Volume Title: Labor in the New Economy

Volume Author/Editor: Katharine G. Abraham, James R. Spletzer, and Michael Harper, editors

Volume Publisher: University of Chicago Press

Volume ISBN: 978-0-226-00143-2; 0-226-00143-1

Volume URL: http://www.nber.org/books/abra08-1

Conference Date: November 16-17, 2007

Publication Date: October 2010

Chapter Title: What Do We Really Know About Changes in Wage Inequality?

Chapter Author: Thomas Lemieux

Chapter URL: http://www.nber.org/chapters/c10812

Chapter pages in book: (17 - 59)

What Do We Really Know about Changes in Wage Inequality?

Thomas Lemieux

1.1   Introduction

It is very well known that wage and earnings inequality has grown substantially over the last thirty years. The initial burst of inequality growth in the 1980s attracted a lot attention among labor economists. This resulted in a set of influential papers published in the early 1990s, in particular Katz and Murphy (1992); Bound and Johnson (1992); Levy and Murnane (1992); and Juhn, Murphy, and Pierce (1993). These papers laid down the main facts and possible explanations for the dramatic increase in wage inequality of the 1980s. At the time, the leading explanation that emerged was based on a pervasive increase in the demand for all dimensions of skills that was mitigated, in part, by swings in relative supply linked to the baby boom cohort. A number of papers later argued that the leading source of increase in the relative demand for skills was skill-biased technological change (SBTC) linked, in large part, to the computer and information technology revolution.[1]

Over the last fifteen years, however, further research has cast some doubt on the basic view that inequality growth is driven by a combination of demand changes linked to SBTC and the computer revolution. This "revisionist" view, to borrow the term suggested by Autor, Katz, and Kearney (2008), is mostly based on the observation that the bulk of broad-based inequality growth was concentrated in the 1980s. In particular, Card and DiNardo (2002) argue that this "episodic" aspect of inequality growth is

1. See, in particular, Krueger (1993); Berman, Bound, and Griliches (1994); Autor, Katz, and Krueger (1998); Autor, Levy, and Murnane (2003).

inconsistent with a simple supply and demand explanation that should instead predict an unabated growth in inequality throughout the 1990s. The episodic view of inequality changes is corroborated in recent work by Lemieux (2006a), who argues that the growth in within-group inequality is also concentrated in the 1980s. This somehow contradicts the earlier work of Juhn, Murphy, and Pierce (1993), who document a continuing growth in this dimension of inequality throughout the 1970s and 1980s.[2] Lemieux (2006a) argues that various measurement issues and composition effects account for these different views about the timing of the growth in within-group inequality.

In response to these recent findings, Autor, Katz, and Kearney (2008) point out that inequality in the upper end of the distribution (top-end inequality) has kept growing steadily throughout the 1990s and early 2000s, a trend that they attribute to relative demand shifts induced by technological change of the type proposed by Autor, Levy, and Murnane (2003). The steady and continuing growth in top-end inequality has also been well documented by Piketty and Saez (2003) using tax data which, unlike data from the Current Population Survey (CPS) or the Census, are not topcoded.

In light of these seemingly diverging views, the main goal of this chapter is to identify in the clearest possible way, using CPS data, what have been the main changes in inequality and the wage structure since the early 1970s. The purpose of this exercise is to establish the basic facts that are robust to the variety of measurement problems frequently encountered in the litera-ture, that is, identify what it is that we really know about changes in wage inequality. These measurement problems include topcoding, the growing nonresponse to earnings item in the CPS, and differences in wage measures in the March and outgoing rotation group (ORG) supplements of the CPS. A first substantive conclusion of the chapter is that, with the exception of the growth in within-group inequality for men in the 1970s, all the main trends appear to be highly robust to these measurement issues. More generally, the results confirm the view of both the "revisionists," who find that changes in broad-based measures of inequality are concentrated in the 1980s, and of Autor, Katz, and Kearney (2006, 2008), who find smooth growth in top-end inequality.

The related goal of the chapter is to assess what these trends tell us about the underlying sources of changes in inequality. I discuss what challenges these findings pose for existing explanations and propose alternative expla-nations linked to broadly defined wage setting institutions (minimum wage, unions, and performance pay) to help reconcile these often contradictory sets for facts.

---

2. Katz and Autor (1999) and Acemoglu (2002) also find that within-group inequality kept increasing in the 1990s.

One difference between this chapter and most of the existing literature is that the majority of basic trends in inequality reported here have been adjusted for changes in the skill (experience and education) composition of the workforce. A more standard approach is to report basis trends without these adjustments for composition effects and then perform decompositions where composition effects are one of the sources of overall change in inequality (see, for example, Juhn, Murphy, and Pierce 1993). One drawback of these approaches is that what is often presented as the basic inequality trends end up mixing up composition effects and true underlying changes in the wage structure. Lemieux (2006a) shows this has important consequences in the case of within-group, or residual, inequality where composition effects are large. Because my goal here is to document the main trends in the wage structure, and in summary measures of inequality induced by changes in the wage structure, I focus on an approach where composition effects are systematically adjusted for.

The wage measure used throughout the chapter is the hourly wage rate that purely reflects the "price" of different types of labor, as opposed to earnings that mix up the hourly wage rate with hours decisions. The primary source of data used to do so is the 1979 to 2006 ORG supplements of the CPS, supplemented with similar wage data from the 1973 to 1978 May CPS. All inequality trends are also presented separately for men and women. Using hourly wage rates, as opposed to weekly earnings for full-time workers (or other earnings measures), is particularly important for women, who are less likely to work full-time and generally exhibit more variation in hours of work than men.

Section 1.2 briefly discusses the data and presents the measurement framework used to compute the various measures of inequality adjusted for composition effects. The basic trends are presented in section 1.3, and the robustness of these trends to a number of measurement issues is discussed in section 1.4. The main findings are summarized in section 1.5, while section 1.6 concludes by discussing the implications of these findings for different explanations about the sources of change in wage inequality.

1.2   Data and Measurement Framework

1.2.1   Data Issues

Data issues are discussed in detail in the data appendix, which explains the construction of wage measures for the May-ORG and March CPS Supplements. I only briefly discuss how the May and ORG supplements of the CPS are processed here. As mentioned in the preceding, the wage measure used is the hourly wage rate. The main advantage of this measure is that theories of wage determination typically pertain to the hourly wage rate.

For example, the interplay of demand and supply considerations has direct implications for the hourly price of labor. By contrast, the impact of these factors on weekly or annual earnings also depends on the responsiveness of labor supply to changes in the hourly wage rate.

The Dual Jobs Supplement of the May CPS for 1973 to 1978 asks questions about wages on the main job held during the survey week to all wage and salary workers. For workers paid by the hour, the May CPS asks workers directly about their hourly rate of pay. This is the hourly wage measure that I use for this group of workers (about 60 percent of the workforce). For the other workers, I compute an hourly wage rate by dividing usual weekly earnings by usual weekly hours of work. I use the same procedure for the 1979 to 1993 ORG supplements that ask the same wage questions as the May CPS. The wage questions in the 1994 to 2006 ORG supplements are similar except that workers not paid by the hour can choose the periodicity at which they report earnings. I compute their hourly wage rate by dividing earnings by hours over the corresponding time period. The merged outgoing rotation group (MORG) files combine this information for all twelve months of the year.

One important advantage of the MORG supplement is that it is roughly three times as large as the May or March supplements of the CPS.[3] Another advantage over the March CPS is that we know the union status of workers in the May-ORG CPS, but not in the March CPS. A potential disadvantage is that wage data in the May-ORG CPS only goes back to 1973, while it is possible to go back to the 1960s using the March CPS. This is of little consequence here, however, because most of the relevant movements in wage inequality and the wage structure only started in the 1970s.

Unlike in the ORG and March supplements of the CPS, in the 1973 to 1978 May CPS, wages were not allocated for workers who refused to answer the wage questions. To be consistent, I only keep workers with nonallocated wages in the 1979 to 2006 ORG supplements in most of the analysis. As a consequence, observations are for 1994 and the first eight months of 1995, in which the workers with missing wages are dropped from the sample when only nonallocated observations are used. Following most of the literature, I trim extreme values of wages (less than \$1 and more than \$100 in 1979\$) and keep workers aged sixteen to sixty-four with positive potential experience.

In the main results presented in the chapter, I adjust for topcoding by

3. The May 1973 to 1978 and March supplements are administered to all (eight) rotation groups of the CPS during these months. By contrast, only one-quarter of respondents (in rotation groups four and eight) are asked the questions from the ORG supplement each month. Combining the twelve months of data into a single MORG file yields wage data for twenty-four rotation groups compared to eight in the May or March supplements (plus the hispanic and Medicare [post-2000] oversamples in the March CPS).

multiplying topcoded earnings by a factor of 1.4. In the case of the March CPS, I also compare this rudimentary adjustment to a more sophisticated stochastic imputation procedure based on the assumption that top earnings follow a Pareto distribution with a parameter estimated separately for each year in the tax data of Piketty and Saez (2003).

