha_i + \beta_i(a_{ct}-20) + u_{it}, \\
u_{it} &= u_{i,t-1} + \pi_{it}, \\
\alpha_i &\sim \text{IID}(0, \sigma^2_{\alpha_i}), \\
\beta_i &\sim \text{IID}(0, \sigma^2_{\beta_i}), \\
\text{cov}(\alpha, \beta) &= \sigma_{\alpha,\beta}, \\
\pi_{it} &\sim \text{IID}(0, \sigma^2_{\pi_{t-c}} = \sigma^2_{\pi}), \\
E(u_{i,t-1}, \pi_{it}) &= 0.
\end{aligned} \tag{14}$$

α_i captures individual-specific intercepts and β_i the individual-specific growth rates of the earnings profiles (α_i incorporates $u_{i(c+20)}$). Their variance and covariance are denoted $\sigma^2_{\alpha_i}$, $\sigma^2_{\beta_i}$ and $\sigma_{\alpha,\beta}$. A negative $\sigma_{\alpha,\beta}$ would signal the presence of Mincerian crossovers (Mincer, 1974). This is referred to as restricted model 1.

In the second model (henceforth restricted model 2), we approximate the age variation in the variance of the permanent component by a fourth-order polynomial in age:

$$\sigma^2_{\mu_{it}} = \xi_0 + \xi_1(a_{ct} - 20) + \xi_2(a_{ct} - 20)^2 + \xi_3(a_{ct} - 20)^3 + \xi_4(a_{ct} - 20)^4. \tag{15}$$

(We thank a referee for this suggestion.)

In the third model (henceforth restricted model 3) we underline the importance of capturing cohort effects in both components of earnings. Controlling for cohort effects is a feature that is often neglected in the inequality literature, but it is of empirical importance as shown by Kalwij and Alessie (2007) and discussed in Blundell and Preston (1996, 1998). We show that ignoring cohort heterogeneity by assuming that the cohort shifters are equal to 1 greatly underestimates estimation of persistent inequality. Our conclusion, as we shall see, reiterates the findings of Kalwij and Alessie (2007) about the importance of accounting for age, time and cohort effects in the estimation of permanent and transitory inequality.

5.3. Permanent versus transitory variance components

The earnings dynamics model determines a theoretical autocovariance structure of earnings which enables separating out persistent and transitory components of inequality.

In our base model, at the first period, and for cohort $c = c(i)$ of initial age $a_0 = 1988 - c$, the variance of log-earnings is

$$\operatorname{var}(Y_{i0}) = E(r_{i0} r_{i0}) \tag{16}$$

$$= \underbrace{\sigma^2_{\mu_{20}} + \sum_{a=21}^{a_0} \sigma^2_{\pi_a}}_{\text{persistent inequality}} + \underbrace{\operatorname{var}(\upsilon_{i0})}_{\text{transitory inequality}}. \tag{17}$$

In subsequent years, the theoretical covariance structure is

$$\operatorname{var}(Y_{it}) = E(r_{it} r_{it}) \tag{18}$$

$$= \underbrace{\gamma^2_{1c}\lambda^2_{1t}\left(\sigma^2_{\mu_{20}} + \sum_{a=21}^{a_t} \sigma^2_{\pi_a}\right)}_{\text{persistent inequality}} + \underbrace{\gamma^2_{2c}\lambda^2_{2t}\{\rho^2 \operatorname{var}(\upsilon_{i,t-1}) + \sigma^2_{\epsilon_t}(1 + 2\rho\theta + \theta^2)\}}_{\text{transitory inequality}}$$

and

$$\operatorname{cov}(Y_{ict}, Y_{i,c,t-s}) = E(r_{ict} r_{i,c,t-s}) \tag{19}$$

$$= \gamma^2_{1c}\lambda^2_{1t}\left(\sigma^2_{\mu_{20}} + \sum_{a=21}^{a_{t-s}} \sigma^2_{\pi_a}\right) + \gamma^2_{2c}\lambda_{2t}\lambda_{2,t-s}\{\rho \operatorname{cov}(\upsilon_{i,t-1}\upsilon_{i,t-s})\} \qquad \text{if } s > 1,$$

$$\operatorname{cov}(Y_{ict}, Y_{i,c,t-1}) = E(r_{ict} r_{i,c,t-1}) \tag{20}$$

$$= \gamma^2_{1c}\lambda^2_{1t}\left(\sigma^2_{\mu_{20}} + \sum_{a=21}^{a_{t-1}} \sigma^2_{\pi_a}\right) + \gamma^2_{2c}\lambda_{2t}\lambda_{2,t-s}\{\rho \operatorname{var}(\upsilon_{i,t-1}) + \theta\sigma^2_{\epsilon_{t-1}}\} \qquad \text{if } s = 1.$$

From equations (17) and (18) we can therefore decompose total earnings variance for any cohort into a permanent and a transitory component and track their respective share over time. Restricted models can be estimated similarly.

5.4. Estimation

Estimation of the model parameters is based on the theoretical autocovariance matrix. The full model specification determines a theoretical autocovariance structure where each cell of the autocovariance matrix is a function of model parameters. Parameters can then be estimated by fitting the theoretical covariance matrix onto the empirical covariance structure by using minimum distance methods. If θ is the set of parameters to be estimated, the minimum distance estimator selects $\hat{\theta}$ to minimize the distance function

$$D(\hat{\theta}) = (M - f(\hat{\theta}))W(M - f(\hat{\theta}))', \tag{21}$$

where M is a column vector of moments of dimension 7513×1. We take W to be the identity matrix, following Altonji and Segal (1996) and Clark (1996) and most empirical applications. For estimating the asymptotic standard errors of the parameter estimates, we apply the delta method, following Chamberlain (1984). This method-of-moments approach does not require additional modelling assumptions and is now the workhorse for estimation of such error components models.

The method-of-moments estimator gives us model parameters for each of our m replications of the data filled in with multiply imputed top-coded incomes. We finally combine the m vectors of estimates by using Reiter's (2003) combination formula as per equation (3) and calculate the corresponding sampling variance of the averaged parameters by using equation (4). This procedure ensures that we account for the variability that is introduced by the stochastic nature of our imputation model for top-coded earnings.

5.5. Assessing subgroup contributions

By estimating the error components model parameters separately for subgroups of workers—nationals, immigrants and cross-border workers—we can allow for different variances within each of the subgroups and identify different trends. Applying simple variance decomposition arithmetic by subgroup, we use the model estimates to track the contribution of each of the subgroups to overall inequality.

