Rising Wage Inequality:
The Role of Composition and Prices*

David H. Autor
MIT and NBER

Lawrence F. Katz
Harvard University and NBER

Melissa S. Kearney
The Brookings Institution and NBER

August 2005
Revised from December 2004

Abstract

During the early 1980s, earnings inequality in the U.S. labor market rose relatively uniformly throughout the wage distribution. But this uniformity gave way to a significant divergence starting in 1987, with upper-tail (90/50) inequality rising steadily and lower tail (50/10) inequality either flattening or compressing for the next 16 years (1987 to 2003). This paper applies and extends a quantile decomposition technique proposed by Machado and Mata (2005) to evaluate the role of changing labor force composition (in terms of education and experience) and changing labor market prices to the expansion and subsequent divergence of upper- and lower-tail inequality over the last three decades We show that the extended Machado-Mata quantile decomposition corrects shortcomings of the original Juhn-Murphy-Pierce (1993) full distribution accounting method and nests the kernel reweighting approach proposed by DiNardo, Fortin and Lemieux (1996). Our analysis reveals that shifts in labor force composition have positively impacted earnings inequality during the 1990s. But these compositional shifts have primarily operated on the lower half of the earnings distribution by muting a contemporaneous, countervailing lower-tail price compression. The steady rise of upper tail inequality since the late 1970s appears almost entirely explained by ongoing between-group price changes (particularly increasing wage differentials by education) and residual price changes.

Introduction

During the early 1980s, earnings inequality in the U.S. labor market rose relatively uniformly throughout the wage distribution. Between 1979 and 1987, the male 90/50 log hourly earnings ratio rose by 8.5 log points and the 50/10 earnings ratio rose by 13.0 log points (see Figure 1). This simultaneous expansion of upper and lower-tail inequality gave way to a significant divergence thereafter. During each of the two subsequent 8 year intervals, 1987 to 1995 and 1995 to 2003, 90/50 inequality rose by an additional 5.2 and 9.7 log points respectively, while 50/10 inequality contracted to levels roughly comparable to 1979.[1] In fact, fully 90 percent of the net increase in male 90-10 earnings inequality between 1979 and 2003 is accounted for by the rise in the 90-50 wage gap. Residual inequality – that is, wage dispersion <u>within</u> demographic and skill groups defined by sex, education, and experience (age) – followed a roughly similar pattern, rising both above and below the median between 1979 and 1987, and then diverging thereafter (Figure 2).[2] As discussed in Autor, Katz and Kearney (2005), this divergence in inequality trends presents a puzzle for canonical explanations for rising wage inequality.[3] Hypotheses such as falling real minimum wages, declining unionization, and monotonically rising demand for 'skill' (often viewed as driven by Skill Biased Technical Change 'SBTC') do not generally predict steadily increasing upper-tail inequality paired with fluctuating lower tail inequality.

In this paper, we explore one potential explanation for the disparate patterns of upper and

---

[1] These figures describe hourly wages for all male wage and salary workers from the merged monthly Outgoing Rotation Group (ORG) files of the Current Population Survey (CPS) for years 1979 to 2003. Similar calculations performed using the March CPS reveal comparable results (also shown in Figure 1). Appendix Table 1 demonstrates qualitatively similar patterns for full-time workers in March and ORG samples. Details of all samples and calculations are found in the Data Appendix.

[2] The overall rise in the variance of male log hourly wages in the May/ORG CPS was 7.3 log points between 1979 and 1988 and 2.3 log points between 1988 and 2003. Estimated using a flexible Mincerian wage regression (see equation (1)), the residual component of this rise was 2.8 log points between 1979 and 1988 and 0.2 log points between 1988 and 2003. The very modest rise in inequality in the May/ORG CPS after 1988 is at odds with the March CPS data, which show an overall rise of log wage inequality of 6.1 log points between 1979 and 1988 and 6.1 log points between 1988 and 2003. The contribution of the residual to this rise is also substantial in both periods: 3.1 log points between 1979 and 1988 and 3.2 log points between 1988 and 2003. In section V, we analyze the discrepancies between the May/ORG and March data. All analyses in the paper are performed using both samples.

[3] The differences in the growth of the 90-50 and 50-10 wage differentials have previously been emphasized by Mishel, Bernstein and Boushey (2002).

lower-tail inequality observed in the last two decades: the changing age and educational

composition of the U.S. labor force. As shown in Figure 3, the education and experience of the

U.S. labor force rose substantially over the last 25 years as the large 1970s college cohorts

reached mid-career during the 1990s. The full-time equivalent employment share of male workers

with a college degree rose from 18 to 32 percent of the U.S. male labor force between 1973 and

2003 and the employment share of workers with high school or lower education fell from 62 to

41 percent. Gains in potential experience were similarly pronounced: the mean potential

experience of males of high school or greater education rose by 2 to 5 years between 1973 and

2003, with the largest gains experienced by the most educated groups.[4] Increases in education and

potential experience among female workers were similarly large.

These large shifts in labor force composition have the potential to contribute to the divergent

behavior of upper and lower tail inequality. As highlighted in an insightful paper by Thomas

Lemieux (2005), the canonical Mincer (1974) earnings model implies that earnings trajectories

may tend to fan out (become more dispersed) as workers gain labor market experience (see also

Heckman, Lochner and Todd 2003). And hourly wage dispersion is typically higher for college

graduates than for less-educated workers. Consequently, changes in the distribution of education

or experience of the labor force may give rise to changes in earnings dispersion. These

compositional effects are distinct from the standard price effects that are often invoked to explain

fluctuations in earnings inequality. Holding market prices constant, changes in labor force

composition can mechanically raise or lower residual earnings dispersion simply by changing the

employment share of worker groups that have more or less dispersed earnings. Similarly, changes

in workforce composition can also raise or lower overall earnings dispersion by increasing or

reducing heterogeneity in observed skills (education and experience), a point emphasized by

Juhn, Murphy and Pierce (1993). These observations suggest that measured earnings dispersion

---

[4] The average potential experience of male high school dropouts actually fell by 2.6 years during 1973 to 2003. This decline reflects the increased share of employed teenagers in the "dropout" group as adult educational attainment has risen (as our sample covers ages 16 to 64 years). An influx of younger, less educated immigrants has also contributed.

may change due to the mechanical impact of composition without any underlying change in market prices.[5]

Building on this observation, Lemieux (2005) argues that the bulk of the growth in residual wage dispersion in the U.S. between 1973 and 2003 – and <u>all</u> of the growth after 1988 – is explained by mechanical effects of changes in workforce composition rather than shifts in residual inequality within defined skill groups (what we call price effects). After adjusting for these compositional effects, Lemieux concludes that the widely discussed rise in residual earnings inequality in the U.S. labor market is primarily attributable to institutional factors during the 1980s – in particular, the declining real minimum wage – and to the mechanical effect of composition during the 1990s.[6]

In this paper, we consider the Lemieux composition hypothesis in an analysis of the divergent trends in the upper and lower tails of the aggregate and residual wage distributions over the last three decades. Our work differs from Lemieux's analysis in two key substantive dimensions. First, in keeping with our focus on trends in upper and lower tail inequality, we separately analyze the contribution of composition to changes in wage inequality above and below the median of the distribution.[7] Second, we assess the contribution of compositional shifts to changes in both overall and residual inequality since composition may impact wage dispersion through either channel.

To perform the analysis, we apply and extend a quantile decomposition technique proposed by Machado and Mata (2005).[8] This decomposition uses quantile regressions to partition the

---

[5] Labor force composition will also affect wage dispersion though general equilibrium effects of quantities on prices. We focus here on the partial equilibrium effects and discuss general equilibrium considerations later in the paper.

[6] By contrast, Juhn, Murphy, and Pierce (1993) find that the contribution of labor force composition to overall male full-time weekly earnings inequality growth over 1964 to 1988 is negligible. They do not consider – and their decomposition method does not readily admit – an effect of composition on residual inequality.

[7] This contrasts to Lemieux's focus on the overall dispersion of wage residuals, measured by the variance or the 90/10 ratio of log wage residuals. We also discuss results for these measures below.

[8] The simple extension of the Machado-Mata method to residual inequality proposed and implemented here is independently developed by Melly (forthcoming), who applies this tool to analyzing sources of US wage inequality growth over 1973 through 1989. Because the Melly analysis only extends to 1989 and does not distinguish between

observed distribution of earnings into 'price' components (wage coefficients) and 'quantity' components (labor force composition) and calculate, through simulation, the impact of each on changes in overall wage dispersion. We implement this decomposition using representative earnings data from the Current Population Survey (CPS). Samples include hourly earnings data from the May CPS and Outgoing Rotation Groups (May/ORG) for 1973 to 2003 and hourly and weekly earnings data from the March CPS Annual Earnings Files for 1976 to 2004 (earnings years 1975 to 2003).

The Machado-Mata method for calculating counterfactual densities is closely related to the kernel reweighting approach proposed by DiNardo, Fortin and Lemieux (1996; 'DFL' hereafter) and extended by Lemieux (2002, 2005). It has two features that we believe make it particularly attractive. First, it provides a precise link between the 'full variance accounting' technique for analyzing inequality introduced by Juhn, Murphy and Pierce (1993; 'JMP' hereafter) and the kernel reweighting proposed by DFL. As we show below, both can be interpreted as applications of the canonical Oaxaca-Blinder decomposition, extended to the analysis of conditional full distributions rather than (exclusively) conditional means. Second, the Machado-Mata approach can be readily extended (as we do here) to provide a uniform, consistent treatment of both overall inequality and residual inequality. By comparison, alternative approaches for analyzing counterfactual wage distributions apply a hybridized set of methods and conceptual models (OLS regressions, parametric probability models, and kernel reweighting) to separately develop counterfactuals for overall and residual inequality.

Our main results are as follows. We find that changes in labor force composition do not contribute to an explanation for the diverging path of upper and lower tail inequality (either aggregate or residual) in the past two decades. The composition hypothesis fails for two reasons. First, we show that the impact of changes in labor force composition on wage dispersion occurs

trends in upper and lower-tail inequality, its substantive conclusions are almost entirely different from those here (though similar to Lemieux 2005).

4

almost entirely below the median of the earnings distribution (i.e., in the lower tail) during the sample period This in turn implies that the steady growth of upper-tail inequality during the 1980s and 1990s is due to changing labor market prices, not mechanical effects of composition.

Second, we show that the puzzle of rising and then stagnating or falling lower tail inequality in the 1980s and 1990s is actually made more puzzling by adjustments for composition. During the 1980s, increasing lower tail inequality appears explained by changing labor market prices, augmented slightly by shifts in composition. In the 1990s, by contrast, changing market prices generated considerable compression in lower tail inequality, but these price effects were in substantial part offset by compositional shifts (which would otherwise have caused lower tail inequality to increase). On net, we find that changes in the education and experience composition of the labor force plays only a secondary role in the evolution of overall and residual inequality since the early 1970s; the persistent rise of upper-tail inequality and the rise and then stagnation of lower-tail inequality are both primarily accounted for by changing labor market prices.

These results lead us to differ from the interpretations offered by Lemieux (2005). Lemieux's conclusion that the rising education and experience of the labor force can mechanically explain the plateauing of residual inequality during the 1990s appears to aggregate over two countervailing forces. The first is the contraction in lower-tail inequality in the 1990s, which is due to compressing residual 'prices' partially offset by changing composition. The second is the rise in upper-tail inequality, which appears entirely explained by price changes. Because composition over-explains the former phenomenon and under-explains the latter, it is technically accurate to say that composition can "fully explain" the aggregate trend in residual inequality during the 1990s. But when upper and lower-tail inequality are considered separately – as seems appropriate given their substantial divergence – composition does not appear a satisfying explanation for either. We discuss possible alternative explanations for the divergence of upper and lower tail inequality in the conclusion (also see Autor, Katz and Kearney 2005).

In the next section, we summarize trends in U.S. overall and residual earnings inequality for

1973 to 2003. In section II, we develop the Machado and Mata methodology for forming counterfactual wage distributions and discuss our extension to it for analysis of residual wage distributions. Sections III and IV present our decomposition of trends in aggregate and residual inequality into price and quantity components. Section V analyzes discrepancies in residual inequality trends between March and May/ORG CPS data sources. Section VI concludes.

I. Data sources and basic trends in overall and residual inequality

A. *March and May/ORG Current Population Survey data*

Our analysis draws on two data sources for measuring trends in wage inequality: the March CPS and the combined CPS May and Outgoing Rotation Group. These sources are described briefly here, with additional details provided in the Data Appendix. The March CPS data have been widely used in studies of wage inequality and provide reasonably comparable data on prior year's annual earnings, weeks worked, and hours worked per week. We use the March CPS of 1976 to 2004, covering earnings from 1975 to 2003, to form a sample of real log hourly wages of wage and salary workers.[9] Hourly earnings in the March data are calculated as the ratio of prior year's annual earnings to prior year's annual hours.

We complement the March series using the May CPS for 1973 through 1978 and CPS Outgoing Rotation Group samples for 1979 through 2003 (CPS May/ORG) to construct hourly wage data for all wage and salary workers employed during the CPS sample survey reference week. Unlike the retrospective annual earnings data in the March CPS, the May/ORG data provide point-in-time measures of usual hourly or weekly earnings. We weight both March and May/ORG data by hours worked to provide a measure of the entire distribution of hours paid.[10]

Both March and May/ORG CPS surveys have features that limit their consistency over the period studied. As price and earnings levels have changed, each survey has applied successive

---

[9] Prior to the 1976 survey, usual hours worked per week in the earnings year are not available.

[10] The March data are weighted by the product of weeks worked and hours per week in the prior year. The May/ORG data are weighted by hours worked during the survey reference week. Weighting by weeks is implicit in the May/ORG sample since the probability that an individual is observed working during the sample reference week is proportional to weeks in the labor force.

increases to the maximum top-coded earnings values. These changes probably present a minor issue for our analysis since we focus on earnings quantiles between the 10th and 90th percentiles.[11]

A potentially more significant issue is posed by the redesign of the CPS ORG survey in 1994. This redesign changed the format of most survey questions, centralized the interviewing process and introduced the use of Computer Assisted Personal Interviewing (CAPI). Perhaps most important, the 1994 redesign revised the procedure used to collect earnings and hours information with the goal of increasing the accuracy of earnings reports and reducing the undercount of workers with part-time schedules (Polivka and Rothgreb 1993; Cohany, Polivka and Rothgreb 1994). Prior to 1994, CPS May/ORG respondents directly reported their usual weekly and hourly earnings to the survey taker. Commencing with the 1994 redesign, respondents were allowed to select a preferred pay frequency for earnings reporting (hourly, weekly, biweekly, monthly) and to report the corresponding earnings amount. Using the respondent's self-reported usual hours, the surveyor calculates an implied hourly wage and verifies the figure with the respondent.[12] The effect of the revised procedure on the <u>level</u> of reported hourly wages has received careful study (see Polivka 2000), but its effect on the precision of earnings reports has not been explored. It seems plausible that the enhanced procedure raised the precision of earnings reports (as designed) and so may have reduced measured earnings dispersion.

Adding to the uncertainty surrounding the CPS redesign, the CPS ORG did not identify ('flag') imputed earnings observations in the first 21 months following the survey redesign of January 1994 to August 1995. Consequently, we are unable to exclude these imputed observations, although we exclude imputed earnings observations from all other March and May/ORG samples. Earnings imputations are potentially a non-trivial issue: the imputation rate in the CPS ORG rose substantially after the redesign, from 15.3 percent in 1993 to 23.3 percent in

---

[11] One unfortunate exception is that the topcoded value for secondary wage and salary earnings in the March CPS fell from $99K to $25K in 1996 (before rising to $37K in 2004). This does not appear a major issue since fewer than 1 percent of workers have top-coded secondary wage/salary earnings at the <u>lower</u> topcode value.

[12] For respondents who report that they are paid hourly, the self-reported wage is always used without further calculations, regardless of the preferred reporting interval.

the last quarter of 1995 (the first quarter in which allocation flags are available) and continued to rise thereafter, reaching 31 percent by 2001 (Hirsch and Schumacher 2004). Since the redesign coincides with a sizable and sustained rise in the earnings imputation rate and a 21 month hiatus in the identification of imputed observations, it is particularly difficult to gauge how changes to the survey affected the comparability of the earnings series before and after 1994. [13]

A final issue affecting the consistency of the earnings reports is a secular change in the composition of pay reporting seen in the May/ORG samples. As documented by Hamermesh (2002), the share of workers who report being paid by the hour increased substantially from 1973 forward, and this increase is found within education and experience groups (Lemieux 2005). This matters for inequality measurement because direct hourly earnings reports in the May/ORG CPS appear to have greater precision than non-hourly wage reports (i.e., those calculated as the ratio of earnings to hours). Rising hourly pay reporting may tend to mask any underlying ('true') rise in residual variation in the May/ORG inequality series. The March survey uses an identical earnings measure (total annual earnings) for all respondents regardless of the pay period. Nevertheless, Lemieux (2005) presents evidence using matched March-May/ORG samples suggesting that measurement error in computed hourly wages is greater for hourly than for non-hourly workers in the March sample. This implies the rise in hourly pay reporting could lead the March CPS to overstate the growth in residual (hourly) wage inequality in a period of an increased share of hourly workers. We further explore these issues in Section V.

The essential trade-off between the March and May/ORG wage data is one of temporal consistency versus point-in-time precision. The data collection procedures in the March data are more stable since the mid-1970s, due to the shift from May to monthly ORG samples in 1979, the major CPS ORG redesign in 1994, and the changing frequency of hourly pay reporting. But, as

---

[13] Earnings imputations are not an issue confined to the May/ORG CPS. The earnings imputation rate in the March sample rose during the same period from 11.8 percent in 1993 to 13.5 percent in 1996. This increase appears less sizable and discontinuous than the increase in the CPS May/ORG from 15.3 in 1993 to 23.3 percent in the last quarter of 1995 (and 22.1 percent in calendar year 1996).

stressed by Lemieux (2005), the May/ORG, is likely to provide a more precise hourly earnings measure at each point in time. We use both March and May/ORG series throughout and compare the sensitivity of our results to the choice of sample.

B. *Trends in overall inequality*

Table 1 presents several summary measures of trends in hourly earnings inequality for years 1973 to 2003 from the CPS March and May/ORG samples (plotted in Figure 1). Overall male hourly 90/10 earnings inequality rose by 30.2 log points between 1973 and 2003 in the CPS May/ORG. Twenty-one log points of this rise occurred between 1973 and 1988 and an additional 9.1 log points between 1988 and 2003. Earnings inequality did not grow uniformly throughout this period. Between 1973 and 1988, the growth in upper and lower-tail inequality was approximately symmetric above and below the median; after 1988, it was entirely concentrated in the upper-tail. In fact, the May/ORG data show a 4.7 log point compression in lower-tail inequality between 1988 and 2003. Rising upper-tail inequality accounts for over three-quarters of the overall expansion of the male 90/10 log earnings ratio between 1973 and 2003.

Subsequent rows of Table 1 confirm that these patterns are also evident for male hourly earnings samples from the March CPS. As is the case with the CPS May/ORG, lower-tail inequality rises sharply from 1979 to 1988 then halts or reverses thereafter. Upper-tail inequality grows steadily throughout, ultimately accounting for over three-quarters of total male inequality growth from 1975 to 2003. One discrepancy between the data series is that the plateau of lower-tail inequality appears far more pronounced in the May/ORG than March samples (see also Figure 1). Whereas the May/ORG data show lower-tail compression after 1988, the March data indicate almost no movement between 1988 and 2003. This is an important issue to which we return in Section V, where we consider potential biases in the May/ORG versus March samples.

The asymmetric growth of earnings inequality is also evident for females (lower panel of Table 1). The rise in female 50/10 earnings inequality is somewhat larger than for males, however, and roughly half of the overall growth in female inequality occurs in the lower tail

(relative to two-thirds for males). As is the case for males, the March hourly data show less compression of lower tail inequality for females after 1988 than do the CPS May/ORG data. Whereas female lower-tail inequality ceased rising after 1988 in the CPS May/ORG data, the March series indicates a continued rise of lower-tail inequality during the 1990s, though at a much slower pace than during 1975 to 1988.

C. Trends in residual inequality

To measure trends in residual earnings inequality, we fit the following earnings model separately by sample, year and gender:

(1) $\ln w_i = \alpha + s_i + x_i + (s_i \times x_i) + e_i$.

In this equation, w_i is the log hourly or weekly wage of individual i , s_i is a vector of five schooling completion categories corresponding to less than high school, high school graduate, some college, bachelor's degree, and post-college schooling, x_i is a vector of thirteen potential experience categories ranging from 0 to 38 years in 3-year increments, and e_i is an error term. The inclusion of a complete set of interaction terms ($s_i \times x_i$) among schooling and experience categories allows for a very flexible earnings equation.[14] When equation (1) is fit to the May/ORG hourly data, we also include a dummy variable equal to one for workers who report their pay on an hourly basis, as well as a set of interactions between this dummy and the schooling dummies. Changes in the frequency of hourly pay reporting in the May/ORG CPS may affect measured residual inequality. Inclusion of the hourly pay control in equation (1) permits partial adjustment for this factor in the subsequent variance decomposition.

The right-hand panel of Table 1 and Figure 2 display the residual wage inequality series by sex. Residual wage inequality increased sharply in the 1980s for men and women regardless of

---

[14] In a subsequent section, we re-estimate the residual inequality model in equation (1) with a more conventional specification in which the 65 education-by-potential-experience interaction terms are replaced with a quartic in potential experience interacted with four education groups (high school dropout, high school graduate, some college, college or greater). The results for trends in residual inequality and its decomposition into 'price' and 'quantity' components are essentially identical.

the wage measures or sample used. Residual wage inequality began rising in the 1970s for males in the March CPS but shows little change in the 1970s for males in the May/ORG CPS.[15]

Consistent with the trends for overall inequality, trends in residual inequality did not move in parallel in the upper and lower tails of the earnings distribution. Residual lower tail inequality rose sharply prior to 1988 and slowed or declined thereafter. Upper tail inequality rose steadily from the mid-1970s forward for males and females in the March CPS samples. In the CPS May/ORG series, there is a parallel rise in upper-tail inequality from the mid-1970s forward. This rise plateaus or reverses over 1994 to 1999 and then resumes its upward trend in 1999 to 2003.

Tables 2 and 3 present the evolution of earnings inequality within narrow gender, education and experience groups. We tabulate inequality metrics for those with exactly college and exactly high school degrees at potential experience levels of 5, 10, 15 and 25 years. Because these groups are (a subset of) the gender-education-experience categories directly controlled (with dummies) in our 'residualizing' specification above, inequality within these cells corresponds precisely to the residual inequality construct used throughout our analysis.[16] We limit the comparison to the period of 1975 to 2003 when both data series are available. We pool three years of data at each point to enhance precision (hence comparisons are centered on 1976, 1988 and 2002).

