Earnings Inequality Trends in the United States: Nationally Representative Estimates from Longitudinally Linked Employer-Employee Data

John M. Abowd	Department of Economics, Cornell University, and NBER
	john.abowd@cornell.edu
Kevin L. McKinney	Center for Economic Studies, U.S. Census Bureau
	kevinm@ccrdc.ucla.edu
Nellie Zhao	Department of Economics, Cornell University
	nlz6@cornell.edu

November 2015

Corrected conference draft. Please contact the authors for final draft before citing.

1 Introduction

The theme of the papers in this conference is the analysis of the role of the employer in earnings inequality trends. Administrative data are used in virually all analyses of the employer effects on earnings inequality. We begin our paper with an analysis of the relations among inequality measures based on frames constructed from administrative data as compared to those based on conventional household frames. This comparison is instructive for three reasons. First, it illuminates quite clearly the limitations of using statutory employment as the primary earnings indicator when dealing with low-income or marginally-attached components of the labor force. Second, it suggests that attempts to attribute components of earnings inequality to employers among the low-income earners can be easily confounded by factors that American administrative earnings data based on statutory employment definitions cannot easily address. Third, it confirms the better-understood phenomenon that administrative data indicate greater earnings inequality for higher-income individuals than household survey data.

Our analysis suggests that measures based on administrative data will generally show more earnings inequality than household survey data in recessions for three reasons. First, the top quantiles of the distribution will be greater in the administrative data. These are probably more accurate and certainly more business-cycle resistant. Second, the bottom quantiles will be lower in the administrative data, especially if the zeroes are properly counted. These are likely less accurate, especially the zeroes, because the survey data indicate income that most likely did not come from statutory employment.

The plan of our paper is as follows. Section 2 presents our motivation and the background for clarifying the measurement of earnings and wage inequality using administrative data. Section 3 discusses the Unemployment Insurance (UI) based individual- and job-level data we use in our analysis. Section 4 presents the trends in earnings inequality based on the different national frames we constructed. In Section 5 we compare the trends using national worker frames based on UI with those found in the Current Population Survey and American Community Survey. In Section 6 we switch to a frame based on all private non-agricultural employers. Section 7 analyzes inequality trends based on our establishment frame. Section 8 concludes.

2 Background and Motivation

For the purposes of studying earnings inequality considering only the wage rate, and not the amount of time a person actually works, is seriously incomplete–especially in the U.S. where there is very little statutory employment security except in the public sector. An important determinant of increased earnings inequality in our analyses is changes in labor force attachment (weeks worked during the year, hours worked per week), which is incompletely measured in American administrative earnings data, whether based on the state Unemployment Insurance system or the federal tax system (W-2 data).

The administrative records database developed at the Census Bureau has basic demographic information on all officially issued SSNs as of a known date (2012 for this version of the paper). Internally, all SSN format values, valid or not, are encrypted as another nine digit number commonly referred to as a PIK. Using the list of PIKs associated with an officially issued SSN a researcher can tell if an earnings record is associated with a valid SSN or not associated with a valid SSN. Records not associated with a valid SSN may be associated with an alternative official person identifier such as an IRS issued TIN, although the researcher can only differentiate between valid and not valid SSNs. If the PIK is associated with a valid SSN we can access the demographic details and employment history associated with the underlying SSN. If the PIK is not associated with a valid SSN then we only have access to an employment history.

In our analysis, there are two kinds of suspect PIKs: ones that are not valid (this means that the employer reported earnings in a state's UI system for an SSN that was never issued) and ones associated with demographic characteristics and an employment history that imply it is unlikely the owner of the SSN used it for some or all of the reported earnings history (leading case: the SSN was issued to a person who was less than 10 years old in the year during which the SSN was used to report UI eligible earnings).

For our anlysis we use two frames: the first frame (all workers) contains all earnings records reported on the UI data, while the second frame attempts to reconstruct the worker-eligible population (18-70) for each year from 1990-2013. For the first frame, our working hypothesis is: (1) the use of a PIK not associated with a valid SSN reflects the work of a single undocumented immigrant and (2) a PIK associated with a valid SSN active when the individual appears legally

too young (or too old) to work also represents only one person in the population. For the second frame (worker-eligible) we restrict the sample to only PIKs associated with a valid SSN. Unlike the first frame, this frame includes periods of observed inactivity. A worker is elgible to work in a given year if they are age 18-70, not reported dead, the SSN is active, and the PIK is not associated with more than 12 employers. If the previous conditions are met then a worker receives a zero for the year unless positive earnings are found for that PIK from at least one of the available state UI earnings files.

Thus, we use related frames to study individual earnings inequality. The "all workers" frame treats every SSN that appears in the administrative record database as a unique individual, but makes no assumption that the inidividual had zero earnings when a record does not appear. The "work-eligible" frame uses only those SSNs that meet the criteria above, but assumes that observed earnings zeroes are true zeroes and includes them in the inequality analysis. We compare estimates based on both of these frames to those derived from surveys with conventional household frames.

To study the role of the employer as a source of earnings inequality, we use the statistical technique introduced to the labor economics literature in Abowd et al. (1999, AKM, hereafter). When this technique has been used in Europe, for example in Card et al. (2013, CHK,hereafter), the biggest contributor to the increase in earnings inequality appears to be increased employer-level heterogeneity (called the firm effect in AKM). Using the same administrative data as we used for the individuaal analysis combined with administrative data for the universe of private non-agricultural employers for 1990-2013, we show that, in the U.S, the CHK result does not hold. There has been very little change in employer-level earnings heterogeneity in the U.S. when one compares wage measures similar to the ones used to analyze the European data. European administrative databases allow one to approximate a wage rate directly, usually, the amount that would be earned if an individual worked full-time full-year. The American data do not directly allow that. We use a statistical adjustment to the earnings data to approximate the full-year full-time wage rate. We show the contribution of this component to the earnings inequality separately. After making this correction, it is clear that the employer contribution to wage rate inequality has not changed over our time period. The labor force attachment component does contribute to changes in earnings inequality. But the largest driver of the change in earnings inequality is the

covariance component, which captures interactions between the different AKM effects.

3 Individual- and Job-level Data

In this section we describe our "all workers" frame. This frame includes data for every unique person identifier and every job in the LEHD infrastructure from 1990 (quarter 1) through 2013 (quarter 4). In this system a "job" is the statuory employment of a worker by a statutory employer as part of the Unemployment Insurance (UI) system in a given state. Mandated reporting of UI-covered wage and salary payments between one statutory employer and one statutory employee is governed by the state's UI system. Reporting covers private employers and state and local government. There are no self-employment earnings unless the proprietor drew a salary, which is indistinguishable from other employees in this case. Federal data, included from 2000 (quarter 1) forward, were supplied separately by the Office of Personnel Management (OPM). The federal data were edited in the LEHD infrastructure processing to produce records that contain quarterly earnings reports comparable to those reported directly in the UI wage and salary payments. When we say "UI-covered" employment, we mean "statutory" employment as defined by the UI system or a statutory federal employee. (See Abowd et al. 2009 for a detailed summary of the construction of the LEHD infrastructure.)

States and the federal government joined the federal/state partnership that supplies input data to the LEHD program at different dates. When a state or the federal government joined, the data custodians were asked to produce historical data for as many quarters in the past, back to 1990 (quarter 1) as could be reasonably recovered from their information storage systems. As a result, the date that a data-supplying entity joined the system is not the same as the first quarter in which that entity's data appear in the system. The start date for any state or the federal government depends primarily on the amount of historical data the state or federal government could recover at the time it joined. This ignorability (in the sense of Rubin 1987 or Imbens and Rubin 2015) is the basis for our methods of constructing nationally representative estimates back to the early 1990s. (See Abowd and Vilhuber 2011 for a description of how these modeling assumptions were used to construct national gross worker and job flow estimates.)

In this paper, we present the basic evidence on earnings inequality for workers using unadjusted that we analyze for four regimes that correspond to different epochs of data availability in the system:

- Regime 1: 1990-2013 [19% of QCEW] (AK CO MD ID IL IN KS LA MO WA WI);
- Regime 2: 1995-2013 [68%] (+ AZ CA FL GA MN NC NY OR MT PA RI SD TX WY);
- Regime 3: 1998-2013 [86%] (+ CT HI KY MI ND NJ NM NV SC TN VA WV);
- Regime 4: 2004-2013 [100%] (+ AL AR DC DE IA MA MS NE NH OH OK UT VT).

The coverage of each sampling regime is indicated by the percentage of all private employment reported in the BLS's Quarterly Census of Employment and Wages (2011, quarter 1) that is represented by the states in that regime. The construction of the different sampling regimes is shown graphicaly in Figure 1.

We define annual earnings (and ln earnings) for a job as the sum of all reported payments for all four quarters in a given year for the worker-employer pair that make up the job. Annual worker earnings are the sum of all job earnings reported for that person in the indicated year. All earnings measures are in year 2000 dollars using the headline Consumer Price Index.

4 Inequality Trends

4.0.1 Inequality Measures from the "All Workers" Frame

Figure 2 shows the 10^{th} through 95^{th} percentiles of the earnings distribution for workers in each of the four regimes summing over all jobs in the LEHD infrastructure for that worker in the indicated year. For the 10^{th}, 25^{th} and 50^{th} percentiles, the sampling regimes are indistinguishable. Above the median, and especially in the 95^{th} percentile, the inclusion of CA and NY data in regime 2 increases the estimated quantile value. Overall the differences are relatively minor, and the trends are very similar across regimes. State entry does not appear to significantly affect earnings inequality measures, especially after 1994. This is very strong evidence that the date of entry of a

state into the LEHD infrastructure can be modeled as ignorably missing (Rubin 1987; Imbens and Rubin 2015), as we do in Section 6.

Figure 3 plots the time series for the 90-10 ratio of the earnings distribution for all workers summing over all jobs in the database by sampling regime. Once again, the entry of CA and NY affects the level of the regime 2 data, but the overall trends are very similar. As shown in Figure 2, the LEHD data do have a large number of workers with very low earnings. Ten percent of the workers have earnings of less than $\$2,000$. Real earnings growth in dollars is higher above the median. Relative real earnings growth is higher below the median, implying that the 90-10 ratio is falling. Reconciling this characteristic of the LEHD data with household survey estimates is what we address next.

4.1 Defining the "Eligible Workers" Frame

We now consider an alternative individual frame designed to estimate the population of individuals at risk to have reported earnings in a given year. LEHD data contains earnings records for workers both with and without a valid SSN. Our "eligible workers" frame is restricted to persons with a valid SSN as reflected on the Census Bureau's edited version of the Social Security Administration's master SSN database, known as the "Numident".˙We include an individual in the "eligible worker" frame when that person was between the ages of 18 and 70 at some sometime during 1990-2013. For an "eligible worker" we use every year of data in which that individual was between the ages of 18 and 70, the year was greater than or equal to the SSN year of issue, less than or equal to the year of death (if available) and the SSN was associated with less than 12 jobs. A person is labeled as "active" in the labor market in a given year when positive UI earnings are found for that year and "inactive" otherwise. We do not use zero earnings for eligible workers in inactive years except when studying the employment-to-population ratio. The valid SSN, age-range, SSN active, date of death, and number of jobs restrictions removed about 7% of the worker-year records in the "all workers" frame from the "eligible workers" frame.

The purpose of creating the "eligible workers" frame is to include individuals for whom the available data suggest that the SSN is being used by its owner for UI wage record reporting purposes. The reuse of SSNs not issued to the person using the SSN has been documented and studied

by Brown et al. (2013) and others. We identify the two suspect uses of SSNs discussed in Section 2 in creating our "eligible workers" frame: invalid SSNs–PIKs that do not match to the Census Numident–and reused SSNs–numbers apparently being used by multiple persons and/or those for whom the age of the person issued the SSNs is inconsistent with labor market activity. We call these records "immigrant candidates." Figure 4 shows the count of earnings records excluded from the "eligible workers" frame in each year. It also indicates the start dates for the sampling regimes. Even accounting for the increasing coverage of each sampling regime, there is a clear, strong upward trend in the number of immigrant candidate records until the great recession, and a strong downward trend thereafter. Because there are data in the system for each of these records, the associated SSN represents at least one worker, but may represent many. A great deal of supplemental research would be required to estimate the relation between how many jobs are reported for an SSN in the immigrant candidate records and the number of individuals employed. Even more research would be required to estimate their characteristics. We do not attempt this research here. We note that when we use the "All Workers" frame, each SSN counts as one individual but there is no adjustment for being "inactive;" that is, we do not assume zero earnings when the SSN is age-eligible but has no reported earnings. When we use the "Eligible Workers" frame, all records for immigrant candidates are excluded and eligible inactive workers are assumed to have zero earnings when there is no reported activity in any job.

Figure 5 shows the percentiles of the earnings distribution for the eligible workers frame. By comparison to Figure 2 and Appendix Figure 24 it is clear that the main consequence of shifting the frame from all workers to eligible workers is to increase the quantile values at the bottom of the earnings distribution. [Insert some discussion of values.]

Figure 6 shows selected inequality measures relative to 2000 for all workers by year. Figure 7 shows the same measures for the eligible worker frame. The trends differ when using the all workers frame as compared to using only eligible jobs for the eligible workers frame. Except for the 99-1 ratio, earnings inequality is relatively stable after 1998 when using the all workers frame. Earnings inequality increases after 2000 when using only eligible jobs for "eligible workers".

5 Comparisons to the Current Population Survey and American Community Survey

5.1 Using Survey Data to Construct Estimates from Household Frames

In order to better understand the differences between administrative record-based analyses of earnings inequality and the more familiar household-based estimates. We constructed a time series of household inequality measures from the Current Population Survey and American Community Survey files that have been linked to the LEHD data. From each of these surveys we used the following records:

- Current Population Survey-Annual Social and Economic Supplement (CPS-ASEC): all persons with positive earnings from 1989-2003;
- American Community Survey (ACS): all persons with positive earnings from 2000-2013.