Finally, I weight all wage observations by hours of work (in addition to the usual CPS weights). In the case of the May-ORG CPS, weekly hours of work are used, while annual hours are used for the March CPS. Doing so has two main advantages. First, it effectively provides a distribution over all hours worked in the economy that do not put excessive focus on workers who only supply very few hours to the labor market. For instance, DiNardo, Fortin, and Lemieux (1996) argue that failing to do so would put excessive weight on the bottom end of the distribution where many workers around the minimum wage provide relatively fewer hours to the labor market than most other workers. Another advantage is that weighting by hours of work makes the March and May-ORG wage distributions more directly comparable (Lemieux 2006a).[4]

1.2.2 Measurement Model

As discussed earlier, unadjusted trends in wage inequality may either reflect underlying changes in the wage structure or composition effects that confound the changes in the wage structure. A simple way of adjusting for composition effects is to reweight the data so that the distribution of education and potential experience remains constant over time (e.g., DiNardo, Fortin, and Lemieux 1996). Doing so is straightforward in cases where the data can be divided up in a finite number of cells. In such cases, the weight attached to each cell can be held constant by multiplying the sample fraction in year t by the average fraction of observations in this cell for all years combined or for an arbitrary base year. Because results can be sensitive to the choice of base year, I follow Juhn, Murphy, and Pierce (1993) and hold the sample composition constant at the average fraction for all years combined. One additional reason for using this average weight over all years combined is that it yields results similar to those obtained using a more sophisticated chain-weighted approach.[5]

After various experimentations, I divided the data into 130 cells defined on the basis of six education groups (elementary, high school dropouts, high

---

4. Abraham, Speltzer, and Steward (1998) also weight by hours for the same reason (i.e., make different data sources comparable) in their study comparing the evolution of average real wages rates from different data sources, including the March and May-ORG CPS.

5. I thank Lawrence Katz for this suggestion. In the case of the within-group variance discussed in the following, the change in variance between year t and $t-1$ holding the workforce composition at its year $t-1$ level is $\Delta W_{t,c} = \Sigma_{jk}\, \theta_{jkt-1}(v_{jkt} - v_{jkt-1})$. The chain-weighted composition adjusted variance in year t, $W_{t,c}$, is equal to the base period variance, W_0, plus the sum of year-to-year changes defined in the preceding: $W_{t,c} = W_0 + \Sigma_{s=1\,\text{to}\,t}\, \Delta W_{s,c}$.

school graduates, college graduates, and college postgraduates) and twenty-two two-year experience groups.[6] The R-square of a wage regression using this set of cell dummies is very close to the R-square for an unrestricted model, and using these broader cells limits the problem of very small cells that can generate excessive variability in the reweighting procedure.[7]

In the first set of results presented in the chapter, the measure of wage dispersion used is the variance of log wages (between and within components) as well as standard wage differentials based on differences across education and experience groups. All these measures can be directly computed as functions of the mean and variance of wages in each cell. Let y_{jkt} denote the average wage of workers with education j and experience k in year t. The variance within this group is v_{jkt}, the share of workers in this group in year t is θ_{jkt}, and the average share over all years is θ_{jk}. The composition unadjusted within- ($W_{t,u}$) and between-group ($B_{t,u}$) variance in year t are

$$B_{t,u} = \Sigma_{jk}\, \theta_{jkt}(y_{jkt} - y_{t,u})^2, \text{ and}$$

$$W_{t,u} = \Sigma_{jk}\, \theta_{jkt}v_{jkt},$$

where $y_{t,u} = \Sigma_{jk}\, \theta_{jkt}\, y_{jkt}$ is the unadjusted mean (grand mean) over all groups. The corresponding expressions adjusted for composition effects are obtained by simply replacing the year t weights, θ_{jkt}, with the average weights θ_{jk}:

$$B_t = \Sigma_{jk}\, \theta_{jk}(y_{jkt} - y_t)^2, \text{ and}$$

$$W_t = \Sigma_{jk}\, \theta_{jk}v_{jkt},$$

where $y_t = \Sigma_{jk}\, \theta_{jk}\, y_{jkt}$ is the composition adjusted mean over all groups.

A standard approach for estimating experience and education wage differentials is to run ordinary least squares (OLS) regressions. I use a slightly different approach by computing separate measures of education differentials for each experience category and vice versa. The experience-group-specific education wage differentials are then aggregated up into a single differential by averaging up the experience-group-specific education differentials using the average sample fraction in each experience group over all years combined.

---

6. Note that the twenty-second experience category includes workers with forty-two years and more of potential experience because some of the two-year experience cells with very high level of experience (forty-three to forty-four or forty-five to forty-six) were too small. Note also that there are no observations in (a) the cell with the lowest level of education (eight years or less) and the lowest level of experience (zero to one years of experience), and (b) the cell with the highest level of education (eighteen year or more) and the highest level of experience (forty-two years or more of experience) because observations in these cells do not fall into the age range (sixteen to sixty-four) used in this paper. This explains why 130, as opposed to 132 (twenty-two experience times six education groups) cells are used in the empirical analysis.

7. In the case of men, the average R-square in 1973, 1979, 1989, 1999, and 2006 in the model with 120 cells is .3466, compared to .3511 in models with an unrestricted set of experience-education dummies based on single years of experience and education. The corresponding average R-square for women are .3110 and .3135, respectively.

Let θ_j represent the fraction of workers (of all experience groups) with education j in all years combined, and θ_k represent the fraction of workers (of all education groups) with experience k in all years combined. The composition-adjusted wage differential between education group j' and j in year t is defined as

$$D_{j',jt} = \Sigma_k \, \theta_k \, (y_{j'kt} - y_{jkt}).$$

Similarly, the composition-adjusted wage differential between experience group k' and k in year t is

$$D_{k',kt} = \Sigma_j \, \theta_j \, (y_{jk't} - y_{jkt}).$$

One important advantage of the variance as a measure of wage inequality is that it can be exactly decomposed as the sum of the between- and within-group components W and B. The decomposition can be directly linked to various "price effects" once composition effects have been adjusted for using the preceding procedure. Using the terminology of Juhn, Murphy, and Pierce (1993), the between-group component solely reflects "observable price effects," that is, differences in mean wages among experience and education groups. By contrast, the within-group component solely captures "unobservable price effects" under the assumption that the distribution of unobservables (ability, school quality, etc.) within a fixed experience-education group does not change over time.

One important disadvantage of the variance, however, is that it is only a summary measure of inequality that does not indicate what happens where in the distribution. This is a major problem in light of recent evidence that inequality is changing very differently at different points of the wage distribution. A simple and popular way of showing what happens at different points of the distribution is to look separately at each wage percentile and compute summary measures such as the 90-50 gap (the difference between the 90th percentile and the median of log wages) and the 50-10 gap (the difference between the median and the 10th percentile of log wages). One important drawback of this approach is that these alternative inequality measures can no longer be decomposed as the sum of a within- and between-group component that add up to the overall inequality measure. For example, the 90-50 gap is not equal to the sum of the 90-50 gap in group means and the 90-50 gap in residuals. As a result, it is not completely clear how to compute the contribution of observable and unobservable prices to changes in the 90-50 or related gaps.

Following Juhn, Murphy, and Pierce (1993); Lemieux (2002); and Autor, Katz, and Kearney (2005), a number of approaches can nonetheless be used to get some indications on the contribution of the various price components to changes in inequality at different points of the distribution. One approach is to look at the distribution of residuals. Consider an individual wage observation y_{ijkt}. The residual u_{ijkt} can be readily computed as the difference

between the individual wage observation and the cell mean y_{jkt} so that $u_{ijkt} = y_{ijkt} - y_{jkt}$.[8] Looking at changes in the distribution of residuals provides some information on changes in unobservable prices.

Another possible approach is to perform a decomposition in spirit of Juhn, Murphy, and Pierce (1993). First remember that the focus here is to first control for composition effects to then see what is explained by changes in the wage structure. Adjusting unconditional quantiles for composition effects is straightforward in the cell-by-cell case considered here. Observation i with education j and experience k in year t can simply be reweighted by the factor θ_{jk}/θ_{jkt} so that the distribution of education and experience remains constant over time (DiNardo, Fortin, and Lemieux 1996; Lemieux 2002). Changes in observable prices can then be controlled for by replacing the actual conditional mean of wages, y_{jkt}, with the average conditional mean for the cell over all years combined, y_{jk}. This yields a counterfactual wage $y'_{ijkt} = y_{ijkt} + (y_{jk} - y_{jkt})$. Remaining changes in the counterfactual wage y'_{ijkt} should then only depend on changes in unobservable prices.

1.3   Basic Results from the May-ORG Data

In this section, I present the basic descriptive facts as well as the decomposition results using the 1973 to 2006 May-ORG data. In this first set of results, I only keep observations with unallocated wages and fix for topcoding using the 1.4 correction factor. I later discuss in section 1.4 what happens when (a) the March CPS is used instead of the May-ORG CPS, (b) allocated wage observations are included (for 1979 on), and (c) other assumptions are used to deal with topcoding. The results are presented using a set of figures and are also summarized in table 1.1. The table shows various measures of wage dispersion at five points of time. At each of these points, three years of data are pooled together (1973 to 1975, 1978 to 1980, 1988 to 1990, 1998 to 2000, and 2004 to 2006) to increase the precision of the results.