Tables 2 and 3 reveal that the trends in aggregate lower and upper-tail residual inequality shown in Table 1 are robustly visible within detailed gender-experience-education cells. For example, in CPS May/ORG and March series, upper-tail inequality rises markedly for both genders and within most education and experience groups during both halves of the sample (the one exception being female high school graduates).

The data also confirm noticeable compression (or plateauing) of lower-tail inequality within detailed gender, education and experience groups. This compression is much more pronounced

---

[15] Our inability to remove allocated observations from the May/ORG sample in January 1994 through August 1995 likely accounts for the visible upward jump in residual inequality in 1994 followed by a downward jump in 1996.

[16] Standard errors in these tables are bootstrapped using 100 sample-weighted replications for each statistic. Appendix Tables 2a and 2b present the levels (rather than the changes) in these inequality measures for narrow gender-education-experience cells and shows higher wage dispersion for more experienced and educated groups.

among less-educated workers; in fact, the 50/10 earnings ratio among college males and females widened in the latter half of the sample, though less rapidly than in the earlier decade.

These detailed comparisons underscore several substantive points for our analysis. First, because the rise in upper-tail residual inequality is evident within detailed gender-experience-education cells, it cannot be ascribed exclusively (or even primarily) to changing labor force composition. Second, this rise is not likely to be a survey design artifact since it is confirmed by all data sources. Third, the fact that the compression of lower-tail inequality is concentrated among high school graduates and high school dropouts (not shown) has an important implication for our subsequent decomposition exercise: the roles attributed to prices and quantities will depend sensitively on the weight given to less educated and experienced groups (which experienced more compression) versus more educated and experienced groups (which experienced more expansion). This observation is especially important because, as above, the experience and education of the U.S. labor force rose substantially after 1980. Consequently, demographic groups that experienced the most residual compression after 1980 become increasingly uncommon as the sample progresses, and vice-versa for groups that experienced greater expansion. We explore these points in our decomposition.

II. Labor force composition and wage inequality: A quantile decomposition

To evaluate the contributions of changing labor market prices and quantities to the shifts in overall and residual inequality documented in Tables 1 through 3, we apply and extend a quantile decomposition technique proposed by Machado and Mata (2005, 'MM' hereafter). The MM technique uses quantile regressions to partition the observed distribution of earnings into 'price' components (wage coefficients) and 'quantity' components (labor force composition) and calculate, through simulation, the impact of each on changes in aggregate wage dispersion.

A. Theory

Let $Q_\theta(w \,|\, x)$ for $\theta \in (0,1)$ denote the θth quantile of the distribution of the log wage (w)

given the vector x of covariates. We model these conditional quantiles as

(2) $\qquad Q_\theta(w \mid x) = x'\beta(\theta)$,

where x is a $k \times 1$ vector of covariates and $\beta(\theta)$ is a conformable vector of quantile regression (QR) coefficients. For given $\theta \in (0,1)$, $\beta(\theta)$ can be estimated by minimizing in β,

(3) $\qquad n^{-1} \sum_{i=1}^{n} \rho_\theta(w_i - x_i'\beta)$ with

$$\rho_\theta(u) = \begin{cases} \theta u & \text{for } u \geq 0 \\ (\theta - 1)u & \text{for } u < 0 \end{cases},$$

where the latter expression is referred to as the 'check function' (see Koenker and Bassett 1978, Buchinsky 1994, Koenker and Hallock 2001, and Angrist, Chernozhukov and Fernández-Val forthcoming).

As discussed by Machado and Mata, if equation (2) is correctly specified, the conditional quantile process – that is, $Q_\theta(w \mid x)$ as a function of $\theta \in (0,1)$ – provides a full characterization of the conditional distribution of wages given x. In this case, realizations of w_i given x_i can be viewed as independent draws from the function $x_i'\beta(\theta)$ where the random variable θ is uniformly distributed on the open interval $(0,1)$. The conditional quantile model will hold exactly in a case where both location and scale depend linearly on the covariates (for example in the classical location shift model $w = x'\beta + \varepsilon$ where ε is a normal, *iid* error term). In more general cases, the conditional quantile model may provide a reasonable approximation to the true conditional quantile, and this approximation can generally be improved by specifying flexible functions of x when estimating $\beta(\theta)$. We show below that, in our application, the fit of the quantile model to the observed data is quite accurate.

Having fit the conditional quantile function, the estimated parameters can be used to simulate the conditional distribution of w given x via an application of the probability integral transformation: If U is a uniform random variable on $[0,1]$, then $F^{-1}(U)$ has the density F.

Thus, if $\theta_1, \theta_2, ..., \theta_j$ are drawn from a uniform $(0,1)$ distribution, the corresponding j estimates of the conditional quantiles of wages at x_i, $\hat{w}_i \equiv \{x_i'\hat{\beta}(\theta_i)\}_{i=1}^j$, constitute a random sample from the (estimated) conditional distribution of wages given x_i.

This simulation procedure characterizes the conditional quantiles of the data for any given x. It does not provide the marginal density of w. This is because the marginal density depends upon both the conditional quantile function, $\hat{\beta}(\theta)$, and the distribution of the covariates, $g(x)$.[17] To generate a random sample from marginal density of w, we can draw rows of data from $g(x)$ and, for each row x_i, draw a random θ_i from the uniform $(0,1)$ distribution. We then form $\hat{w}_i \equiv x_i\hat{\beta}'(\theta_i)$, which is a draw from the wage density implied by the model. By applying this procedure repeatedly, we can draw an arbitrarily large random sample from the desired distribution. This procedure – successively drawing from $g(x)$ and θ to form $\hat{w}$ – is equivalent to numerically integrating the estimated conditional quantile function $\hat{Q}_\theta(w \mid x)$ over the distribution of x and θ to form $f(\hat{w}) = \iint_{x,\theta} \hat{Q}_\theta(w \mid x)g(x)\partial x\partial\theta$.[18]

For the purposes of decomposition, the MM procedure has two attractive properties. First, the conditional quantile model conveniently divides the observed wage distribution into 'price' and 'quantity' components.[19] This is similar to a standard Oaxaca-Blinder procedure using OLS regression coefficients, with the key difference that the OLS model only characterizes the central tendency of the data (i.e., the conditional mean function, describing 'between-group' inequality). By contrast, the conditional quantile model characterizes both the central tendency of the data (in this case, the median) and the <u>dispersion</u> of the outcome variable conditional on x, i.e., the wage

---

[17] For simplicity, we treat $g(x)$ throughout as known.

[18] Because θ is uniformly distributed on the unit interval, $f(\theta) = 1$.

[19] As with all measures of 'residual' inequality, this division depends upon which $x's$ are included in the conditioning set. We return to this point below.

'residuals.' This feature is critical for estimating the impact of composition on the shape of the residual wage distribution.

Second, under the convenient (but economically unappealing) partial equilibrium assumption that aggregate quantities of skills in the labor market do not affect skill prices, we can use the conditional quantile model to simulate the impact of changing composition or prices on distribution of wages.[20] In particular, by applying the labor force composition data $g_t(x)$ from a given time period t to the price matrix $\hat{\beta}_\tau(\theta)$ from any other time period τ, we can simulate the counterfactual distribution of wages that would prevail if labor force composition were given as in time period t and labor market prices were given as in time period τ. Note that because the $\hat{\beta}_\tau(\theta)$ matrix describes the conditional <u>distribution</u> of wages for given values of x, this simulation captures the effects of composition on both between-group and residual inequality.

Although the Machado-Mata approach was not specifically developed for estimating counterfactual measures of <u>residual</u> inequality, it can be extended to this application. We define the coefficient vector $\hat{\beta}(50)$ as our measure of <u>between-group</u> inequality, and we refer to it as $\hat{\beta}^b \equiv \hat{\beta}(50)$. $\hat{\beta}^b$ serves a role akin to $\hat{\beta}_{OLS}$ in a conventional Oaxaca-Blinder decomposition. In the conventional application, $\hat{\beta}_{OLS}$ provides a measure of between-group inequality because it estimates the central tendency of the data conditional on x. In our application, the central tendency measure is the conditional median, estimated by $\hat{\beta}^b$. The difference between conditional means and medians is not empirically important in our application.

Following this logic, we define a measure of within-group inequality as the difference between the estimated coefficient vector $\hat{\beta}(\theta)$ and the median coefficient vector $\hat{\beta}^b$:

---

[20] In their analysis of the contributions of price and quantities to rising inequality during the 1980s, DiNardo, Fortin and Lemieux (1996) perform supply and demand adjustments for observed wages by education by experience cells (in addition to applying a kernel reweighting procedure). These supply-demand adjustments are shown by DFL to be quantitatively important, explaining 21 to 33 percent of the growth in male 90/10 log hourly earnings inequality between 1979 and 1988 (Tables III and V). This indicates that the independence assumption (i.e., that aggregate skill supplies do not affect skill prices) is likely to be far from innocuous.

(4) $\hat{\beta}^w(\theta) \equiv [\hat{\beta}(\theta) - \hat{\beta}^b]$ for $\theta \in (0,1)$.

Notice that, by construction, $\hat{\beta}^w(50) = 0$. Hence, the residual quantile coefficient matrix is

purged of 'between-group' inequality, and measures the expected dispersion of w at any given

value of x, holding the conditional median at zero. By applying the coefficient matrix $\hat{\beta}^w$ to the

distribution of covariates, $g(x)$, we can calculate the (estimated) dispersion of w that is

exclusively attributable to residual inequality. If, for example, $\hat{\beta}(\theta) = \hat{\beta}^b \;\forall\; \theta$, then residual

inequality is zero in this model.

　　To summarize, the (correctly specified) conditional quantile model provides a complete

characterization of the distribution of w as a function of three components: the distribution of

covariates, $g(x)$, the vector of between-group prices, $\hat{\beta}^b$, and the matrix of within-group

(residual) prices $\hat{\beta}^w$. We write $f_t(\hat{w}_t) \equiv f(g_t(x), \hat{\beta}_t^b, \hat{\beta}_t^w)$.

B. *Implementation and proof of concept*

　　We implement the augmented Machado and Mata quantile decomposition procedure in three

steps. We first estimate quantile coefficient vectors for each time period and gender. Using the

May/MORG and March samples we estimate a wage model equivalent to equation (1) above, fit

using quantile rather than OLS regressions. For each sample, year, and gender, we estimate

models for quantiles [0.1, 0.3,...,99.7,99.9] at intervals of one-fifth of a centile, with one

additional model fit for the median (quantile 50.0). These QR coefficients, $\hat{\beta}(\theta)$, provide the

'prices' for our simulation exercise. Here, $\hat{\beta}(\theta)$ is a $k \times m$ matrix of quantile regression

coefficients, where k is the number of elements (65 including the intercept in equation (1)) in x

and m is the number of quantiles (501) estimated in θ.

　　In the second step, we calculate the residual price vector $\hat{\beta}_t^w$ using equation (4). This yields

$\hat{\beta}_t^b$, a $k \times 1$ vector of between-group prices, and $\hat{\beta}_t^w$, a $k \times (m-1)$ matrix of within-group prices.

In the final step, we draw simulated data from the distribution $f(g_t(x), \hat{\beta}_t^b, \hat{\beta}_t^w)$ by applying the price matrices $\hat{\beta}_t^b, \hat{\beta}_t^w$ to $g_t(x)$.[21]

Under the maintained assumption that labor market prices and quantities can be treated as independent, this procedure may be used to simulate any counterfactual distribution of overall or residual earnings inequality simply by pairing the quantity series from one time period with the price series (overall and/or residual) from another.

Before applying this technique to simulate counterfactual distributions, we first check the performance of the model in replicating observed (actual) distributions of overall and residual inequality. To benchmark the model, we apply the QR coefficients $\hat{\beta}_t^b, \hat{\beta}_t^w$ to the quantity series, $g_t(x)$ from the contemporaneous time period to generate simulated wage distributions. The simulated series are shown in Appendix Figure 1. If the QR model were a perfect fit to the conditional wage distributions, these series would exactly overlap. In practice, the discrepancy between the actual statistic and the model-based simulations is small enough to be almost undetectable. Other distributional statistics for the simulated data, such as the 90/50, 50/10, and variance of wages, are also extremely close to the corresponding values for the source data.

To benchmark the performance of our simulation procedure for residual wage inequality, we compare simulated and actual 90/10 residual wage dispersion by year and gender in Appendix Figure 2. The series labeled 'Observed Median Reg Residual' is formed from median regressions of log hourly earnings on the covariates above. The series 'Simulated Median Reg Residual' presents the corresponding statistics for the simulated series formed using $\hat{\beta}_t^w$ and $g_t(x)$. The simulated residual series also fit the observed data quite accurately.

---

[21] Machado and Mata recommend drawing rows from $g(x)$ with replacement. An analysis by Angrist, Autor and Chernozhukov (2004) finds that greater precision can be obtained simply by multiplying the entire $g(x)$ distribution by the transposed $\beta(\theta)$ coefficient matrix. In practice, this means that our simulated data set is many times larger than the original $g(x)$ data (500 times larger, to be exact). Angrist, Autor and Chernozhukov (2004) experiment with using finer and coarser bandwidths for the quantile estimation and find that 500 quantile intervals provides excellent verisimilitude in CPS samples, while finer bandwidths produces little marginal gain.

Since almost all prior residual decompositions analyze OLS regression residuals, we also plot for comparison the 90/10 residual inequality series for an identically specified OLS model. As is visible in Appendix Figure 2, OLS and Median Regression residuals have nearly identical dispersion in our application, indicating that the distinction between mean and median regression is not substantively import for interpreting our residual decomposition results. These benchmarks provide assurance that our approach is able to capture relevant features of earnings distributions.

We next apply this decomposition to the inequality data. We first propose and implement a quantile regression version of the well-known Juhn, Murphy and Pierce (1993) decomposition. Subsequently, we use the quantile model to implement a variant of the extended DiNardo, Fortin and Lemieux (1996) kernel density reweighting approach. As we show below, these two approaches are nested in the quantile model.

III. The contributions of between-group prices, residual prices, and composition: A quantile JMP decomposition

As a compact method of summarizing the contributions of prices and quantities to rising earnings inequality, we implement a quantile analog of the well known Juhn, Murphy and Pierce (1993, "JMP") decomposition. In Appendix 2, we provide a more thorough discussion of the differences between our quantile approach and the canonical JMP decomposition. As before, we write the observed distribution of wages at time t as a function of three components: the distribution of worker characteristics, $g_t(x)$, the between-group prices for those characteristics, β_t^b, and the within-group prices for those characteristics, β_t^w (suppressing "hats" on the estimated β_t vectors).

The observed change in inequality between any two periods, t and τ can be decomposed into three components using the following sequential decomposition. Let $\Delta Q_\theta = Q_\theta(f_\tau(w)) - Q_\theta(f_t(w))$ equal the observed change in the θth wage quantile between periods t and τ. We define

(5) $\Delta Q_\theta^x = Q_\theta(f(g_\tau(x),\beta_t^b,\beta_t^w)) - Q_\theta(f(g_t(x),\beta_t^b,\beta_t^w))$

as the contribution of changing quantities (labor force composition) to ΔQ_θ. We define

(6) $\Delta Q_\theta^b = Q_\theta(f(g_\tau(x),\beta_\tau^b,\beta_t^w) - Q_\theta(f(g_\tau(x),\beta_t^b,\beta_t^w)$

as the marginal contribution of changing between-group prices to ΔQ_θ. And, we finally define

(7) $\Delta Q_\theta^w = Q_\theta(f(g_\tau(x),\beta_\tau^b,\beta_\tau^w) - Q_\theta(f(g_\tau(x),\beta_\tau^b,\beta_t^w)$

as the marginal contribution of changing within-group prices to ΔQ_θ.

Notice that this decomposition sums to the total observed change: $\Delta Q_\theta^w + \Delta Q_\theta^b + \Delta Q_\theta^x = \Delta Q_\theta$. This is an important advantage over the JMP procedure, in which the 'residual price and quantity component' must be estimated as a remainder term after the other two components are calculated.[22]

As is well known, the order of a sequential decomposition implicitly corresponds to a set of weights reflecting which characteristics are held at their start or end period values as other components are changed, with the share of any observed change attributed to each component $(\Delta Q_\theta^w, \Delta Q_\theta^b, \Delta Q_\theta^x)$ depending on which other components are varied first. We perform the decomposition above using two orderings. In the first, we first vary labor force characteristics, then between-group prices and finally within-group prices. In the second, we vary within-group prices, then between-group prices and finally labor force composition.

A. Quantile JMP results

Table 4 presents results of the quantile JMP decomposition for the CPS May/ORG hourly inequality data. As shown in the upper left panel of the table, the rise of male earnings dispersion between 1973 and 1988 is almost entirely accounted for by prices rather than quantities. Of 21.1 log points rise in male 90/10 inequality over this period, the decomposition attributes a small offsetting roll (-1.0 log points) to quantities, versus 16.8 points to between-group prices and 5.3

---

[22] The JMP decomposition will not generally "add up" unless the conditional distribution of residuals is homoskedastic – that is, residual dispersion does not depend on the *x's*. As Tables 2 and 3 demonstrate, this assumption is amply rejected by the data. See Appendix 2 for discussion.

points to within-group prices.[23] When we reverse the order of the decomposition, we find that the estimated contribution of labor force composition to inequality is within rounding error of zero, with only a slightly larger role attributed to within-group prices.[24] Panel B for females also finds a comparatively small role for composition in the rise of earnings inequality. But this role is larger than for males and also more pronounced in the lower tail of the earnings distribution. Similar patterns prevail when upper and lower tail inequality are decomposed separately: composition never explains more than a third of the growth of any 90/50 or 50/10 inequality metric during 1973 to 1988; the remainder is explained by prices.[25]

When we perform this decomposition for the growth of inequality between 1988 and 2003, an important contrast with the earlier period emerges. Both prices and composition play an economically significant role in explaining changes in lower-tail earnings inequality after 1988. Male 50/10 inequality contracted by 4.8 log points during 1988 to 2003. Holding prices constant, the change in composition alone would have been expected to increase lower tail inequality by 4.3 log points, reflecting, as noted by Lemieux (2005), the increasing education and experience of the labor force. Offsetting this force, between and within-group price changes reduced lower-tail inequality by -1.2 and -7.9 log points respectively during the same interval. On net, the 4.8 log points of observed lower-tail compression reflects the countervailing impacts of 'falling' prices and an increase in the prevalence of characteristics associated with higher earnings dispersion.

But these discrepant forces are not responsible for trends in upper tail inequality. In fact, the

---

[23] Small discrepancies between the tabulated changes in overall inequality in Table 1 and the changes shown in Tables 4 and 5 reflect simulation error, that is, instances where the quantile simulation does not perfectly capture the observed distribution. These discrepancies are quite small (never more than 5 percent of the observed change) and we do not make further adjustments for them.

[24] Note that when we decompose the change in inequality for the full 1973 to 2003 period (third panel of Tables 4), the subcomponents of the quantile decomposition do not sum to the subcomponents of the decomposition for the sub-periods 1973 to 1988 and 1988 to 2003. This is because the implicit decomposition weights differ between panels. In the final panel, we apply 1973 quantities to 2003 prices and vice versa; in the prior two panels, the dates for the quantities and prices do not differ by more than 15 years (i.e., 1973 vs. 1988 and 1988 vs. 2003).

[25] It is common in a sequential decomposition for the factors that are varied first (e.g., quantities or prices) to be credited with a larger contribution to the change in the outcome. That is not the case here; in fact the role of composition appears more important when we vary it third rather than first. The explanation seems to be that between-education and between-experience group gaps in wage variance became more pronounced during 1973 to 1988. Hence, when we first vary allow these 'prices' to vary, the subsequent quantity adjustments become more important.

contribution of changing labor force composition to rising upper tail inequality is negligible for both males and females. The rise of upper-tail inequality after 1988 is almost entirely explained by rising between and within-group prices, as was the case for the period of 1973 – 1988. These price increases are comparable in magnitude during 1973 through 1988 and 1988 through 2003 with the estimated rise in within-group prices for males is somewhat larger in the latter period.

Table 5 presents the comparable decomposition for hourly wage inequality in the March CPS data. The conclusions are qualitatively similar to those from Table 4. In particular, in the earlier period, the March data show a small role for composition (though larger for females than males) and a predominant role for rising between and within-group prices in explaining the parallel growth of upper and lower tail inequality. Like the May/ORG data, the March data reveal countervailing effects of prices and quantities in the lower half of the distribution and almost no role for quantities (i.e., composition) in the upper half of the distribution. Consistent with our descriptive results in Table 1, we find less price compression in the lower tail of the earnings distribution in the March data than in the May/ORG data.[26]

As a robustness test, we also repeated the estimation exercise in Tables 4 and 5 using a more conventional Mincer model featuring a quartic in potential experience – replacing the 13 dummy experience categories in our main model – interacted with four education categories. The substantive findings of this alternative model, reported in Appendix Tables 3a and 3b, are qualitatively identical to our main estimates.

These results demonstrate that a focus only on trends in 90/10 inequality (or in the variance of log wages) may lead to an incomplete view of the roles of prices and quantities in some periods. For example, the overall rise in male 90/10 hourly wage inequality in the CPS May/ORG over 1988 to 2003 was 7.9 log points. Adjusting for changes in composition during this time period and holding all prices at their 1988 level, we calculate that 90/10 inequality would

---

[26] A comparable analysis for the March full-time data (not shown) echoes these conclusions, though in the full-time data there is even less lower-tail compression than in either hourly sample.

counterfactually have risen by 4.8 log points. One could conclude from this exercise that two-thirds of the rise in inequality during 1988 to 2003 is "explained by" compositional shifts.

This conclusion, however, ignores a substantial contemporaneous rise in between-group prices. Holding residual prices constant, we estimate that rising between-group prices would have counterfactually raised 90/10 inequality by an <u>additional</u> 7.0 log points. Thus, between-group prices could be said to "explain" <u>90 percent</u> of the rise in 90/10 inequality. What reconciles these claims is that within-group prices compressed significantly over the same interval – reducing 90/10 inequality by 3.8 log points ("explaining" minus 48 percent of the phenomenon).

Although none of these statements is factually incorrect, a more precise statement of the findings in Tables 4 and 5 is that: (1) upper-tail inequality rose steadily over 1988 to 1993, a rise almost entirely explained by rising between and within-group prices; (2) lower-tail inequality plateaued or compressed during the same time interval due to the countervailing effects of changing composition (leading to more expansion) and changing between and within-group prices (leading to more compression).

IV. Quantile decomposition: An extended DFL approach

In this section, we extend our analysis to the decomposition of residual inequality. Methodologically, we demonstrate how the quantile decomposition technique may be used to unify alternative models for wage density decompositions (e.g., JMP, DFL and extensions). Substantively, we demonstrate how our core findings – i.e., that prices, not quantities are the predominant force explaining the growth of aggregate and residual inequality – can be reconciled with the conclusion of Lemieux (2005) that changes in labor force composition essentially explain the entire rise of residual inequality after 1988.