In the overlapping years (2000-2003), we interpolated the CPS-ASEC and ACS data series are interpolated between 2000-2003 using the method in Abowd and Vilhuber (2011). In the CPS-ASEC the respondent reports for the previous calendar year, and that is the dating we have used. In the ACS, the respondent reports earnings for the past 12 months, and the survey is in the field continuously throughout the calendar year. We dated the earnings with the calendar year that contained the majority of the months covered by the response with ties going to the more recent year. We designated a survey worker as a "covered worker" if their employment should have appeared in the LEHD adminstrative record data according to their responses on the survey. To compute real earnings we used the same CPI measure as used for the LEHD data. For the linked LEHD data we used annual real wage and salary earnings for the "eligible worker" sample.

Figure 8 shows the ratio of the percentiles of the earnings distribution measured using the LEHD eligible workers frame for workers with positive earnings in the year to the same percentiles from combined CPS/ACS estimates. In the UI data for individuals below the median, real earnings grew prior to 2000 but changed little post 2000 except for the dip after the great recession. In contrast, CPS/ACS covered workers below the median show a larger pre 2000 increase, but give

up all of the gains by 2013. In the administrative data for individuals above the median, there is an almost continuous increase in real earnings with higher rates for higher percentiles. The administrative data also show faster earnings growth relative to the CPS/ACS for the 95th percentile after the great recession. At the same point in the earnings distribution ACS/CPS covered workers earn about $\$4,000$ more than those in the UI data. This implies that the adminstrative UI data have more low earning jobs. For example, the 20^{th} percentile earnings in the UI date are similar to the 10^{th} percentile on the CPS/ACS.

Figure 9 shows that the estimated employment to population ratio is higher in the CPS/ACS, but converges to the UI value by 2009 when comparing UI-covered workers in both data sources. All series show a large drop in the employment to population ratio during the great recession. Employment to population ratio is higher on the CPS/ACS covered workers sample relative to the UI. Although employment levels have recovered, a significant proportion of workers employed prior to the great recession still do not have a UI covered job in 2013.

Figures 10, 11 and 12 show the 95-5, 90-10 and 80-20 inequality ratios for each household and UI eligible worker frames, respectively. In these figures the label "SA" refers to the estimates from the combined CPS/ACS for all workers with positive earnings and the label "SC" uses only the covered workers in the household samples. Inequality ratios are much higher on the UI data compared to the CPS/ACS. This result is primarily due to the relatively large number of low-earning workers on the UI. Trends are similar. All series show pre 2000 decrease with a subsequent increase post 2000. UI data show a relatively large increase in inequality during the great recession.

5.2 Matching the UI and ACS Records

The LEHD UI records and the individual responses in the ACS have been linked by the Census Bureau. The person identifier, an encrypted SSN called a PIK, is attached to ACS records using birth date, sex, name, and geography. Over 93% of ACS individual records have an associated PIK. We matched ACS records to the UI for 2005-2013. For these years, both the ACS and LEHD (as shown in sampling regime 4) are fully national. Match results are shown in Figure 13.

Inequality measures typically focus only on persons with positive earnings. Combining the Census Numident with the UI data allows us to measure persons that are eligible to work but have

no reported UI earnings. Shocks to the labor market or structural changes may also lead to a change in the employment to population ratio.

Figure 14 analyzes the 5^{th} percentile of the earnings distribution. In the records that appear in the ACS only, the ACS estimate is greater than the similar point on the UI earnings distribution (approximately $\$2,500$ v. $\$1,000$). For the records that appear on both the ACS and UI, the ACS estimate is greater than the UI estimate by approximately $\$500$. There are also noticable differences in the trends.

Figure 15 analyzes the $95th$ percentiles of the earnings distribution for the same matched samples. These look more similar, but there are still some differences. The difference between the ACS and UI estimates of this quantile are apparently due to lower earnings in the ACS-only group than either measure in the both group.

6 Establishment Job Data

6.1 Defining an Employer Frame

In the next two sections we focus on the trends in earnings inequality as measured at private non-agricultural firms. While this has been one of the usual frames for analyzing employer contributions to earnings inequality, it has not always been possible to separate the components of earnings inequality and maintain a frame that is representive of employment in private firms. We attempt to do this in our analysis by using the Census Bureau's Employer Business Register (BR), which is the frame for the economic censuses, the Longitudinal Business Database (LBD), and the Business Dynamics Statistics (BDS). While there are small differences in the use of the BR for these three activities, we have used the definition of an active establishment that underlies the LBD and BDS as our frame.

We use all establishments with positive annual payroll and employment in the Business Register that also appear on the LBD for the economic census years: 1992, 1997, 2002, 2007, 2012. We did final edits using the 2013 LBD to maximize comparability with the Business Dynamics Statistics frame. We only used variables from the economic census years. Those variables were employment (LBD version), NAICS, and geography (state and county). We make a very minor

adjustment to final weights so that employment totals in our data always match those on the published BDS data for the economic census years. We used the economic census responses reported on the BR for total revenue. Real total revenue was computed using the same CPI as we used in the earning analysis. When total revenue was missing, this indicated that an establishment was not sampled into the economic census. We muliply imputed total revenue for these establishments using a model based on the factors that affect the sampling rate in the economic census: establishment employment, industry and geography. We adjusted total revenue for estimated materials costs using sector-level data from the Bureau of Economic Analysis.

6.1.1 Linking Employer-level LEHD Measures to the Establishment Frame

We linked the BR establishments to the LEHD Employer Characteristics File using the methodology from Abowd and Zheng (2014), which links UI-account level information in LEHD (SEIN) to all establishments associated with that UI account in the BR.

We fit the AKM decomposition, with modifications described below, then compute all components of the AKM decomposition for real annual ln earnings aggregated to the SEIN level. These are linked directly to each establishment in the BR (all parts of the mean and variance decomposition). Any establishment that is missing data had all missing components of the AKM decomposition completed using an ignorable multiple-imputation missing data model estimated separately for each economic census year.

We fit a standard AKM decomposition to the all dominant jobs in the "all workers" frame 1990-2013, restricted to individuals ages 18-70 in the estimation year. The AKM decomposition was enhanced in two important ways. First, we added an extensive set of controls for labor force attachment. Some of these controls are available for every job-year–those that are based in which quarters during the year the individual had observed earnings–and some of the controls were imputed from linkages to the CPS and Census 2000 and ACS–weeks worked/year and usual weekly hours. Second, following CHK, we used five-year windows to fit separate employer effects. These windows were 1990-1994, 1995-1999, 2000-2004, 2005-2009 and 2010-2013. Except for the last window, which is only for years wide, each window is centered on an economic census year.

Let y_{ijt} be individual i's earnings at employer j in year t. We decompose the variance of

ln earnings, $\mathrm{Var}_t\left[y_{ijt}\right]$, into a between firm and a within firm component, which we estimate separately for $t = 1992, 1997, 2002, 2007,$ and 2012 using all the data that occurs in the 1990-1994 window to estimate the 1992 variance, and so forth through 2012.

The between variance is based on $\mu_{jt} - \mu_t$, where $\mu_{jt} = \mathrm{E}\left[y_{ijt}\right]$. The within variance is based on $y_{ijt} - \mu_{jt}$. Both components of the variance decomposition are measured at the employer level. Using the estimated AKM model, we further decomposed the within and between variances into components representing the worker effect, firm effect, labor-force attachment effect, labor-force experience effect, residual, and the covariance contribution.

7 Inequality Analysis among All Private Employers

Figure 16 presents the main result from the ln earnings variance decomposition for each period. The total variance of ln earnings for private non-agricultural businesses declined from 1992 to 2002, rose in 2007, then declined again in 2012. Each of these summaries should be interpreted as applying to the five-year window surrounding the indicated economic census year. The contribution of worker effect has been declining over time. The contribution of the firm effect declines to 2002 then rises through 2012, but its variation is small compared to some other contributions. The contribution of the labor-force attachment effect declined through 2007 and is approximately the same in 2012 as it was in 2007. The contribution of the residual variance also declined through 2007 and has approximately the same value in 2012 as it had in 2007. The covariance term, which captures contributions from the interactions of all the components, increased dramatically from 1992 to 2007, and continues to increase in 2012.

Figure 17 decomposes just the between-employer components of the variance of ln earnings. Differences in the firm-average worker effect and labor-force experience effect contribute relatively little to the between variance. The firm effect, differences in the firm-average labor-force attachment effect and the covariance component are the major contributors to the between firm component of variance. The very small contribution of the residual, as opposed to its expected zero contribution, occurs because of the difference between the establishment frame used in this analysis and the job frame used for estimation. As can be seen in the figure, the firm effect makes its

largest contribution at the two end points (1992 and 2012). The covariance contribution increases monotone from 1992 to 2012.

Figure 18 decomposes just the within-employer components of the variance of ln earnings. The worker effect, labor-force attachment effect and residual all make large but declining contributions that are about the same in 2007 as in 2012. While the covariance contribution is increacing over time, it is relatively small even in 2012. There is virtually no change in the contribution of the labor force experience effect over time. The firm effect cannot contribute to the within-firm variance by construction.

Figures 19 to 23 present the total variance decomposition stratified by the establishments position in the revenue per worker (adjusted for material costs) distribution. The establishment-level distribution of material-cost-adjusted revenue per worker is divided into quintiles in 2002. The real value of the 2002 quintiles in 1992, 1997, 2007, and 2012 is used to stratify the establishments in those years.

As Figure 19 shows, the biggest contributors to ln earnings variation are the lowest revenue/worker employers. But we are concerned here with the contribution of the firm effect, which declines monotonically as we move from the bottom of the revenue/worker distribution (quintile 1) to the top (quintile 5). This result is surprising since the firms at the top of the revenue/worker distribution have the most scope to induce variance in the earnings distribution.

8 Conclusions

State entry year to LEHD does not materially affect the trends in earnings inequality, but LEHD data contain a relatively large number of workers with very low earnings. Using all jobs, earnings inequality for those employed is shown to have declined over time (declining prior to 1998, relatively stable after this period). Earnings inequality has increased for "eligible" workers–those whose SSNs are valid and appear to be used by the person to whom they were issued. Increases in inequality are larger during the great recession when you include "eligible" workers with zero earnings. Using the our establishment frame and all LEHD jobs, the between-establishment component declines and then increases.The within-establishment component declines continuously.

The firm effect component is relatively stable. Firms with greater revenue per worker contribute less than firms with lower revenue per worker to total earnings variance.

Acknowledgements

We have benefitted from discussion with David Card, Patrick Kline, Francis Kramarz, and Lars Vilhuber.

Abowd acknowledges direct support from NSF Grants SES-0339191, CNS-0627680, SES-0922005, TC-1012593, and SES-1131848. This paper was written while the first author was Distinguished Senior Research Fellow and the third author was Pathways Intern at the U.S. Census Bureau. Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed. This research uses data from the Census Bureau's Longitudinal Employer-Household Dynamics Program, which was partially supported by the following National Science Foundation Grants SES-9978093, SES-0339191 and ITR-0427889; National Institute on Aging Grant AG018854; and grants from the Alfred P. Sloan Foundation. An archive of the code used to prepare analysis samples and conduct estimation along with the released results is available directly from http://digitalcommons.ilr.cornell.edu/ldi/nn.

Bibliography

Abowd, J. M., Kramarz, F. and Margolis, D. N. (1999). High wage workers and high wage firms, *Econometrica* **67**(2): 251–333.

Abowd, J. M. and Vilhuber, L. (2011). National estimates of gross employment and job flows from the Quarterly Workforce Indicators with demographic and industry controls, *Journal of Econometrics* **161**: 82–99.

Brown, J. D., Hotchkiss, J. L. and Quispe-Agnoli, M. (2013). Does employing undocumented workers give firms a competitive advantage?, *Journal of Regional Science* **53**: 158–170.

Card, D., Heining, J. and Kline, P. (2013). Workplace heterogeneity and the rise of West German wage inequality, *The Quarterly Journal of Economics* **128**(3): 967–1015.