1.3.1   Variance over All Experience and Education Groups

The evolution of the overall variance of wages (sum of the between and within component) is first reported in Figure 1.1 for men and women, respectively. A couple of clear patterns emerge from these figures. First, the bulk of the growth in the variance is concentrated in the 1980s. Second, adjusting for composition effects noticeably reduces the long-term growth in the variance. In the case of men, the variance is very stable in the 1970s and 1990s but starts increasing again after 1999. The same pattern can roughly

---

8. It is not clear whether the residual should be defined relative to the conditional mean or median. Juhn, Murphy, and Pierce (1993) and Lemieux (2002) use the conditional mean, while Autor, Katz, and Kearney (2005) use the conditional median. A few experimentations suggested that this choice had little impact on the results.

Table 1.1 **Summary measures of wage inequality for men and women, May-outgoing rotation group Current Population Survey**

	Inequality level					Inequality change		
	1974	1979	1989	1999	2005	1974–1989	1989–2005	1974–2005
A. Men								
Unadjusted variance								
Total	0.251	0.250	0.324	0.340	0.359	0.073	0.035	0.108
Within	0.175	0.173	0.205	0.209	0.218	0.030	0.013	0.044
Between	0.076	0.077	0.119	0.131	0.141	0.043	0.022	0.065
Adjusted variance								
Total	0.262	0.256	0.328	0.330	0.344	0.066	0.016	0.082
Within	0.183	0.179	0.206	0.198	0.202	0.024	–0.005	0.019
Between	0.080	0.076	0.122	0.133	0.143	0.042	0.021	0.063
Other (adjusted) inequality measures								
90-10 gap	1.309	1.285	1.447	1.463	1.498	0.139	0.051	0.190
50-10 gap	0.700	0.696	0.732	0.668	0.660	0.032	–0.072	–0.039
90-50 gap	0.609	0.590	0.715	0.795	0.838	0.106	0.123	0.229
Education wage gaps								
High school-dropout	0.171	0.172	0.218	0.222	0.220	0.047	0.002	0.049
College-high school	0.367	0.320	0.461	0.526	0.548	0.093	0.087	0.181
Post graduate-college	–0.007	0.040	0.098	0.193	0.220	0.105	0.122	0.227
Experience wage gap								
20–29 years-0–9 years	0.390	0.404	0.474	0.399	0.421	0.083	–0.052	0.031
B. Women								
Unadjusted variance								
Total	0.189	0.173	0.261	0.285	0.307	0.072	0.046	0.118
Within	0.137	0.130	0.177	0.180	0.193	0.041	0.016	0.057
Between	0.052	0.043	0.084	0.105	0.114	0.032	0.030	0.062
Adjusted variance								
Total	0.202	0.183	0.262	0.271	0.287	0.061	0.025	0.086
Within	0.143	0.137	0.177	0.166	0.172	0.034	–0.004	0.030
Between	0.059	0.046	0.085	0.105	0.115	0.027	0.030	0.056
Other (adjusted) inequality measures								
90-10 gap	1.069	1.054	1.310	1.315	1.338	0.241	0.028	0.269
50-10 gap	0.503	0.465	0.627	0.550	0.559	0.124	–0.068	0.056
90-50 gap	0.567	0.589	0.683	0.765	0.779	0.117	0.096	0.213
Education wage gaps								
High school-dropout	0.193	0.172	0.244	0.232	0.245	0.051	0.001	0.052
College-high school	0.370	0.320	0.458	0.539	0.560	0.088	0.101	0.189
Post graduate-college	0.189	0.181	0.166	0.230	0.233	–0.023	0.067	0.044
Experience wage gap								
20–29 years-0–9 years	0.130	0.155	0.264	0.275	0.302	0.134	0.038	0.171

Notes: Inequality measures computed by pooling groups of three years centered on the year listed in the table. For example, "1974" corresponds to year 1973 to 1975, and so on. The "adjusted" variance (and other inequality measures) are adjusted for composition effects using the procedure described in the text.
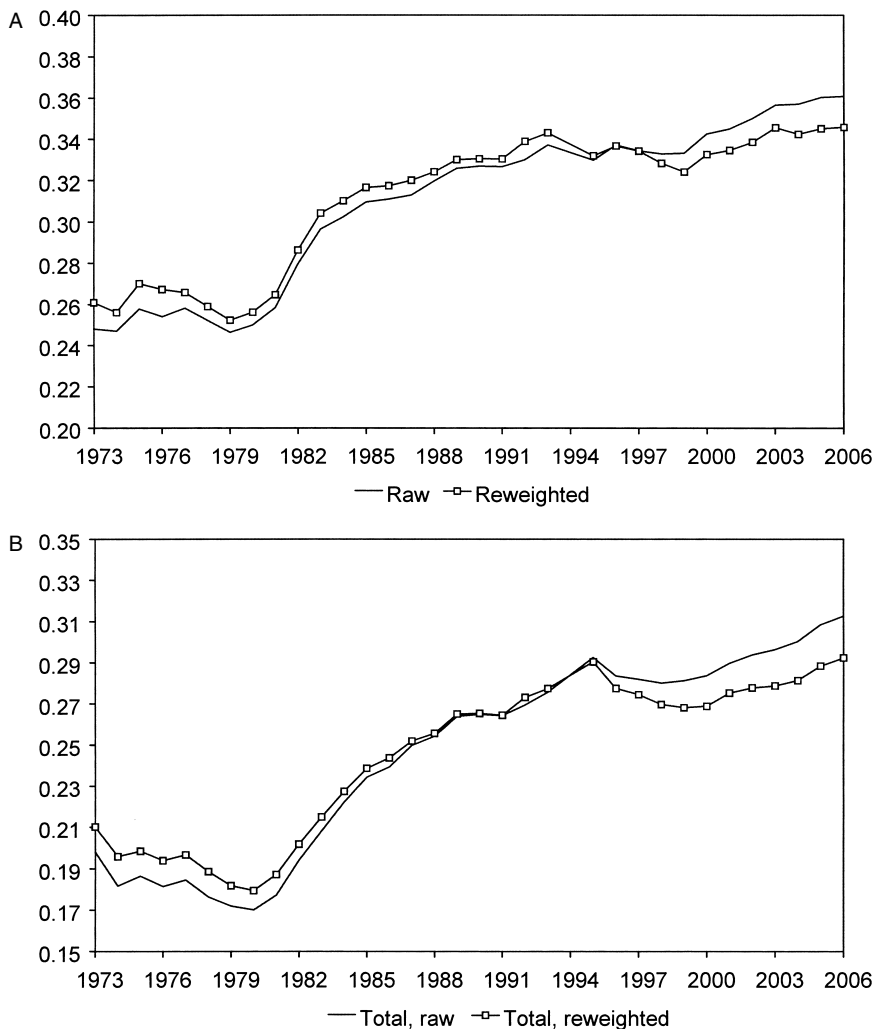
Fig. 1.1   Total variance: *A,* Men; *B,* Women

be observed for women, with the notable difference that the variance clearly declines in the 1970s.

   Figure 1.2 presents a first decomposition by showing separately the evolution of the between- and within-group components of the variance. Three interesting patterns emerge from these figures. First, the pattern of change in the between- and within-group components is remarkably similar over time. For both men and women, both the within- and between-group components grow sharply in the 1980s, grow less in the 2000s, and remain stable or decline (for women in the 1970s) in the 1970s and 1990s. For men, both components

A

B
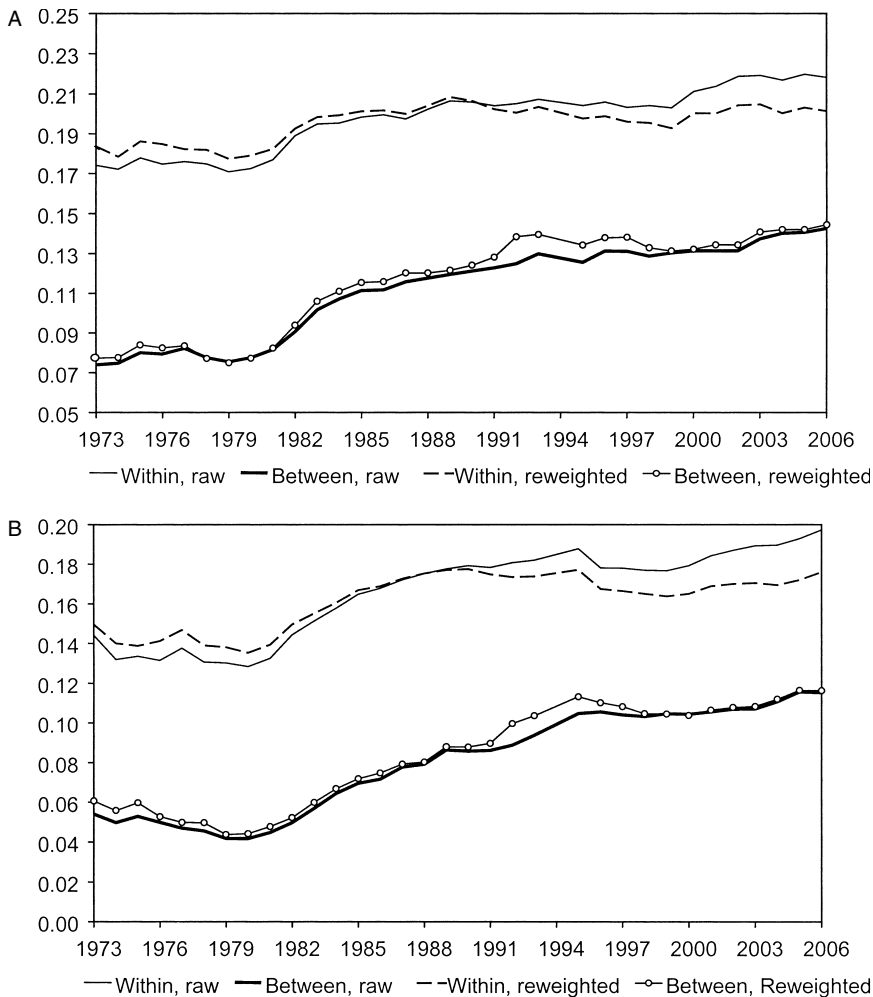
Fig. 1.2   Within- and between-group variances: *A,* Men; *B,* Women

decline slightly in the 1990s once composition effects are adjusted for. For women, the 1990s is the only period where the between- and within-group components move in opposite directions. I argue in the following, however, that part of the growth in the between-group variance for women likely captures a different type of composition effect linked to the growth in actual labor market experience given potential experience. So, on balance, the pattern of change in the within- and between-group components is very similar.