We begin with a quantile implementation of the extended DiNardo, Fortin and Lemieux (1996) procedure proposed by Lemieux (2002 and 2005). The original DFL procedure reweights observed wage densities to account for compositional (and other) shifts. Lemieux (2002 2005) extends this technique to the analysis of counterfactual <u>residual</u> wage densities. The extension

involves estimating flexible wage models similar to our equation (1) (using OLS rather than quantile regressions), assigning each observation its regression residual, and reweighting the distribution of $x's$ to account for compositional shifts between any two time periods. To the extent that reweighting generates a counterfactual residual distribution that approximates the observed change in the residual distribution, the extended DFL procedure ascribes the observed changes to compositional shifts.[27]

In our notation, the Lemieux approach (superscripted by *L* below) estimates the following quantity:

$$(8) \qquad \Delta Q_\theta^L = Q_\theta(f(g_\tau(x), \beta^b = 0, \beta_t^w)) - Q_\theta(f(g_t(x), \beta^b = 0, \beta_t^w)) .$$

Here, the counterfactual contribution of changing labor force composition to the change in the θth quantile of the residual distribution is estimated by replacing the x distribution from period t with x distribution from period τ while holding residual prices at their base (period t) level and setting between-group prices to zero.

Inspection of equation (8) reveals that the Lemieux/DFL calculation is a subcomponent of the full quantile Juhn-Murphy-Pierce decomposition.[28] More specifically, it is identical to the first component (ΔQ_θ^x) of the three component quantile JMP decomposition above (and hence straightforward for us to implement using the quantile tool) except that the calculation sets β^b to zero in both time periods, thus suppressing the role of between-group prices. The utility of this extension is that it exclusively analyzes the contribution of labor force composition to changing residual inequality, abstracting from the role played by between group prices and their interaction with labor force composition.[29] A potential limitation of this partial-decomposition is that the

---

[27] Lemieux also proposes a variance decomposition that explicitly models the conditional heteroskedasticity of regression residuals and applies this to the U.S. residual distribution. This decomposition is not suited for analyzing inequality measures other than the wage variance (e.g., 90/10, 90/50, etc.) and we do not implement it here.

[28] It is not a subset of the original JMP decomposition, however, which does not account for the heteroskedasticity of the residual wage distribution (see Appendix 2).

[29] The DFL approach does not explicitly model the role of prices in shifting the overall or residual wage distribution. Rather, the contribution of prices is calculated as a residual – that is the component that remains after the quantity

allocation of total wage variation into 'between-group' and 'residual' components is inherently arbitrary since it depends upon the set of *x's* included in the conditioning set. The sum of between and within-group inequality is, however, invariant to the conditioning set.

A. *Extended DFL decomposition: Results for overall inequality*

For comparison with the prior results, we first implement the Lemieux-DFL analysis for overall inequality using the May/ORG data. We evaluate the importance of compositional shifts by applying the labor force composition data, $g_t(x)$, from each sample year to the price series (β_t^b, β_t^w) from four different years: the contemporaneous year – thus producing that year's observed level of inequality – and the price series for years 1973, 1988 and 2003. Aggregate inequality statistics (90/10, 90/50, 50/10) for these counterfactual densities are plotted in Figure 4 (see also Table 6). The differences in the vertical height of each series within a given year in the figure reflect the effect of between and within-group prices on earnings inequality, holding labor force composition at the appointed year's level. The over-time change in the level of each series (moving along the *x*-axis) reflects the effect of changes in labor force composition, holding prices at their 1973, 1988 or 2003 level.

Two results stand out in Figure 4. First, there is a large rise in overall earnings inequality that is not attributable to compositional changes. The magnitude of the compositional effects – reflected in the shallow upward slope of the counterfactual series – is small relative to the substantial rise in inequality observed when comparing the price changes over the three decades of the sample. For example, the male 90/10 log hourly earnings ratio rose by 29 log points between 1973 and 2003. Holding composition at the 1973, 1988 or 2003 levels respectively, we estimate that this earnings ratio would counterfactually have risen by 22 to 25 log points.

Second, compositional and price shifts play distinct roles in the upper and lower tails of the

---

reweighting is implemented. Leibbrandt, Levinsohn and McCrary (2005) propose an extension to the DFL approach that reweights the wage distribution for changes in prices as well as quantities, though it also does not differentiate between 'residual' and 'between-group' prices.

wage distribution. As is visible in the middle two panels of Figure 4, the rise of upper tail inequality over the entire three decades is almost exclusively a consequence of changing between and within-group prices. When evaluated at any point along the *x*-axis (i.e., using 1973 to 2003 $x's$), the counterfactual rise in upper-tail inequality – measured by the vertical difference between the three price series – is (about) equally pronounced during both halves of the sample.

The comparable analysis for lower-tail inequality, shown in the lower panels of Figure 4, also reveals an important role for prices in the plateauing of 50/10 inequality after 1988. Here, the role of prices is the opposite of that above; price compression substantially reduced lower tail inequality after 1988 – seen by the lower height of the 2003 counterfactual price series relative to the 1988 series – a fact that is also robust to the choice of the quantity series for any year. Labor force composition also played a role – by offsetting lower-tail price compression. Using any of the three counterfactual price series (1973, 1988, 2003), the simulated 50/10 log earnings ratio slopes gradually upward from the late 1980s, indicating that changing labor-force representation raised observed 50/10 inequality above what would have occurred had prices alone compressed.

B. *Extended DFL decomposition: Results for residual inequality*

We now use the extended MM quantile tool to implement the Lemieux/DFL decomposition for residual inequality. The results of this decomposition for the May/ORG sample at once confirm the Lemieux findings and, we believe, substantially alter their interpretation.

The top panel of Figure 5 shows that holding composition constant at the 1973, 1988 or 2003 level, 90/10 residual inequality rose sharply between 1973 and 1988 (compare the height of the 1973 versus 1988 series). Between 1988 and 2003, however, residual 90/10 inequality contracted by about 15 to 30 percent of its original rise, holding composition constant. This confirms the finding of Lemieux (2005) that residual inequality plateaued or contracted after 1988. But the vertical differences among these three counterfactual series at each point along the *x*-axis reveal that changing 'residual prices' are primarily responsible for the rise and then contraction in residual inequality in the first and second halves of the sample. These shifts in residual inequality

occur with composition constant. What role did composition play? As may be seen by shallow upward slopes of each counterfactual series (moving along the *x*-axis), compositional shifts also contributed to rising residual inequality (holding residual prices constant), particularly after 1988. But these compositional shifts are modest relative to the price effects.

How can these modest effects be reconciled with the claim that compositional changes explain the full rise in residual inequality after 1988? The answer may be seen by studying the actual net rise in estimated residual inequality from 1988 to 2003 (series labeled "observed residual"). The observed rise in residual inequality over 1988 to 2003 is quite small; we estimate it at 0.4 log points (Table 7). Accordingly, the relatively modest contribution of compositional shifts to rising residual inequality can be said to explain all (in fact, more than all) of the observed rise in residual inequality for this period. But what Figure 5 reveals is that the offsetting effects of price contractions were economically as or more important than these compositional shifts.

These points apply with even greater force to upper and lower tail residual inequality treated separately. As shown in the lower two panels of Figure 5, composition-constant lower-tail inequality unambiguously contracted between 1988 and 2003. This is entirely a price effect; composition effects work in the opposite direction. By contrast, upper-tail residual inequality continued to rise for both genders between 1988 and 2003 – and substantially so for males, as it did in the prior half of the sample. In fact, only when labor force weights from 1973 are used in place of labor force weights from 1988 or 2003 does this rise appear muted for males and reversed for females.

What explains the sensitivity to the choice of weights? Recall from Tables 2 and 3 that 90/10 inequality rose among better educated workers after 1988 but fell among less educated workers during the same interval. Since the labor force in 1973 was composed of considerably fewer highly educated workers and considerable more high school graduates and dropouts than in 1988 or 2003, the use of 1973 labor force characteristics puts substantial weight on the groups that experienced falling inequality and puts proportionately less weight on the groups that experienced

26

rising inequality, suggesting a smaller rise in counterfactual residual inequality than the same comparison made using 1988 or 2003 characteristics. Under the maintained assumption that prices and quantities are independent, no one weighting scheme is correct. But the sensitivity of conclusions about trends in inequality to the choice of weights cautions researchers against focusing exclusively on one set of labor force characteristics.

In summary, we find that composition plays only a secondary role in explaining the time patterns of overall and residual inequality in the CPS May/ORG; the ongoing rise of upper-tail inequality and the rise and then stagnation of lower-tail inequality are both primarily accounted for by changing labor market prices.

C. Comparison with the March CPS

The right-hand panels of Tables 6 and 7 present counterfactual simulations for March hourly earnings inequality, and Figures 6 and 7 plot these counterfactuals. The pattern of inferences is quite similar for March versus May/ORG data. But the estimated contribution of composition to the rise of overall or residual inequality is considerably smaller in these data – though not negligible – as would be expected from earlier tables. The principle reason for this discrepancy is that lower-tail inequality (both overall and residual) contracted by considerably less in the March than May/ORG data. Placing greater weight on less educated workers in the March data (by holding composition constant at an early period level) has a comparatively lesser offsetting effect on the observed rise of overall or residual inequality during this period.

This finding is qualitatively consistent with Lemieux (2005), who also reports that the March data reveal a smaller role for composition and a correspondingly larger role for prices in the growth of residual earnings. Interestingly, in Lemieux's analysis, the discrepancy between the May/ORG and March results for the role of composition is far greater than we report here. We believe the source of the difference is that Lemieux's analysis decomposes the variance of wage residuals while ours decomposes wage quantiles (90/10) and their subcomponents. As shown in Appendix Tables 2a and 2b, there are substantially greater discrepancies between the March and

May/ORG data for variance than quantile measures, and this difference becomes particularly pronounced between 1988 and 2002. The particularly high variance of the wage residual measures in the March data appear likely to reflect the fact that hourly wages are calculated in these data using prior year's earnings divided by weeks and earnings. The quantile measures, which exclude the tails of the distribution, are likely less sensitive to outliers generated by the construction of hourly wages in the March CPS.

Is there reason to believe that one series is more reliable than the other? Lemieux argues that the May/ORG CPS sample provides a more accurate measure of the hourly wage distribution and hence is a preferred data source for analysis of U.S. wage inequality. As discussed in Section I, we believe the March and May/ORG data both have advantages and disadvantages and provide complementary sources of information on the evolution of U.S. wage inequality. .

V.   Bias in the May/ORG versus March CPS data: Levels or trends?

A key issue in comparing wage inequality trends is whether the differences in measurement error in the May/ORG and March CPS have remained stable over time. As noted earlier, the share of workers who report being paid by the hour has increased substantially since 1973, both in aggregate and within education/experience groups (Hamermesh 2002; Lemieux 2005).

Lemieux (2005) points out that a growth in the share of hourly workers will tend to lead to a growth in measured residual hourly wage inequality in the March CPS relative to the May/ORG as the share of workers with well-measured hourly wages increases in the May/ORG. Although this growth in hourly wage reporting may make the May/ORG CPS a more accurate measure of wage inequality at a point in time, this does not necessarily mean it becomes a preferred measure of trends in wage inequality relative to the March CPS.

A simple variance decomposition can help clarify these issues. Let m index the March CPS, o index the ORG/May CPS, j index sex-education-experience group, t index year, h index hourly workers (those paid by the hour), and n index non-hourly workers. We define V_{jt}^m and V_{jt}^o

as the observed variances of hourly wages in the March and May/ORG CPS respectively for

workers in group j in year t. We also define σ_{jt}^2 as the true variance in hourly wages for group j

in year t; σ_{hm}^2 as the variance of measurement error in hourly wages for hourly workers in March

CPS; and σ_{nm}^2 as the variance of measurement error in hourly wages for non-hourly workers in

March CPS. The measurement error variances for hourly and non-hourly workers in the

May/ORG CPS are analogously given by σ_{ho}^2 and σ_{no}^2. We also let H_{jt} equal the share of hourly

workers in group j in year t.

Under the assumption of white noise measurement error in hourly wages, the observed

variances of hourly wages for group j in year t in the March and May/ORG CPS can be written

as:

$$
\begin{aligned}
(9) \quad & V_{jt}^m = \sigma_{jt}^2 + H_{jt}\sigma_{hm}^2 + (1-H_{jt})\sigma_{nm}^2, \\
& V_{jt}^o = \sigma_{jt}^2 + H_{jt}\sigma_{ho}^2 + (1-H_{jt})\sigma_{no}^2.
\end{aligned}
$$

Based on the estimates from matched March-May/ORG samples presented in Lemieux (2005),

we assume that hourly wages are more accurately measured in the May/ORG CPS than in the

March CPS: $\sigma_{hm}^2 > \sigma_{ho}^2$. We also assume that measurement error in the hourly wages for non-

hourly workers is essentially equivalent in the March and May/ORG CPS: $\sigma_{nm}^2 = \sigma_{no}^2$. These

assumptions imply, as Lemieux notes, that the March CPS generates higher (and less accurate)

measures of residual hourly wage inequality in levels than does the May/ORG CPS: $V_{jt}^m > V_{jt}^0$. It

is also reasonable to assume that hourly wages are more accurately measured for hourly workers

than non-hourly workers in the May/ORG samples: $\sigma_{no}^2 > \sigma_{ho}^2$. Although there is little reason to

believe in differential measurement error for the hourly wages of non-hourly and hourly workers

in the March CPS, Lemieux's estimates imply $\sigma_{hm}^2 > \sigma_{nm}^2$.

We next consider the accuracy of the different CPS samples for measuring changes in

residual wage inequality from period t to period τ when the share of workers paid by the hour is

rising – as has been the case since the 1970s. Let $\Delta x_t = x_t - x_\tau$. The changes in measured residual

hourly wage variance for group j from t to τ in the different CPS samples are given by

$$
\begin{aligned}
\text{(10)} \quad & \Delta V_{jt}^m = \Delta \sigma_{jt}^2 + \Delta H_{jt} (\sigma_{hm}^2 - \sigma_{nm}^2), \\
& \Delta V_{jt}^o = \Delta \sigma_{jt}^2 + \Delta H_{jt} (\sigma_{ho}^2 - \sigma_{no}^2),
\end{aligned}
$$

so that

$$
\text{(11)} \quad \Delta V_{jt}^m - \Delta V_{jt}^o = \Delta H_{jt} [(\sigma_{hm}^2 - \sigma_{ho}^2) + (\sigma_{nm}^2 - \sigma_{no}^2)].
$$

Under the assumptions given above, this difference between March and May/ORG is positive

is rising in the share of workers paid by the hour: $\Delta V_{jt}^m - \Delta V_{jt}^0 = \Delta H_{jt}(\sigma_{hm}^2 - \sigma_{ho}^2) > 0$ if $\Delta H_{jt} > 0$.

Thus, as emphasized by Lemieux (2005), a rise in the share of hourly workers should

systematically lead to a rise in measured residual wage inequality in the March CPS relative to

the May/ORG CPS.

But in contrast to the conclusion drawn by Lemieux (2004), this divergence between the

series does not imply that the May/ORG CPS provides a more accurate measure of changes in

actual residual wage inequality. Rather, equation (11) implies that the May/ORG CPS provides a

downward biased measure of changes in actual residual wage inequality ($\Delta V_{jt}^o < \Delta \sigma_{jt}^2$) when the

share of hourly workers is rising (since $\sigma_{ho}^2 < \sigma_{no}^2$). In contrast, the March CPS provides an

upward biased measure of changes in actual residual wage inequality if $\sigma_{hm}^2 > \sigma_{nm}^2$.

To provide a rough direct assessment of the bias that the rise of hourly pay reporting may

induce in the May/ORG CPS, we plot in Figure 8 observed trends in residual wage inequality in

the May/ORG CPS alongside counterfactual trends in residual inequality, holding the frequency

of hourly wage reporting at its base (1974) level.[30] This counterfactual series is formed using the

extended MM quantile simulation procedure, reweighting the data in each year to maintain the

frequency of hourly pay reporting at its start of period level within the gender by education by

---

[30] We increase the precision of this simulation by using the mean frequency of hourly pay reporting within education
by experience by gender cells for 1973, 1974 and 1975.

experience cells used in our analysis. No other adjustments are made for compositional changes in the counterfactual series. Hence, the contrast between the observed and counterfactual series is exclusively due to hourly pay reporting.

To provide a more detailed look at the hourly reweighting bias, Figure 9 plots the gap (by gender) between estimated 90/10 residual inequality and the hourly-reporting-corrected series alongside a chain-weighted index of the evolution of hourly pay reporting in our sample.[31] Figure 9 indicates that the rise in hourly pay reporting has probably led to a non-negligible downward bias in measured trends in residual inequality in the May/ORG CPS. For males, this bias cumulates over 1973 to approximately 1989, plateaus to 1994, and then appears to increase again between 1997 and 2002. For females, the bias is relatively stable after 1979 but then begins a sharp rise in 1991.[32] The further rise in the hourly-pay bias in the MORG samples after 1994 is particularly puzzling since the frequency of hourly pay reporting in the May/ORG plateaued in the early 1990s. This suggests that the difference between the precision of hourly and non-hourly wage reports became more pronounced after the 1994 CPS redesign.

Notably, the hourly-pay bias in the CPS May/ORG can potentially be 'controlled' by making adjustments for hourly pay composition comparable to the ones we apply for education and experience. In fact, our simulation procedure for the May/ORG data makes this adjustment throughout. Specifically, we include an indicator variable for hourly pay reporting, interacted with each education category, in the conditioning set used for the quantile analysis (both overall and residual) and we reweight this covariate along with other $x's$ in the simulation. It is therefore possible that our analysis of residual inequality may correct for the hourly-pay bias in residual inequality trends in the May/ORG sample. However, our procedure is unlikely to provide an accurate adjustment if the relationship between hourly pay and earnings precision in the CPS

[31] In the base year, this index is equal to the observed frequency of hourly pay reporting. In each subsequent year, we add to it the weighted sum of year-over-year change in hourly pay-reporting in each education-by-experience group cell, where the weights are the employment share of each cell over the prior two years.

[32] Notably, the bias for females is particularly pronounced in the lower-tail of the distribution. This suggests that part of the plateau in female 50/10 inequality after 1989 is an artifact of changes in hourly earnings reports.

survey undergoes a discontinuous change – as may have occurred in 1994.

In summary, although the May/ORG CPS provides a more accurate measure of the <u>level</u> of residual hourly wage variance than the March CPS, our analysis suggests that there are reasons to suspect the March CPS may provide an upward biased measure of <u>changes</u> (and trends) in residual hourly wage inequality while the May/ORG may provide a downward biased estimate.[33]

Lemieux further argues and presents evidence using the matched March-May/ORG samples that measurement error in hourly wages within hourly and non-hourly workers increased over the past three decades in both the March and May/ORG samples with the rise being larger in the March CPS for males and of similar magnitude in the two samples for females. These estimates assume white noise measurement error that is uncorrelated across the two samples and assume that both are attempting to capture the same true measure of hourly earnings. But there are real changes in the labor market and structure of compensation that could generate actual divergences in earnings inequality measures based on a comprehensive measure of annual earnings from the March samples versus a straight-time hourly wage rate or usual weekly earnings measure from the May/ORG samples. For example, the growth of bonuses and other contingent pay in the 1990s is more likely to be captured in the March data than the May/ORG earnings measure (Lebow, Sheiner, Slifman, and Starr-McCluer 1999; Lemiuex, MacLeod, and Parent 2005). Changes in the importance of transitory versus permanent components of earnings could also reduce the true covariance between annual and point-in-time measures of earnings. Furthermore, most of Lemieux's estimated growth in the relative measurement error in the March versus the ORG samples comes in the 1990s in the period overlapping the major 1994 revision of the CPS that more greatly affected the ORG data.

We therefore do not concur with Lemieux's (2005) suggestion to researchers to use the

---

[33] The divergences of inequality trends in the March and May/ORG samples are smaller for samples limited to full-time workers and to analyses of full-time, weekly earnings as indicated in Appendix Table 1. The persistent rise in upper-tail wage inequality combined with the post-1988 stagnation of lower-tail inequality is also apparent for full-time weekly earnings. A substantial rise in residual wage inequality after 1988 is apparent in both the March and May/ORG samples when examining full-time weekly earnings.

May/ORG CPS rather than the March CPS to study the evolution of U.S. wage inequality over recent decades. To be clear, both CPS samples have strengths and limitations. And there are many actual labor market changes and data processing changes – ranging from top coding, imputations of missing wage information, changes in the wage and hour questions – that may differentially the actual earnings and the measurement error in wage measures in the different CPS samples. Researchers would be wise to combine information from both sources for measuring changes in the U.S. wage structure.

VI.   Conclusion

This paper set out to evaluate one potential explanation for the diverging trends of U.S. inequality during the 1990s. Our analysis tested whether the rising experience and education of the U.S. labor force may have led to a mechanical (i.e., not due to price changes) rise in earnings dispersion from 1990 forward. We fail to confirm this hypothesis. Shifts in labor force composition <u>have</u> affected earnings inequality during the 1990s. But these compositional shifts have primarily operated on the lower tail of the earnings distribution and, moreover, have primarily served to <u>mute</u> a contemporaneous, countervailing lower-tail price compression. By contrast, the steady rise of upper tail inequality since the early 1980s appears almost entirely explained by ongoing between-group and residual price changes.

The source of the asymmetric rise in earnings inequality with a steady rise in upper-tail wage inequality and some evidence of flat or declining lower-tail wage inequality (particularly after making adjustments for composition) since the mid-1980s remains poorly understood. This pattern is suggestive of a "polarization" of the labor market with a particularly strong market for workers in the top part of the skill distribution, deterioration in market conditions for workers in the middle, and reasonably steady market conditions for those near the bottom. Autor, Levy, and Murnane (2003) postulate that computerization could have such non-monotonic impacts on the demand for skill: sharply raising demand for the cognitive and interpersonal skills used by educated professionals and managers; reducing demand for clerical and routine analytical skills

that comprised many middle-educated white collar jobs; and reducing demand for routine manual skills of many previously high-paid manufacturing production jobs. But computerization has probably had little impact on the demand for the non-routine manual skills used in many 'low-skilled' (in-person) service jobs such as health aides, orderlies, cleaners, servers, etc. This framework suggests that computerization – as well as the increased international outsourcing (off-shoring) of routine cognitive and manual tasks – may have raised demand for skill among high-skill workers, depressed demands for 'middle-skill' workers, and left the bottom of the wage distribution comparatively unscathed. Goos and Manning (2003) call such a process the "polarization of work," and argue that it may have contributed to a similar hollowing out of the wage distribution in the United Kingdom during 1975 to 2000

We finally stress that the informativeness of the counterfactual exercises undertaken in this study depends critically on the maintained partial-equilibrium assumption that prices and quantities can be treated as independent. If this assumption is substantially violated, the counterfactual inequality series developed will not correspond to any potential state of the world since altering labor force characteristics will necessarily alter the prices associated with those characteristics. Lending weight to this concern, the careful analyses of earnings premiums by experience, age and education group performed by Card and Lemieux (2001) and Borjas (2003) strongly affirm that the relative wages paid to education, experience and age groups are quite responsive to own supply as well as aggregate supply. Hence, shifts in composition and shifts prices will generally negatively covary, particularly over long time intervals – such as the three decades above. This underscores that the impact of changing composition on prices could potentially be a first order source of bias in these counterfactual exercises. Assessing this bias remains an important item for analysis.