Let $\bar{V}$ denote the average within-group log-earnings variance at time t (Chakravarty, 2001):

$$\bar{V} = \sum_{g=1}^{k} n_g V_g \tag{22}$$

where n_g and V_g are the population share and the permanent variance of group g. A basic decomposition takes the difference between the observed total variance V and $\bar{V}$ as a measure of the 'between-group' contributions:

$$B = V - \bar{V}. \tag{23}$$

The evolution of V can then mechanically be linked to the evolution of the subgroup shares n_g, the subgroup variances V_g and the residual measure of between-group contributions B. (Semantics are important here since B is not a measure of between-group 'inequality', i.e. it is not equal to the overall variance of log-earnings that would be observed if all earnings were set to equal to their subgroup means—the typical definition of a between-group inequality component (Shorrocks, 1984). The latter cannot be recovered from our model parameters since it is based on modelling the *logarithm* of earnings.)

In Section 6, we apply these simple mechanics to both the transitory variance and the permanent variance on the basis of model-based predictions for V_g as per equations (17) and (18) and a model-based prediction for overall V estimated from the overall pooled population of the three worker subgroups. It must be borne in mind, however, that, as explained in Section 2, we observe employment *in* Luxembourg only. Cross-border workers with employment spells outside Luxembourg have partially observed earnings trajectories. For model estimation, such workers contribute to the calculation of contemporaneous variances and to the calculation of lag autocovariances only for pairs of years in which their earnings are observed. It follows that the share of workers with continuous employment in Luxembourg contributing to the calculations of lag auto-correlations is larger than their share in the calculation of contemporaneous variances. Although this appropriately reflects the autocovariances of Luxembourg earnings, it can be expected to push the observed persistence of earnings of cross-border workers upwards, to the extent that workers with continuous employment in Luxembourg have more stable earnings profiles. However, as we show below, cross-border workers still turn out to exhibit lower persistent earnings inequality than resident workers. We do not expect this issue to affect *trends* in inequality.

6. Inequality trends in Luxembourg 1988–2009

6.1. Base model estimates

We estimated our most detailed base model on the entire population of (male) workers and then separately on subgroups of workers: Luxembourg nationals, immigrants (foreign residents) and cross-border (non-resident) workers. (Tables in the on-line appendix E report estimates and associated standard errors for all model parameters.) Parameter estimates were then used to decompose the variance of log-earnings in each year into its permanent and transitory components to examine the nature of inequality trends in our 22 years of data.

Fig. 6 displays the trends in inequality (observed and as predicted by the model parameters) and the absolute and relative contributions of the persistent and transitory components for all workers. As in Baker and Solon (2003), to account for different age compositions over time, the variance of log-earnings is as predicted at the age of 40 years, which is approximately the middle of the active career. (Predictions at each year are based on the relevant combinations of period and birth cohort parameters.)

Note, first, the close coincidence of the trends in the observed and predicted variances for 40-year-old men—an indication of the good fit of our model. Note, second, that the trends in predicted variance at age 40 years roughly follow the patterns that are outlined in Fig. 2 for all age groups combined: inequality remained approximately constant from 1988 to 1993, after when it drifted upwards (although with temporary ups and downs—the increase in inequality at age 40 years is not as marked between 1993 and 1999). (Predicted trends at ages 30 and 50 years exhibit similar patterns (see Fig. F.10 in the on-line appendix F).)

The modest increase in inequality since the middle of the 1990s turns out to be mainly driven by an increase in the persistent component of the model, alongside a reduction in the transitory component. We observe a fanning out of the two components from the mid-1990s, which is a period which coincided with the acceleration of the development of the financial sector and the contraction of the steel industry. Overall, persistent inequality increased by 23.4%, whereas transitory inequality decreased by 25.1% between the late 1980s and the late 2000s. These offsetting trends led to a modest increase in overall inequality of 7.5%. The share of persistent inequality in total inequality rose from over 60% in 1988 to close to 80% in 2009. These trends contrast with what has been observed elsewhere, e.g. in the USA, the UK, Italy and Germany where the transitory variance increased faster than persistent inequality (Kalwij and Alessie, 2007; Moffitt and Gottschalk, 2012; Cappellari and Leonardi, 2013; Bönke *et al.*, 2015). (We return to cross-national comparisons in Section 8.)

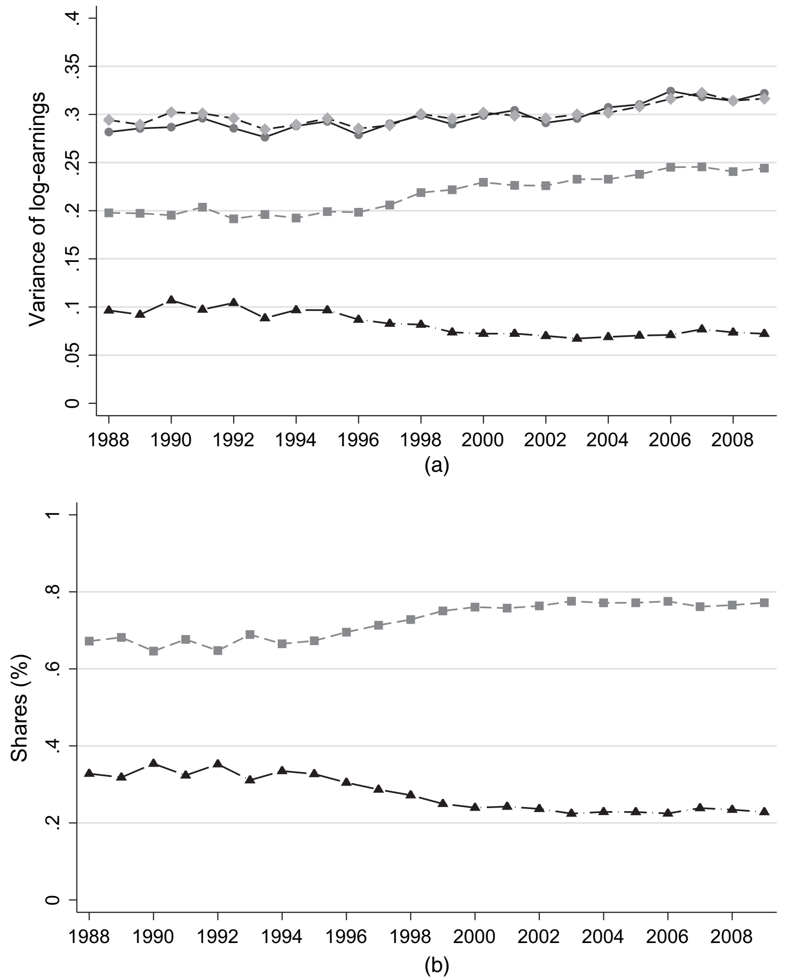


Fig. 6. Inequality decomposition for all men at age 40 years—base model: (a) variance (●, actual variance; ◆, predicted variance; ■, permanent component; ▲, transitory component); (b) shares (■, permanent component as a percentage of the overall variance; ▲, temporary component as a percentage of the overall variance)

6.2. Alternative model specifications

To assess how much estimation of a flexible model is important to capture the fine details of the structure of earnings dynamics and how much the inequality decomposition is affected by potential restrictions that are imposed on this structure, we estimated the three restricted model specifications described above and derived the corresponding inequality decompositions. Fig. 7 shows the predicted variances and the decomposition of inequality that were obtained with restricted specifications of the model.