A second important finding is that, consistent with Lemieux (2006a), composition effects account for a very substantial part of the growth in the within-group variance. By contrast, composition effects play little role in

long-run changes in the between-group component. This result holds for both for men and women. The third finding is that the between-group component accounts for most of the growth in the overall variance, especially once composition effects are adjusted for. This particular result is sensitive to the data used (March versus May-ORG CPS) and will be further discussed in section 1.4.

I also show in figure 1B.3 (for men) that using a chain-weighted procedure (see footnote 5) instead of the fixed-weighted approach used in figures 1.1 and 1.2 has very little impact on the results. This suggests that using the fixed-weighted procedure with average weights computed over the whole sample period provides a simple and accurate way of controlling for composition effects and is not as arbitrary as using either base-period or end-period weights.[9]

1.3.2   Relative Wages and Variances by Education and Experience Groups

Because the groups (cells) used from the decomposition are solely based on education and experience, the source of the growth in the (composition-adjusted) between-group component must either come from a growth in education or experience wage differentials. Figure 1.3 shows the evolution of education wage differentials over time, while figure 1.4 shows the evolution of the returns to experience. As is well known, education wage differentials increased for both men and women in the 1980s. Relative to high school graduates, panel A of figure 1.3 shows that the wage advantage of men with some college or more increased, while the wage disadvantage of high school dropouts or those with eight years of education or less (elementary category in the figure) also expanded, albeit more modestly. By contrast, in the 1990s and 2000s, most education wage differentials remained stable except for college graduates and postgraduates, who kept gaining relative to high school graduates.

Looking at the whole 1973 to 2006 period, the striking fact is that most of the expansion in wage differentials has been happening at the top of the education distribution. In particular, the gap between college postgraduates and high school graduates more than doubled over time. College graduates and people with some college also experienced substantial gains, while relative wages among workers with a high school diploma or less remained very stable over time. This mirrors the finding by Lemieux (2006b) of a growing convexification in the relationship between wages and education. In other words, wage differentials at the top end of the education distribution have increased steadily over time, while wage differentials at the bottom end remained more or less stable. As I will show in the following, this fits with

9. The adjustment for composition effects in the within-group variance are substantially larger when base-period are used instead of end-period weights (Lemieux 2006a).
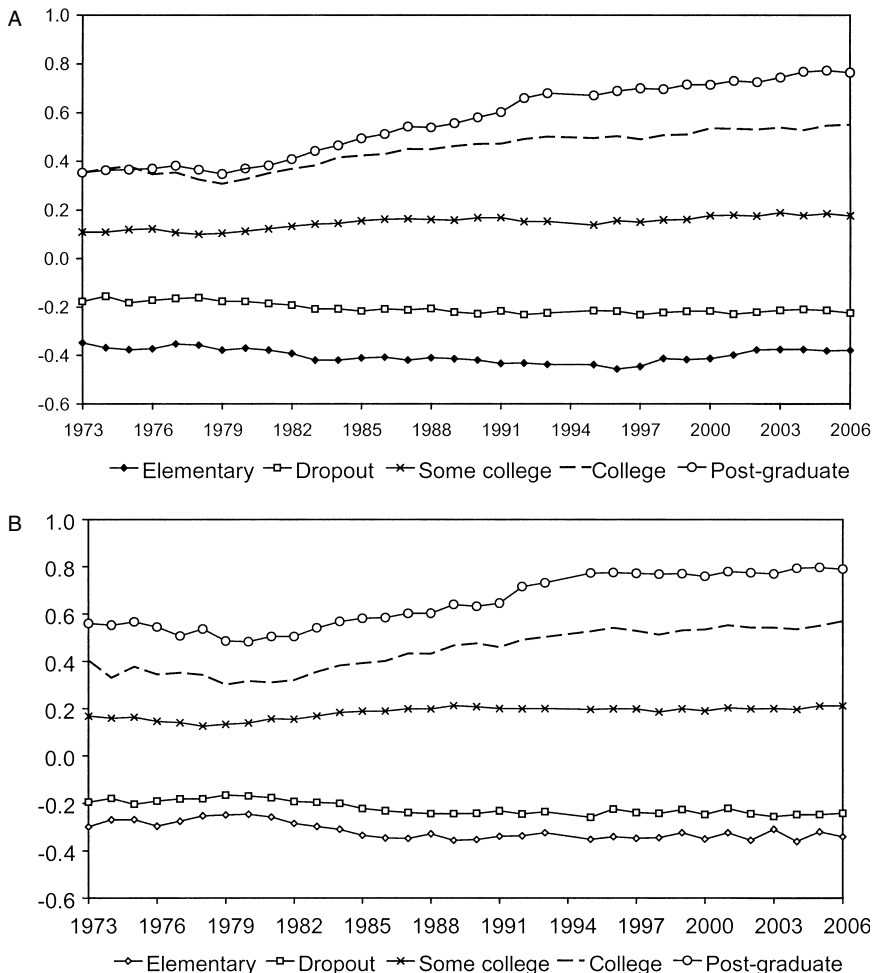
A

B

Fig. 1.3   Education wage differentials (relative to high school graduates): *A,* Men; *B,* Women

the general pattern that inequality has kept growing at the top end but not at the low end of the distribution (Autor, Katz, and Kearney 2005). Note also that the results for women in panel B of figure 1.3 are qualitatively similar to those for men.

The evolution of experience wage differentials is shown in figure 1.4. Because it would not be very informative to present the differentials for each of the twenty-one experience groups, I have regrouped the two-year experience groups into in four major experience groups (zero to nine, ten to nineteen, twenty to twenty-nine, and thirty to thirty-nine years of experience) by computing a weighted mean of the wages differentials for the
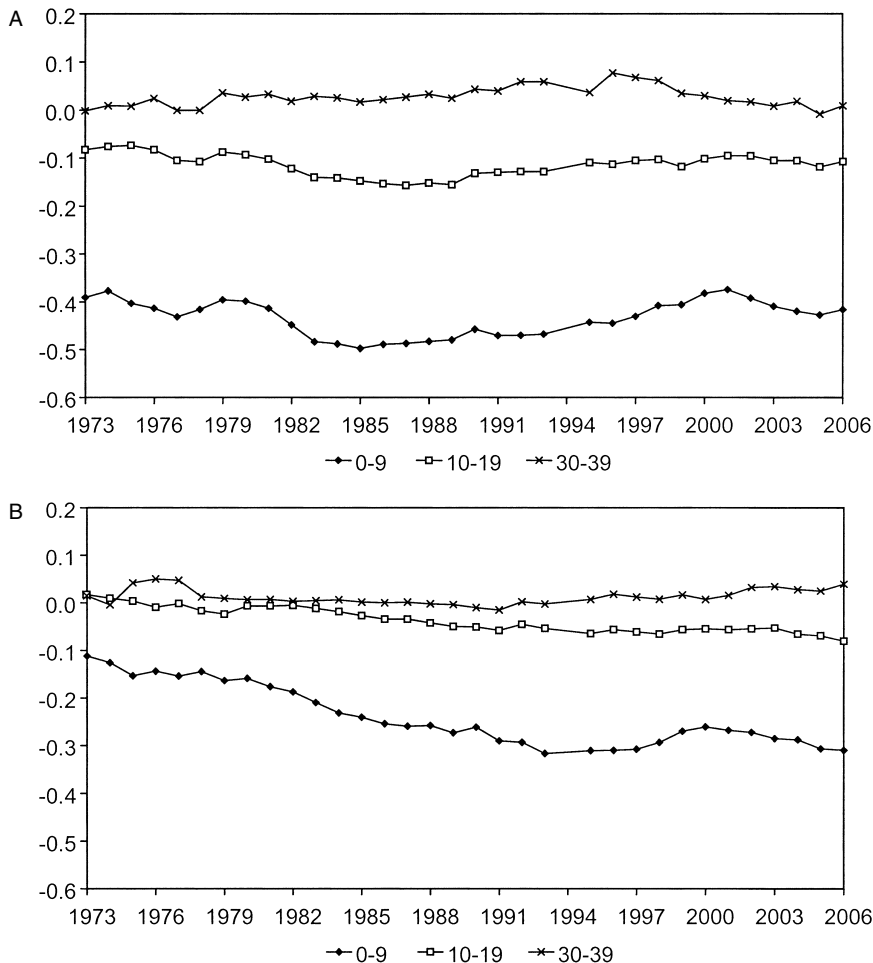
Fig. 1.4   Experience wages differentials (relative to 20–29 years of experience):
A, Men; *B,* Women

smaller groups. The differentials reported in the figure show the gap relative to workers with twenty to twenty-nine years of experience. Consistent with earlier studies such as Katz and Murphy (1992), there is a clear expansion in experience wage differentials in the 1980s. The differentials start contracting again, however, in the late 1980s, and by the end of the sample period, the differentials are more or less back to their initial 1973 levels. The situation is quite different for women (panel B of figure 1.4). Differentials keep expanding over time except in the late 1990s when inexperienced (zero to nine years of experience) women gain ground again, probably because of the very strong labor market at that time. By 2006, the gap

between experienced (twenty to twenty-nine years of experience) and inexperienced (zero to nine years of experience) women is three times as large as back in 1973 and has almost caught up with the level of the corresponding gap for men.