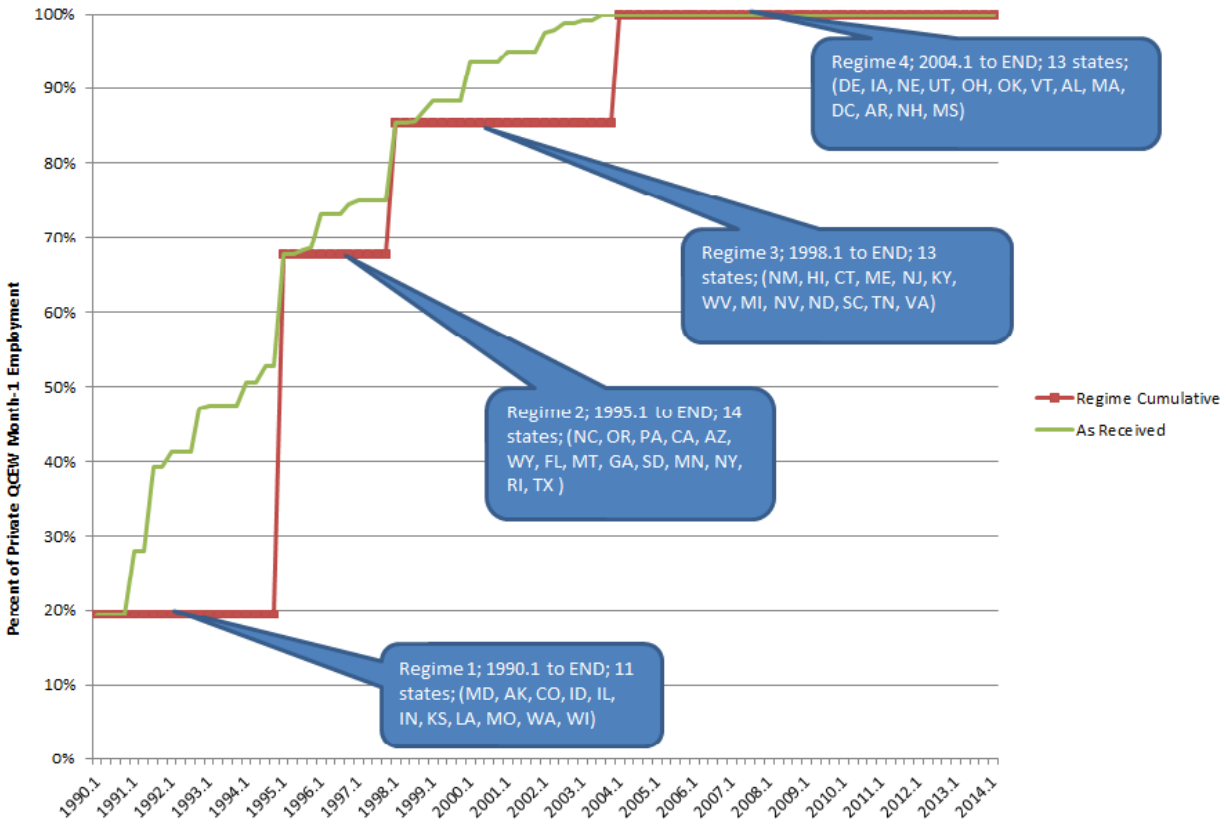


Figure 1: Years covered by each of the four sampling regimes

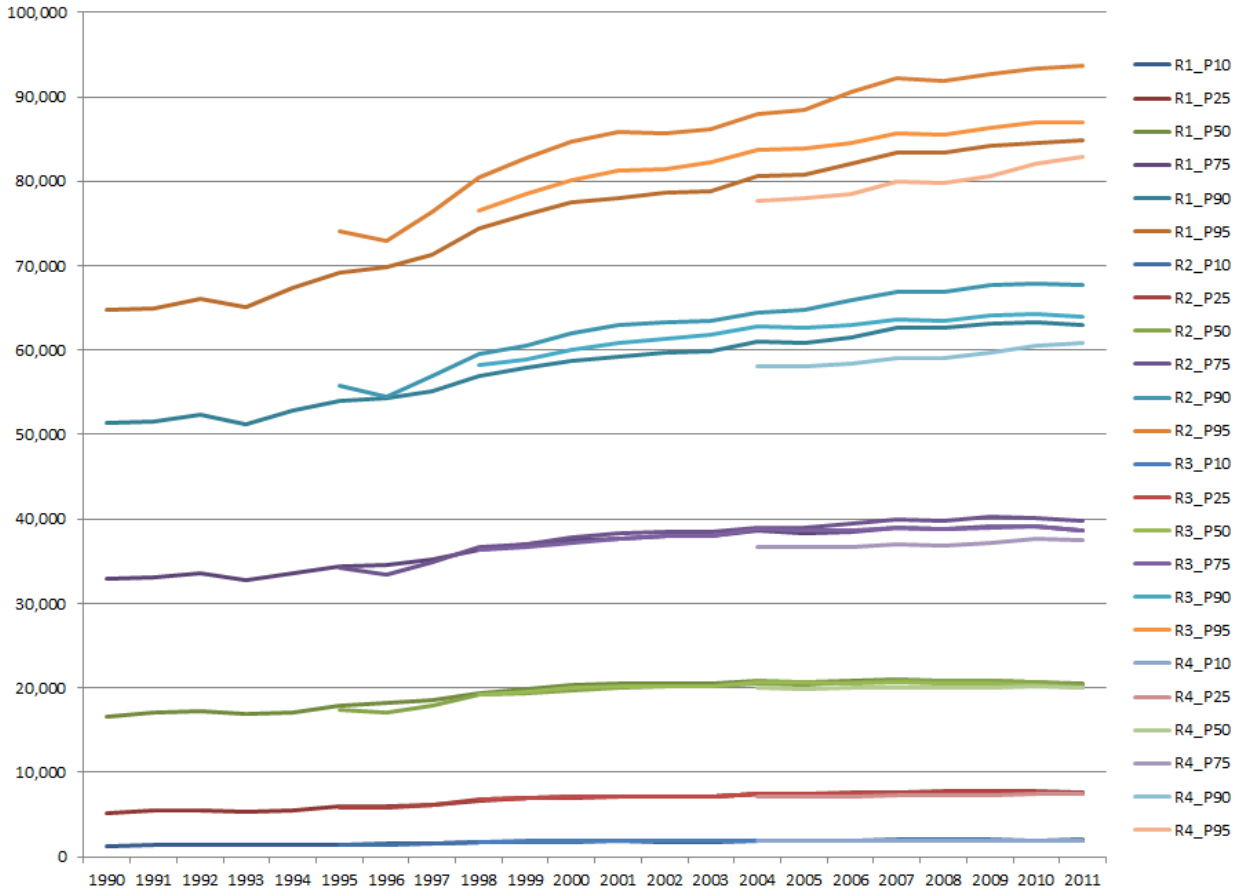


Figure 2: Percentiles of the Earnings Distribution for All Workers by Sampling Regime and Year

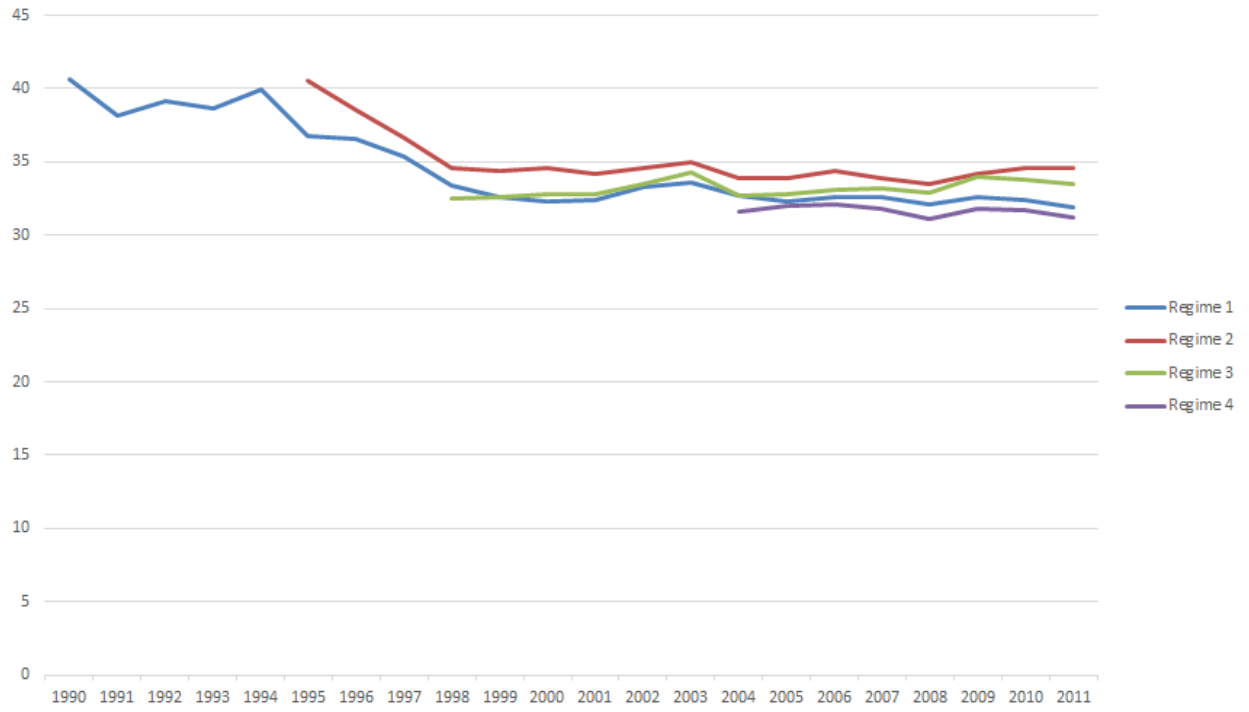


Figure 3: Ratio of the 90th to the 10th Percentile of the Earnings Distribution for All Workers by Sampling Regime and Year

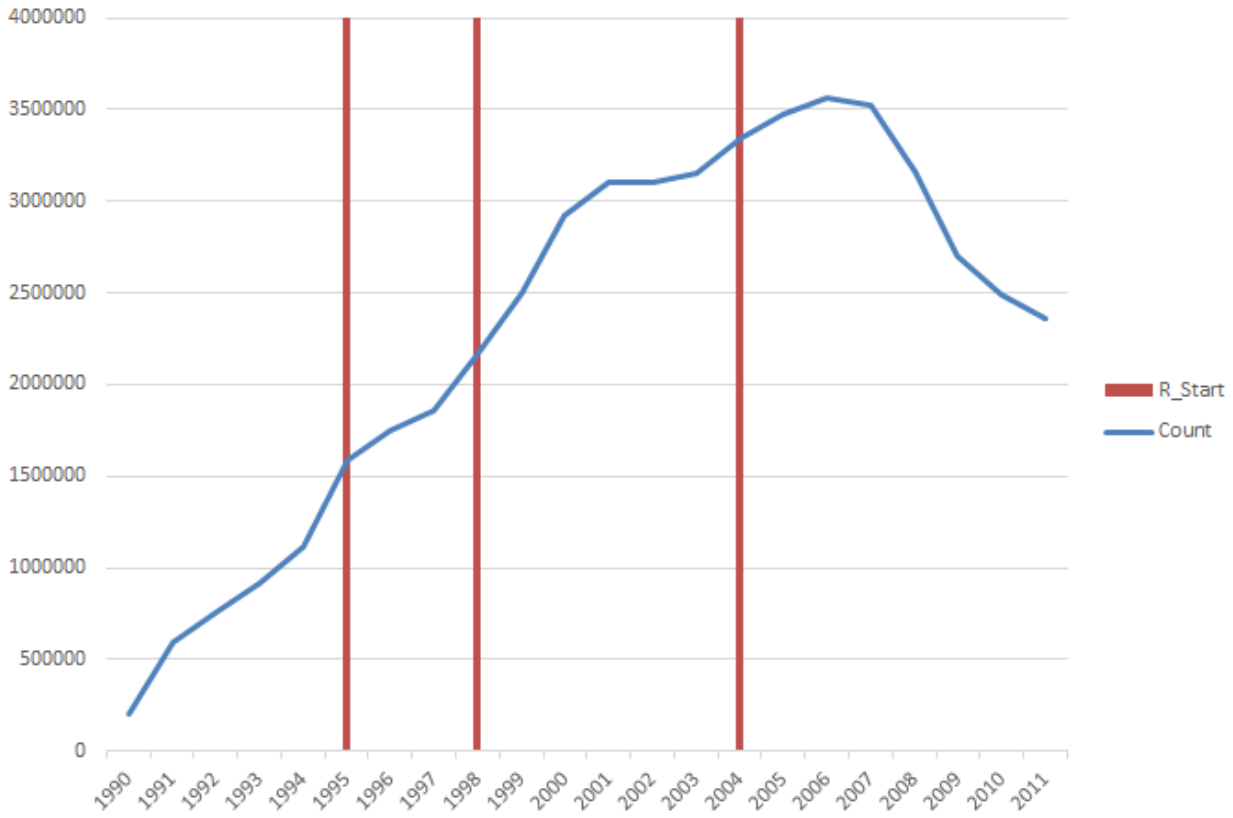


Figure 4: Earnings Records Not Included in Eligible Workers Analysis: Immigrant Candidates

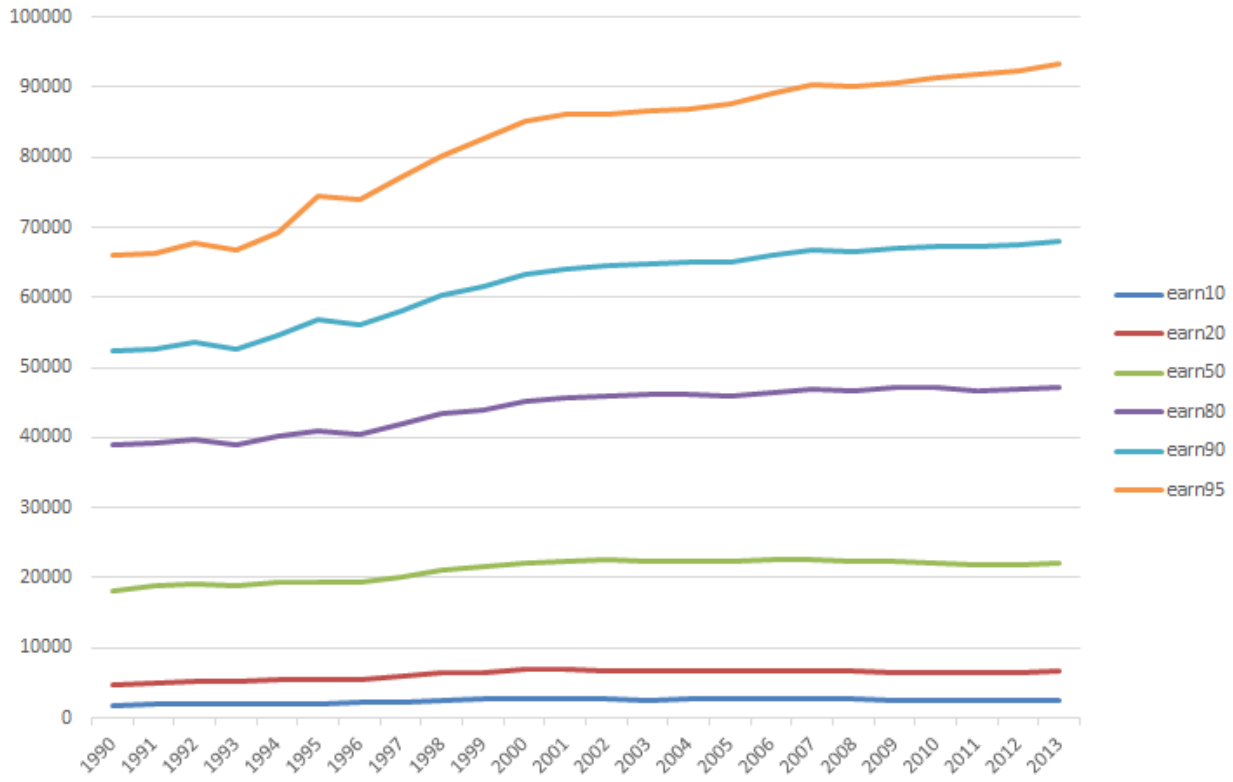


Figure 5: Percentiles of the Earnings Distribution for Eligible Workers All Jobs by Year