Replacing the age-specific innovation variances by a constant innovation variance complemented by a random-growth component (following Baker and Solon (2003) and more recently Moffitt and Gottschalk (2012), restricted model 1) leads to a noticeably worse fit to the actual variances than does the base model. This is confirmed by the sum of squared residuals (0.340 *versus* 0.422) and by a Bayes information criterion (5.63×10^{-5} *versus* 6.73×10^{-5} for restricted

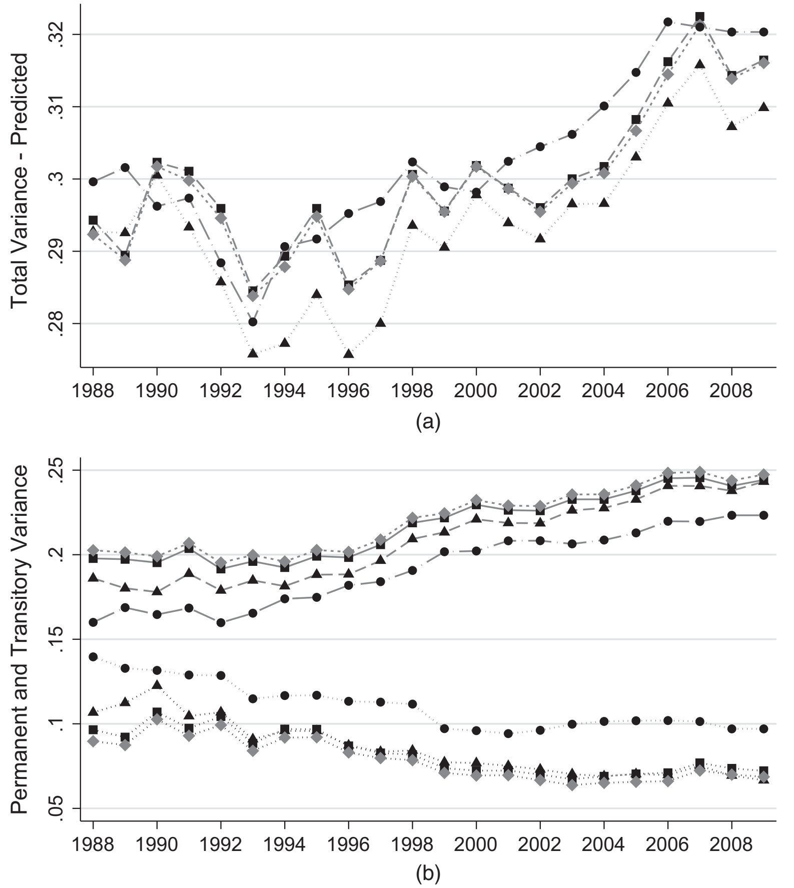


Fig. 7. Predicted total variance of earnings for all men at age 40 years: (a) total minus predicted variance (■, predicted total variance, base model; ▲, predicted total variance, restricted model 1; ◆, predicted total variance, restricted model 2; ●, predicted total variance, restricted model 3); (b) permanent and transitory variance (—■—, permanent variance, base model; --▲--, permanent variance, restricted model 1; ◆, permanent variance, restricted model 2; —●—, permanent variance, restricted model 3; ■, transitory variance, base model; ▲, transitory variance, restricted model 1; ◆, transitory variance, restricted model 2; ●, transitory variance, restricted model 3)

model 1) according to which the improvement in fit outweighs the complexity that is introduced by the higher the number of parameters. The restrictions affect the decomposition, mostly by affecting persistent inequality estimates. With restricted model 1, the level of permanent inequality is lower at any point in time, whereas the transitory inequality is roughly at the same level, except for the late 1980s, when it is higher. Thus permanent inequality makes up relatively less of the total inequality than predicted by our base model: the difference is on average over 5 percentage points in the first half of the period and around 2 percentage points in the second half of the period. This in turn implies a higher contribution of transitory inequality to overall inequality. Thus a coarse specification of the persistent components loads a higher persistence onto the transitory component, predicting more persistent transitory shocks. Another difference is in the slope of the predicted permanent inequality: the trend increase is greater under restricted model 1 than under the base model.

Second, we assess whether the complexity of our model for the permanent components could be approximated by a fourth-order polynomial in age. We find that restricted model 2 provides a very good approximation of the decomposition captured by our base model. The polynomial in age can capture precisely the trend and the level of permanent inequality, which in turn assures that transitory inequality does not capture parts of the persistent component. Both the sum of squared residuals and the Bayes information criterion are very close to those for the base model (0.341 and 5.43×10^{-5} respectively). This somewhat more parsimonious specification should therefore be of interest when estimation of the fully flexible base model is difficult, e.g. because of sample size limitations.

Third and finally, we examine another feature that is neglected in many studies, namely cohort heterogeneity in both components of earnings. To assess the implication of such a restriction, we estimate our base model without any of the cohort shifters. The fit is considerably worse than our base model as confirmed by the sum of squared residuals (0.340 *versus* 0.672) and by a Bayes information criterion (5.63×10^{-5} *versus* 1×10^{-4}). The decomposition is greatly affected by the restriction. The restriction affects the level and the trends of both components. Permanent inequality is underestimated, whereas transitory inequality is overestimated at every point in time. Thus permanent inequality makes up much less of the total inequality than predicted by our base model: the difference is on average over 10 percentage points in the first half of the period and around 8 percentage points in the second half of the period. Moreover, the trend increase in permanent variance is steeper under restricted model 3. These findings reiterate the importance of accounting for the age, time and cohort effects in such decompositions, as highlighted by Kalwij and Alessie (2007).

6.3. Distinguishing trends by workers' origin

The rise in persistent inequality from the mid-1990s may be related to the changing structure of employment and the massive inflow of foreign labour and cross-border workers in particular. The mid-1990s marks the period when the share of cross-border workers in the labour force overtook the share of nationals. To see this, we decompose the permanent inequality component by using the parameter estimates reported in the on-line table appendix E to predict persistent inequality within each of the native, immigrant and cross-border worker subgroups. The predicted persistent inequalities for each group are then aggregated to obtain an estimate of overall within-group persistent inequality as $\mathrm{PV} = \Sigma_{g=1}^{k} n_g \mathrm{PV}_g$ where n_g and PV_g are the population share and the permanent variance of group g and the residual difference between overall persistent inequality and PV is a measure of the between-group contribution to permanent inequality (see Section 5).