The most plausible explanation for growing experience gaps for women is that the large increases in female employment rates have lead to a corresponding increase in the level of actual labor market experience conditional on potential experience.[10] For instance, women with twenty years of potential experience now have much more actual experience than they used to. As a result, their wage advantage over inexperienced women should have increased substantially even if the return to actual experience did not grow over time. If actual experience was measured in the CPS, changes in actual experience could be corrected for by holding the distribution of education and actual (as opposed to potential) experience constant over time. Because only potential experience is available in the CPS, the change in experience wage differentials documented in panel B of figure 1.4 should be interpreted with caution. Trends in experience differentials for men are more reliable and suggest that "real" returns to experience played little role in the growth in wage dispersion.

Taken together, the results in figures 1.2 to 1.4 suggest that, at least for men, changes in the between-group variance induced by growing top-end education wage differentials account for most of the growth in the variance of wages, a results also found in more formal decompositions by Lemieux (2006b); Firpo, Fortin, and Lemieux (2007); and Goldin and Katz (2007).

Turning to the within-group component, figures 1.5 and 1.6 show the evolution of the within-group variance by education and experience groups. Like the wage differentials, the within-group variance by education groups are computed as fixed-weighted averages across experience groups and vice versa. The within-group variance for education group j at time t is

$$V_{jt} = \Sigma_k \, \theta_k \, v_{jkt},$$

while the within-group variance for experience group k at time t is

$$V_{kt} = \Sigma_j \, \theta_j \, v_{jkt}.$$

As before, the different experience groups are combined into four broader groups for the ease of exposition.

Figure 1.5 shows that the within-group variance moves quite differently over time for different education groups. Dispersion increases among all education groups in the 1980s, keeps growing for the more-educated groups (college graduates and postgraduates) after 1990, but remains stable or even

---

10. Using Panel Study of Income Dynamics (PSID) data, Blau and Kahn (1996) show that actual experience has indeed increased a lot over time for given levels of potential experience.
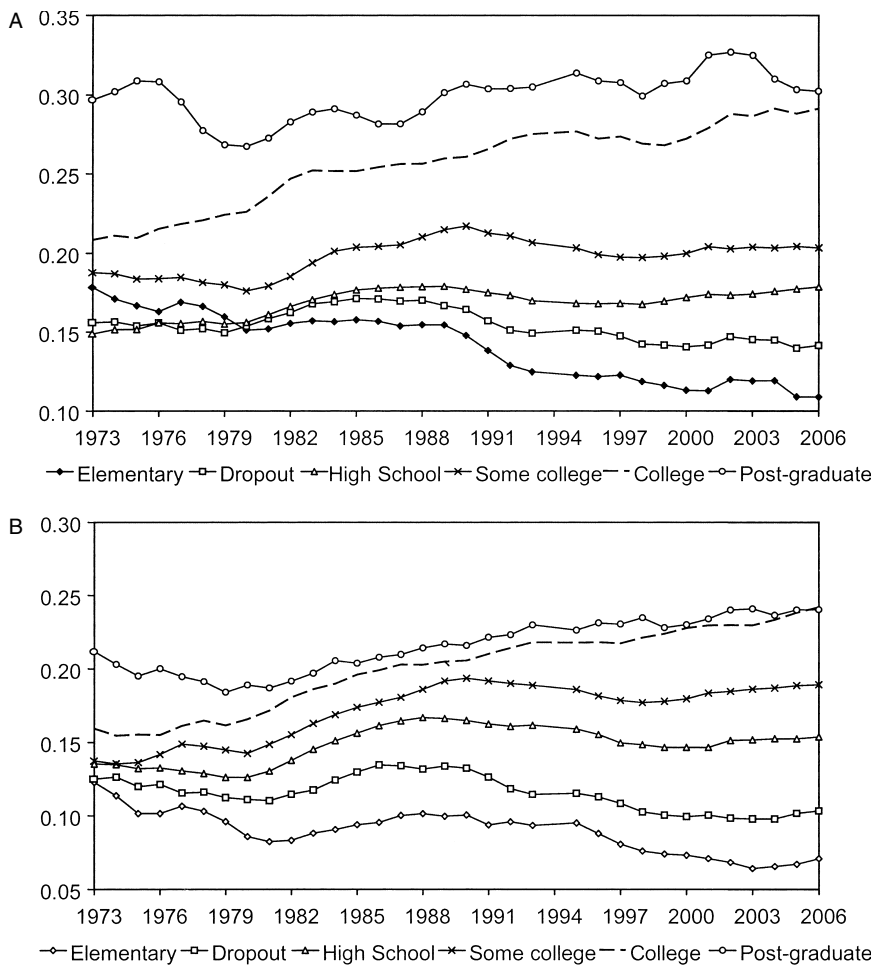
A

B

Fig. 1.5   **Within-group variance by education groups:** *A,* **Men;** *B,* **Women**

declines for less-educated groups. Interestingly, the evolution in within-group dispersion closely mirrors the evolution of relative wages by education groups, which reinforces the earlier conclusion that inequality growth in concentrated at the top end of the distribution.

By contrast, changes in within-group wage dispersion by experience groups are more homogenous across groups. Roughly speaking, the within-group variance increases for all experience groups during the 1980s but remains more or less constant after 1990. The only exception is inexperienced women for whom the within-group variance declines after reaching a peak in the late 1980s.
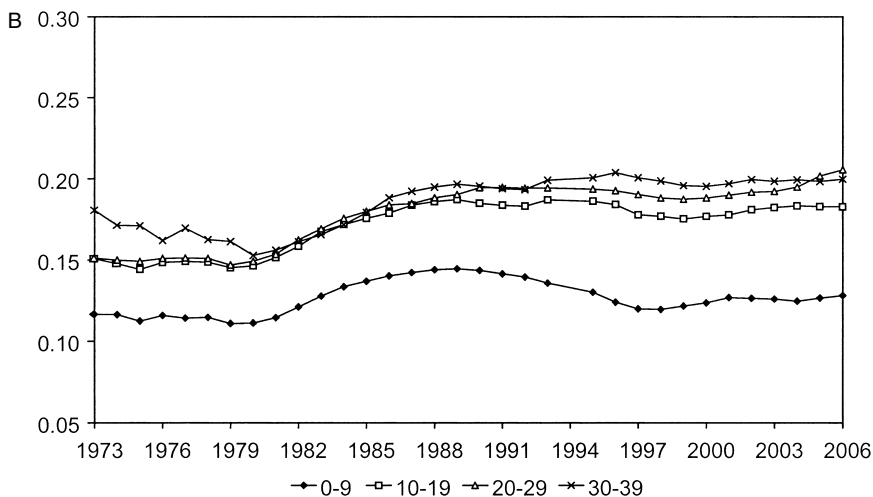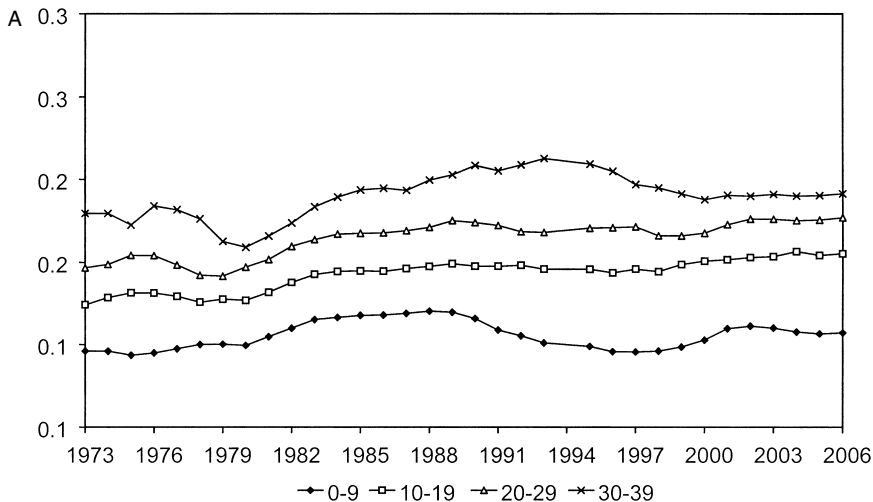
Fig. 1.6   Variance by experience groups: *A,* Men; *B,* Women

Note also that education has a much larger effect on the within-group variance than experience. In figures 1.5 and 1.6, the range of variation in within-group variances (in a single year) is about twice as large between the lowest and the highest variance groups for education as it is for experience. This suggests that secular changes in the distribution of education potentially account for more of the composition effects than changes in the distribution of experience. Panels A and B of figure 1B.4 indicate that this is indeed the case and that about two-thirds of

the composition effects can be linked to changes in the distribution of education.[11]

Taken together, the results in figures 1.2 to 1.6 suggest that changes in the relative wages and variances of highly educated workers is a key element in the growth of wage inequality since the late 1970s. There is indeed a very intriguing parallel between what is happening to the between- and within-group components of wage dispersion. For both components, changes are concentrated in the 1980s, and long-run growth is concentrated among college graduates and postgraduates. This suggests that both components may be reflecting the same underlying changes in the labor market, an issue to which I return in section 1.6. This also suggests, consistent with Autor, Katz, and Kearney (2005), that changes in the top end of the distribution are very different from changes at the low end. I now explore this issue in more detail by looking at what happened at different percentiles of the wage distribution.