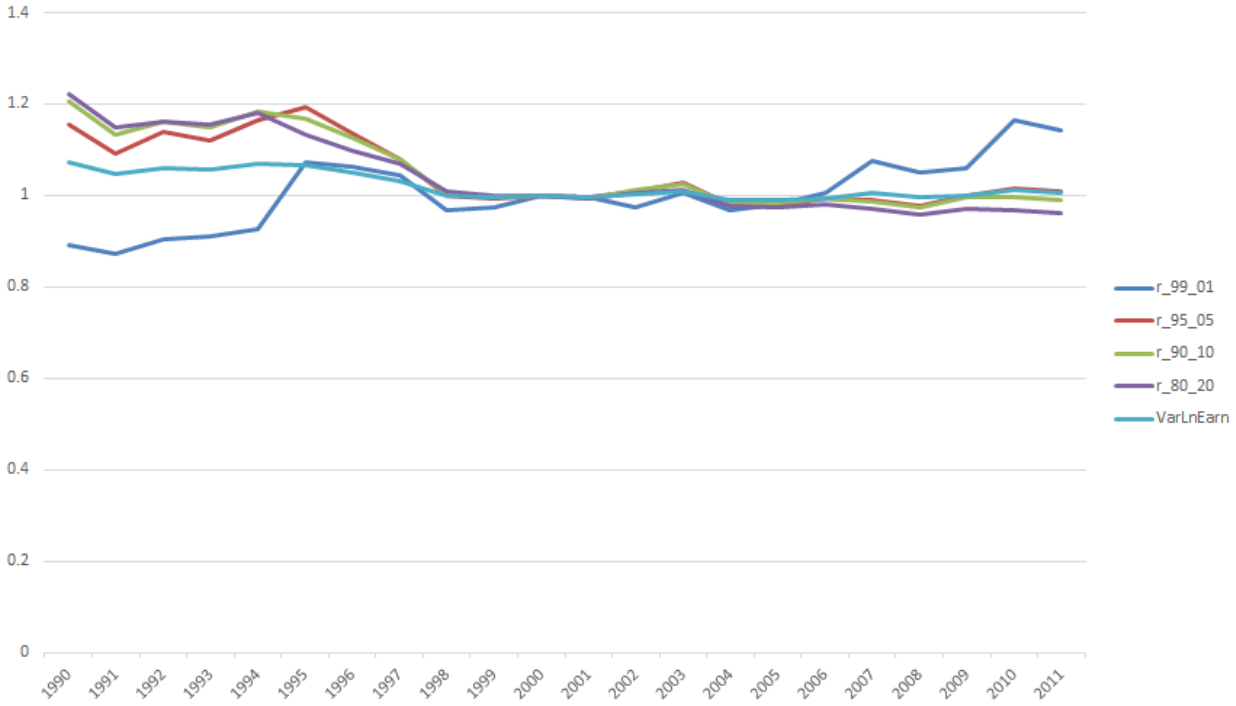


Figure 6: Selected Inequality Measures Relative to 2000 for All Workers by Year

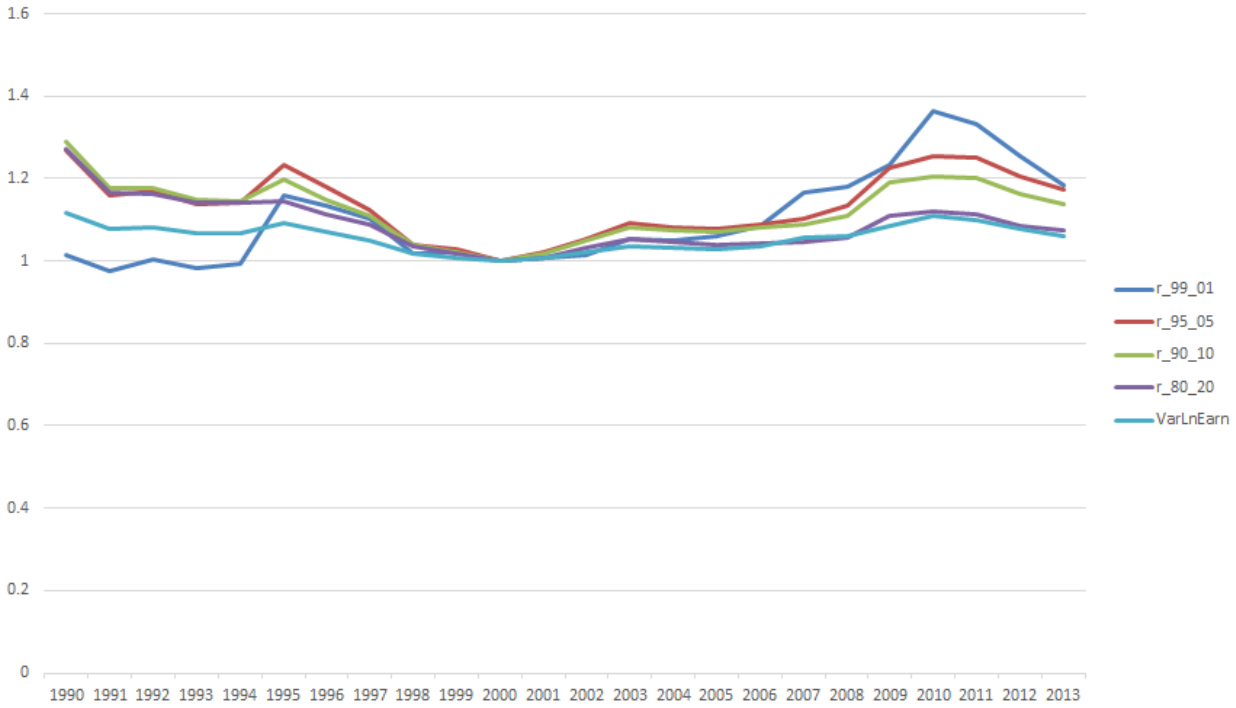


Figure 7: Selected Inequality Measures Relative to 2000 for Eligible Workers by Year

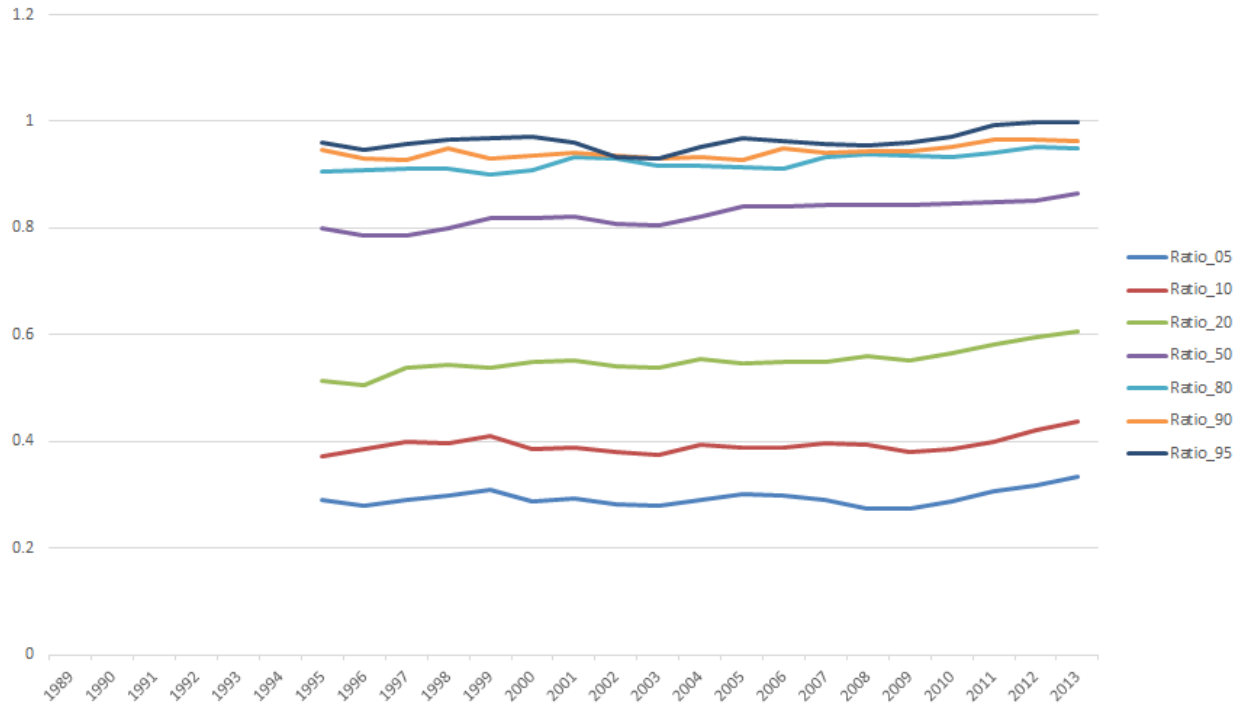


Figure 8: Ratio of the Percentiles of the Earnings Distribution Measured Using Unemployment Insurance Records to the Percentiles from Current Population Survey/American Community Survey

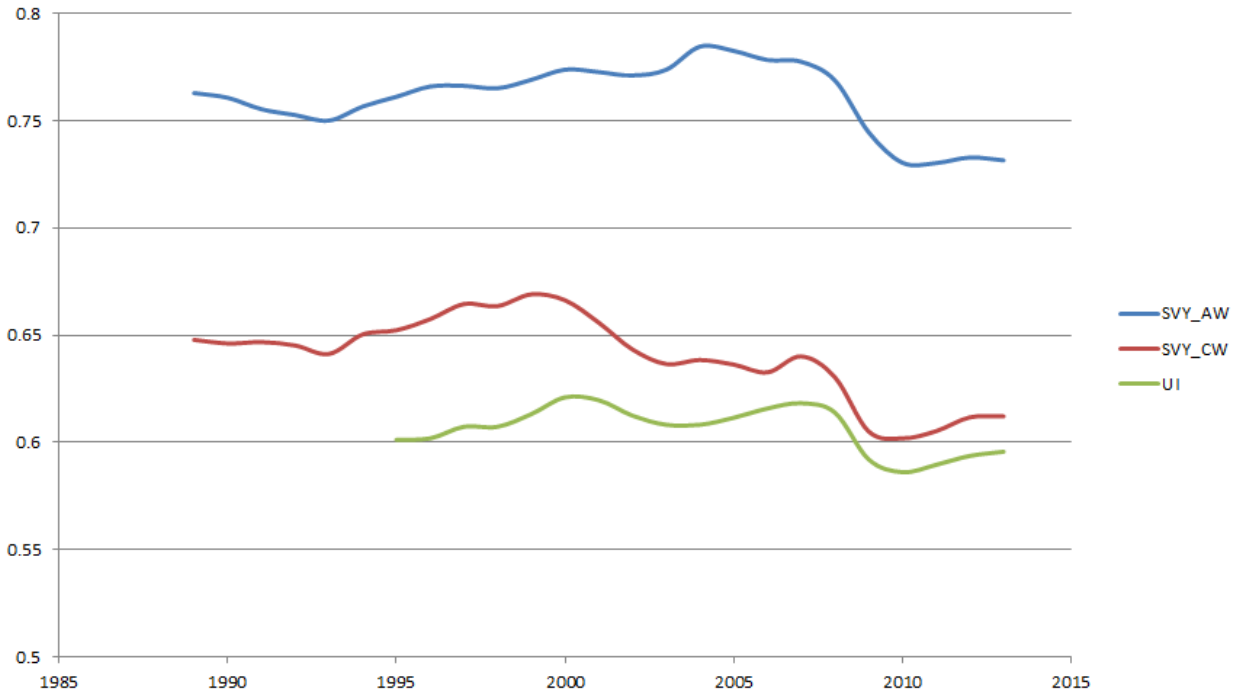


Figure 9: Comparison of the Employment to Population Ratio Between Survey and UI Sources by Year