Fig. 8 shows the decomposition of the trends in persistent inequality. Cross-border workers exhibit the lowest persistent earnings inequality throughout the period, signalling a more homogeneous group in terms of persistent earnings capacity than immigrants and nationals. Immigrants display the highest persistent differentials from the 2000s. This is consistent with the argument that Luxembourg immigrants have become concentrated at both ends of the skill distribution (see, for example, Amétépé and Hartmann-Hirsch (2011), Choe and Van Kerm (2014) and Fusco *et al.* (2014)).

Trends in permanent inequality are more sharply marked within subgroups than for the total population. Cross-border workers recorded the largest relative increase in persistent inequality (91.4%), followed by immigrants with a relative increase of 80.6%. Overall permanent inequality did not increase in similarly large proportions (23.5%) because

(a) persistent differentials decreased by 22.3% among nationals (in particular between 1988 and 1996 after when it started to trend upwards also),

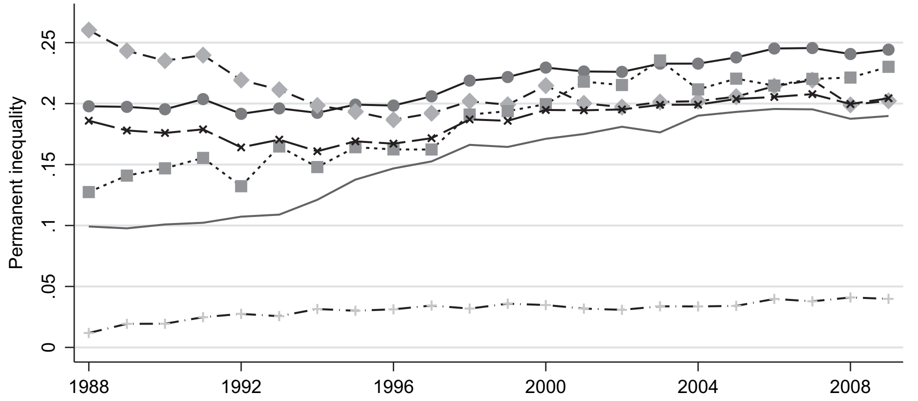


Fig. 8. Permanent inequality decomposition by worker type: ●, permanent inequality, all; ◆, permanent inequality, nationals; ■, permanent inequality, immigrants; ——, permanent inequality, cross-border workers; ✕, aggregated within-group permanent inequality; +, between-group (nationals *versus* immigrants *versus* cross-border workers) permanent inequality

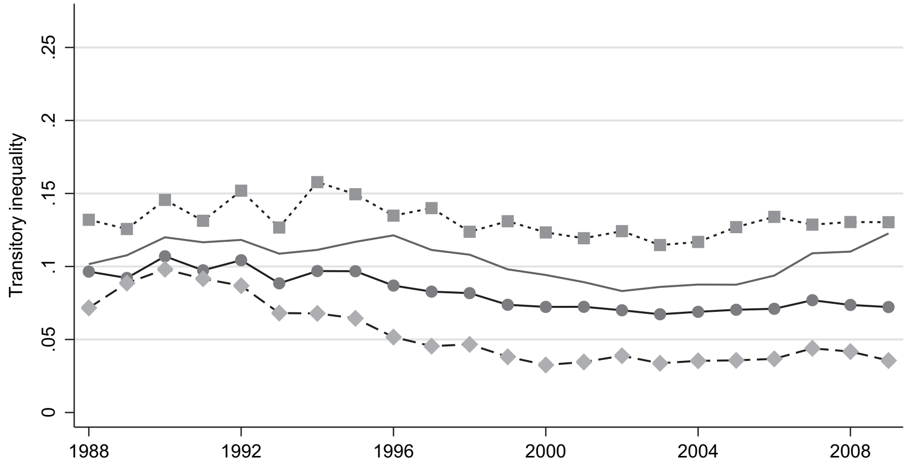


Fig. 9. Transitory inequality decomposition by worker type: ●, transitory inequality, all; ◆, transitory inequality, nationals; ■, transitory inequality, immigrants; ——, transitory inequality, cross-border workers

(b) the weight of nationals decreased during the period and
(c) the weight of cross-border workers—which, despite the increase, still have less inequality than the other groups—increased during the period.

The contribution of persistent differentials among cross-border workers to overall permanent inequality is compounded by the sharp increase in their share in the labour market. In 1988, persistent inequality among cross-border workers (weighted by their population shares) accounted for 10.9% of overall persistent inequality, against 19.1% by immigrants and 64% by nationals. The remaining 6% are claimed by persistent earnings differences between the three groups. By 2009, persistent inequality among cross-border workers accounts for the largest share in the overall persistent inequality (37.7%), followed by immigrants with 27.5% and by nationals with 18.5%. 16.3% are claimed by persistent earnings differences between the three groups.

An increase in between-group differentials also contributed to the overall growth in permanent inequality. Whereas it contributed to about a tenth of the overall permanent inequality in 1988, it contributes to close to a fifth by 2009.

We finally turn to trends in the transitory components of inequality—earnings instability. As illustrated in Fig. 9, earnings instability at age 40 years changed little until the mid-1990s, and then decreased until the mid-2000s. Again, this relative stability hides contrasted levels and trends for population groups. Immigrants had the highest transitory fluctuations in earnings throughout the period, followed by cross-border workers and nationals. Earnings instability for nationals decreased substantially over the whole period whereas the earnings instability of cross-border workers appears to increase sharply from 2005 and, by 2009, almost converged to the level that was observed for immigrants and was higher than in any previous year.

7. Business cycles, employment and inequality trends

Our results reveal a somewhat surprising stability of earnings inequality in the face of the large changes in the size and structure of employment and the rapid growth that the country experienced in the period under scrutiny. The changes are modest in comparison wih trends observed in other countries such as Germany where both persistent and transitory inequality increased considerably (see below). These results possibly hint at the role of Luxembourg's relatively strict labour market regulations and collective bargaining institutions in holding back earnings inequality—yet not so much in holding back *persistent* inequality.

Although it is beyond the scope of this paper to pin down the contribution of particular labour market institutions or regulations in this rather peculiar context, we examine in this section the significance of the estimated changes over time and check whether these changes over 22 years are associated with economic cycles, the sectoral composition of employment or basic labour market regulation indicators.