1.3.3   Changes at Different Percentiles

Figure 1.7 plots the changes in real wages at each percentile over the 1974 to 1989, 1989 to 2004, and the whole 1974 to 2004 period. I use fifteen-year changes for both periods for the sake of comparability and also pool three years of data around 1974 (1973 to 1975), 1989 (1988 to 1990), and 2004 (2003 to 2005) to increase the precision of estimates at each percentile. Similar results for wages residuals are presented in figure 1.8. The results essentially reproduce the findings of Autor, Katz, and Kearney (2005). The main point is that, consistent with Juhn, Murphy, and Pierce (1993), wage changes are more or less a linear and positively sloped function of percentiles during the 1974 to 1989 period, suggesting similar changes in wage dispersion at all points of the wage distribution. The situation is radically different for the 1989 to 2004 period, however. While wage dispersion keeps growing above the median, wages become more compressed below the median as real wage gains at the bottom end exceed those around the median, a phenomena Autor, Katz, and Kearney (2006) refer to as the polarization of the labor market. Note that this phenomenon can be observed both for wages and wage residuals (figure 1.8).

Figure 1.9 then performs a Juhn, Murphy, and Pierce (1993) decomposition on the 90-50 and 50-10 gap using the procedure described in section 1.2. The results are, once again, very similar to those of Autor, Katz, and Kearney (2005). Basically, the figures show that at the top end, composition effects explain little of the growth in the 90-50 gap, while prices of observables account for more than half of the growth in the gap. As shown in the preceding, the relevant observable prices that likely account for most of the

---

11. The lines labeled "education fixed" in panels A and B of figure 1B.4 show what happens when the distribution of education is fixed to its average value over the whole sample period while the distribution of experience conditional on education remains as observed in the data.
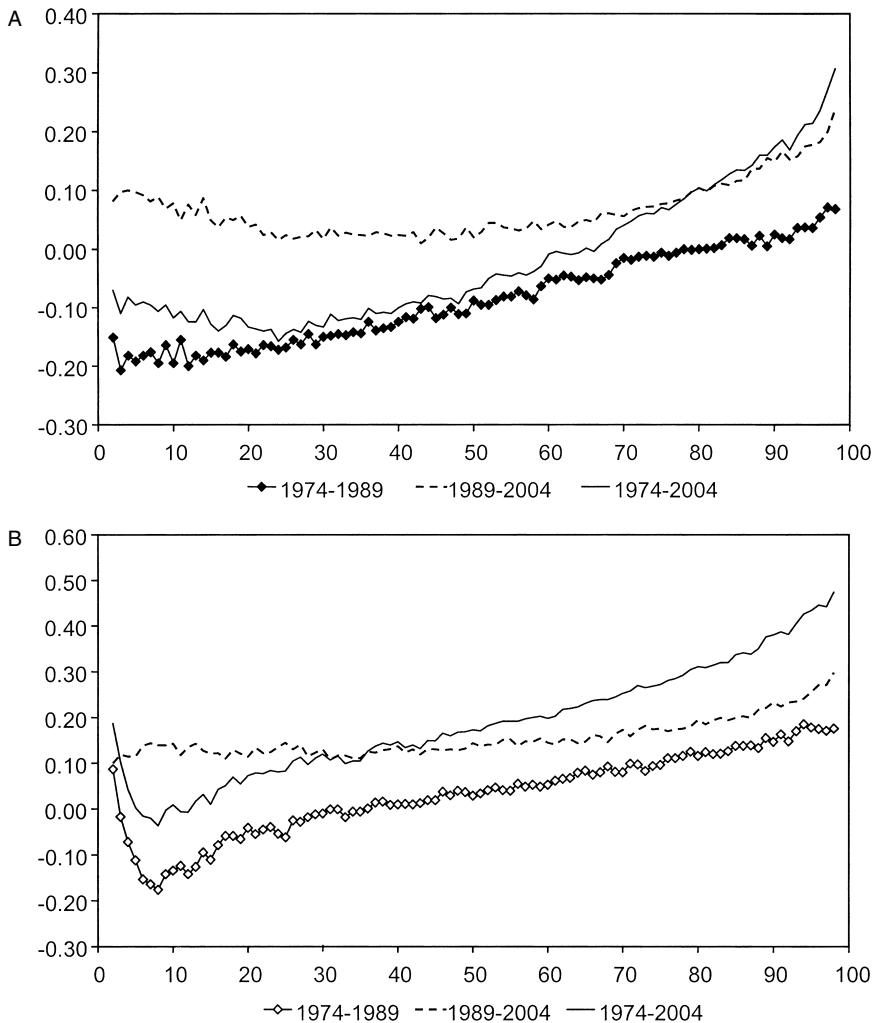
A

B

Fig. 1.7   **Change in real wages by percentile: *A,* Men; *B,* Women**

growth in the 90-50 gap are the wage differentials between college graduates and postgraduates relative to high school graduates. The remaining change in the 90-50 gap is due to changes in unobservable prices that are also likely driven by changes (growing within-group variance) happening among highly educated workers.[12]

12. Using a more sophisticated quantile decomposition, Firpo, Fortin, and Lemieux (2007) indeed find that education accounts for the bulk of the growth in the 90-50 gap. Their education "effects" include both the between- and within-group components.
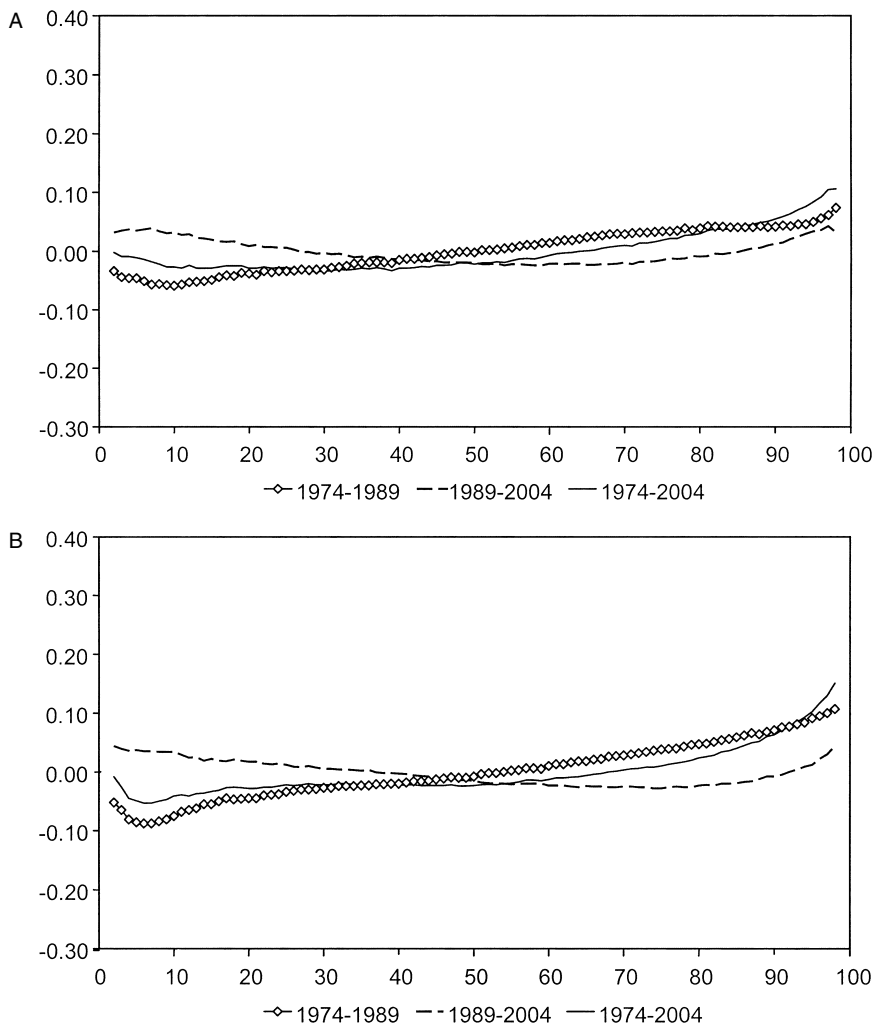
A