VII. References

Angrist, Joshua, David Autor and Victor Chernozhukov. 2004. "Quantile Wage Decompositions: Estimation and Inference, with an Application to the U.S. Minimum Wage." Work in Progress, Massachusetts Institute of Technology.

Angrist, Joshua, Victor Chernozhukov and Iván Fernández-Val. Forthcoming. "Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure."*Econometrica*.

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney. 2005. "Trends in U.S. Wage Inequality: Re-Assessing the Revisionists." MIT Mimeograph, September.

Autor, David H., Frank Levy, and Richard J. Murnane. 2003. "The Skill Content of Recent Technological Change: An Empirical Investigation." *Quarterly Journal of Economics* 118 (November): 1279-1333.

Blau, Francine D. and Lawrence M. Kahn. 1997. "Swimming Upstream: Trends in the Gender Wage Differential in 1980s." *Journal of Labor Economics*, 15(1), Part 1, January, 1 – 42.

Blinder, Alan. 1973. "Wage Discrimination: Reduced Forms and Structural Estimation." *Journal of Human Resources*. 8, 436 – 455.

Borjas, George J. 2003. "The Labor Demand Curve *Is* Downward Sloping: Reexamining the Impact of Immigration on the Labor Market." *Quarterly Journal of Economics* 118 (November): 1335-74.

Buchinsky, Moshe. 1994. "Changes in the U.S. Wage Structure 1963 – 1987: Application of Quantile Regression." *Econometrica*, 62(2), 405 – 458.

Card, David and John E. DiNardo. 2002. "Skill-Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles." *Journal of Labor Economics* 20 (October): 733-83.

Card, David and Thomas Lemieux. 2001. "Can Falling Supply Explain the Rising Return to College for Younger Men?" *Quarterly Journal of Economics* 116 (May): 705-46.

Cohany, Sharon R., Anne E. Polivka and Jennifer M. Rothgreb. 1994, "Revisions in the Current Population Survey Effective January 2004," Current Population Survey Technical Documentation CPS-94, Attachment 5, Bureau of the Census: Washington, DC.

Cutler, David M. and Lawrence F. Katz. 1992. "Rising Inequality? Changes in the Distribution of Income and Consumption in the 1980s." *American Economic Review* 82 (May): 546-51.

DiNardo, John, Nicole Fortin and Thomas Lemieux. 1996. "Labor Market Institutions, and the Distribution of Wages, 1973-1992: A Semiparametric Approach." *Econometrica* 64 (September), 1001-1044.

Goos, Maarten and Alan Manning. 2003. "Lousy and Lovely Jobs: The Rising Polarization of Work in Britain." Unpublished paper, Center for Economic Performance, London School of Economics, September.

Hamermesh, Daniel. 2002. "12 Million Salaried Workers are Missing." *Industrial and Labor Relations Review,* 55(4), July, 649 – 666.

Heckman, James J., Lance J. Lochner and Petra E. Todd. 2003 "Fifty Years of Mincer Earnings Regressions." NBER Working Paper No. 9732, May.

Hirsch, Barry T. and Edward J. Schumacher. 2004. "Match Bias in Wage Gap Estimates due to Earnings Imputation." *Journal of Labor Economics*, 22(3), July, 689 – 722.

Jaeger, David A.. 1997. "Reconciling the Old and New Census Bureau Education Questions: Recommendations for Researchers," *Journal of Business and Economics Statistics*, 15, 300–309.

Juhn, Chinhui, Kevin M. Murphy and Brooks Pierce. 1993. "Wage Inequality and the Rise in Returns to Skill." *Journal of Political Economy*, 101(3), 410 – 442.

Katz, Lawrence F. and David H. Autor. 1999. "Changes in the Wage Structure and Earnings Inequality." In O. Ashenfelter and D. Card, eds. *Handbook of Labor Economics*, volume 3, North Holland.

Katz, Lawrence F. and Kevin M. Murphy. 1992. "Changes in Relative Wages, 1963-87: Supply and Demand Factors." *Quarterly Journal of Economics* 107 (February): 35-78.

Karoly, Lynn and Gary Burtless. 1995. "Demographic Change, Rising Earnings Inequality, and the Distribution of Well-Being, 1959-1989." *Demography* 32 : 379-405.

Koenker, Roger and Kevin F. Hallock. 2001. "Quantile Regression." *Journal of Economic Perspectives*, 15(4), Fall, 143 – 156.

Koenker, Roger and Gilbert Bassett Jr. 1978. "Regression Quantiles," *Econometrica*, 46(1), 33 – 50.

Lebow, David E., Louise Sheiner, Larry Slifman, and Martha Starr-McCluer. 1999. "Recent Trends in Compensation Practices." FEDS Working Paper No. 99-32, Federal Reserve Board, July.

Lemieux, Thomas. 2002. "Decomposing Changes in Wage Distributions: A Unified Approach," *Canadian Journal of Economics*, 35(4), November, 646-688.

Lemieux, Thomas. 2005. "Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill?" University of British Columbia, May.

Lemieux, Thomas, W. Bentley MacLeod, and Daniel Parent. 2005. "Bonus Pay and Wage Inequality." University of British Columbia, April.

Leibbrandt, Murray, James Levinsohn, and Justin McCrary. 2005. "Incomes in South Africa Since the Fall of Apartheid." NBER Working Paper No. 11384, May.

Machado, José and José Mata. 2005. "Counterfactual Decompositions of Changes in Wage Distributions Using Quantile Regression." *Journal of Applied Econometrics*, 20(4), 445-65.

Melly, Blaise. Forthcoming. "Decomposition of Differences in Distribution Using Quantile Regression," *Labour Economics.*

Mincer, Jacob. 1974. *Schooling, Experience, and Earnings.* New York: NBER.

Mishel, Lawrence, Jared Bernstein, and Heather Boushey. 2002. *The State of Working America: 2002-03*. Ithaca, ILR Press.

Murphy, Kevin M. and Finis Welch. 1992. "The Structure of Wages." *Quarterly Journal of Economics* 107 (February): 285-326.

Oaxaca, Ronald. 1973. "Male-Female Wage Differentials in Urban Labor Markets." *International Economic Review*, 14, 693 – 709.

Park, Jin Heum, "Estimation of Sheepskin Effects and Returns to Schooling Using the Old and the New CPS Measures of Educational Attainment," Princeton University, Industrial Relations Section Working Paper No. 338, December 1994.

Polivka, Anne. 2000. "Using Earnings Data from the Monthly Current Population Survey." Mimeograph, Bureau of Labor Statistics, October.

Polivka, Anne E. and Jennifer M. Rothgreb. 1993. "Redesigning the CPS Questionnaire," *Monthly Labor Review*, September, 10 – 28.

Appendix 1: Data

A. *Basic processing of May/ORG CPS data*

We use the May CPS for 1973 to 1978 and the CPS Merged Outgoing Rotation Groups for years 1979 to 2003. All samples include wage/salary workers ages 16 to 64 with 0 to 39 years of potential experience in current employment. Earnings weights are used in all calculations. Full-time earnings are weighted by CPS sampling weights. Hourly earnings are weighted by the product of CPS sampling weights and hours worked in the prior week. Full-time earnings are the logarithm of reported usual weekly earnings. In all years, hourly wages are the logarithm of reported hourly earnings for those paid by the hour and the logarithm of usual weekly earnings divided by hours worked last week (not usual weekly hours) for non-hourly workers. We use hours last week instead of usual weekly hours because usual weekly hours is not consistently available: starting with the CPS redesign in 1994, workers who report that their weekly hours vary are not asked to report usual weekly hours, yielding a non-report rate of 7.0 to 8.5 percent of workers in 1994 to 2003. To check sensitivity, we have tabulated and plotted overall and residual inequality measures using imputed usual weekly hours in place of hours last week in all years 1973 – 2003. This has little impact on our results.

Topcoded earnings observations are multiplied by 1.5. Full-time earnings of below \$67/week in 1982\$ (\$112/week in 2000\$) and hourly earners of below \$1.675/hour in 1982 dollars (\$2.80/hour in 2000\$) are dropped, as are hourly wages exceeding 1/35th the topcoded value of weekly earnings. All earnings numbers are deflated by the chain-weighted (implicit) price deflator for personal consumption expenditures. Allocated earnings observations are excluded in all years, except where allocation flags are unavailable (January 1994 to August 1995). As discussed by Hirsch and Shumacher (2004), only about 25 percent of allocated observations in the MORG CPS are flagged as allocated between 1989 and 1993. Following Lemieux (2004), we identify and drop non-flagged allocated observations by using the unedited earnings values provided in the source data.

B. *Basic processing of March CPS data*

We use the March Current Population Survey for earnings years 1975 to 2003 for workers age 16 to 64 (during the earnings year) with 0 to 39 years of potential experience whose class of work in their longest job was private or government wage/salary employment. Hourly earnings are calculated as annual earnings divided by the product of weeks worked and usual hours in the prior year. Full-time, full-year workers are those who work 35 hours per week (using the Census Bureau's full-time worker flag) and worked 40+ weeks in the previous year. Full-time weekly earnings are calculated as the logarithm of annual earnings over weeks worked for the full-time, full-year sample. Allocated earnings observations are excluded using family earnings allocation and individual earnings allocation flags. Weights are used in all calculations. Full-time earnings are weighted by the product of the CPS sampling weight and weeks worked. Hourly earnings are weighted by the product of the CPS sampling weight, weeks worked, and hours worked in the prior year.

Prior to March 1989, all wage and salary income in the March CPS was reported in a single variable, which was topcoded at values between \$50,000 and \$99,999 in years 1975 to 1988. For these cases, we multiply the topcoded earnings value by 1.5, following Katz and Murphy (1992). Commencing in 1989, wage and salary incomes were collected in two separate earnings variables, corresponding to primary and secondary labor earnings. After adjusting for topcoding, we sum these values to calculate total wage and salary earnings. Topcodes after 1988 are handled

as follows. For the primary earnings variable, topcoded values are reported at the topcode maximum up to 1996. We multiply these values by 1.5. Starting in 1996, topcoded primary earnings values are assigned the mean of all topcoded earners. In these cases, we simply reassign the topcoded value and, again, multiply by 1.5. For the secondary earnings value, the topcoded maximum is set at 99,999 from 1989 to 1995 and then falls to 25,000 in 1996 to 2003, and rises to 37,000 in 2004. For lack of a better alternative, we again use the topcoded value multiplied by 1.5.

After making adjustments for topcoding, full-time earnings of below $67/week in 1982$ ($112/week in 2000$) and hourly earners of below $1.675/hour in 1982 dollars ($2.80/hour in 2000$) are dropped, as are hourly wages exceeding 1/35th the topcoded value of weekly earnings.

C. *Coding of education and potential experience in CPS samples*

To attain comparable educational categories across the redefinition of Census Bureau's education variable introduced in 1992 in the CPS, we use the method proposed by Jaeger (1997). In CPS samples coded with the pre-1992 education question, we defined high school dropouts as those with fewer than 12 years of completed schooling; high school graduates as those having 12 years of completed schooling; some college attendees as those with any schooling beyond 12 years (including an incomplete 13th year) and fewer than 16 completed years; college graduates as those with 16 or 17 years of completed schooling; post-college as those with more than 17 years of completed schooling. In CPS samples coded with the revised education question, we define high school dropouts as those with fewer than 12 years of completed schooling; high school graduates as those with either 12 completed years of schooling and/or a high school diploma or G.E.D.; some college as those attending some college or holding an Associate's Degree; college as those with exactly a baccalaureate degree; and post-college as those with a post- baccalaureate degree.

To calculate potential experience in data years coded with the revised education question, we use figures from Park (1994) to assign years of completed education to each worker based upon race, gender and highest degree held. Years of potential experience were calculated as age minus assigned years of education minus 6, rounded down to the nearest integer value.

Appendix 2: Comparison of traditional and quantile JMP decompositions

JMP propose a model for wage inequality as a function of the distribution of covariates, $g(x)$, the vector of between-group prices ($\beta's$) and the cumulative distribution of the residual, $F(\theta)$, where $\theta \in [0,1]$. JMP implement this model by fitting OLS regressions to estimate β and using the resulting wage residuals to form $F^{-1}(\theta)$, which is the inverse cumulative empirical residual wage distribution. In JMP's notation, individual $i's$ wage, w_{it} can be written as:

$w_{it} = x_{it}'\hat{\beta}_t + \hat{F}_t(\theta_i)$ where $\theta_{it} = F_t^{-1}(e_{it})$ and e_{it} is $i's$ wage residual from the regression. JMP use this methodology to decompose changes in the wage distribution over 1963 to 1989 into components due to changes in prices ($\Delta\beta$), changes in quantities ($\Delta g(x)$), and changes in the residual distribution ($\Delta F(\theta)$). The principle virtue of the JMP decomposition over the standard Blinder-Oaxaca method is that it can potentially be applied to simulate any feature of a counterfactual wage distribution. Indeed, JMP's principle focus is on earnings inequality among various wage quantiles: 90/10, 90/50 and 50/10.

The JMP methodology has two shortcomings that are solved by the quantile model. The first is that the OLS wage regression at the heart of the JMP technique provides a model for the conditional <u>mean</u> of the wage distribution and its results do not extend naturally to wage quantiles. For example, there is no presumption that $F_t(50) = 0$, and so it need not be the case that $\overline{y}_t = \overline{x}_t'\beta_t + F_t(50)$. This inconsistency is not present in the quantile model, which provides a uniform treatment of between and within-group price measurement. In practice, this inconsistency may not be substantively important, however.

A more significant shortcoming of the method is that the components of the JMP counterfactual decomposition need not sum to the total observed change. Specifically, adjusting the wage distribution in period t for the change in each of the three accounting components between periods t and τ ($\Delta\beta$, $\Delta g(x)$ and $\Delta F(\theta)$), will not generally recover the observed wage distribution in period τ. The reason is that knowledge of the marginal distribution of $x'\beta$ and $F(\theta)$ is not generally sufficient to characterize the marginal distribution of $w = x'\beta + F(\theta)$. Because $x'\beta$ and $F(\theta)$ are random variables, the distribution of their sum, $g(w_i = x_i'\beta + F(\theta_i))$ depends on both their variances and covariance (i.e., their joint distribution). This covariance is not accounted for by the JMP decomposition unless $F(\theta)$ is allowed to depend explicitly on x.[34]

There is one special case where the marginal distribution of w is recoverable from the marginal distributions of $x'\beta$ and $F(\theta)$. This is when the distributions are independent. Under classical OLS assumptions – in particular, a normally distributed, homoskedastic error term – this independence condition will hold. But as underscored by Lemieux (2005), this error structure is at odds with the predictions of human capital theory. Moreover, the evidence in Tables 2 and 3 clearly demonstrates that trends in residual inequality differ markedly by education and experience groups. Hence, the homoskedastic error structure is strongly rejected for the U.S. wage distribution.[35]

---

[34] Lemieux (2002) provides a thorough discussion of these points.

[35] JMP recognize that the 'adding up' property does not hold in their decomposition and recommend calculating the contribution of the 'residual price and quantity component' as a remainder term, after accounting for observed prices and quantities.

Unlike the original JMP decomposition, the quantile JMP decomposition does "add up" because the full conditional distribution of wages is allowed to depend upon x through β^b, β^w. Neither independence nor homoskedasticity is assumed.