We first follow Baker and Solon (2003) and regress the persistent and transitory inequality components on a linear trend and on the growth rate in real GDP. Results are reported in Table 2. The point estimates indicate a significant positive trend for permanent inequality and a less strong negative trend for transitory inequality. Coefficient estimates on GDP growth rate suggest that both permanent and transitory inequality are sensitive to the business cycle but in opposite

Table 2. Trend and cyclical variation of the persistent and transitory components

Model	*Dependent variable*	*Linear trend*		*Real GDP growth rate*		*Adjusted R^2*
		Estimate	*Standard error*	*Estimate*	*Standard error*	
All men	Permanent variance	0.0032	0.0002	0.1417	0.0352	0.9331
	Transitory variance	−0.0019	0.0002	−0.0995	0.0417	0.7876
Nationals	Permanent variance	−0.0011	0.0006	0.1942	0.1063	0.3412
	Transitory variance	−0.0030	0.0004	−0.0873	0.0812	0.7256
Immigrants	Permanent variance	0.0053	0.0004	0.0934	0.0840	0.8820
	Transitory variance	−0.0011	0.0004	−0.1360	0.0670	0.2590
Cross-border workers	Permanent variance	0.0058	0.0003	0.1493	0.0641	0.9360
	Transitory variance	−0.0012	0.0004	−0.1888	0.0745	0.2668

Table 3. Trend, cyclical variation and labour market drivers of the persistent and transitory components

Dependent variable	*Linear trend estimate (standard error)*	*Real GDP growth rate estimate (standard error)*	*Finance, insurance, real estate, business services estimate (standard error)*	*Other services estimate (standard error)*	*Industry estimate (standard error)*	*Minimum wage relative average estimate (standard error)*	*Union density estimate (standard error)*	*Adjusted R^2*
Permanent variance	0.0053	0.1016	0.3806	0.2499	0.8088			0.9728
	(0.0005)	(0.0256)	(0.1566)	(0.2829)	(0.1853)			
Transitory variance	−0.0008	−0.0604	−0.7727	−1.2453	−0.5108			0.8812
	(0.0007)	(0.0342)	(0.2089)	(0.3772)	(0.2475)			
Permanent variance	0.0035	0.1220				−0.1226	0.0006	0.9397
	(0.0004)	(0.0388)				(0.0691)	(0.0011)	
Transitory variance	−0.0033	−0.0280				−0.1219	−0.0039	0.8567
	(0.0005)	(0.0407)				(0.0726)	(0.0012)	

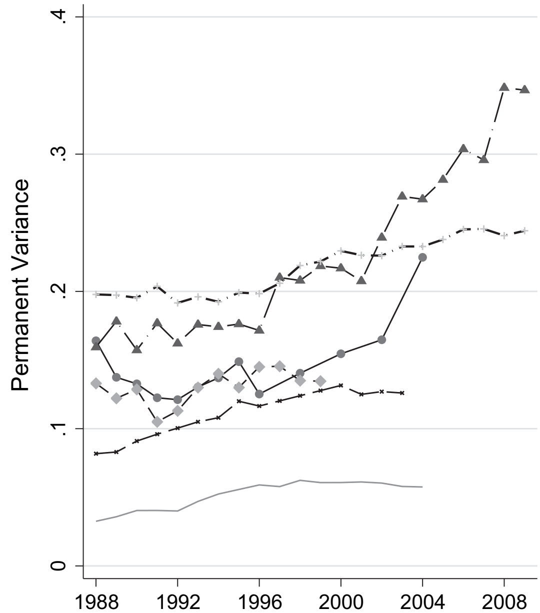
directions, with higher growth rates linked to increasing permanent inequality and decreasing transitory inequality. These findings are consistent for cross-border workers and immigrants taken separately, yet cross-border workers appear to be most sensitive to the business cycle. The picture for nationals is, by contrast, one of a significant negative trend in both permanent *and* transitory inequality with low sensitivity to the business cycle. Foreign labour therefore appears to act as a buffer absorbing business cycles whereas Luxembourg nationals are protected from macroeconomic fluctuations.

Underlying GDP growth was the increase in the contribution of the service sector—banking and financial services in particular—to both aggregate value added and employment. Table 3 shows regression results when we add the shares of the financial, insurance, real estate and business services, of the non-financial services and of industry to the model for all workers. Controlling for a linear trend and GDP growth, the share of the financial services is positively correlated with permanent inequality and negatively correlated with transitory inequality. This result suggests that the growth in the financial sector did not increase earnings inequality, unlike what has been found elsewhere; Bell and Van Reenen (2013), Godechot (2012), Kaplan and Rauh (2010) and Lin (2015). However, the change in employment made inequality more persistent, consistently with an interpretation of persistent inequality as reflecting higher returns to skills and human capital.

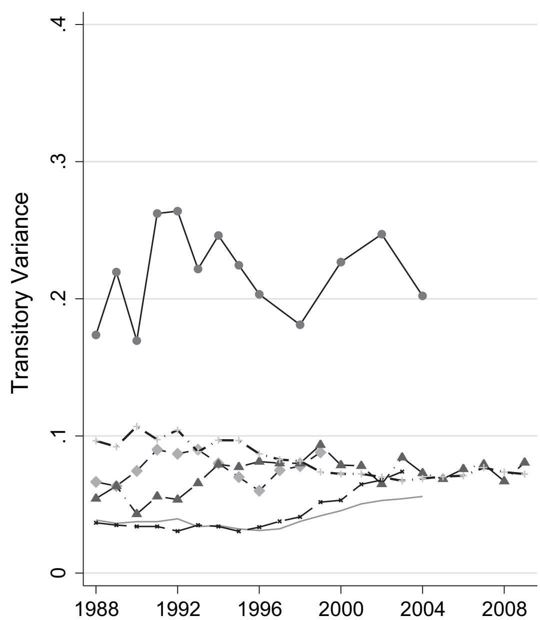
The increase in the persistence of inequality may, however, also be associated with relatively strict labour market regulation and collective bargaining. According to Organisation for Economic Co-operation and Development (2012), collective bargaining and trade union density are above the Organisation for Economic Co-operation and Development average (by 3 percentage points for collective bargaining and over 11 percentage points for union density). Employment protection legislation regarding temporary contracts is strict compared with the Organisation for Economic Co-operation and Development average and the minimum wage is high by international standards (Organisation for Economic Co-operation and Development, 2010). The only quantitative time series indicators linked to wage setting mechanisms that are available for Luxembourg between 1988 and 2009 are minimum wages (relative to mean wages) and trade union density. Table 3 shows regression results when we add those indicators to our model. The level of the minimum wage is negatively correlated with both components of inequality whereas union density is negatively correlated with transitory inequality but is not correlated with permanent inequality. This is consistent with the view that labour market institutions have contributed to mitigate upward pressures on inequality. Note that most coefficients have large standard errors, however, and that many components of labour market regulations are not captured by our regressions.