B

Fig. 1.8   Change in wages residuals by percentile: *A,* Men; *B,* Women

The situation is very different at the lower end of the distribution (50-10 gap). Unlike the case of the 90-50 gap, there are also substantial differences between men and women. I thus discuss these two cases separately. For men, the unadjusted 50-10 gap reported in panel B of figure 1.9 grows in the 1980s but more or less returns to its 1970s level by the late 1990s or early 2000s. After controlling for composition effects, however, the 50-10 declines substantially and is lower in the early 2000s than in the 1970s. While observable prices explain a substantial part of the growth in the 1980s and decline in the 1990s, they have little impact on long-run changes between 1973 and 2006.
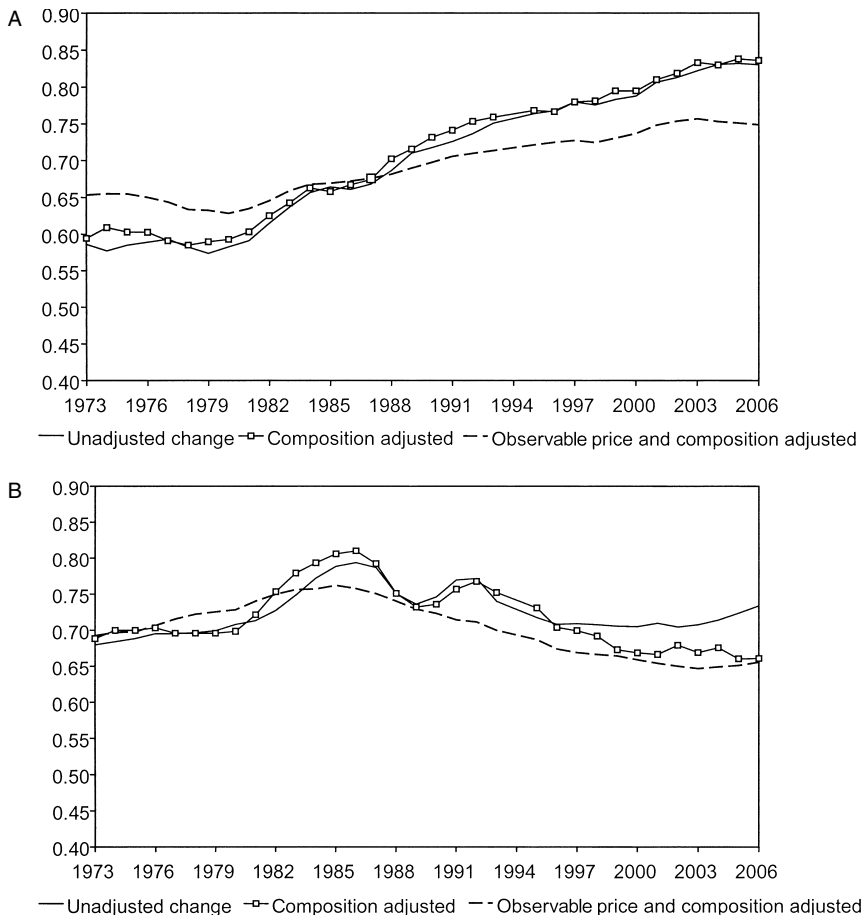
A

B

Fig. 1.9   Decomposition of changes in 90-50 and 50-10 gaps: *A,* Decomposition of changes in 90-50 gap, men; *B,* Decomposition of changes in the 50-10 gap, men; *C,* Decomposition of changes in 90-50 gap, women; *D,* Decomposition of changes in 50-10 gap, women

The pattern of changes suggests that movements in the return to experience plays an important role here. The fact that the relative wages of young workers fell in the 1980s and went back up in the 1990s likely accounts for the swings in the 50-10 gap because young workers are overrepresented at the bottom of the wage distribution. Furthermore, the stability of education differentials at the low end of the education distribution is consistent with the lack of observable price effects in the change in the 50-10 gap. Finally, the "residual" explanation for the decline in the 50-10 after adjusting for composition and observable price effects likely has to do with the decline in the within-group variance among less-educated workers.
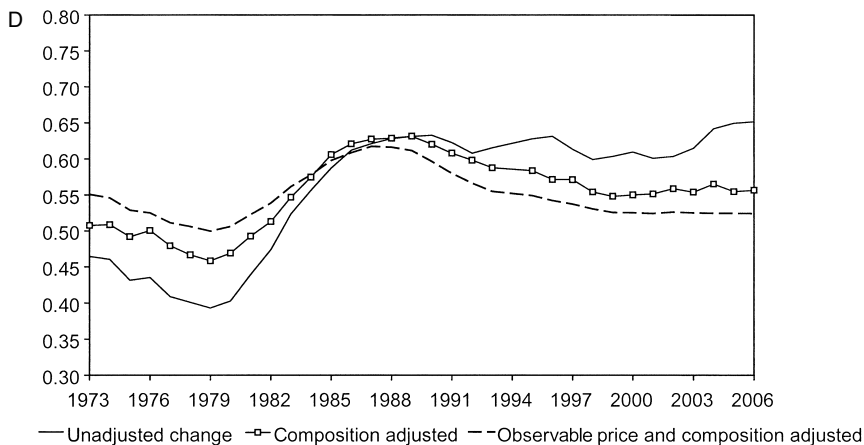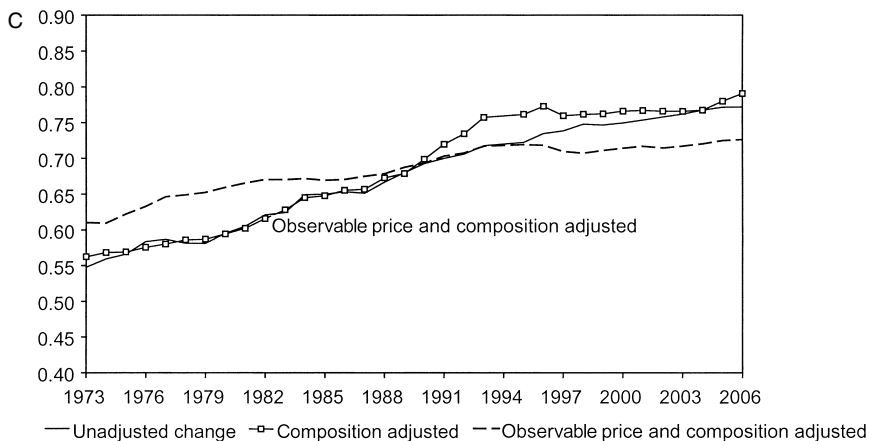
Fig. 1.9   (cont.)

In the case of women, there is a much steeper growth in the 50-10 gap in the 1980s, a phenomena likely linked to the large decline in the real value of the minimum wages over this period (DiNardo, Fortin, and Lemieux 1996).[13] As a result, the 50-10 gap in the early 2000s remains substantially higher than in the 1970s. As in the case of men, composition effects account for a substantial part of the growth in the 50-10 gap. Unlike men, however, changes in observable prices also account for a significant part of the growth in the 50-10 gap, which is consistent with the continuing growth in experience wage differentials documented in panel B of figure 1.4. Remember, however,

13. DiNardo, Fortin, and Lemieux (1996) find that over 60 percent of the increase in the 50-10 gap between 1979 and 1998 can be accounted for by the decline in the real value of the minimum wage.

that composition effects linked to the mismeasurement of actual experience likely accounts for much of these changes. By 2006, the 50-10 gap is a little smaller than back in 1973, suggesting that, as in the case of men, changes in unobservable prices lead to a small reduction in the 50-10 gap over time.

1.4   Measurement Issues: What Are the Robust Facts?

All the findings reported up to now are based on unallocated wage observations from the May-ORG CPS where topcoding is adjusted for using a 1.4 imputation factor. I now look at how robust the main results are to these data processing assumptions, focusing on trends in the (composition-adjusted) within- and between-group variance.

1.4.1   Wage Allocation

As mentioned in section 1.2, wages were not allocated (imputed) for individuals who failed to report their wages and earnings in the 1973 to 1978 CPS. In 1979, the Census Bureau started allocating wages for these individuals using the well-known hot deck matching procedure. Back in 1979, 17.9 percent of male and 14.8 percent of female workers did not report their wages. By 2006, the nonresponse rate had grown to a staggering 35.6 percent for male workers and 34.1 percent for female workers. Unless nonresponse is completely random, excluding workers with allocated wages could bias the trends in wage inequality measures. Of course, correcting for nonresponse is not perfect either, as assumptions have to be made about the determinants of nonresponse. The standard approach used by the Census Bureau to correct for nonresponse is to use a matching procedure where the missing wage is replaced with the wage of a "donor" with similar observed characteristics (location, education, age, race, etc). Note that this matching or hot decking procedure can be thought of as a stochastic imputation procedure. Instead of imputing a wage based on a regression model (e.g., Lillard, Smith, and Welch 1986), such a matching procedure preserves the wage dispersion conditional on characteristics, which is important when looking at wage dispersion, in general, and at within-group wage dispersion, in particular.