Figure 1. Hourly Wage Inequality in May/ORG and March CPS Series, 1973 - 2003
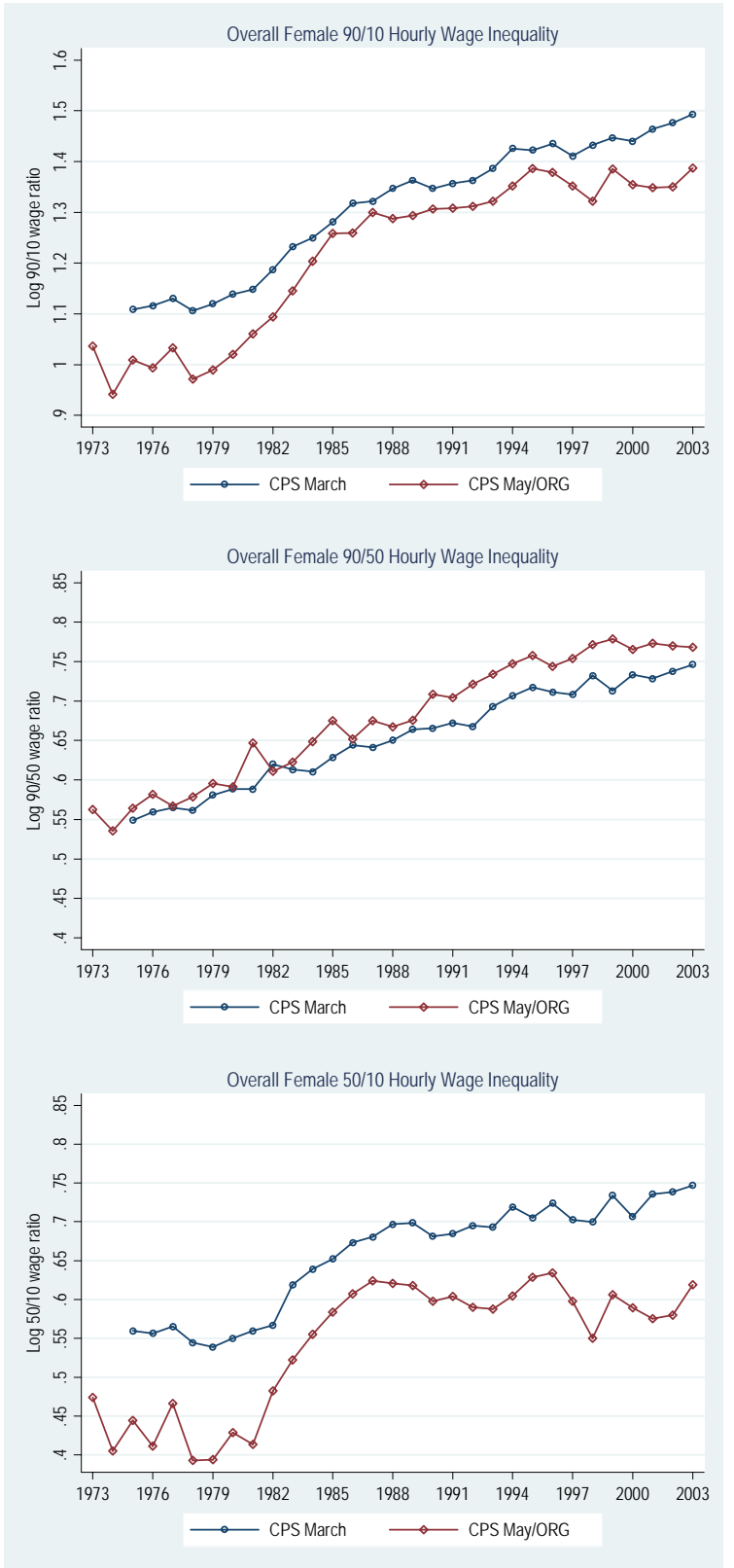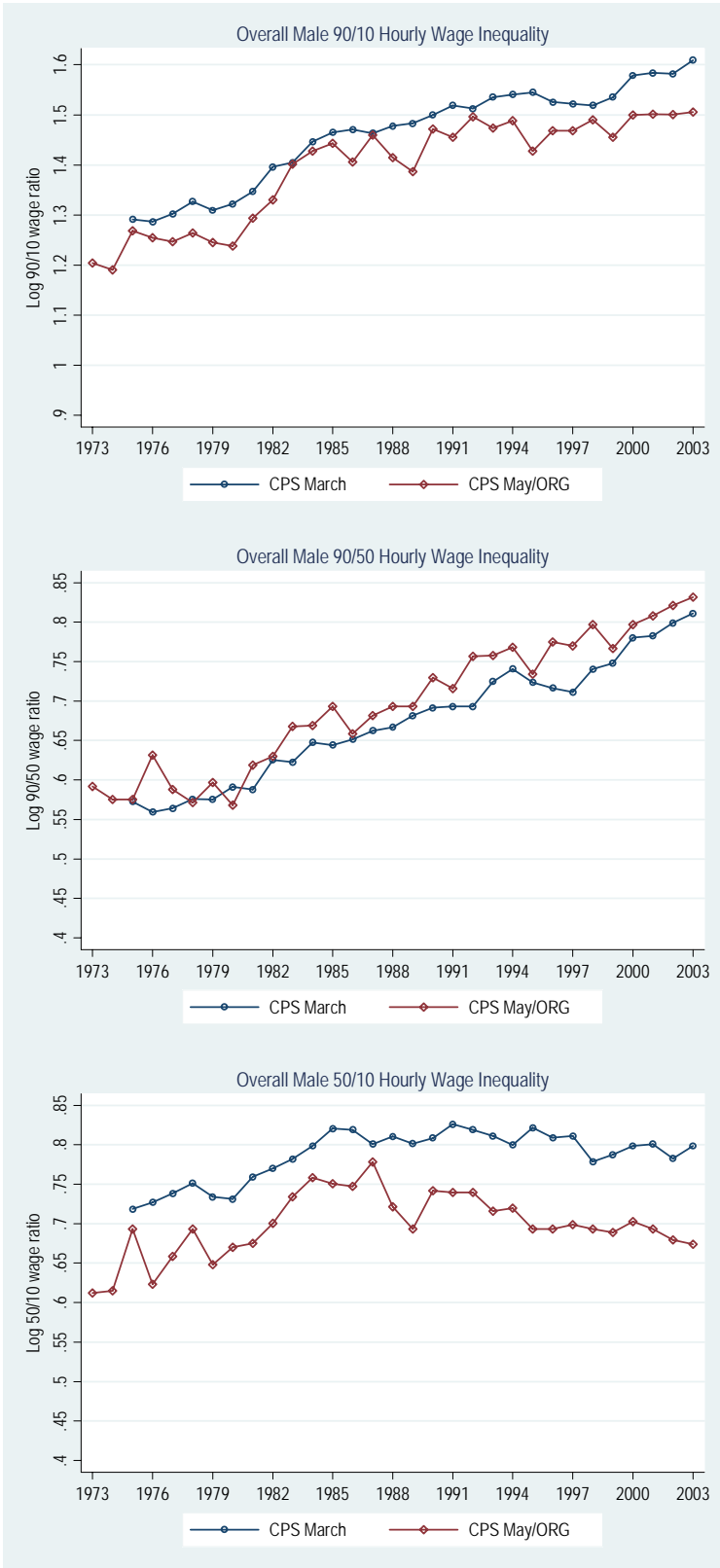
Figure 2. Hourly Residual Wage Inequality in May/ORG and March CPS Series, 1973 - 2003
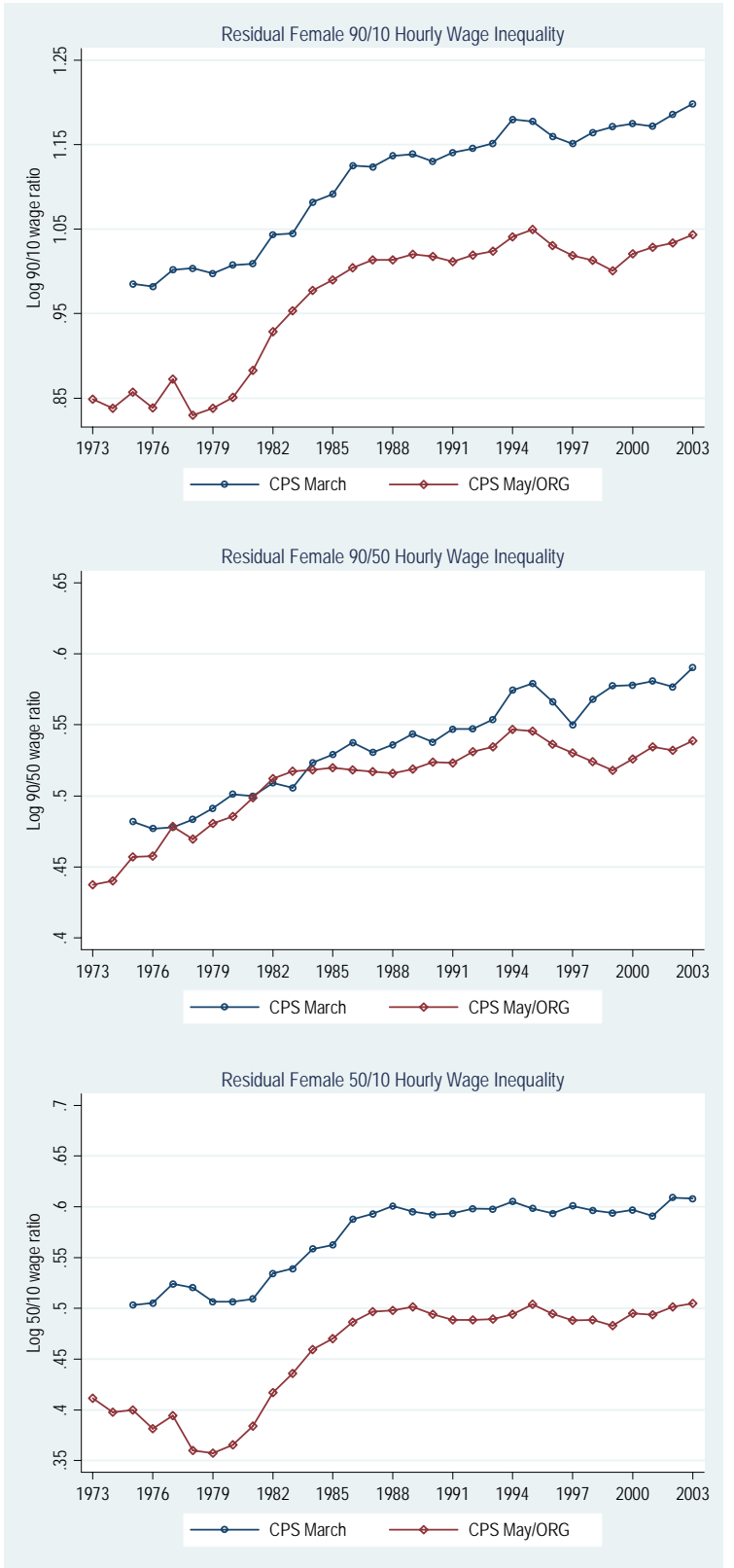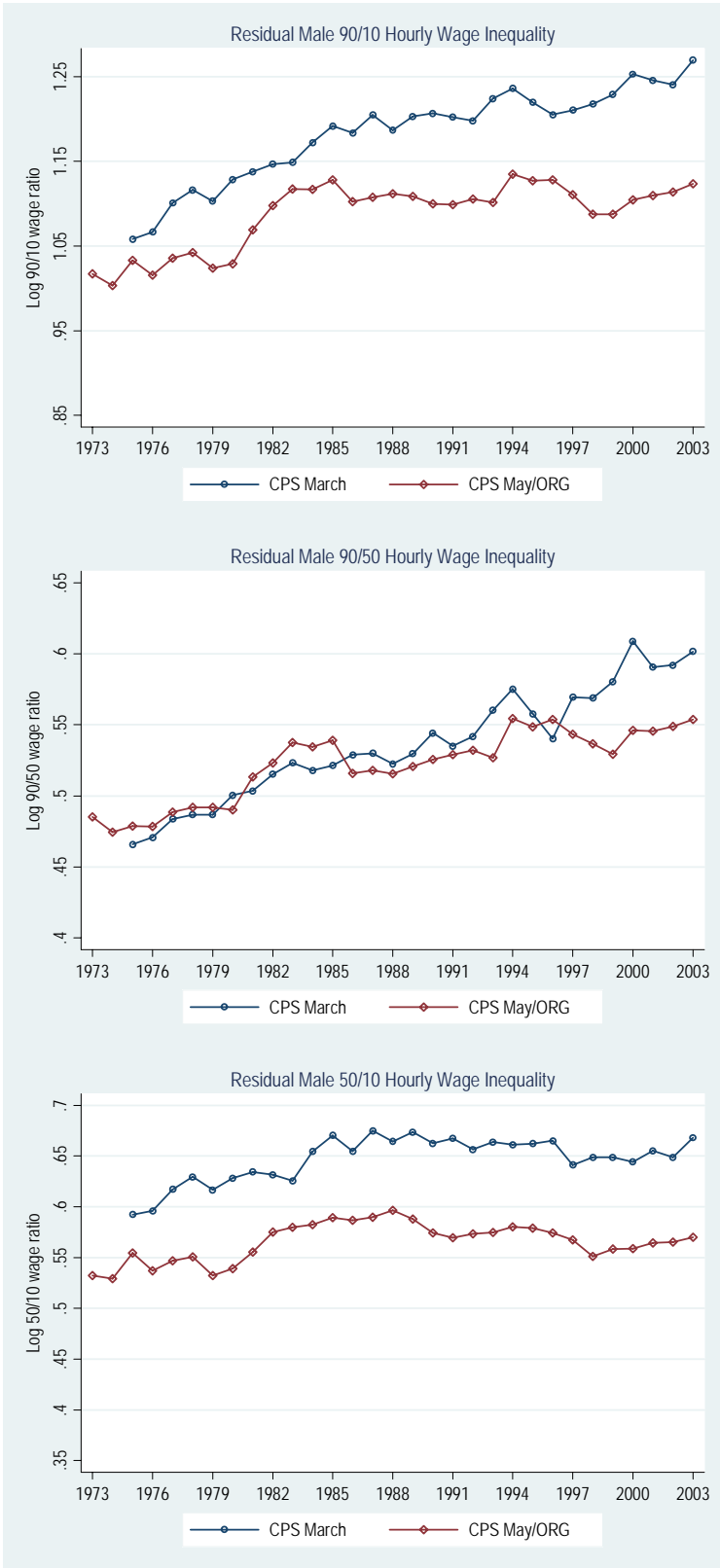
Figure 3. Education and Experience of U.S. Labor Force, 1973 - 2003 (CPS May/ORG)

A. Full-Time Equivalent Employment Shares by Education
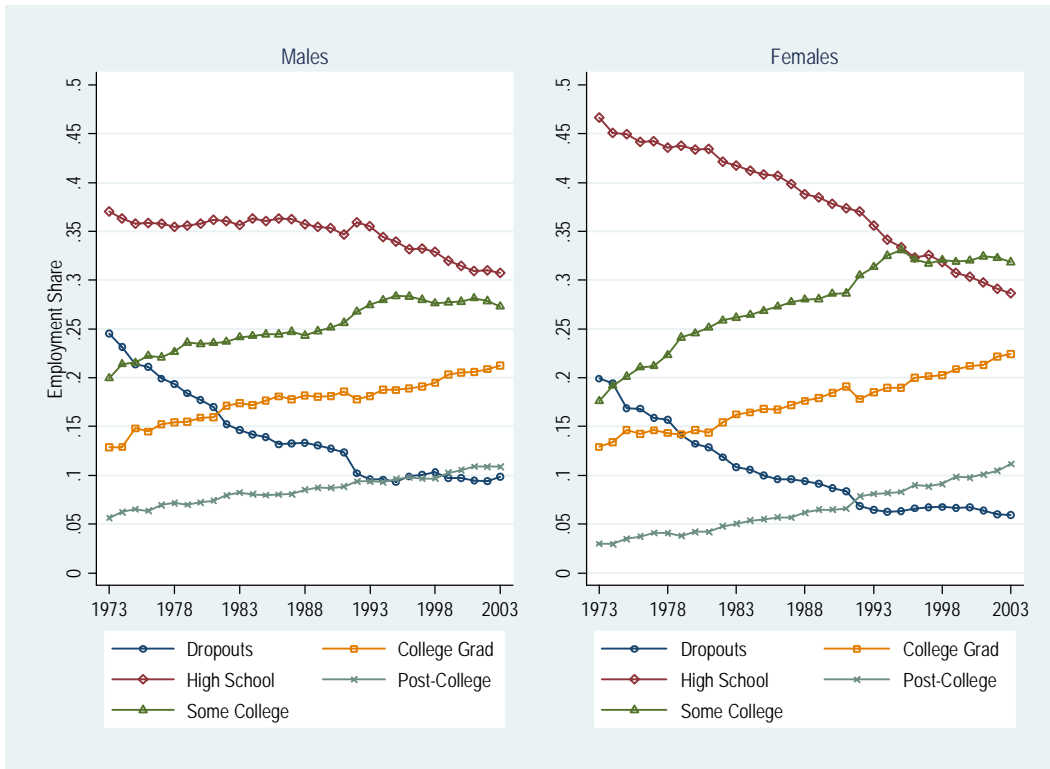
B. Mean Potential Experience of Employed Workers by Education
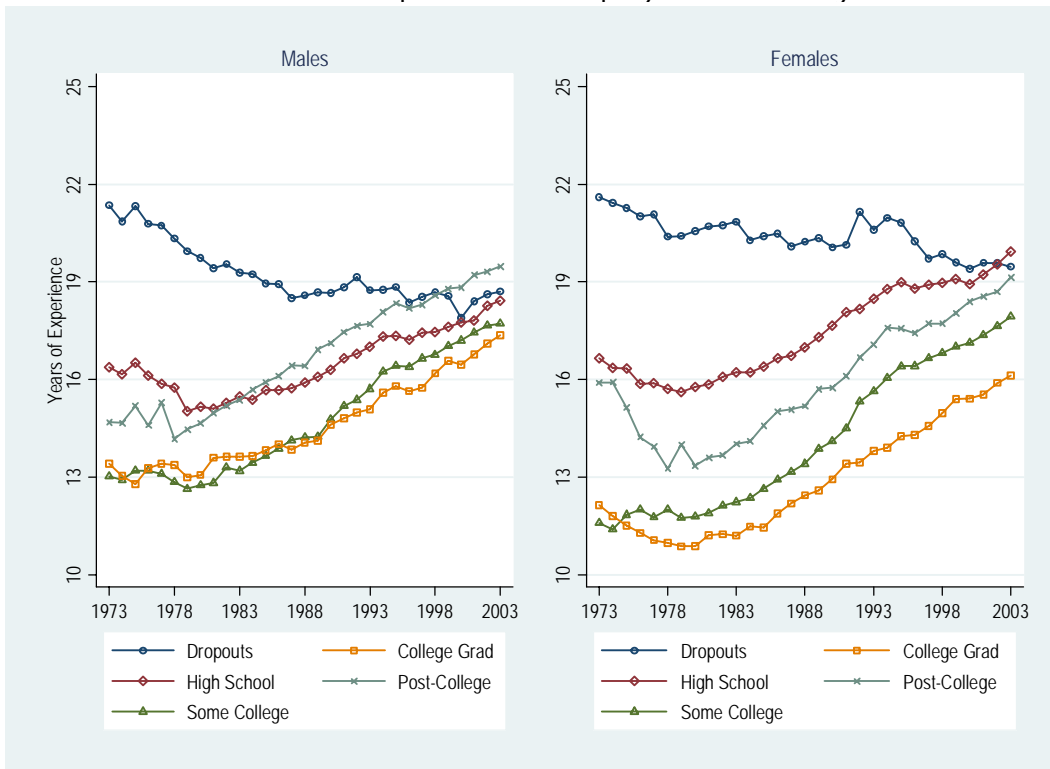
Figure 4. Actual and Counterfactual Overall Hourly Wage Inequality 1973-2003: CPS May/ORG
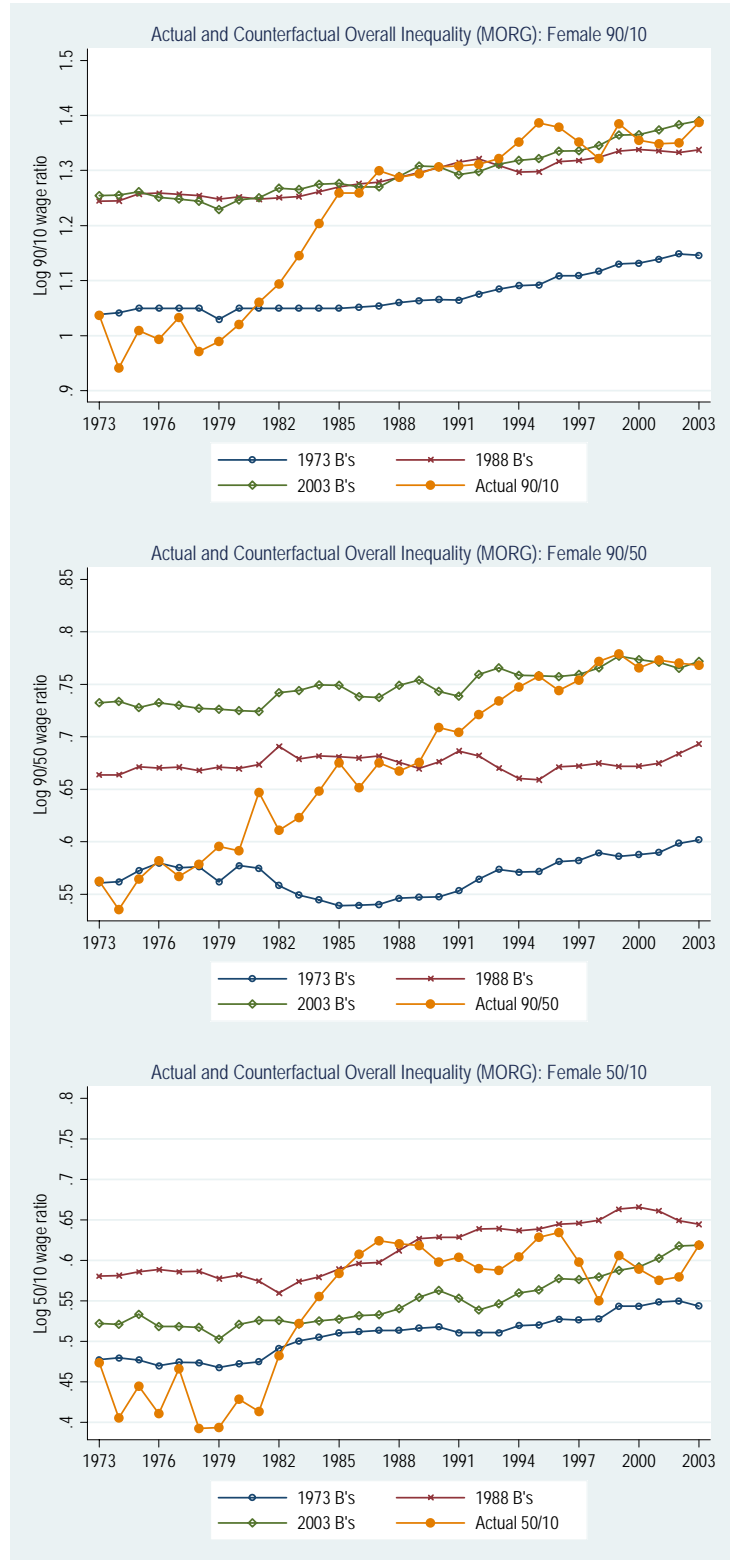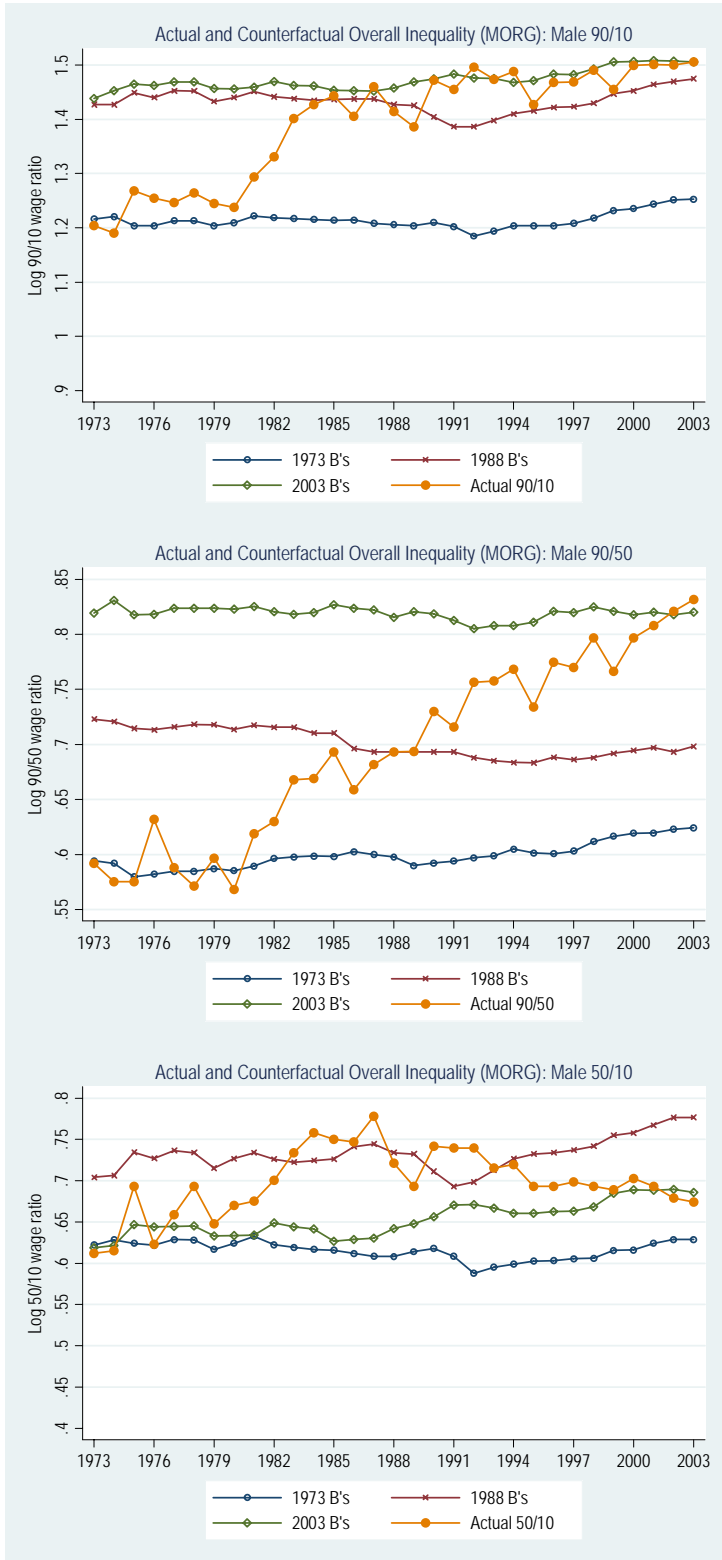
Figure 5. Actual and Counterfactual Residual Hourly Wage Inequality 1973-2003: CPS May/ORG
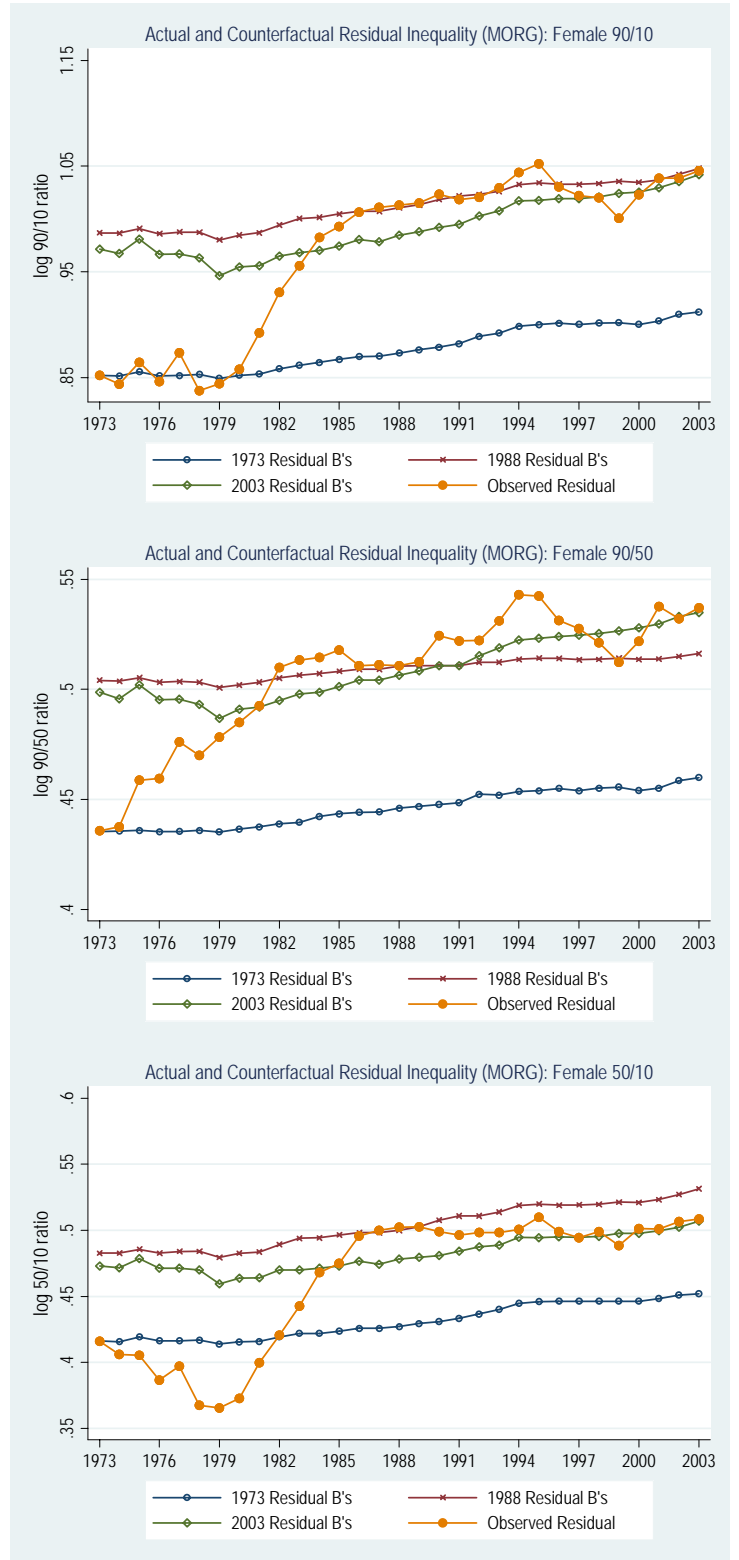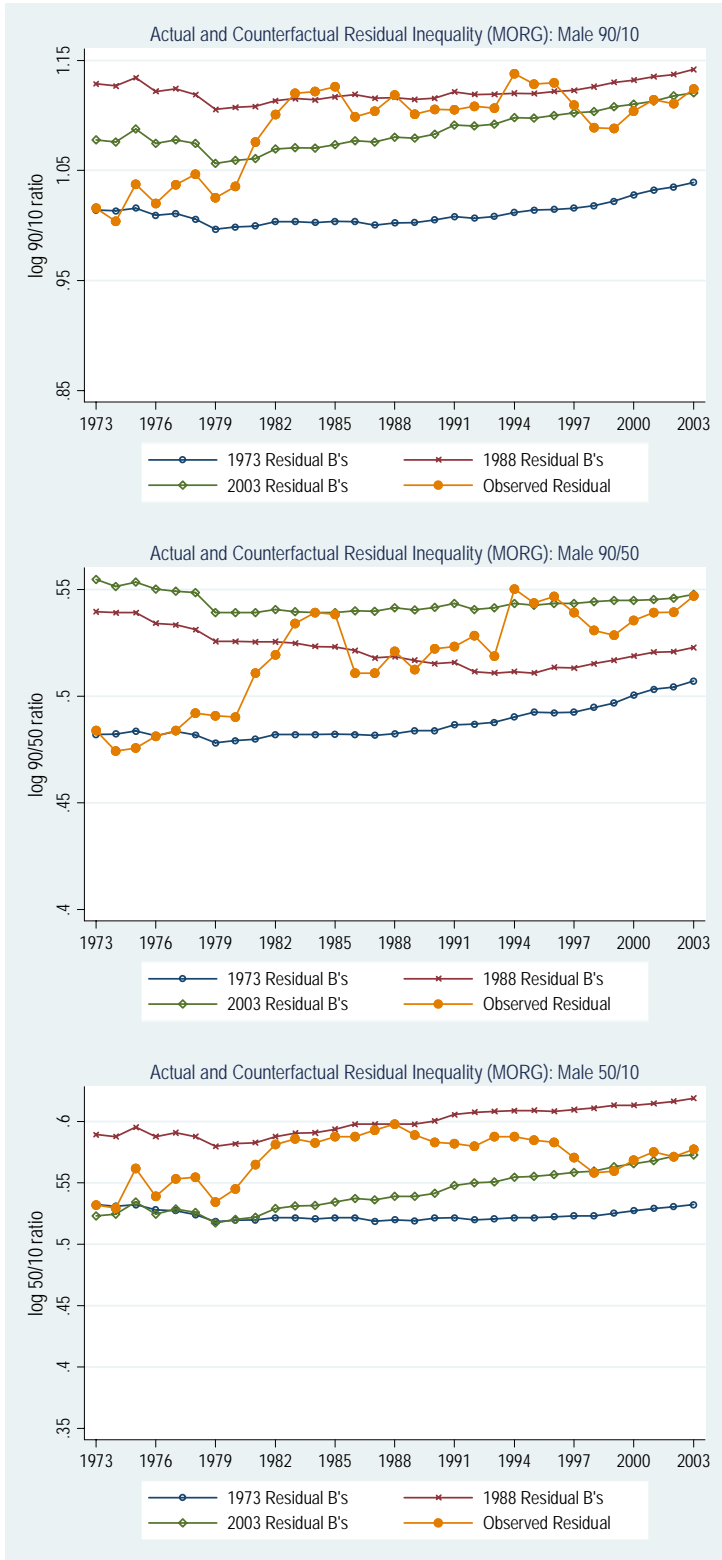
Figure 6. Actual and Counterfactual Overall Hourly Wage Inequality 1975-2003: CPS March
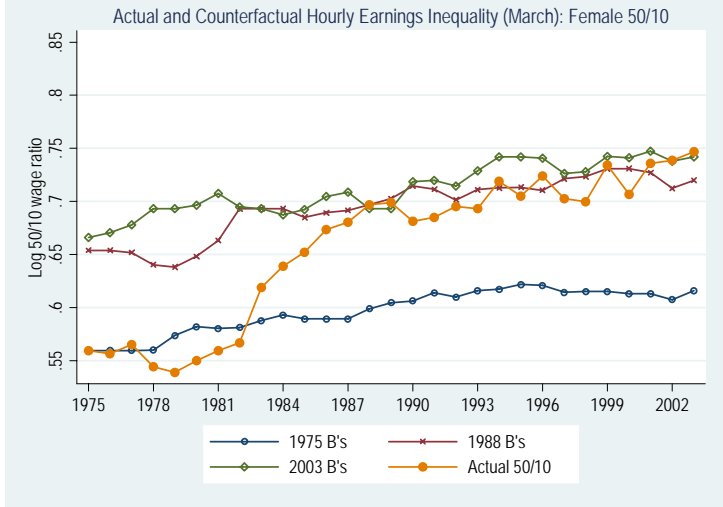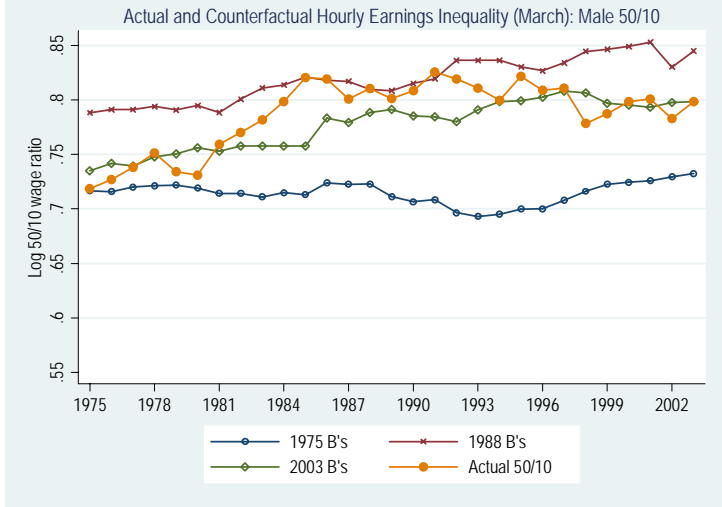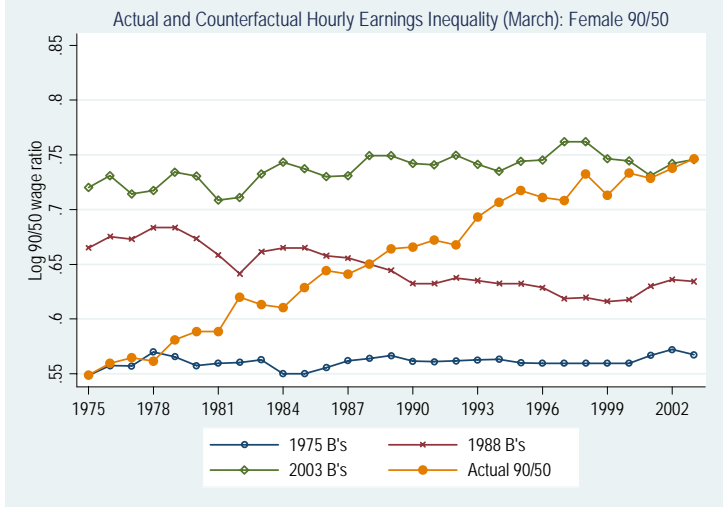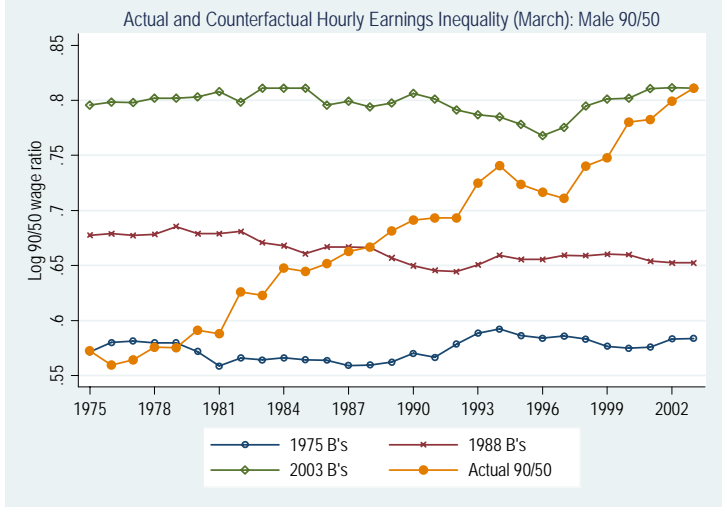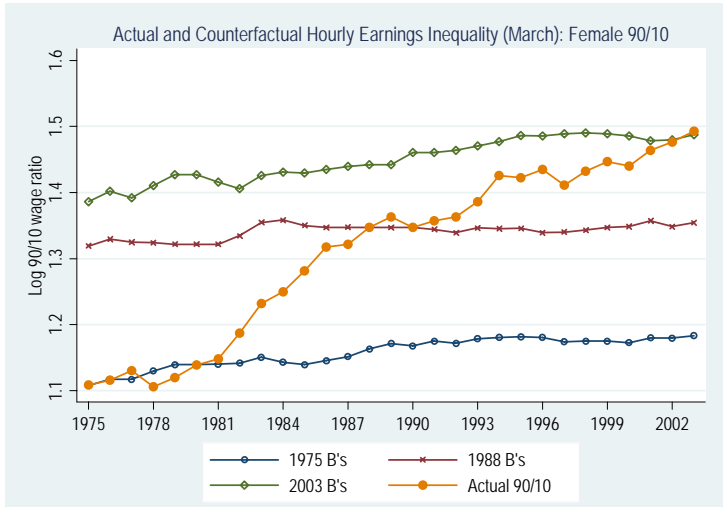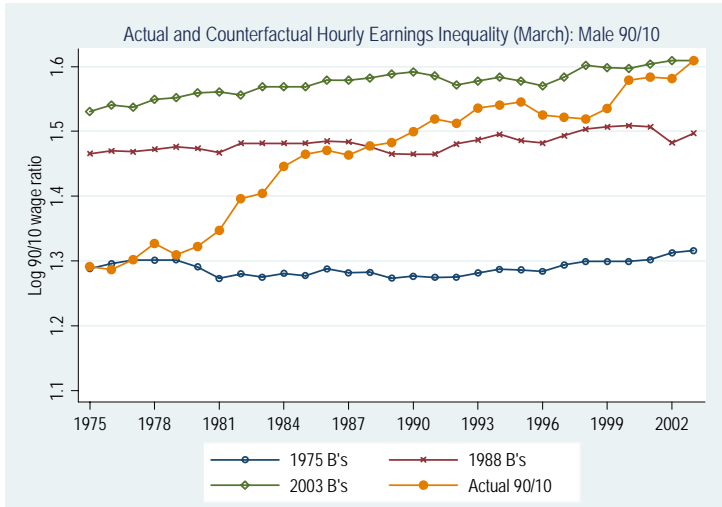
Figure 7. Actual and Counterfactual Residual Hourly Wage Inequality 1975-2003: CPS March