8. Cross-national comparisons

Finally, to put estimates into perspective, we compare the findings for Luxembourg with published estimates for other countries between 1988 and 2009. The benchmarks of our comparison are countries with available information for the longest overlapping period. We report both the cross-national differences in levels (Fig. 10) and the comparison of trends relative to 1988 (Fig. 11). We compare the evolution of persistent and transitory inequality of annual earnings in the USA between 1988 and 2004 based on the results in Moffitt and Gottschalk (2012), of annual earnings in Sweden between 1988 and 1990 based on Gustavsson (2008) and between 1991 and 1999 based on Gustavsson (2007), of monthly earnings in Denmark between 1988 and 2004 based on Bingley *et al.* (2013), of monthly earnings in Germany between 1988 and 2009 based on Bönke *et al.* (2015) and of weekly earnings in Italy between 1988 and 2003 based

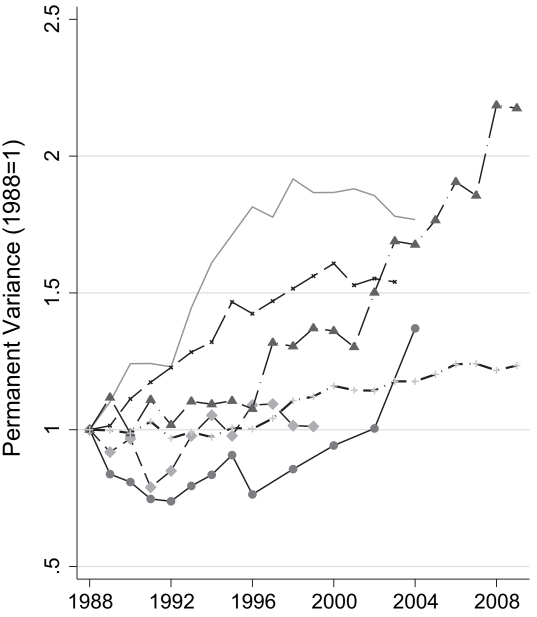


(a)

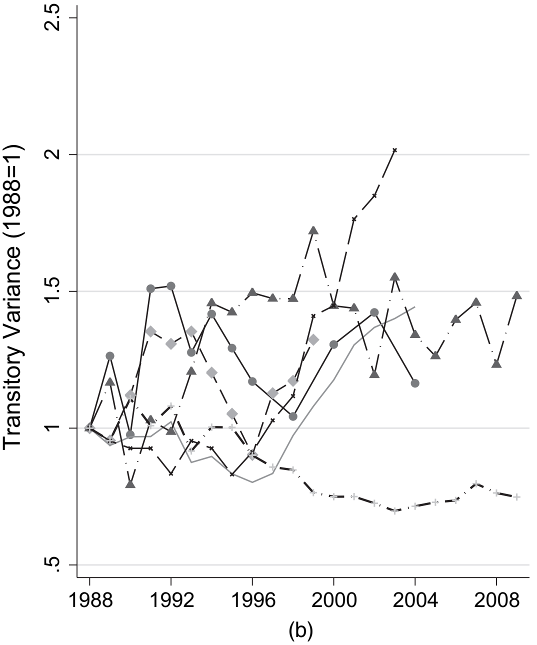


(b)

Fig. 10. Evolution of (a) permanent and (b) transitory variance of log-earnings for men in the USA (1988–2004) (●), Sweden (1988–1999) (◆), Denmark (1988–2004) (———), Germany (1988–2009) (▲), Italy (1988–2003) (●) and Luxembourg (1988–2009) (+): the numbers for the USA are based on Moffitt and Gottschalk (2012) for men age 40–49 years, Table A-3; the numbers for Sweden are based on Gustavsson (2007, 2008) for men age 40 years, Table 2 and Fig. 3; the numbers for Danish men are based on Bingley *et al.* (2013), Fig. 2; the numbers for Italian men are based on Cappellari and Leonardi (2013), Fig. 3; the numbers for Germany are based on Bönke *et al.* (2015) for men age 40 years; the numbers for Luxembourg are based on on-line Tables appendix E and E.5, men age 40 years



(a)



(b)

Fig. 11. Relative evolution (1988) of (a) permanent and (b) transitory variance of log-earnings for men in the USA (1988–2004) (●), Sweden (1988–1999) (◆), Denmark (1988–2004) (———), Germany (1988–2009) (▲), Italy (1988–2003) (●) and Luxembourg (1988–2009) (+): the numbers for the USA are based on Moffitt and Gottschalk (2012) for men age 40–49 years, Table A-3; the numbers for Sweden are based on Gustavsson (2007, 2008) for men age 40 years, Table 2 and Fig. 3; the numbers for Danish men are based on Bingley *et al.* (2013), Fig. 2; the numbers for Italian men are based on Cappellari and Leonardi (2013), Fig. 3; the numbers for Germany are based on Bönke *et al.* (2015) for men age 40 years; the numbers for Luxembourg are based on on-line tables appendix E, men age 40 years

on Cappellari and Leonardi (2013), with the estimates for monthly earnings for Luxembourg between 1988 and 2009 based on the on-line Tables appendix E and E.5. Of course, the comparability of findings is affected by the definition of income, sample designs, sources of data and especially earnings model specifications. Comparisons are therefore indicative and we focus on broad trends rather than more detailed analysis of levels.

According to the model estimates compared, Luxembourg displays a significantly higher persistent inequality than the USA, Germany until the early 2000s, Sweden, Italy and Denmark. However, we do not observe such a strong increase in Luxembourg during the 1990s as in Denmark, Italy and Germany, or during the 2000s as in Germany and the USA. In contrast, transitory inequality appears considerably larger in the USA than in Germany, Luxembourg, Sweden, Denmark and Italy. Moreover, whereas transitory inequality spikes upwards for Germany, Italy and Denmark, it tends to decrease in Luxembourg. This decline in transitory inequality therefore appears particularly at odds with international evidence. According to the simple regression results that were shown in Section 7, this can be related to the speed of economic growth, the change in the employment structure and relatively strict labour market regulations that were observed in Luxembourg throughout the period. But, although total earnings inequality is lower in Luxembourg than in the USA, it is considerably more persistent. Luxembourg is more unequal and more persistent than Germany until the early 2000s, becoming less unequal and less persistent thereafter. In addition, total earnings inequality appears much more stable over time in Luxembourg compared with the other countries. This is quite surprising given the major structural changes which have taken place in the labour market throughout the period that is covered by our analysis. Bear in mind, however, that the comparisons that are shown here are not all based on identical model specifications or data sources. Comparisons, especially of levels, must therefore be taken as indicative.

9. Summary and concluding remarks

This paper exploits longitudinal earnings data from a large extract from the Luxembourg social security administration registers to estimate a flexible model of earnings dynamics and documents the trends and sources of earnings inequality between 1988 and 2009. This has been a time when the country underwent a drastic industrial redevelopment towards the financial sector and sustained, high economic growth. In this process, labour demand soared and the country experienced a massive expansion of its employment through an inflow of foreign labour, especially of cross-border workers residing in neighbouring countries Belgium, France and Germany. Relatively strict labour market regulations were maintained (Fusco *et al.*, 2014).