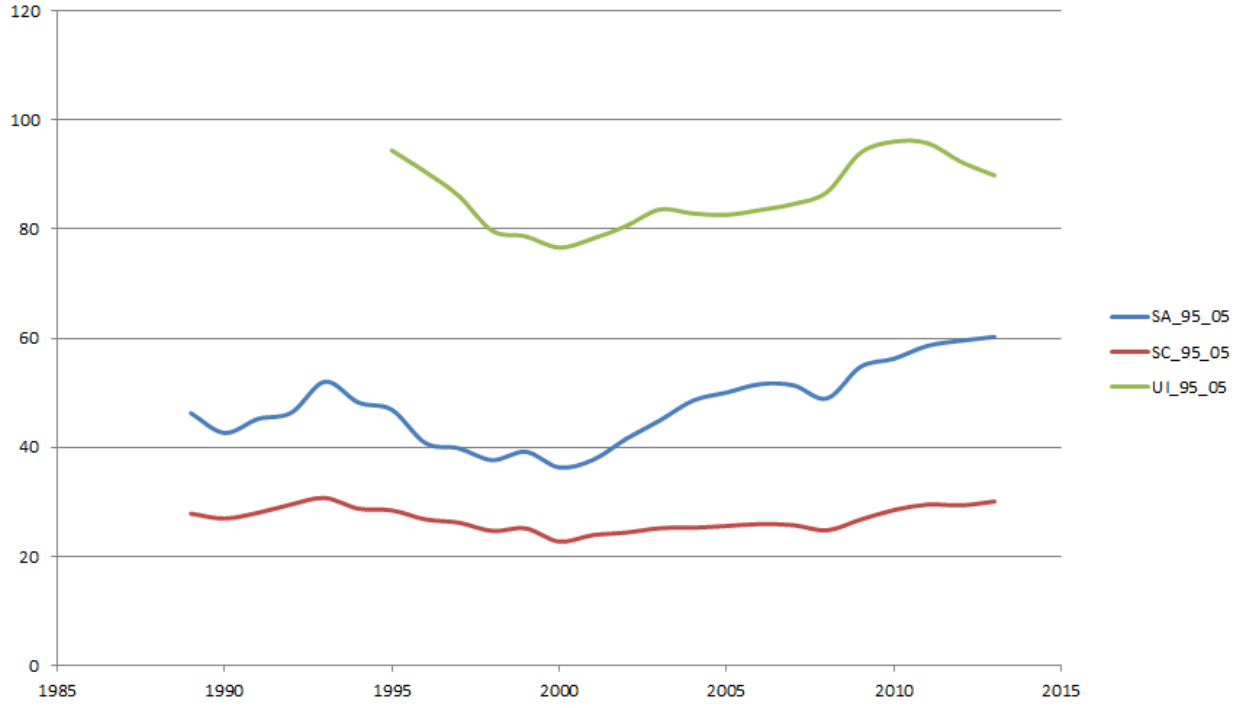


Figure 10: Ratio of the 95^{th} to the 5^{th} Percentile of the Earnings Distribution by Source and Year

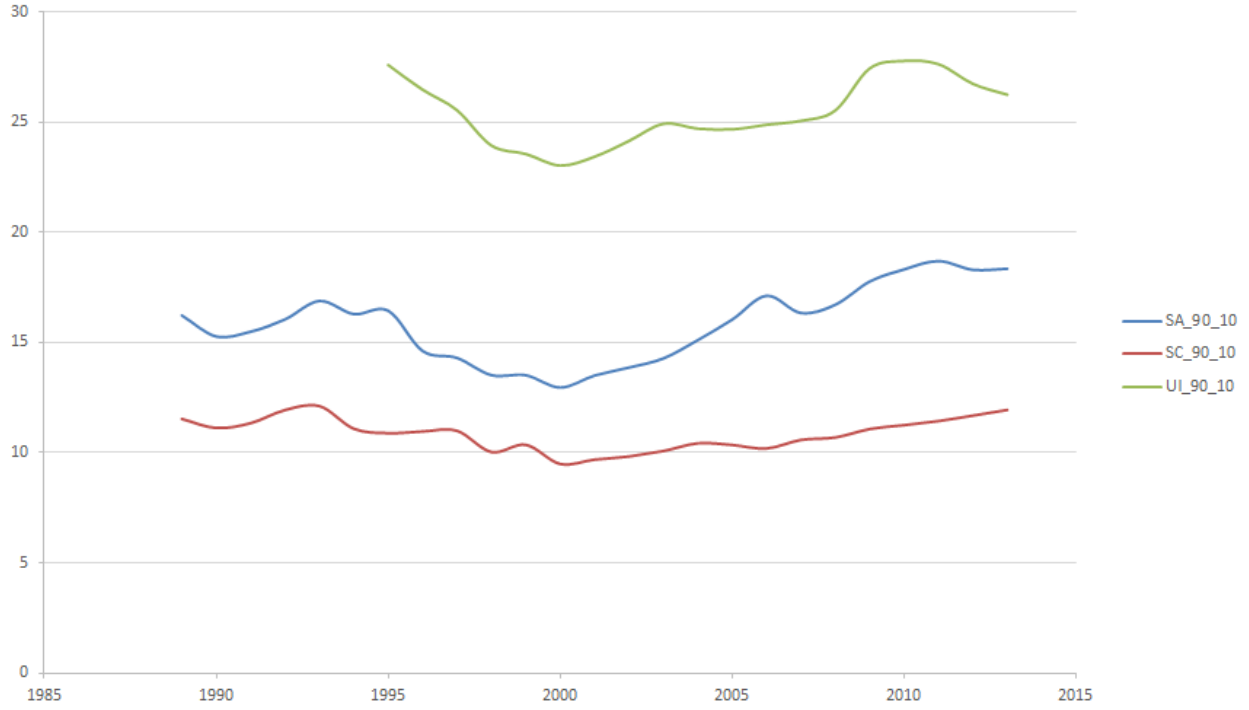


Figure 11: Ratio of the 90^{th} to the 10^{th} Percentile of the Earnings Distribution by Source and Year

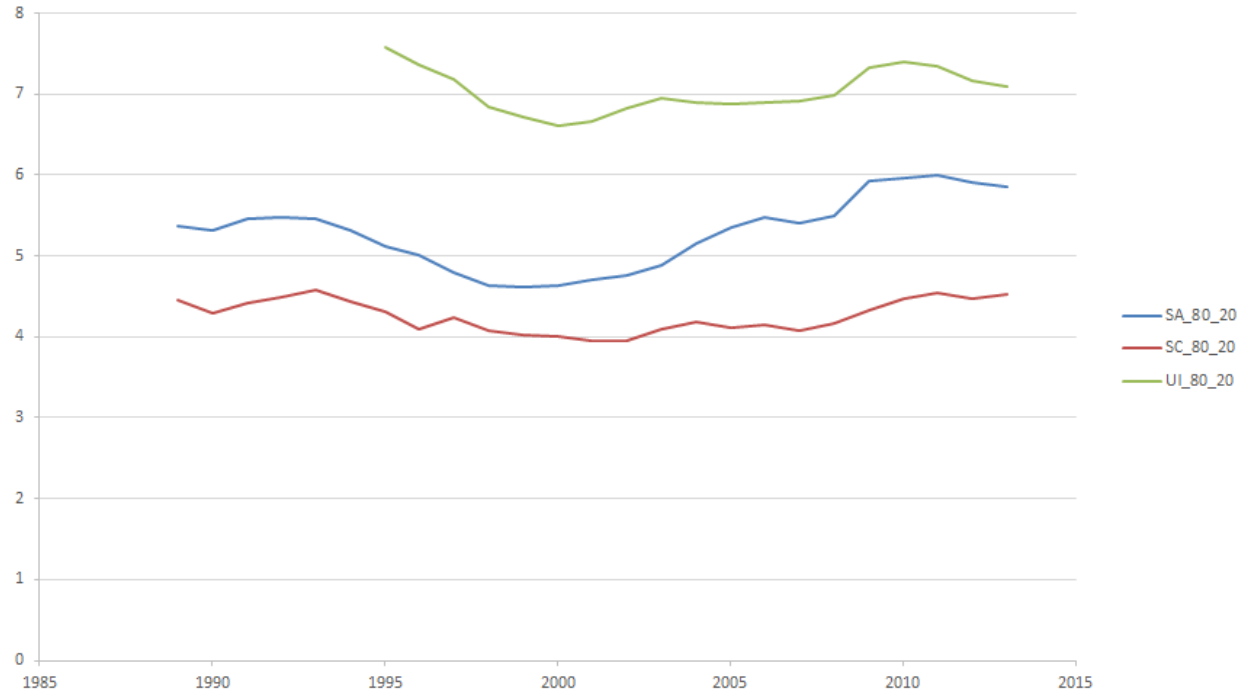


Figure 12: Ratio of the 80^{th} to the 20^{th} Percentile of the Earnings Distribution by Source and Year

All Workers				Covered Workers		
ACS	UI	Pct		ACS	UI	Pct
E>0	Not(E>0)	21		E>0	Not(E>0)	15
Not(E>0)	E>0	4		Not(E>0)	E>0	~0
E>0	E>0	75		E>0	E>0	85

Figure 13: Characteristics of the CPS/ACS Matched to the UI Earnings Data

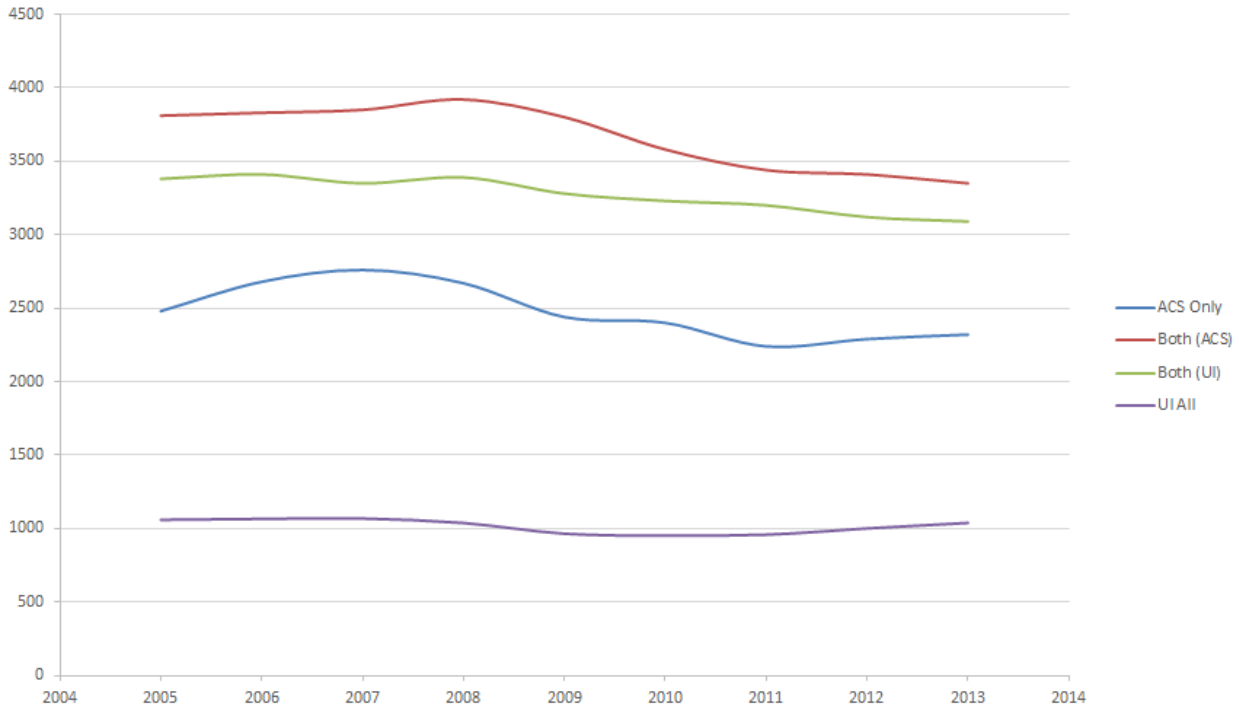


Figure 14: Comparison of the 5^{th} Percentile of the Earnings Distribution for UI Covered Workers by Match Type

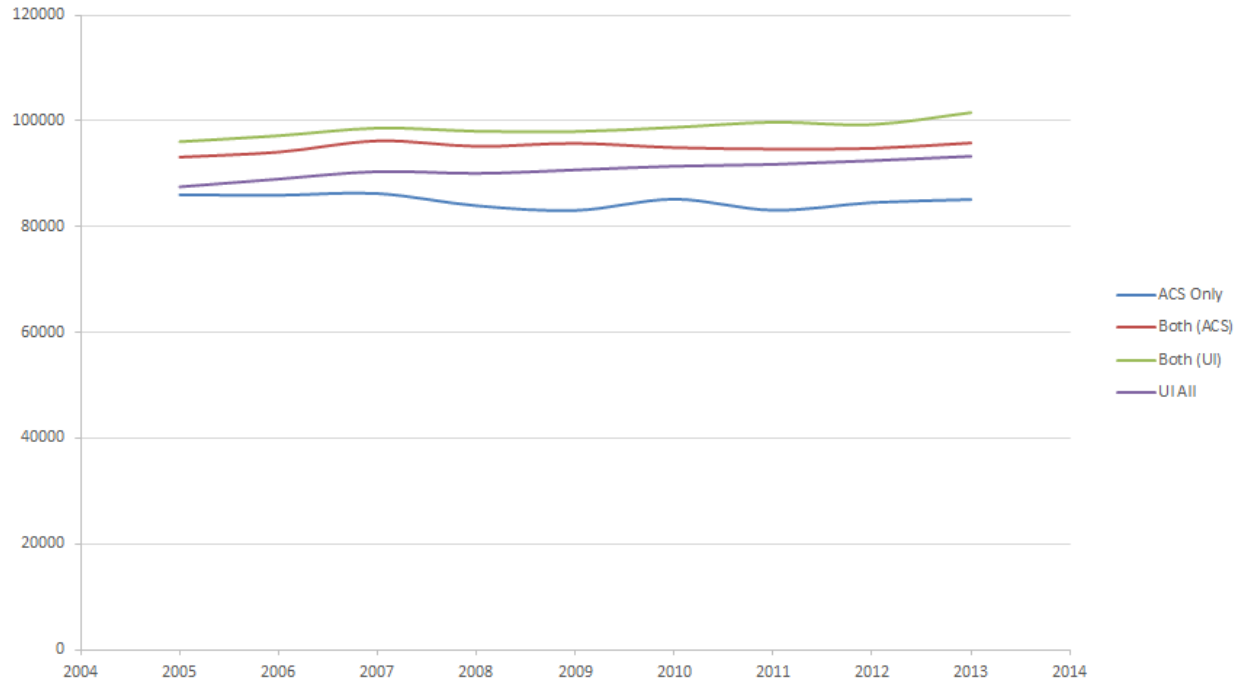


Figure 15: Comparison of the 95^{th} Percentile of the Earnings Distribution for UI Covered Workers by Match Type