Actual and Counterfactual Hourly Residual Inequality (March): Male 90/10

Actual and Counterfactual Hourly Residual Inequality (March): Female 90/10

Actual and Counterfactual Hourly Residual Inequality (March): Male 90/50

Actual and Counterfactual Hourly Residual Inequality (March): Female 90/50

Actual and Counterfactual Hourly Residual Inequality (March): Male 50/10

Actual and Counterfactual Hourly Residual Inequality (March): Female 50/10

Figure 8. The Growth of Residual Inequality in the CPS May/ORG data 1973-2003:
Accounting for the Rising Frequency of Hourly Pay
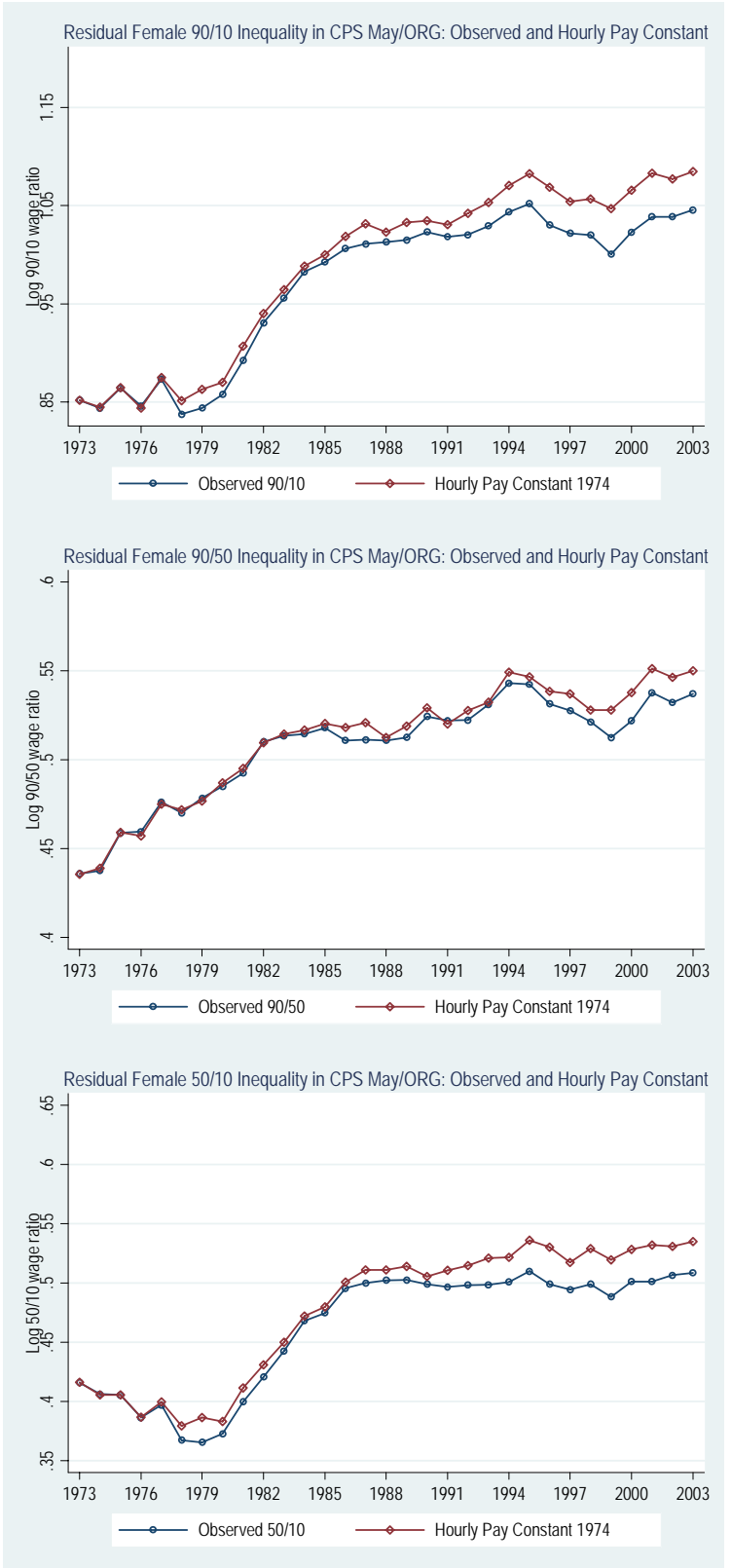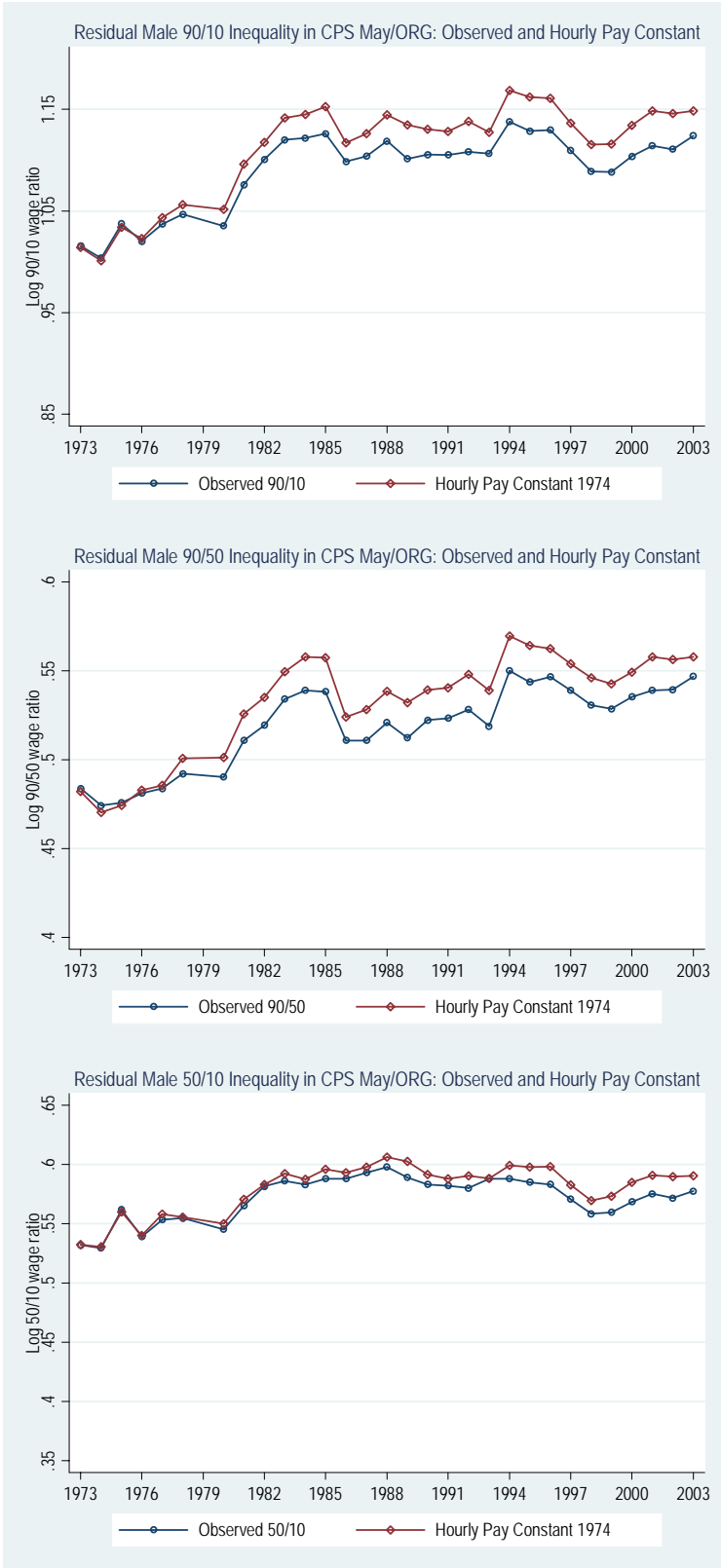
Figure 9. Estimated Hourly Pay Bias for Residual 90/10 Inequality by Gender in CPS May/ORG, 1973 - 2003

Hourly Pay Bias in Male 90/10 Residual Inequality in CPS May/ORG: Corrected-Observed

Understatement of 90/10 Residual Inequality (left axis)
Chain-Weighted Share of Males Paid Hourly (right axis)

Hourly Pay Bias in Female 90/10 Residual Inequality in CPS May/ORG: Corrected-Observed

Understatement of 90/10 Residual Inequality (left axis)
Chain-Weighted Share of Females Paid Hourly (right axis)

Appendix Figure 1. Comparison of Observed and Simulated Overall Hourly Wage Inequality