In spite of these major structural and employment changes, we observe only a small overall increase in earnings inequality. This surprising stability appears, however, to be the net result of somewhat more complex underlying changes. Taking advantage of the large scale of our data, we estimate a rich model of earnings dynamics to distinguish first between persistent and transitory components of inequality. This shows how inequality became remarkably more persistent (whereas earnings instability decreased), suggesting an overall increase in returns to skills and human capital throughout the 22 years of our data. Second, we distinguish trends for native workers, immigrants and cross-border workers to capture the contribution of changes in employment composition better. This reveals that (persistent) inequality did grow significantly *within* the cross-border and immigrant worker groups and *between* the three worker groups, but that *overall* inequality growth was contained by

(a) a reduction in persistent inequality among nationals,
(b) the decreasing employment share of nationals and

(c) the increasing employment share of cross-border workers—the group exhibiting the lowest, yet most rapidly rising, 'within-group' persistent inequality.

Whereas earnings instability declined overall, immigrants still have higher transitory earnings variance whereas the transitory earnings variance for cross-border workers sharply increased in the late 2000s.

Overall our results show favourable trends for Luxembourg nationals among whom both persistent and transitory inequality declined throughout the period. Foreign labour, especially cross-border workers, appeared to buffer macroeconomic fluctuations. The increase in the persistence of inequality can be related to the increase in the contribution of the financial and related services in employment—but the latter did not lead to any massive increase in overall inequality, unlike what has been documented in other countries. Labour market regulation indicators appear to be generally negatively correlated with both components of inequality and seem to have mitigated otherwise upward pressures on earnings inequality. Although it may be difficult to transpose the situation of Luxembourg to bigger economies, our results reveal the privileged situation of native workers and how foreign workers can buffer macroeconomic fluctuations. They also hint at the role of strict labour market regulations and collective bargaining institutions in holding back earnings inequality, at least in a period of fast economic growth and soaring demand for labour.

On the technical side, we specify and estimate a flexible error components model that captures rich dynamics and provides a better fit to our data than more standard models. We also show that top coding, which is often prevalent in both survey and administrative data sources, can be handled in the estimation of those variance components by using a multiple-imputation approach. Our analysis illustrates the usefulness of access to large-scale administrative registers for detailed analysis of inequality trends. The limited sample size and length of most panel surveys prevent detailed analysis within population subgroups and/or impose restrictions on the sophistication of variance components models that can be fitted and affect the reliability of inference (Doris *et al.*, 2013). In line with Dickens (2000), Kalwij and Alessie (2007), Baker and Solon (2003) or Gustavsson (2008), our model estimates bring evidence against simple restrictions concerning the life cycle and cohort variation in the two components of earnings dynamics and the relevance of these features when exploiting the covariance structure of earnings for inference regarding the evolution of permanent and transitory inequality.

Acknowledgements

This research is part of the 'Earnings dynamics and microsimulation' project cofunded by the 'Fonds National de la Recherche' Luxembourg (grant PDR893613) under the Marie Curie 'Actions of the European Commission'. It is also part of the 'Information and wage inequality: evidence on wage differences between natives, immigrants and cross-border workers in Luxembourg' supported by the 'Fonds National de la Recherche' Luxembourg (contract C10/LM/785657). We are grateful to the Inspection Générale de la Sécurité Sociale of Luxembourg for providing the anonymized extracts from social security registers and to Isabelle Debourges, Tom Dominique, Béatrice Lepêcheur, Thierry Mazoyer and Raymond Wagener for their invaluable support. Comments on earlier drafts by Lorenzo Cappellari, Alessio Fusco, Stephen Jenkins, Gary Solon, Don Williams and participants at the 2012 Luxembourg Organisation for Economic Co-operation and Development–Economic Development and Review Committee seminar on inequalities and social mobility, the 2013 Workshop on Household Finance and Consumption (Banque Centrale du Luxembourg) and the Population Association of America

2014, European Society of Population Economics 2014 and European Association of Labour Economists 2014 annual conferences are gratefully acknowledged.

References

Abowd, J. M. and Card, D. (1989) On the covariance structure of earnings and hours changes. *Econometrica*, **57**, 411–445.

Acemoglu, D. (2002) Technical change, inequality, and the labor market. *J. Econ. Lit.*, **40**, 7–72.

Alfons, A., Templ, M. and Filzmoser, P. (2013) Robust estimation of economic indicators from survey samples based on Pareto tail modelling. *Appl. Statist.*, **62**, 271–286.

Allegrezza, S., Heinrich, G. and Jesuit, D. (2004) Poverty and income inequality in Luxembourg and the Grande Région in comparative perspective. *Socio-Econ. Rev.*, **2**, 263–283.

Altonji, J. G. and Segal, L. M. (1996) Small sample bias in GMM estimation of covariance structures. *J. Bus. Econ. Statist.*, **14**, 353–367.

Amétépé, F. and Hartmann-Hirsch, C. (2011) An outstanding positioning of migrants and nationals: the case of Luxembourg. *Popln Rev.*, **50**, 195–217.

Annaert, J.-L. (2004) A bright spot in the heart of Europe: what can we learn from the Luxembourg success story? *Country Focus 1(15)*. European Commission's Directorate General for Economic and Financial Affairs, Brussels.

Atkinson, A. B. and Piketty, T. (eds) (2010) *Top Incomes: a Global Perspective*. Oxford: Oxford University Press.

Atkinson, A. B., Piketty, T. and Saez, E. (2011) Top incomes in the long run of history. *J. Econ. Lit.*, **49**, 3–71.

Autor, D. H. (2014) Skills, education, and the rise of earnings inequality among the "other 99 percent". *Science*, **344**, 843–851.

Baker, M. and Solon, G. (2003) Earnings dynamics and inequality among Canadian men, 1976-1992: evidence from longitudinal income tax records. *J. Lab. Econ.*, **21**, 289–321.

Bell, B. and Van Reenen, J. (2013) Extreme wage inequality: pay at the very top. *Am. Econ. Rev.*, **103**, 153–157.

Biewen, M. (2005) The covariance structure of East and West German incomes and its implications for the persistence of poverty and inequality. *Ger. Econ. Rev.*, **6**, 445–469.

Bingley, P., Cappellari, L. and Westergard-Nielsen, N. (2013) Unemployment insurance, wage dynamics and inequality over the life cycle. *Econ. J.*, **123**, 341–372.

Blundell, R., Graber, M. and Mogstad, M. (2014) Labor income dynamics and the insurance from taxes, transfers and the family. *Working Paper W14/01*. Institute for Fiscal Studies, London.

Blundell, R. and Preston, I. (1996) Income, expenditure and the living standards of UK households. *Fisc. Stud.*, **16**, 40–54.

Blundell, R. and Preston, I. (1998) Consumption inequality and income uncertainty. *Q. J. Econ.*, **113**, 603–640.

Bönke, T., Giesecke, M. and Lüthen, H. (2015) The dynamics of earnings in Germany: evidence from social security records. *Manuscript*. Free University of Berlin, Berlin.

Cappellari, L. (2004) The dynamics and inequality of Italian men's earnings. *J. Hum. Resour.*, **39**, 475–499.

Cappellari, L. and Leonardi, M. (2013) Earnings instability and tenure. *Working Paper 4145*. CESifo, Munich.