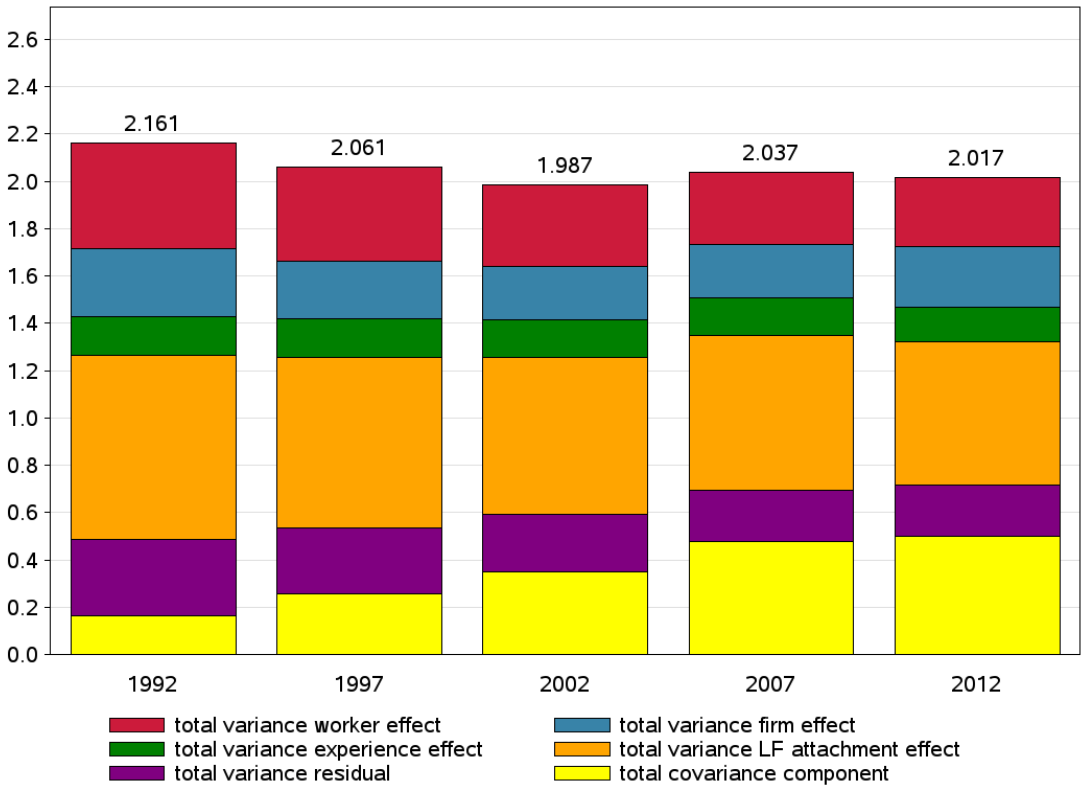


Figure 16: Components of the Total Variance of Ln Earnings by 5-year Window

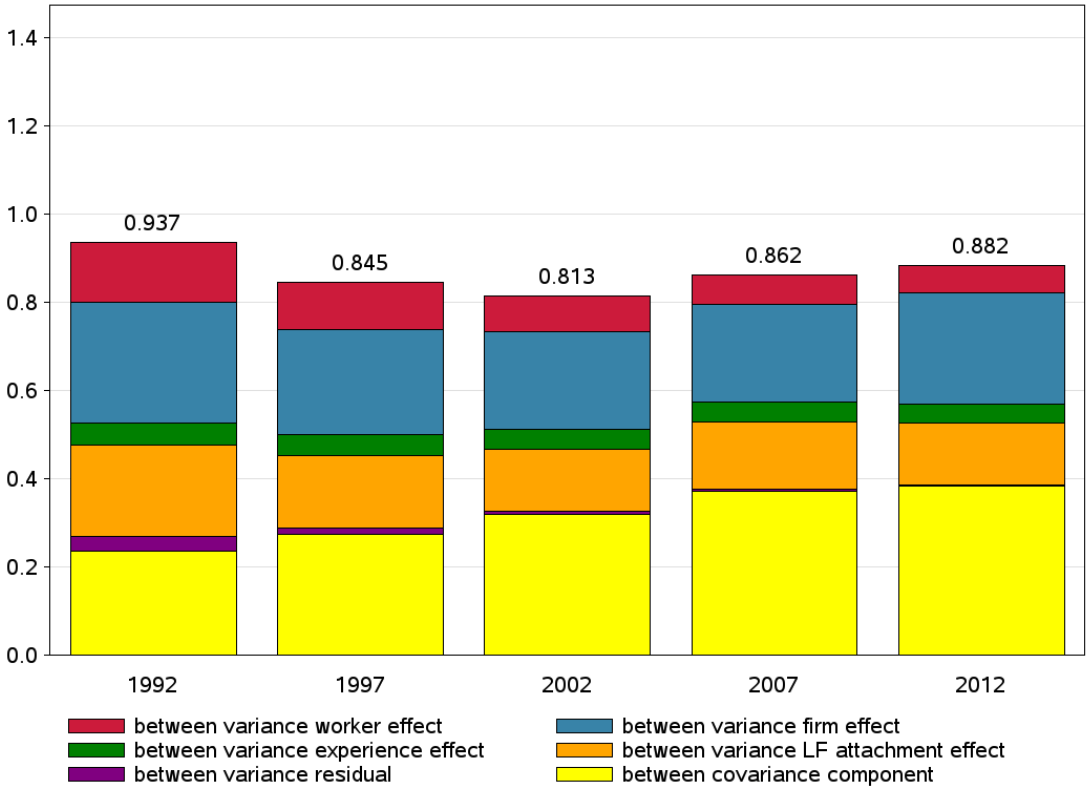


Figure 17: Components of the Variance of Ln Earnings between Establishments by 5-year Window

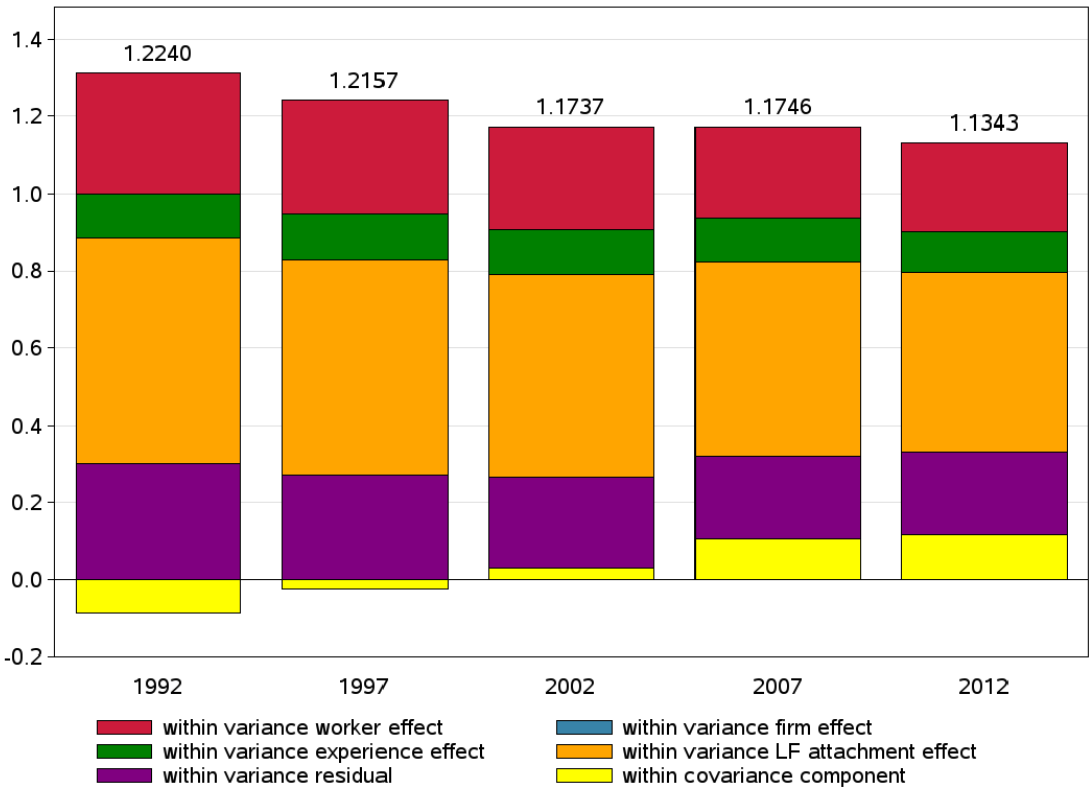


Figure 18: Components of the Variance of Ln Earnings within Establishments by 5-year Window

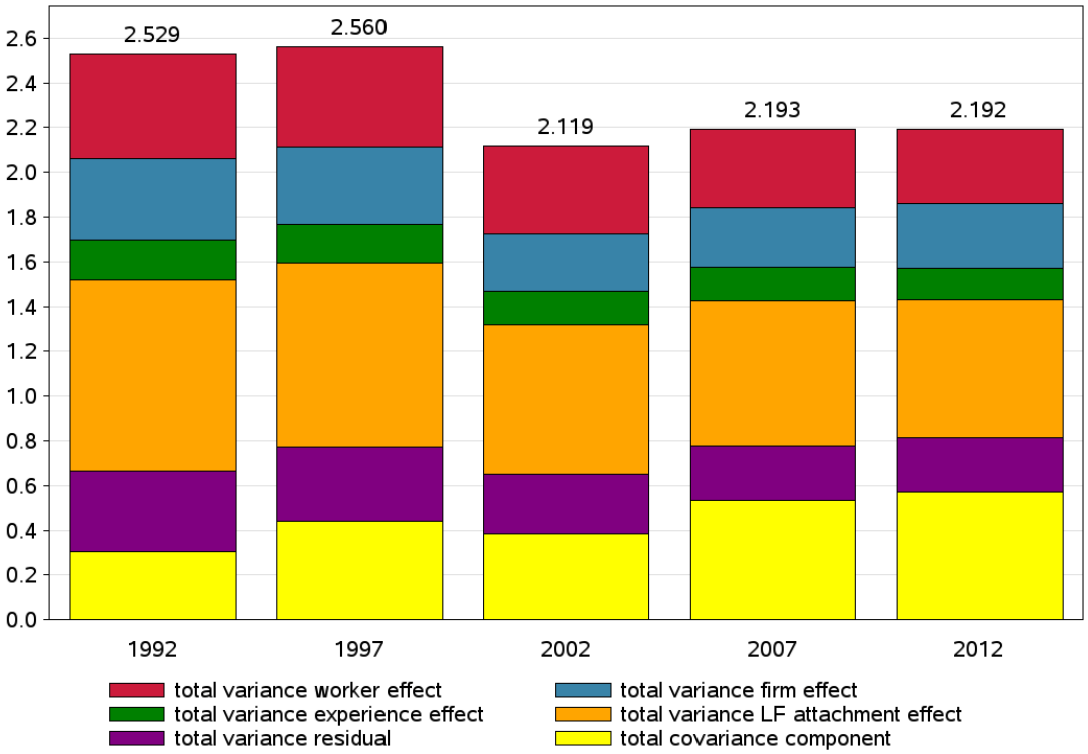


Figure 19: Components of the Total Variance of Ln Earnings by 5-year Window
Quintile 1 Sales per Worker (Adjusted for Material Costs)

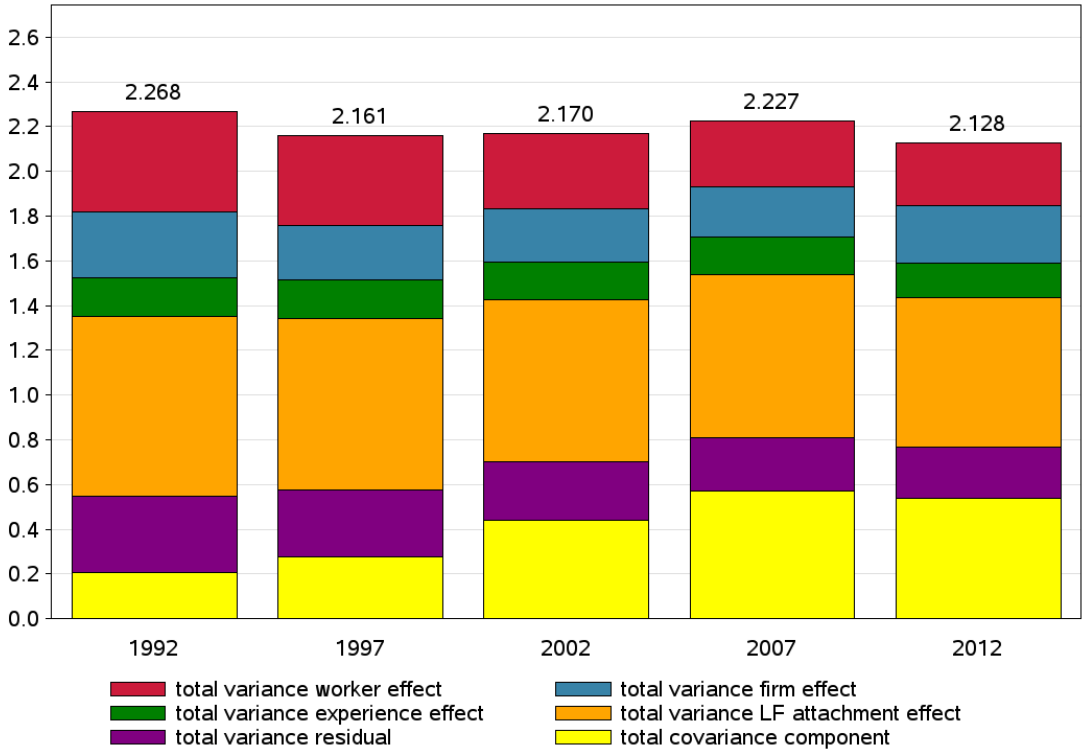


Figure 20: Components of the Total Variance of Ln Earnings by 5-year Window
Quintile 2 Sales per Worker (Adjusted for Material Costs)