Actual and Simulated Overall Inequality (MORG): Male 90/10

Actual and Simulated Overall Inequality (MORG): Female 90/10

Actual and Simulated Overall Hourly Inequality (March): Male 90/10

Actual and Simulated Overall Hourly Inequality (March): Female 90/10
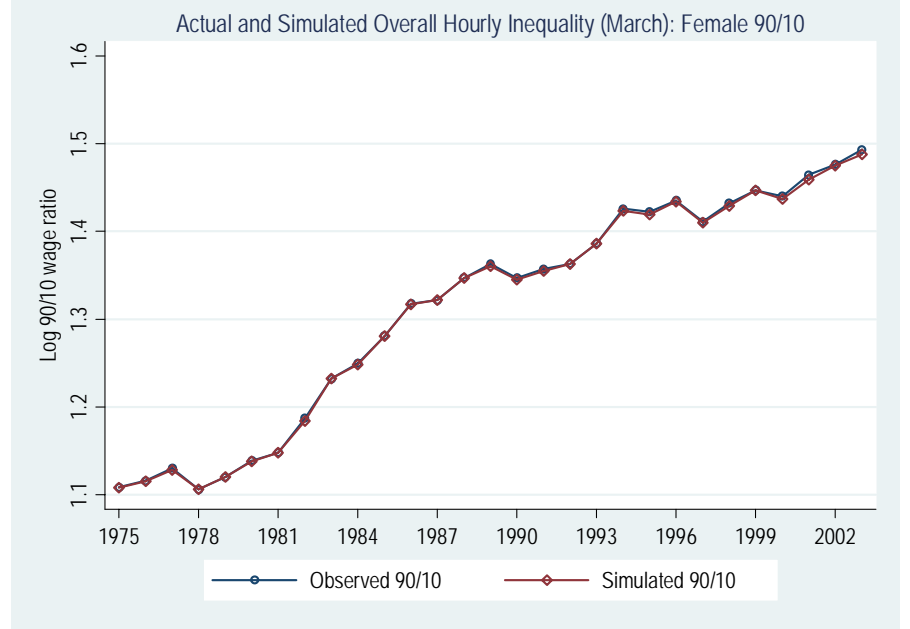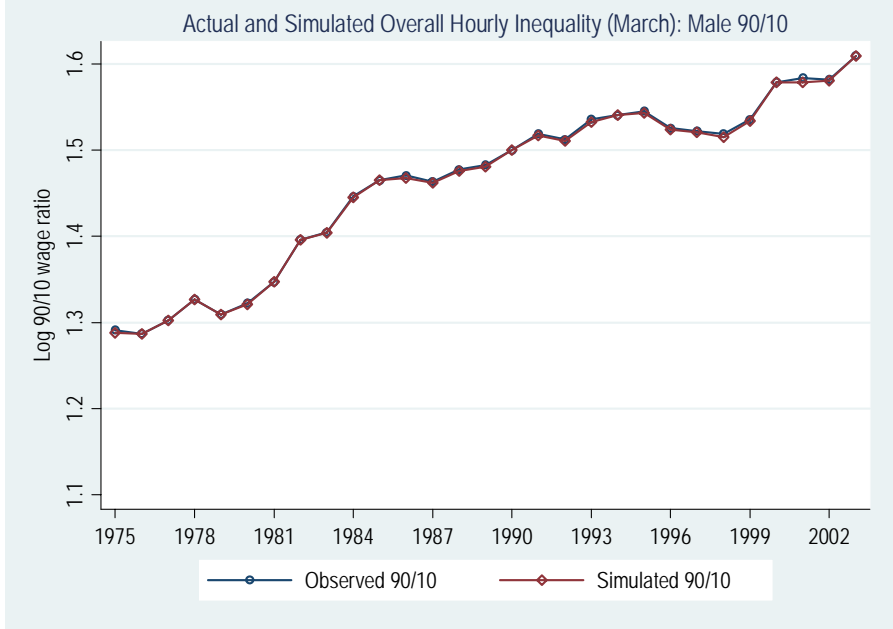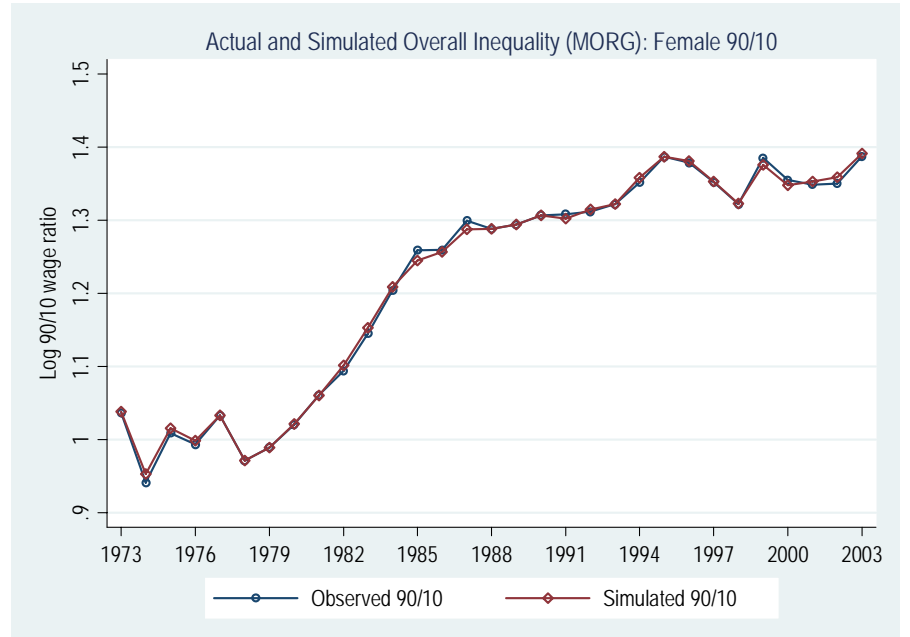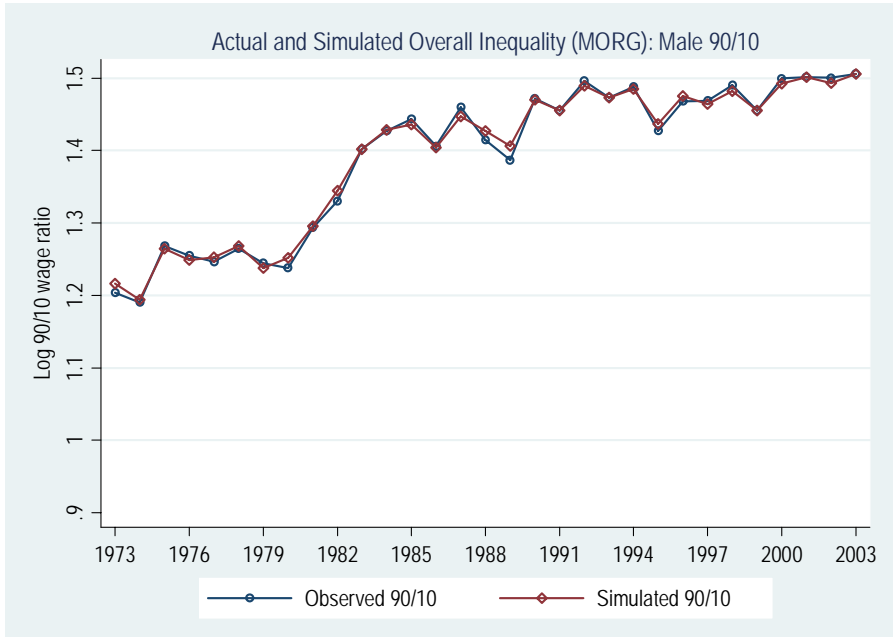
Appendix Figure 2. Comparison of Observed and Simulated Residual Hourly Wage Inequality
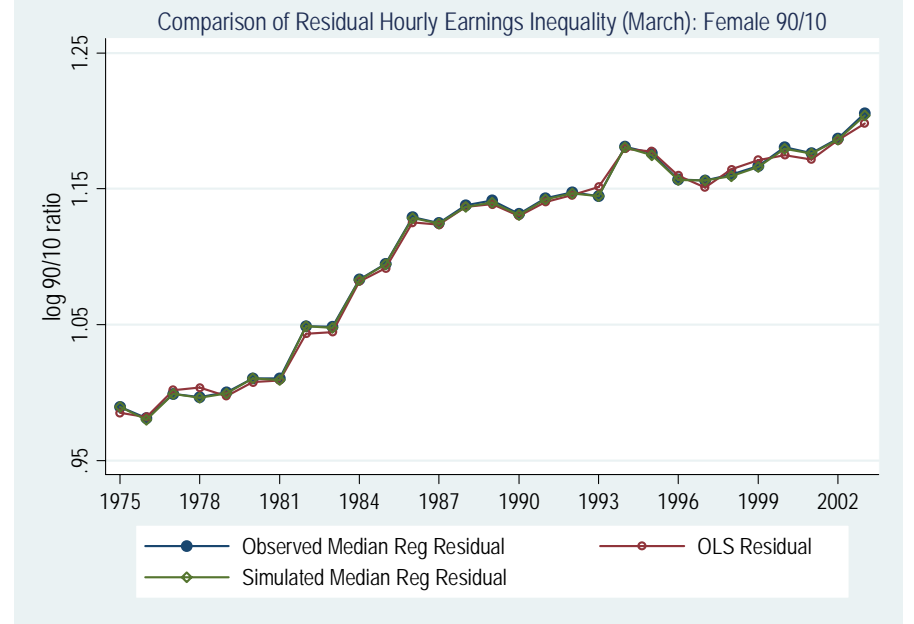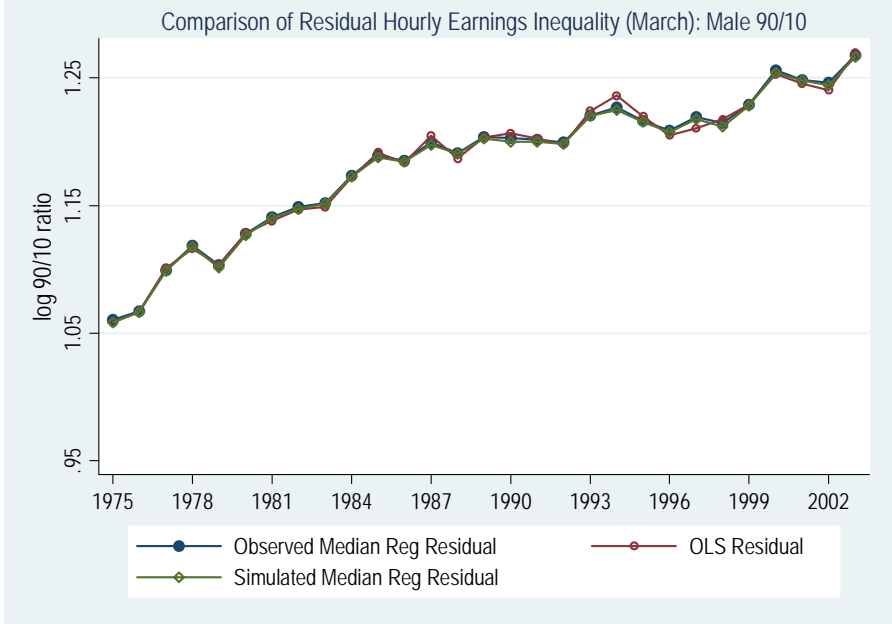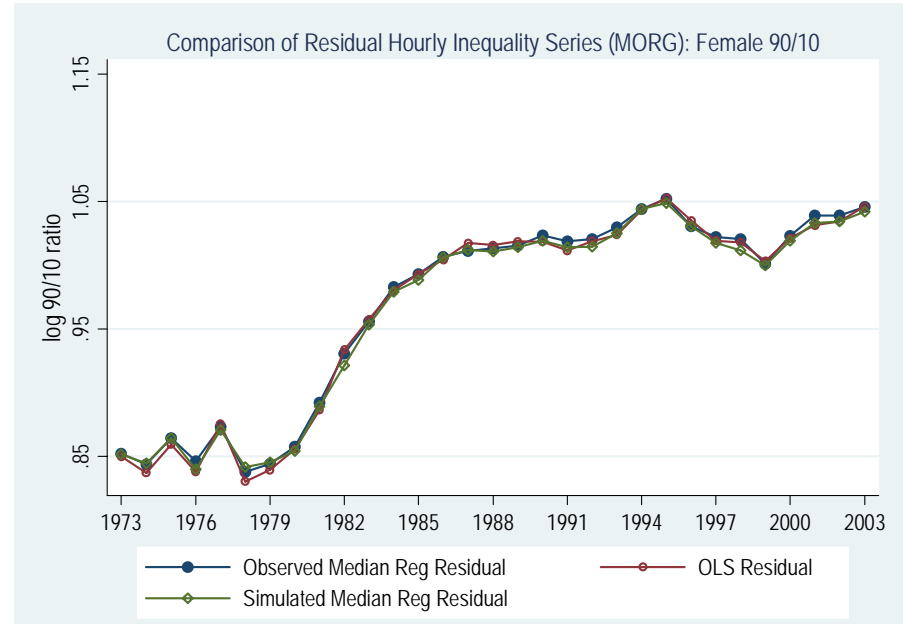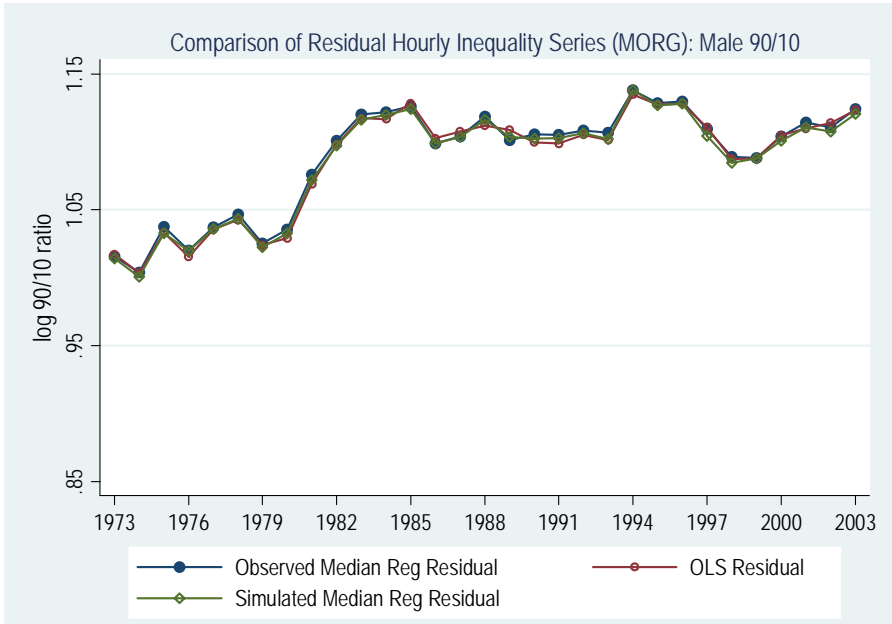
Table 1. 100 x Changes in Overall and Residual Hourly Earnings Inequality: 1973 - 2003
(May/ORG CPS) and 1975 - 2003 (March CPS)

	A. Overall Inequality						B. Residual Inequality					
	1973-1988*	1988-2003	1973-2003*	1979-1989	1989-1999	1979-1999	1973-1988*	1988-2003	1973-2003*	1979-1989	1989-1999	1979-1999
Males												
May/ORG CPS												
90-10	21.0	9.1	30.2	14.2	6.9	21.0	9.5	1.2	10.6	8.5	-2.1	6.4
90-50	10.1	13.8	24.0	9.7	7.3	17.0	3.0	3.8	6.8	2.9	0.9	3.7
50-10	10.9	-4.7	6.2	4.5	-0.4	4.1	6.4	-2.7	3.8	5.6	-3.0	2.6
March CPS												
90-10	19.1	13.2	32.3	17.3	5.3	22.6	12.0	8.3	20.3	10.0	2.6	12.6
90-50	10.7	14.4	25.1	10.6	6.6	17.3	5.2	7.9	13.1	4.3	5.1	9.4
50-10	8.4	-1.2	7.1	6.7	-1.4	5.3	6.8	0.4	7.2	5.7	-2.5	3.2
Females												
May/ORG CPS												
90-10	25.1	9.9	35.0	30.4	9.1	39.6	16.5	3.0	19.5	18.2	-1.9	16.3
90-50	10.5	10.1	20.5	8.0	10.3	18.3	7.8	2.3	10.1	3.8	-0.1	3.7
50-10	14.7	-0.2	14.5	22.4	-1.2	21.2	8.7	0.7	9.3	14.4	-1.8	12.6
March CPS												
90-10	23.1	14.6	37.7	24.3	8.4	32.7	15.5	6.1	21.6	14.1	3.3	17.4
90-50	9.1	9.6	18.7	8.3	4.9	13.2	5.9	5.4	11.3	5.2	3.4	8.6
50-10	14.0	5.0	19.0	16.0	3.5	19.5	9.6	0.7	10.3	8.9	-0.1	8.8

* March hourly series begin in 1975 rather than 1973.
Source data: May/ORG CPS 1973 - 2003 and March CPS 1976 - 2004.

Table 2. 100 x Changes in Male Hourly Earnings Inequality Within Narrow Education and Experience Categories: 1976 to 1988 and 1988 to 2002 (May/ORG and March CPS)

Potn'l Exp	College Graduate Males				High School Graduate Males			
	1976 - 1988		1988 - 2002		1976 - 1988		1988 - 2002	
	ORG	March	ORG	March	ORG	March	ORG	March
A. 90/10								
4-6	17.4	14.3	5.5	10.7	7.6	8.4	-15.3	3.2
	(2.6)	(4.5)	(2.9)	(4.6)	(2.1)	(2.4)	(2.0)	(2.7)
14-16	8.6	-0.8	14.2	21.2	16.6	28.2	-3.6	-5.5
	(6.6)	(6.7)	(3.1)	(6.0)	(2.3)	(3.6)	(1.8)	(3.1)
24-26	5.1	-0.8	7.9	18.2	4.0	25.3	-0.2	-4.0
	(7.9)	(7.3)	(4.6)	(6.4)	(3.9)	(3.9)	(2.0)	(3.4)
34-36	7.6	27.5	-7.1	16.5	14.7	22.6	-6.8	-0.1
	(13.7)	(10.1)	(8.6)	(10.4)	(3.1)	(4.4)	(2.8)	(4.8)
B. 90/50								
4-6	5.0	2.0	6.6	10.7	4.9	4.8	-3.0	7.1
	(1.9)	(2.4)	(1.8)	(2.4)	(1.6)	(1.8)	(1.8)	(2.0)
14-16	8.1	4.7	10.1	18.0	9.2	11.0	7.8	3.7
	(3.7)	(3.5)	(2.4)	(3.3)	(1.4)	(1.8)	(1.4)	(2.0)
24-26	4.8	0.7	1.1	12.6	1.7	5.3	3.9	8.3
	(6.8)	(6.1)	(3.9)	(5.9)	(2.2)	(2.7)	(1.4)	(2.2)
34-36	-1.2	7.6	-5.9	15.0	1.9	6.4	5.5	8.7
	(9.0)	(6.6)	(5.5)	(5.9)	(2.1)	(3.4)	(2.2)	(3.9)
C. 50/10								
4-6	12.3	12.3	-1.1	0.0	2.7	3.5	-12.3	-3.9
	(2.3)	(4.0)	(2.5)	(4.1)	(1.7)	(2.0)	(1.3)	(2.4)
14-16	0.5	-5.4	4.0	3.2	7.3	17.2	-11.4	-9.2
	(5.2)	(6.2)	(2.3)	(4.9)	(1.9)	(3.2)	(1.2)	(2.4)
24-26	0.3	-1.5	6.8	5.6	2.3	20.0	-4.1	-12.3
	(5.2)	(5.9)	(3.2)	(4.7)	(3.2)	(3.1)	(2.0)	(2.8)
34-36	8.8	19.8	-1.2	1.5	12.7	16.2	-12.3	-8.9
	(10.6)	(8.7)	(6.5)	(8.9)	(2.9)	(3.9)	(2.3)	(3.8)

Source data: May/ORG CPS 1975 - 2003 and March CPS 1976 - 2004. Statistics pool three years of data centered on indicated year. Standard-errors in parentheses are bootstrapped using 100 replications. College graduates are those with 16 or 17 years of completed schooling (surveys prior to 1992) or a baccalaureate degree only (1992 forward). See the Data Appendix for further details.

Table 3. 100 x Changes in Female Hourly Earnings Inequality Within Narrow Education and Experience Categories: 1976 to 1988 and 1988 to 2002 (May/ORG and March CPS)

| | College Graduate Females | | | | High School Graduate Females | | | |
| | 1976 - 1988 | | 1988 - 2002 | | 1976 - 1988 | | 1988 - 2002 | |
Potn'l Exp	ORG	March	ORG	March	ORG	March	ORG	March
				A. 90/10				
4-6	16.2	10.6	0.6	1.3	13.6	8.4	-14.2	0.4
	(3.2)	(4.9)	(2.7)	(4.4)	(1.7)	(3.0)	(1.7)	(3.1)
14-16	7.8	-1.4	10.5	25.2	15.1	14.8	-11.7	-4.6
	(5.8)	(7.1)	(2.8)	(5.1)	(1.9)	(3.4)	(1.6)	(3.1)
24-26	14.0	3.3	11.4	8.9	19.4	15.2	-6.1	-0.7
	(5.2)	(8.9)	(2.4)	(6.9)	(2.8)	(4.2)	(1.9)	(3.2)
34-36	33.5	30.6	6.5	12.4	17.0	3.6	-6.4	9.1
	(8.7)	(14.6)	(7.2)	(11.6)	(2.9)	(4.3)	(2.2)	(4.5)
				B. 90/50				
4-6	4.7	9.2	4.5	2.6	5.3	2.5	-5.6	3.3
	(2.4)	(3.1)	(1.9)	(2.7)	(1.4)	(2.4)	(1.6)	(2.7)
14-16	-0.7	-0.5	8.1	14.5	2.4	6.7	-4.9	-3.4
	(3.8)	(5.3)	(2.4)	(3.0)	(2.1)	(2.3)	(1.3)	(2.3)
24-26	12.7	-2.0	8.0	10.2	7.7	8.1	0.1	-0.7
	(4.0)	(4.4)	(2.3)	(3.3)	(2.6)	(2.9)	(1.6)	(2.5)
34-36	11.1	12.1	7.9	14.2	6.8	-2.7	0.5	2.1
	(6.5)	(8.6)	(4.0)	(8.6)	(2.7)	(3.4)	(1.9)	(3.3)
				C. 50/10				
4-6	11.5	1.4	-3.9	-1.3	8.2	6.0	-8.7	-2.9
	(2.1)	(4.0)	(2.0)	(3.5)	(1.1)	(2.3)	(1.2)	(2.2)
14-16	8.6	-0.9	2.4	10.7	12.7	8.1	-6.9	-1.3
	(4.8)	(4.9)	(2.1)	(4.2)	(1.7)	(2.4)	(1.4)	(2.4)
24-26	1.3	5.3	3.4	-1.4	11.7	7.1	-6.2	0.0
	(4.7)	(7.9)	(2.5)	(6.3)	(1.8)	(3.1)	(1.4)	(2.4)
34-36	22.4	18.5	-1.5	-1.8	10.2	6.4	-7.0	7.0
	(6.9)	(12.0)	(6.6)	(8.9)	(2.3)	(3.0)	(1.8)	(2.8)

Source data: May/ORG CPS 1975 - 2003 and March CPS 1976 - 2004. Statistics pool three years of data centered on indicated year. Standard-errors in parentheses are bootstrapped using 100 replications. College graduates are those with 16 or 17 years of completed schooling (surveys prior to 1992) or a baccalaureate degree only (1992 forward). See the Data Appendix for further details.

Table 4. Quantile JMP Decompositions of Hourly Earnings Inequality into Price and Quantity
Components, May/ORG CPS 1973 - 2003
(100 x log point changes)

	1973 - 1988				1988 - 2003				1973 - 2003			
	Total	Quan-tities	Btwn Prices	Within Prices	Total	Quan-tities	Btwn Prices	Within Prices	Total	Quan-tities	Btwn Prices	Within Prices
A. Males												
Order: Quantities, Between B's, Within B's												
90-10	21.1	-1.0	16.8	5.3	7.9	4.8	7.0	-3.8	29.0	3.6	22.7	2.6
90-50	9.9	0.4	9.0	0.6	12.7	0.5	8.1	4.1	22.6	3.0	17.0	2.6
50-10	11.2	-1.4	7.9	4.7	-4.8	4.3	-1.2	-7.9	6.4	0.6	5.7	0.0
Order: Within B's, Between B's, Quantities												
90-10	21.1	0.0	14.4	6.7	7.9	4.9	4.3	-1.3	29.0	6.8	19.2	3.0
90-50	9.9	-3.0	9.9	3.0	12.7	0.5	6.3	5.9	22.6	0.1	14.2	8.3
50-10	11.2	3.0	4.5	3.7	-4.8	4.4	-2.0	-7.2	6.4	6.7	5.0	-5.3
B. Females												
Order: Quantities, Between B's, Within B's												
90-10	25.0	2.2	9.5	13.3	10.3	5.0	5.1	0.2	35.2	10.7	12.6	11.9
90-50	11.5	-2.4	5.3	6.0	9.6	1.8	3.1	4.8	21.1	4.1	6.7	10.3
50-10	13.5	3.6	3.7	6.1	0.7	3.2	2.1	-4.6	14.2	6.6	5.9	1.6
Order: Within B's, Between B's, Quantities												
90-10	25.0	4.3	7.0	13.6	10.3	10.1	3.2	-3.0	35.2	13.6	10.7	10.9
90-50	11.5	1.2	3.2	7.1	9.6	2.3	3.6	3.7	21.1	3.9	8.0	9.2
50-10	13.5	3.2	3.8	6.5	0.7	7.9	-0.4	-6.8	14.2	9.7	2.7	1.7

Source data: May/ORG CPS 1973 - 2003. See text for details of quantile decomposition procedure.