Chakravarty, S. R. (2001) The variance as a subgroup decomposable measure of inequality. *Socl Indic. Res.*, **53**, 79–95.

Chamberlain, G. (1984) Panel data. In *Handbook of Econometrics*, vol. 2 (eds Z. Griliches and M. D. Intriligator), ch. 22, pp. 1247–1318. Amsterdam: Elsevier.

Choe, C. and Van Kerm, P. (2014) Foreign workers and the wage distribution: where do they fit in? *Working Paper 2014-02*. Centre d'Etudes de Population, de Pauvreté et de Politique Socio Economiques, Esch-sur-Alzette.

Clark, T. E. (1996) Small sample properties of estimators of non-linear models of covariance structure. *J. Bus. Econ. Statist.*, **14**, 367–373.

DeBacker, J., Heim, B., Panousi, V., Ramnath, S. and Vidangos, I. (2013) Rising inequality: transitory or persistent?; New evidence from a panel of U.S. tax returns. *Brook. Pap. Econ. Activ.*, **46**, 67–142.

Dickens, R. (2000) The evolution of individual male earnings in Great Britain: 1975-95. *Econ. J.*, **110**, 27–49.

Doris, A., O'Neill, D. and Sweetman, O. (2013) Identification of the covariance structure of earnings using the GMM estimator. *J. Econ. Inequal.*, **11**, 343–372.

Freeman, R. B. and Katz, L. F. (1994) Rising wage inequality: the United States vs. other advanced countries. In *Working under Different Rules* (ed. R. B. Freeman), pp. 29–62. New York: Russell Sage Foundation.

Fusco, A., Van Kerm, P., Alieva, A., Bellani, L., Etienne-Robert, F., Guio, A.-C., Kyzyma, I., Leduc, K., Liégeois, P., Pi Alperin, M. N., Reinstadler, A., Sierminska, E., Sologon, D., Thill, P., Valentova, M. and Voicu, B. (2014) Luxembourg: has inequality grown enough to matter? In *Changing Inequalities and Societal Impacts in Rich Countries: Thirty Countries Experiences* (eds B. Nolan, N. B. Salverda, W. Wiemer, D. Checchi, I. Marx, A. McKnight, I. Tóth and H. Van de Werfhorst). Oxford: Oxford University Press.

Godechot, O. (2012) Is finance responsible for the rise in wage inequality in France? *Socio-Econ. Rev.*, **10**, 447–470.

Gottschalk, P. and Moffitt, R. (1994) The growth of earnings instability in the U.S. labour market. *Brook. Pap. Econ. Activ.*, no. 2, 217–254.

Gustavsson, M. (2007) The 1990s rise in Swedish earnings inequality—persistent or transitory? *Appl. Econ.*, **39**, 25–30.
Gustavsson, M. (2008) A new picture of Swedish earnings inequality: persistent and transitory components, 1960-1990. *Rev. Incm. Wlth*, **54**, 324–349.
Haider, S. (2001) Earnings instability and earnings inequality of males in the United States: 1967-1991. *J. Lab. Econ.*, **19**, 799–836.
Jäntti, M. and Jenkins, S. P. (2015) Income mobility. In *Handbook of Income Distribution*, vol. 2 (eds A. B. Atkinson and F. Bourguignon), ch. 10, pp. 807–935. Amsterdam: North-Holland.
Jaumotte, F., Lall, S. and Papageorgiou, C. (2013) Rising income inequality. *IMF Econ. Rev.*, **61**, 271–309.
Jenkins, S. P., Burkhauser, R. V., Feng, S. and Larrimore, J. (2011) Measuring inequality using censored data: a multiple-imputation approach. *J. R. Statist. Soc.* A, **174**, 63–81.
Kalwij, A. and Alessie, R. (2007) Permanent and transitory wage inequality of British men, 1975–2001: year, age and cohort effect. *J. Appl. Econ.*, **22**, 1063–1093.
Kaplan, S. N. and Rauh, J. (2010) Wall Street and Main Street: what contributes to the rise in the highest incomes? *Rev. Finan. Stud.*, **23**, 1004–1050.
Kopczuk, W., Saez, E. and Song, J. (2010) Earnings inequality and mobility in the United States: evidence from social security data since 1937. *Q. J. Econ.*, **125**, 91–128.
Lillard, L. A. and Willis, R. J. (1978) Dynamic aspects of earnings mobility. *Econometrica*, **46**, 985–1012.
Lin, K.-H. (2015) The financial premium in the US labor market: a distributional analysis. *Socl Forces*, **94**, 1–30.
MaCurdy, T. E. (1982) The use of time series processes to model the error structure of earnings in a longitudinal data analysis. *J. Econmetr.*, **18**, 83–114.
Meghir, C. and Pistaferri, L. (2011) Earnings, consumption, and life cycle choices. In *Handbook of Labor Economics*, vol. 4B (eds O. Ashenfelter and D. Card). Amsterdam: Elsevier.
Mincer, J. (1974) *Schooling, Experience and Earnings*. New York: National Bureau of Economic Research.
Moffitt, R. and Gottschalk, P. (2002) Trends in the transitory variance of earnings in the United States. *Econ. J.* **112**, C68–C73.
Moffitt, R. and Gottschalk, P. (2012) Trends in the transitory variance of male earnings: methods and evidence. *J. Hum. Res.*, **47**, 204–236.
Organisation for Economic Co-operation and Development (2010) *OECD Economic Survey of Luxembourg*. Paris: Organisation for Economic Co-operation and Development.
Organisation for Economic Co-operation and Development (2011) *Divided We Stand: Why Inequality Keeps Rising*. Paris: Organisation for Economic Co-operation and Development.
Organisation for Economic Co-operation and Development (2012) *Economic Policy Reforms 2012—Going for Growth*. Paris: Organisation for Economic Co-operation and Development.
Ostrovsky, Y. (2010) Long-run earnings inequality and earnings instability among Canadian men revisited, 1985-2005. *J. Econ. Anal. Poly*, **10**, 1–34.
Ramos, X. (2003) The covariance structure of earnings in Great Britain, 1991-1999. *Economica*, **70**, 353–374.
Reiter, J. P. (2003) Inference for partially synthetic, public use microdata sets. *Surv. Methodol.*, **29**, 181–188.
Shorrocks, A. F. (1984) Inequality decomposition by population subgroups. *Econometrica*, **52**, 1369–1385.
Sologon, D. M. and O'Donoghue, C. (2012) Cross-national trends in permanent earnings inequality and earnings instability in Europe 1994–2001. *Working Paper 2012-006*. United Nations University–Maastricht Economic and Social Research Institute on Innovation and Technology, Maastricht University, Maastricht.

Supporting information
Additional 'supporting information' may be found in the on-line version of this article:

'On-line supporting information: Modelling earnings dynamics and inequality: foreign workers and inequality trends in Luxembourg, 1988–2009'.