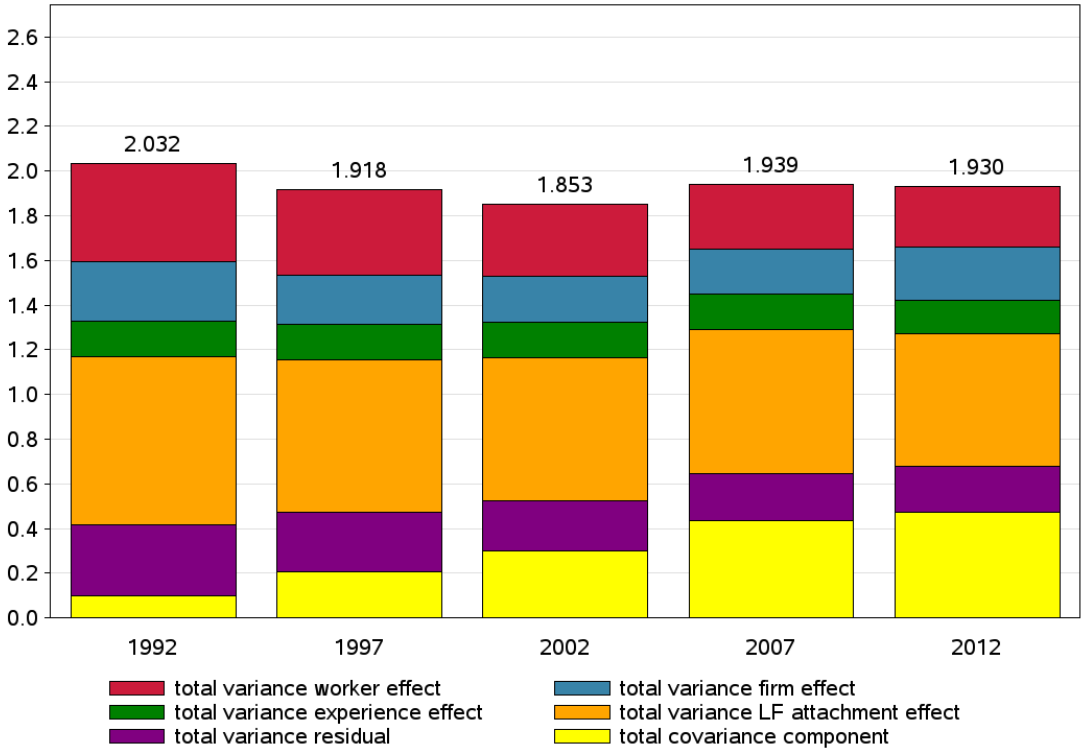


Figure 21: Components of the Total Variance of Ln Earnings by 5-year Window
Quintile 3 Sales per Worker (Adjusted for Material Costs)

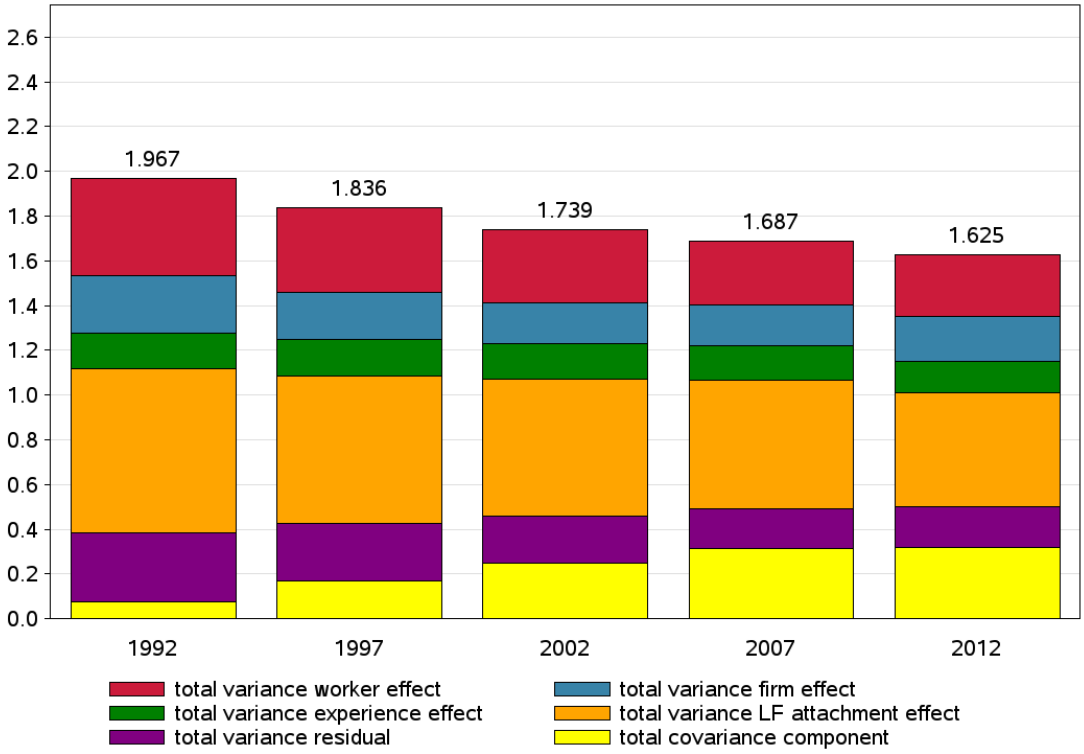


Figure 22: Components of the Total Variance of Ln Earnings by 5-year Window
Quintile 4 Sales per Worker (Adjusted for Material Costs)

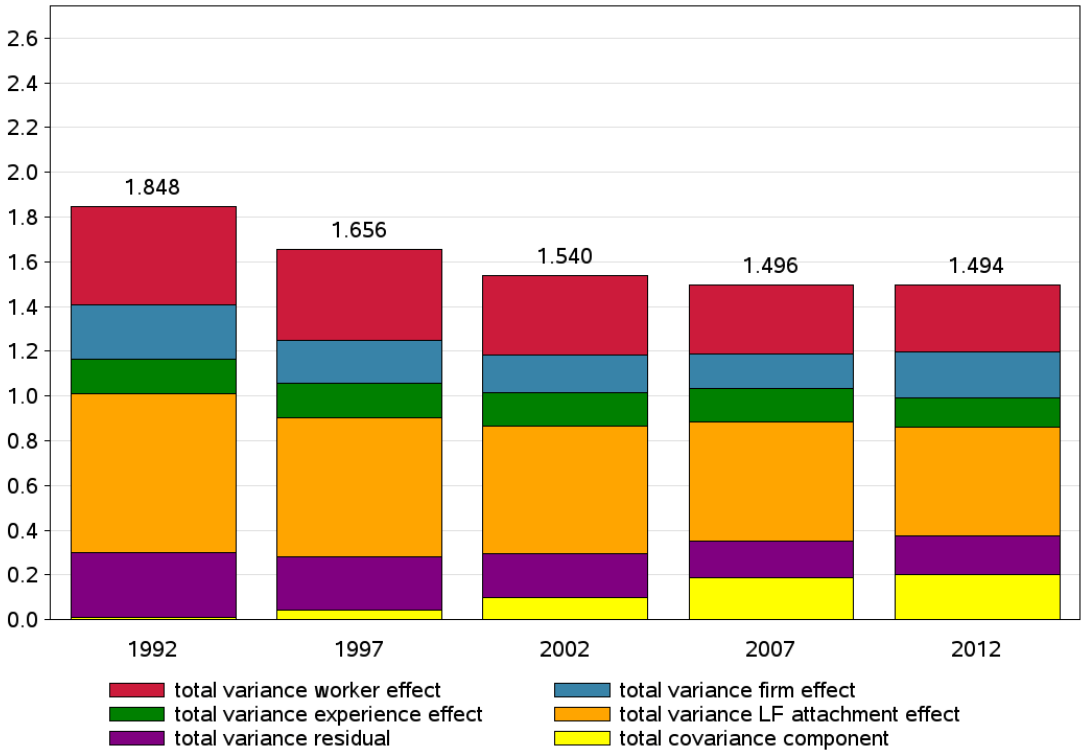


Figure 23: Components of the Total Variance of Ln Earnings by 5-year Window
Quintile 5 Sales per Worker (Adjusted for Material Costs)

Appendices

Figure 24 shows the percentiles of the earnings distribution for the "all worker" frame using all jobs and just the dominant job. The dominant job is the producing the most earnings in a year among all jobs held by the worker that year.

Figures 25 and 26 present the results of the CHK event study test for the male workers in the U.S. in periods 1990-1994 and 2010-2013, respectively. As can be seen from the figures, the movements from low/high paying employers to low/high paying employers as measured in the raw data do not show the same patterns as were seen in the German data. Figures 27 and 28 repeat the analysis after adjusting the observed earnings to the full-year, full-time value implied by the coefficients in the AKM decomposition. As can be seen from the figures, when we analyze wage rates, as was done with the German data in CHK, rather than earnings, the figures are very similar to those fit in the German event studies.

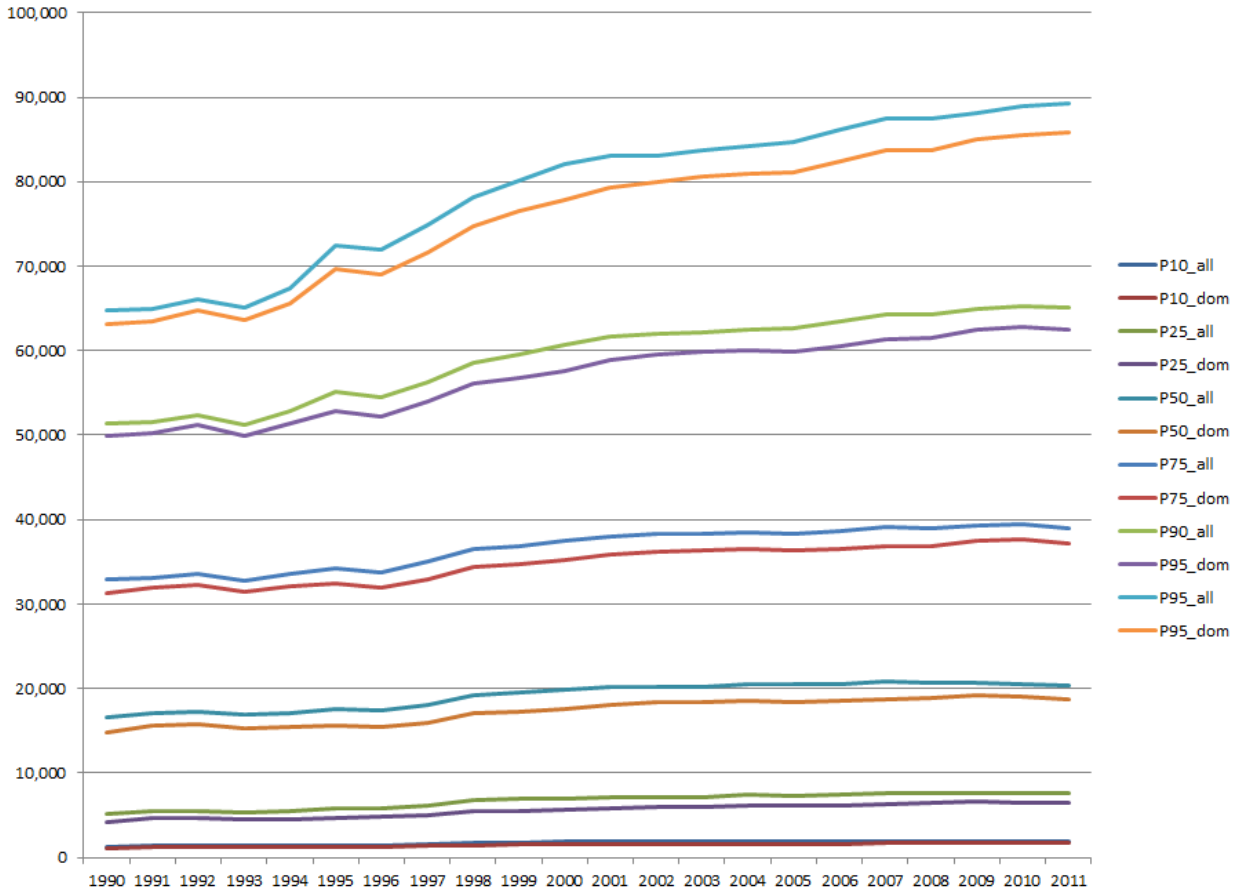


Figure 24: Percentiles of the Earnings Distribution for All Workers All Jobs v Dominant Jobs by Year