Table 5. Quantile JMP Decompositions of Hourly Earnings Inequality into Price and Quantity Components, March CPS 1975 - 2003
(100 x log point changes)

	1975 - 1988				1988 - 2003				1975 - 2003			
	Total	Quan-tities	Btwn Prices	Within Prices	Total	Quan-tities	Btwn Prices	Within Prices	Total	Quan-tities	Btwn Prices	Within Prices
A. Males												
Order: Quantities, Between B's, Within B's												
90-10	18.8	-0.6	8.2	11.1	13.4	2.1	5.6	5.7	32.1	2.8	14.3	15.1
90-50	9.5	-1.2	5.8	4.8	14.5	-1.4	6.3	9.6	24.0	1.2	11.5	11.2
50-10	9.3	0.6	2.4	6.3	-1.1	3.5	-0.7	-3.9	8.2	1.6	2.7	3.9
Order: Within B's, Between B's, Quantities												
90-10	18.8	1.0	6.4	11.4	13.4	2.7	5.9	4.7	32.1	7.9	9.4	14.9
90-50	9.5	-1.1	4.0	6.6	14.5	1.7	5.2	7.6	24.0	1.5	9.0	13.4
50-10	9.3	2.1	2.4	4.7	-1.1	1.0	0.7	-2.8	8.2	6.3	0.4	1.4
B. Females												
Order: Quantities, Between B's, Within B's												
90-10	23.9	5.5	6.3	12.1	14.0	0.7	5.2	8.2	37.9	7.5	11.7	18.7
90-50	10.2	1.6	4.3	4.3	9.5	-1.6	2.1	9.0	19.7	1.9	5.5	12.3
50-10	13.7	3.9	2.0	7.8	4.5	2.3	3.0	-0.8	18.2	5.6	6.2	6.4
Order: Within B's, Between B's, Quantities												
90-10	23.9	2.8	6.5	14.5	14.0	4.5	4.6	4.9	37.9	10.1	11.0	16.8
90-50	10.2	-1.5	4.6	7.1	9.5	-0.4	4.8	5.1	19.7	2.5	7.4	9.7
50-10	13.7	4.3	2.0	7.5	4.5	4.9	-0.2	-0.2	18.2	7.6	3.6	7.1

Source data: March CPS 1976 - 2004. See text for details of quantile decomposition procedure.

Table 6. 100 x Observed and Composition-Constant Changes in
Overall Hourly Inequality Measures

	A. May/ORG CPS			B. March CPS		
	1973-1988	1988-2003	1973-2003	1975-1988	1988-2003	1975-2003

Δ Male 90/10

Observed	21.1	7.9	29.0	18.8	13.4	32.1
1973 X's*	21.1	1.1	22.2	17.7	6.5	24.3
1988 X's	22.1	3.0	25.2	19.3	10.7	30.0
2003 X's	22.2	3.1	25.3	18.1	11.2	29.3

Δ Male 90/50

Observed	9.9	12.7	22.6	9.5	14.5	24.0
1973 X's*	12.9	9.6	22.5	10.6	11.8	22.4
1988 X's	9.5	12.2	21.8	10.7	12.8	23.4
2003 X's	7.4	12.2	19.6	6.9	15.9	22.7

Δ Male 50/10

Observed	11.2	-4.8	6.4	9.3	-1.1	8.2
1973 X's*	8.2	-8.5	-0.3	7.1	-5.3	1.8
1988 X's	12.6	-9.2	3.4	8.7	-2.1	6.6
2003 X's	14.8	-9.1	5.7	11.3	-4.6	6.6

Δ Female 90/10

Observed	25.0	10.3	35.2	23.9	14.0	37.9
1973 X's*	20.6	1.0	21.6	21.1	6.7	27.8
1988 X's	22.8	0.1	22.9	18.4	9.5	27.9
2003 X's	19.2	5.3	24.5	17.1	13.3	30.4

Δ Female 90/50

Observed	11.5	9.6	21.1	10.2	9.5	19.7
1973 X's*	10.3	6.9	17.1	11.7	5.5	17.2
1988 X's	12.9	7.4	20.3	8.6	9.9	18.5
2003 X's	9.1	7.9	17.0	6.7	11.1	17.8

Δ Female 50/10

Observed	13.5	0.7	14.2	13.7	4.5	18.2
1973 X's*	10.3	-5.9	4.5	9.4	1.2	10.7
1988 X's	9.9	-7.2	2.6	9.8	-0.4	9.4
2003 X's	10.1	-2.5	7.5	10.4	2.2	12.6

Source data: May/ORG CPS 1973 - 2003 and March CPS 1976 - 2004.

Table 7. 100 x Observed and Composition-Constant Changes in
Residual Hourly Inequality Measures

	A. May/ORG CPS			B. March CPS		
	1973-1988	1988-2003	1973-2003	1975-1988	1988-2003	1975-2003

Δ Male 90/10

Observed	10.2	0.4	10.6	13.2	7.6	20.8
1973 X's*	11.5	-5.1	6.4	14.3	3.7	18.1
1988 X's	11.4	-3.6	7.8	13.8	6.3	20.1
2003 X's	10.3	-2.1	8.1	12.5	7.0	19.6

Δ Male 90/50

Observed	3.6	2.9	6.6	5.8	7.3	13.1
1973 X's*	5.8	1.5	7.3	6.3	6.0	12.3
1988 X's	3.6	2.3	5.9	6.1	6.8	13.0
2003 X's	1.6	2.5	4.1	5.3	7.6	12.9

Δ Male 50/10

Observed	6.6	-2.5	4.1	7.4	0.3	7.7
1973 X's*	5.7	-6.6	-0.9	8.0	-2.3	5.7
1988 X's	7.8	-5.9	1.9	7.7	-0.5	7.2
2003 X's	8.7	-4.6	4.1	7.2	-0.5	6.7

Δ Female 90/10

Observed	15.9	3.1	19.0	14.7	6.7	21.4
1973 X's*	13.5	-1.5	12.0	13.5	2.0	15.5
1988 X's	13.8	-2.6	11.2	13.4	3.9	17.3
2003 X's	13.6	-0.6	13.0	14.3	5.8	20.1

Δ Female 90/50

Observed	7.5	2.4	9.9	5.9	5.3	11.2
1973 X's*	6.9	-0.5	6.3	6.6	1.6	8.2
1988 X's	6.5	-0.4	6.0	5.4	3.7	9.1
2003 X's	5.6	1.9	7.5	5.0	5.9	10.9

Δ Female 50/10

Observed	8.4	0.7	9.1	8.9	1.4	10.2
1973 X's*	6.6	-1.0	5.7	6.9	0.4	7.3
1988 X's	7.3	-2.2	5.1	8.0	0.2	8.2
2003 X's	7.9	-2.4	5.5	9.3	-0.1	9.1

Source data: May/ORG CPS 1973 - 2003 and March CPS 1976 - 2004.

Appendix Table 1. 100 x Changes in Overall and Residual Full-Time Weekly Earnings Inequality, 1973 - 2003: May/ORG and March CPS

	A. Overall Inequality						B. Residual Inequality					
	1973-1988	1988-2003	1973-2003	1979-1989	1989-1999	1979-1999	1973-1988	1988-2003	1973-2003	1979-1989	1989-1999	1979-1999
	Males											
	May/ORG CPS											
90-10	34.0	9.6	43.6	26.9	5.6	32.5	24.9	13.4	38.3	28.5	10.2	38.7
90-50	16.4	11.6	28.0	17.4	6.8	24.2	13.3	8.2	21.5	11.8	5.6	17.4
50-10	17.6	-2.0	15.6	9.6	-1.2	8.3	11.7	5.2	16.8	16.7	4.6	21.3
	March CPS											
90-10	23.2	17.9	41.2	20.8	10.5	31.3	17.5	10.4	28.0	12.8	6.3	19.0
90-50	9.8	15.7	25.5	11.2	10.5	21.7	7.0	8.8	15.8	6.1	5.0	11.1
50-10	13.4	2.2	15.7	9.6	0.0	9.6	10.5	1.6	12.2	6.7	1.3	8.0
	Females											
	May/ORG CPS											
90-10	22.3	13.7	36.0	24.6	12.6	37.2	16.1	8.4	24.5	15.4	3.2	18.6
90-50	8.6	10.4	19.0	7.4	7.8	15.2	9.6	5.4	15.0	5.1	3.6	8.8
50-10	13.7	3.3	17.0	17.2	4.8	22.0	6.5	3.0	9.5	10.2	-0.4	9.8
	March CPS											
90-10	18.5	8.7	27.2	14.5	4.7	19.2	19.5	5.7	25.2	18.8	2.6	21.4
90-50	6.8	8.0	14.8	6.7	4.4	11.1	9.7	5.3	15.0	5.8	3.3	9.1
50-10	11.7	0.7	12.3	7.9	0.3	8.1	9.8	0.4	10.2	13.0	-0.6	12.3

Source data: May/ORG CPS 1973 - 2003 and March CPS 1974 - 2004.

Appendix Table 2a. 100 x Levels of Male Log Hourly Earnings Inequality Within Narrow Education and Experience Categories: 1976, 1988 and 2002 (May/ORG and March CPS)

Potn'l Exp	A. All Males						B. College Graduate Males						C. High School Graduate Males					
	1976		1988		2002		1976		1988		2002		1976		1988		2002	
	ORG	March	ORG	March	ORG	March	ORG	March	ORG	March	ORG	March	ORG	March	ORG	March	ORG	March
A. 90/10																		
4-6	1.08	1.09	1.28	1.36	1.29	1.45	0.99	0.97	1.16	1.11	1.22	1.22	0.93	1.08	1.01	1.16	0.86	1.19
14-16	1.16	1.12	1.36	1.36	1.50	1.52	1.14	1.22	1.22	1.21	1.37	1.43	0.93	1.00	1.09	1.28	1.06	1.22
24-26	1.27	1.19	1.41	1.39	1.46	1.45	1.28	1.26	1.33	1.25	1.41	1.43	1.06	1.03	1.10	1.29	1.10	1.25
34-36	1.19	1.13	1.51	1.44	1.50	1.43	1.44	1.17	1.53	1.44	1.44	1.61	1.02	1.07	1.17	1.30	1.10	1.30
0-38	1.27	1.28	1.43	1.46	1.50	1.56	1.30	1.33	1.38	1.40	1.43	1.47	1.08	1.21	1.19	1.39	1.15	1.35
B. 90/50																		
4-6	0.52	0.49	0.68	0.65	0.77	0.74	0.48	0.46	0.53	0.48	0.59	0.58	0.46	0.50	0.51	0.55	0.48	0.62
14-16	0.53	0.51	0.66	0.58	0.82	0.80	0.48	0.48	0.56	0.53	0.66	0.71	0.39	0.42	0.48	0.53	0.56	0.57
24-26	0.61	0.55	0.68	0.54	0.75	0.72	0.58	0.54	0.62	0.54	0.63	0.67	0.47	0.43	0.49	0.48	0.53	0.57
34-36	0.56	0.51	0.69	0.61	0.78	0.70	0.69	0.53	0.68	0.61	0.62	0.76	0.46	0.46	0.48	0.52	0.54	0.61
0-38	0.59	0.57	0.69	0.65	0.83	0.77	0.63	0.62	0.66	0.61	0.69	0.70	0.48	0.50	0.56	0.62	0.60	0.64
C. 50/10																		
4-6	0.56	0.60	0.60	0.70	0.52	0.71	0.51	0.52	0.64	0.64	0.62	0.64	0.47	0.58	0.50	0.61	0.37	0.57
14-16	0.63	0.61	0.70	0.77	0.68	0.72	0.66	0.74	0.66	0.69	0.70	0.72	0.54	0.57	0.61	0.75	0.50	0.65
24-26	0.66	0.64	0.74	0.84	0.71	0.73	0.71	0.72	0.71	0.71	0.78	0.76	0.59	0.60	0.61	0.80	0.57	0.68
34-36	0.62	0.62	0.82	0.83	0.72	0.73	0.75	0.63	0.85	0.83	0.82	0.85	0.56	0.61	0.69	0.78	0.56	0.69
0-38	0.68	0.71	0.74	0.81	0.68	0.79	0.67	0.71	0.72	0.80	0.74	0.77	0.60	0.71	0.63	0.76	0.55	0.71
D. Variance																		
4-6	0.18	0.20	0.24	0.28	0.25	0.33	0.17	0.16	0.24	0.22	0.24	0.26	0.14	0.20	0.16	0.22	0.13	0.23
14-16	0.22	0.21	0.28	0.29	0.32	0.38	0.23	0.25	0.28	0.27	0.27	0.35	0.15	0.17	0.19	0.24	0.18	0.24
24-26	0.26	0.24	0.31	0.31	0.32	0.36	0.27	0.27	0.32	0.29	0.30	0.37	0.20	0.19	0.20	0.27	0.21	0.27
34-36	0.24	0.23	0.33	0.34	0.32	0.37	0.37	0.26	0.39	0.37	0.32	0.43	0.18	0.21	0.21	0.28	0.20	0.30
0-38	0.25	0.27	0.32	0.34	0.33	0.39	0.27	0.30	0.31	0.33	0.30	0.37	0.19	0.24	0.22	0.29	0.21	0.30

Source data: May/ORG CPS 1975 - 2003 and March CPS 1976 - 2004. Statistics pool three years of data centered on indicated year.

Appendix Table 2b. 100 x Levels of Female Log Hourly Earnings Inequality Within Narrow Education and Experience Categories: 1976, 1988 and 2002 (May/ORG and March CPS)

Potn'l Exp	All Females						College Graduate Females						High School Graduate Females					
	1976		1988		2002		1976		1988		2002		1976		1988		2002	
	ORG	March	ORG	March	ORG	March	ORG	March	ORG	March	ORG	March	ORG	March	ORG	March	ORG	March
A. 90/10																		
4-6	0.97	1.11	1.19	1.35	1.21	1.48	0.89	1.06	1.05	1.17	1.05	1.18	0.77	1.02	0.90	1.11	0.76	1.11
14-16	1.06	1.20	1.33	1.35	1.42	1.58	1.12	1.21	1.20	1.19	1.30	1.44	0.95	1.11	1.10	1.26	0.98	1.21
24-26	1.04	1.20	1.29	1.36	1.39	1.47	1.03	1.21	1.17	1.25	1.28	1.34	0.92	1.12	1.11	1.27	1.05	1.27
34-36	1.06	1.15	1.26	1.29	1.35	1.45	0.96	0.98	1.29	1.28	1.36	1.41	0.94	1.14	1.11	1.18	1.04	1.27
0-38	1.01	1.15	1.28	1.36	1.37	1.50	1.02	1.16	1.18	1.27	1.27	1.42	0.89	1.09	1.09	1.26	1.03	1.24
B. 90/50																		
4-6	0.52	0.53	0.64	0.70	0.70	0.75	0.44	0.46	0.49	0.55	0.53	0.57	0.41	0.49	0.46	0.52	0.40	0.55
14-16	0.60	0.62	0.68	0.65	0.80	0.80	0.55	0.52	0.55	0.51	0.63	0.66	0.54	0.55	0.57	0.62	0.52	0.59
24-26	0.59	0.62	0.67	0.67	0.75	0.75	0.41	0.54	0.53	0.52	0.61	0.62	0.50	0.57	0.58	0.65	0.58	0.64
34-36	0.60	0.59	0.64	0.66	0.75	0.76	0.44	0.42	0.55	0.54	0.63	0.68	0.49	0.62	0.56	0.59	0.56	0.61
0-38	0.57	0.59	0.68	0.68	0.77	0.78	0.49	0.53	0.55	0.55	0.64	0.69	0.50	0.56	0.57	0.63	0.58	0.63
C. 50/10																		
4-6	0.45	0.58	0.55	0.66	0.51	0.73	0.45	0.60	0.56	0.62	0.52	0.60	0.36	0.53	0.44	0.59	0.36	0.56
14-16	0.46	0.58	0.65	0.71	0.63	0.78	0.56	0.69	0.65	0.68	0.67	0.78	0.40	0.56	0.53	0.64	0.46	0.63
24-26	0.45	0.59	0.63	0.69	0.64	0.72	0.62	0.67	0.64	0.73	0.67	0.71	0.42	0.55	0.53	0.62	0.47	0.63
34-36	0.46	0.57	0.62	0.63	0.61	0.69	0.52	0.56	0.74	0.74	0.73	0.72	0.45	0.53	0.55	0.59	0.48	0.66
0-38	0.44	0.57	0.60	0.68	0.60	0.72	0.53	0.63	0.63	0.72	0.63	0.73	0.39	0.53	0.53	0.63	0.45	0.61
D. Variance																		
4-6	0.15	0.21	0.22	0.29	0.24	0.35	0.14	0.19	0.19	0.23	0.19	0.25	0.11	0.18	0.13	0.22	0.11	0.25
14-16	0.19	0.23	0.27	0.29	0.31	0.40	0.19	0.22	0.23	0.26	0.27	0.37	0.15	0.21	0.19	0.24	0.16	0.26
24-26	0.18	0.24	0.25	0.28	0.29	0.35	0.18	0.27	0.22	0.24	0.26	0.35	0.14	0.21	0.19	0.26	0.18	0.25
34-36	0.18	0.22	0.24	0.28	0.28	0.33	0.16	0.16	0.24	0.32	0.27	0.35	0.15	0.22	0.18	0.23	0.18	0.27
0-38	0.18	0.22	0.25	0.29	0.29	0.36	0.18	0.23	0.22	0.27	0.25	0.34	0.14	0.20	0.18	0.24	0.17	0.27

Source data: May/ORG CPS 1975 - 2003 and March CPS 1976 - 2004. Statistics pool three years of data centered on indicated year.

Appendix Table 3a. Quantile JMP Decompositions of Hourly Earnings Inequality into Price and Quantity Components May/ORG CPS 1973 - 2003: Model Using Quartic in Potential Experience (100 x log point changes)

	1973 - 1988				1988 - 2003				1973 - 2003			
	Total	Quan-tities	Btwn Prices	Within Prices	Total	Quan-tities	Btwn Prices	Within Prices	Total	Quan-tities	Btwn Prices	Within Prices
A. Males												
Order: Quantities, Between B's, Within B's												
90-10	21.9	0.0	15.7	6.2	8.1	4.9	7.0	-3.7	30.0	3.1	23.3	3.6
90-50	11.3	0.5	9.6	1.3	11.0	0.7	7.1	3.1	22.3	2.3	17.4	2.6
50-10	10.6	-0.5	6.1	5.0	-2.8	4.1	-0.2	-6.8	7.7	0.8	5.9	1.0
Order: Within B's, Between B's, Quantities												
90-10	21.9	0.8	13.6	7.5	8.1	5.5	5.1	-2.4	30.0	8.8	17.4	3.9
90-50	11.3	-0.9	8.6	3.7	11.0	0.8	5.5	4.6	22.3	1.2	12.9	8.2
50-10	10.6	1.7	5.0	3.9	-2.8	4.7	-0.5	-7.0	7.7	7.6	4.5	-4.4
B. Females												
Order: Quantities, Between B's, Within B's												
90-10	25.1	1.8	10.8	12.5	9.9	4.8	5.1	0.0	35.0	8.4	14.8	11.7
90-50	12.3	-0.5	5.7	7.1	8.8	1.1	3.3	4.5	21.1	3.1	7.9	10.0
50-10	12.8	2.3	5.1	5.5	1.1	3.7	1.8	-4.5	13.9	5.3	6.9	1.7
Order: Within B's, Between B's, Quantities												
90-10	25.1	3.5	8.5	13.1	9.9	9.0	3.4	-2.6	35.0	12.0	12.4	10.6
90-50	12.3	1.2	4.2	6.9	8.8	2.6	3.2	3.1	21.1	3.2	8.5	9.3
50-10	12.8	2.3	4.3	6.2	1.1	6.5	0.3	-5.7	13.9	8.8	3.8	1.3

Source data: May/ORG CPS 1973 - 2003. See text for details of quantile decomposition procedure.

Appendix Table 3b. Quantile JMP Decompositions of Hourly Earnings Inequality into Price and Quantity Components, March CPS 1975 - 2003: Model Using Quartic in Potential Experience
(100 x log point changes)

	1975 - 1988				1988 - 2003				1975 - 2003			
	Total	Quan-tities	Btwn Prices	Within Prices	Total	Quan-tities	Btwn Prices	Within Prices	Total	Quan-tities	Btwn Prices	Within Prices
A. Males												
Order: Quantities, Between B's, Within B's												
90-10	17.5	-1.0	9.8	8.7	10.3	2.6	3.7	4.0	27.8	1.6	16.2	10.0
90-50	8.3	0.3	5.9	2.1	12.9	-0.5	3.5	9.9	21.2	1.9	11.8	7.6
50-10	9.2	-1.3	3.9	6.6	-2.6	3.1	0.2	-5.9	6.6	-0.2	4.4	2.5
Order: Within B's, Between B's, Quantities												
90-10	17.5	-0.3	8.2	9.6	10.3	2.5	4.8	3.0	27.8	6.6	11.4	9.8
90-50	8.3	-0.8	4.8	4.4	12.9	-0.2	4.4	8.7	21.2	1.4	8.2	11.7
50-10	9.2	0.5	3.4	5.2	-2.6	2.7	0.4	-5.7	6.6	5.2	3.3	-1.9
B. Females												
Order: Quantities, Between B's, Within B's												
90-10	18.6	2.3	7.7	8.6	13.5	1.9	5.7	5.9	32.2	5.6	13.5	13.1
90-50	9.5	0.3	3.8	5.4	8.4	-1.4	3.8	6.1	17.9	0.9	8.1	8.9
50-10	9.1	2.0	3.9	3.2	5.1	3.3	2.0	-0.2	14.2	4.6	5.4	4.3
Order: Within B's, Between B's, Quantities												
90-10	18.6	1.0	6.6	11.0	13.5	5.7	4.9	3.0	32.2	10.0	10.3	11.8
90-50	9.5	-0.7	3.9	6.3	8.4	0.5	4.0	3.9	17.9	1.6	7.5	8.8
50-10	9.1	1.7	2.6	4.7	5.1	5.2	0.9	-0.9	14.2	8.4	2.8	3.0

Source data: March CPS 1976 - 2004. See text for details of quantile decomposition procedure.