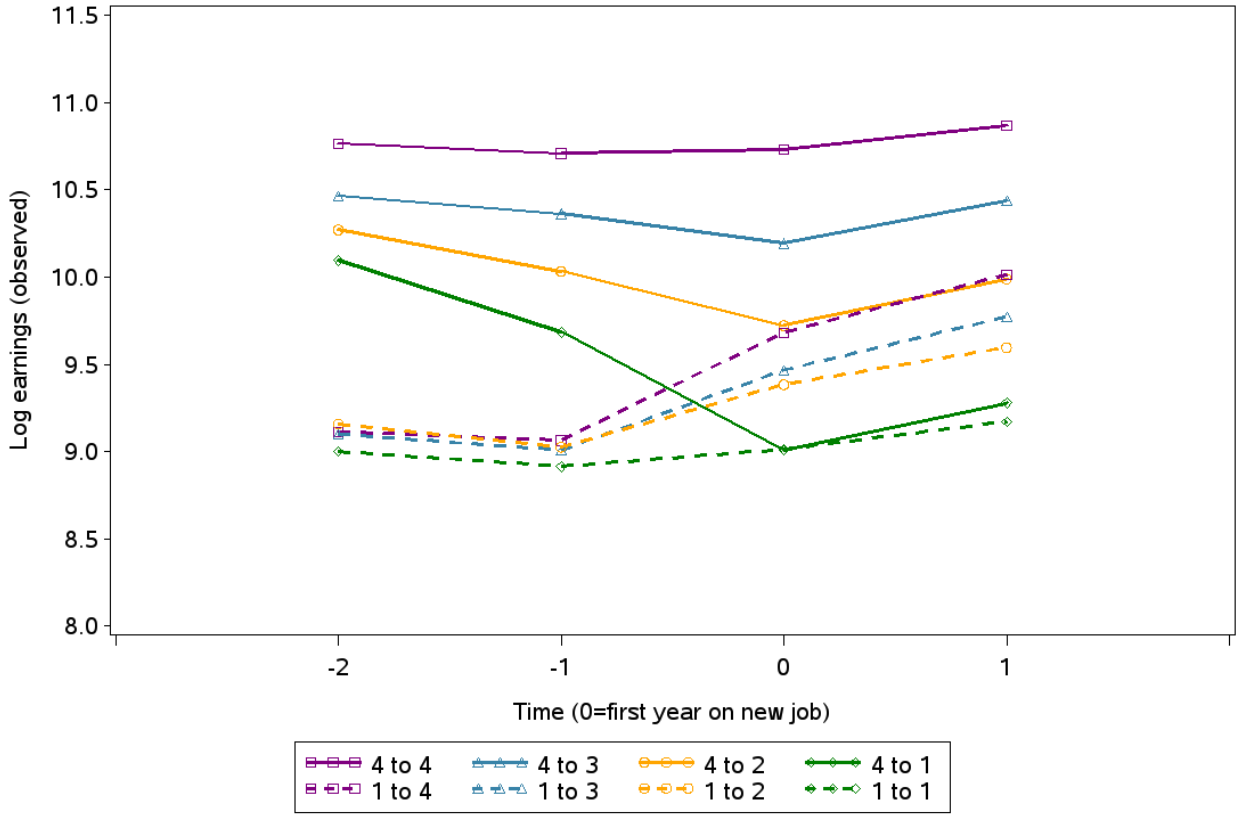


Figure 25: Mean Unadjusted Ln Earnings for Male Job Changers Classified by the Quartile of the Mean Ln Earnings of Coworkers for 1990-1994

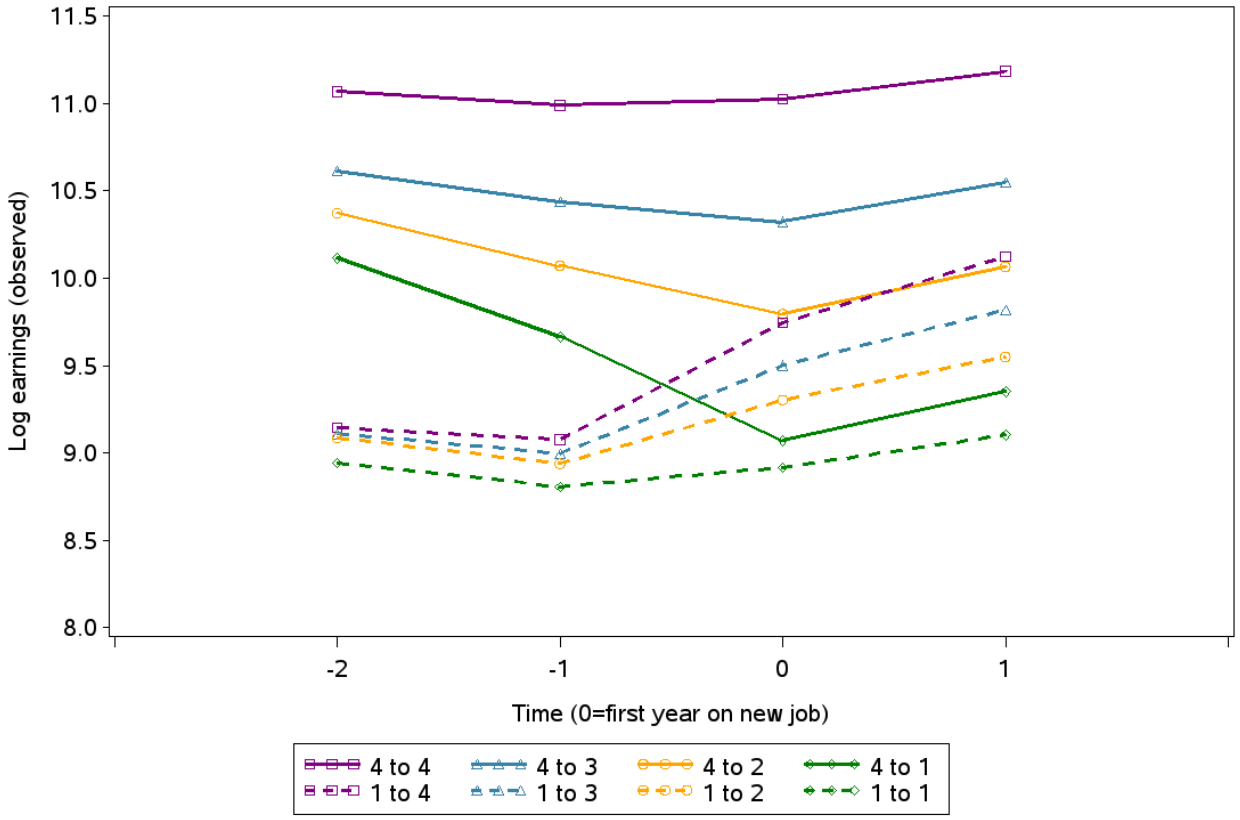


Figure 26: Mean Unadjusted Ln Earnings for Male Job Changers Classified by the Quartile of the Mean Ln Earnings of Coworkers for 2010-2013

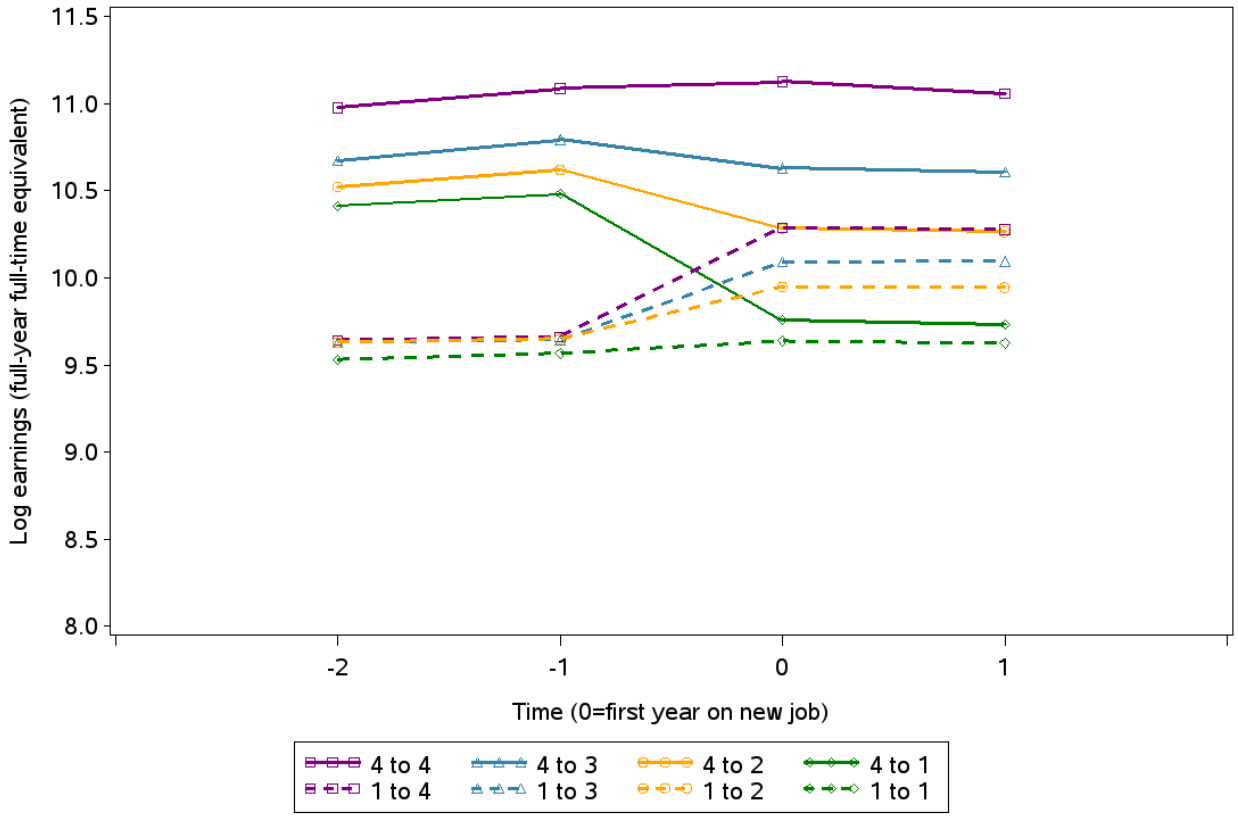


Figure 27: Mean Full-year Full-time Equivalent Ln Earnings for Male Job Changers Classified by the Quartile of the Mean Ln Earnings of Coworkers for 1990-1994

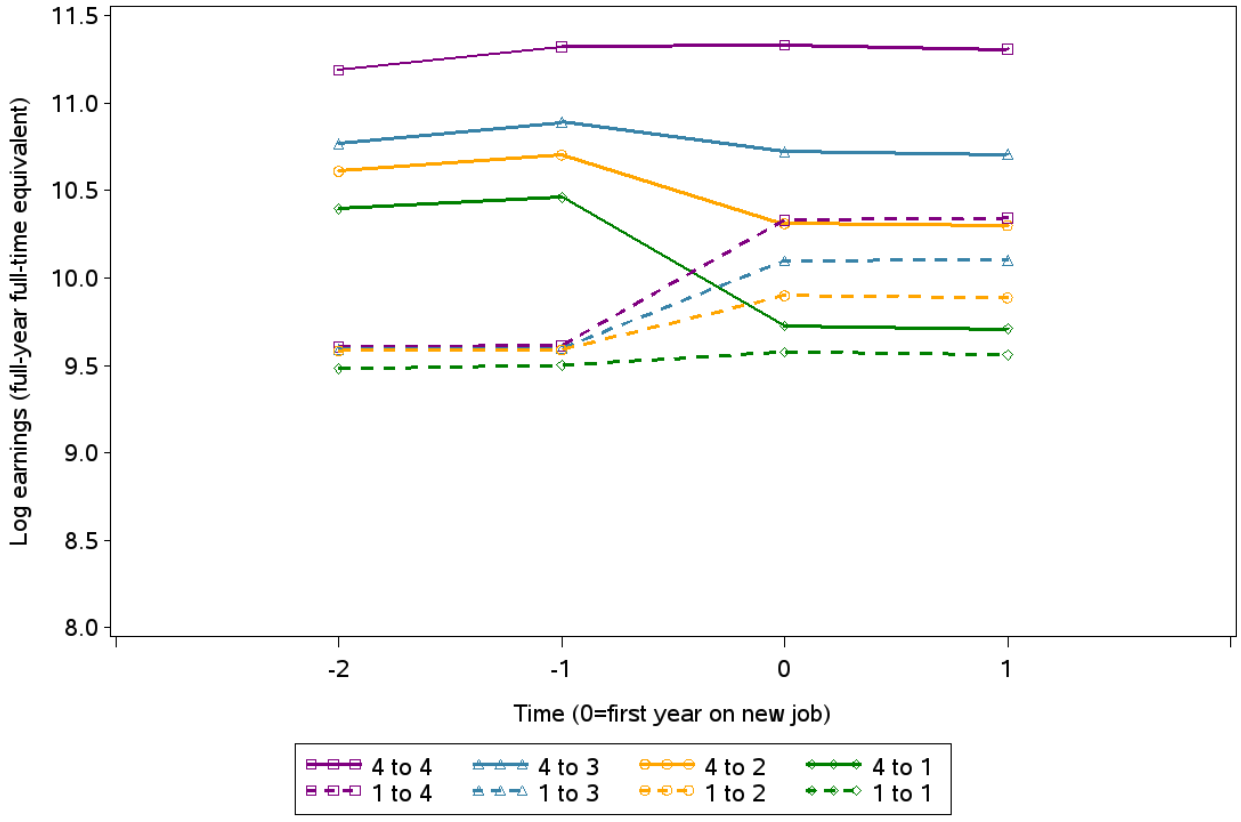


Figure 28: Mean Full-year Full-time Equivalent Ln Earnings for Male Job Changers Classified by the Quartile of the Mean Ln Earnings of Coworkers for 2010-2013