Figure 1.10 shows the difference in the between- and within-group variances computed with and without allocated wage observations. The series with allocated wage observations only starts in 1979 because, as mentioned earlier, the Census Bureau did not provide allocated wage observations in the 1973 to 1978 May CPS (Hirsch and Schumacher 2004). The main message from figure 1.10 is that, fortunately, the trends in wage inequality are fairly robust to the treatment of allocated wage observations. For both men and women, adding back the allocated wage observations reduces the between-group variance and increases the within-group variances. In terms of trends, however, the only noticeable difference is that the between-group component grows a little slower over time when allocated wage observations
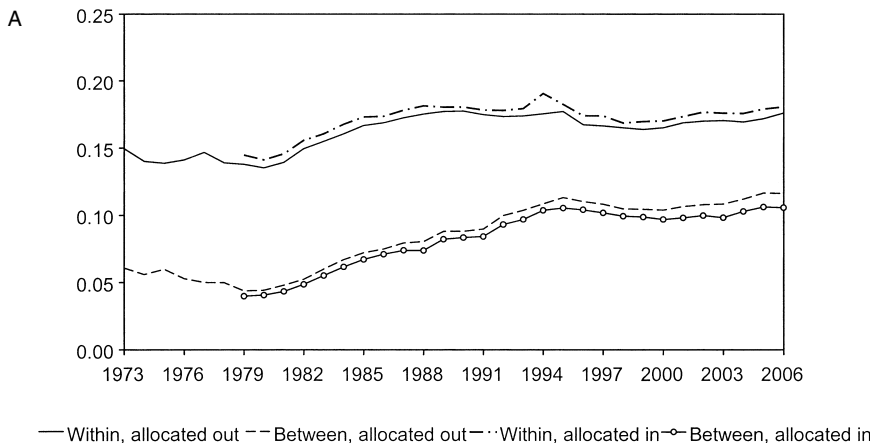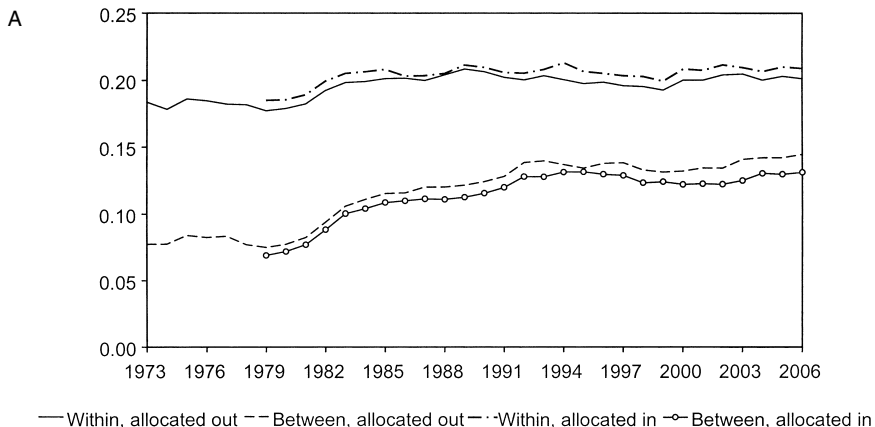
A

— Within, allocated out — — Between, allocated out — ·· Within, allocated in —◦— Between, allocated in

A

— Within, allocated out — — Between, allocated out — ·· Within, allocated in —◦— Between, allocated in

Fig. 1.10   Variance with and without allocated wages controlling for composition effects: *A,* **Men;** *B,* **Women**

are included. Another noticeable difference is the "blip" in the within-group variance in 1994, the year the redesigned and computer-based CPS was introduced. Because allocation flags were not included in the 1994 (and part of 1995; see section 1.2) CPS, I computed the 1994 data points in the series without allocators as a simple interpolation based on the 1993 and 1995 numbers. As such, the two series are not comparable for 1994. Because 1994 seems to be just a one time blip, I did not perform any systematic adjustments to take account of the CPS redesign in 1994.[14]

14. If anything, the sharp increase in the within-group variance between 1993 and 1994 would suggest that the redesign lead to a spurious increase in the within-group variance, which reinforces the conclusion that the within-group variance did not increase after 1990.

1.4.2   Topcoding in the May-ORG CPS

Figure 1.11 compares the variance with and without the 1.4 topcoding adjustment. First notice that, in the case of women (panel B of figure 1.11), adjusting for topcoding has essentially no effect on either the between- or the within-group variance. This is hardly a surprise because only a small fraction of women have earnings at or above the topcode.

The adjustment for topcoding has a more noticeable impact in the case of men (panel B of figure 1.11). The impact of the increase in the topcode in 1986 and 1998 is clearly visible in the case of the within-group variance, where the unadjusted series experience unusual jumps in those two years. The impact of the increases in the topcode is also visible in the case of the between-group variance. Generally speaking, the impact of the correction for topcoding is larger for the within- than for the between-group component. Overall, adjusting for topcoding tends to modestly increase the growth in inequality over time. This pattern is consistent with the finding of Piketty and Saez (2003) that inequality at the very top end of the distribution has increased rapidly since the 1970s. Because an important part of these changes are missed because of the topcoding in the CPS, it is natural to expect the topcoding adjustment to result in more inequality growth. What is not clear, however, is whether a simple and time-invariant adjustment like the 1.4 imputation factor adequately captures all of the inequality growth at the very top end. I will return to this issue in more detail in the case of the March CPS where additional information from the tax data can be used to devise a better imputation procedure.

1.4.3   March versus May-ORG CPS

Differences in inequality trends in the May-ORG and March supplements of the CPS have been well documented by Lemieux (2006a) and Autor, Katz, and Kearney (2008). The most striking discrepancy between the series shown in figure 1.12 is that the within-group variance is substantially higher in the March than in the May-ORG supplement of the CPS. Lemieux (2006a) shows that this gap mostly reflect the fact that wages of workers paid by the hour are less precisely measured in the March CPS relative to the May-ORG CPS. This particular finding is reproduced in panels A (men) and B (women) of figure 1B.1. The problem with the March CPS is that all workers are only asked about their annual earnings and hours. An hourly wage is then obtained by dividing annual earnings by hours. In the May-ORG CPS, however, workers paid by the hour are asked directly about their hourly wage rates, which yields a more precise measure of hourly wages than in the March CPS (Lemieux 2006a).

Because (classical) measurement error cancels out when wages are averaged out at the cell level, mismeasurement of hourly wages in the March CPS should not affect much the between-group variance. This yields the simple
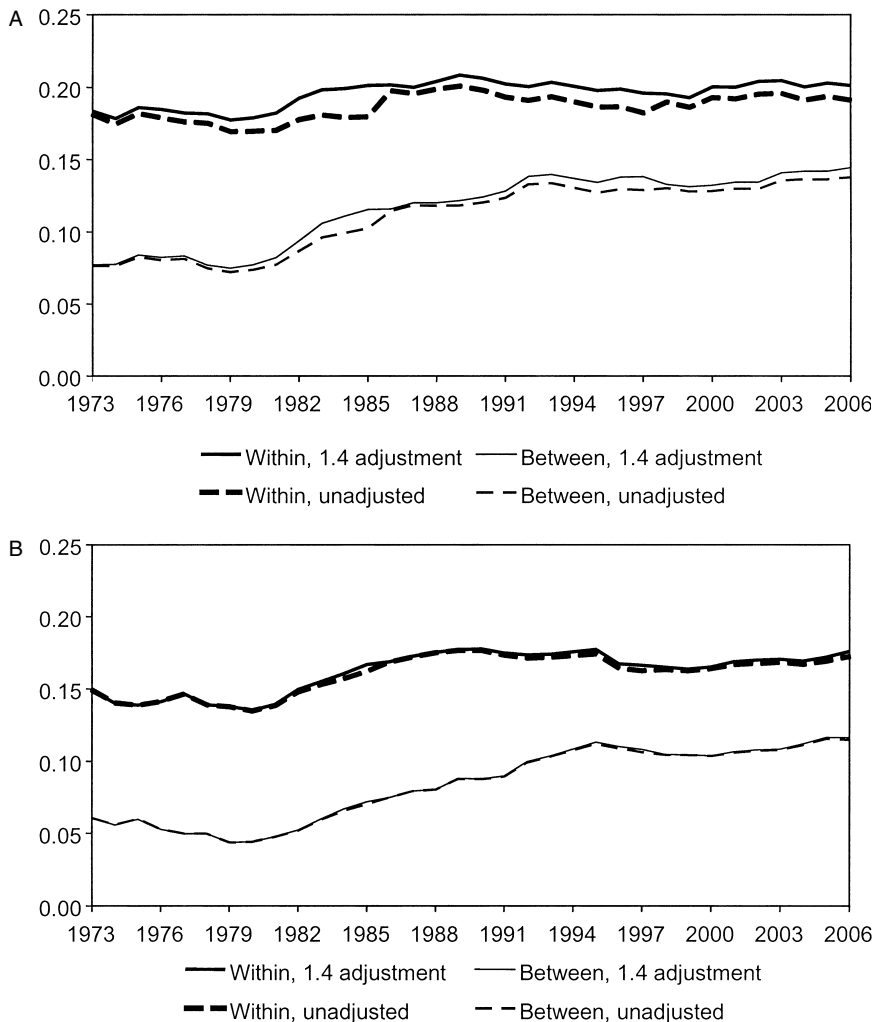
A

B

Fig. 1.11   Effect of topcoding adjustment, controlling for composition effects:
A, Men; *B,* Women

prediction, strongly supported by the data in figure 1.12, that the between-group variance should be about the same in the March and May-ORG CPS, and that the within-group variance should be larger in the March than in the May-ORG CPS.

A more challenging pattern to explain on the basis of these measurement issues is the fact that the trend growth in the within-group variance is higher in the March than in the May-ORG CPS. One possible explanation suggested by Lemieux (2006a) is that the fraction of workers paid by the
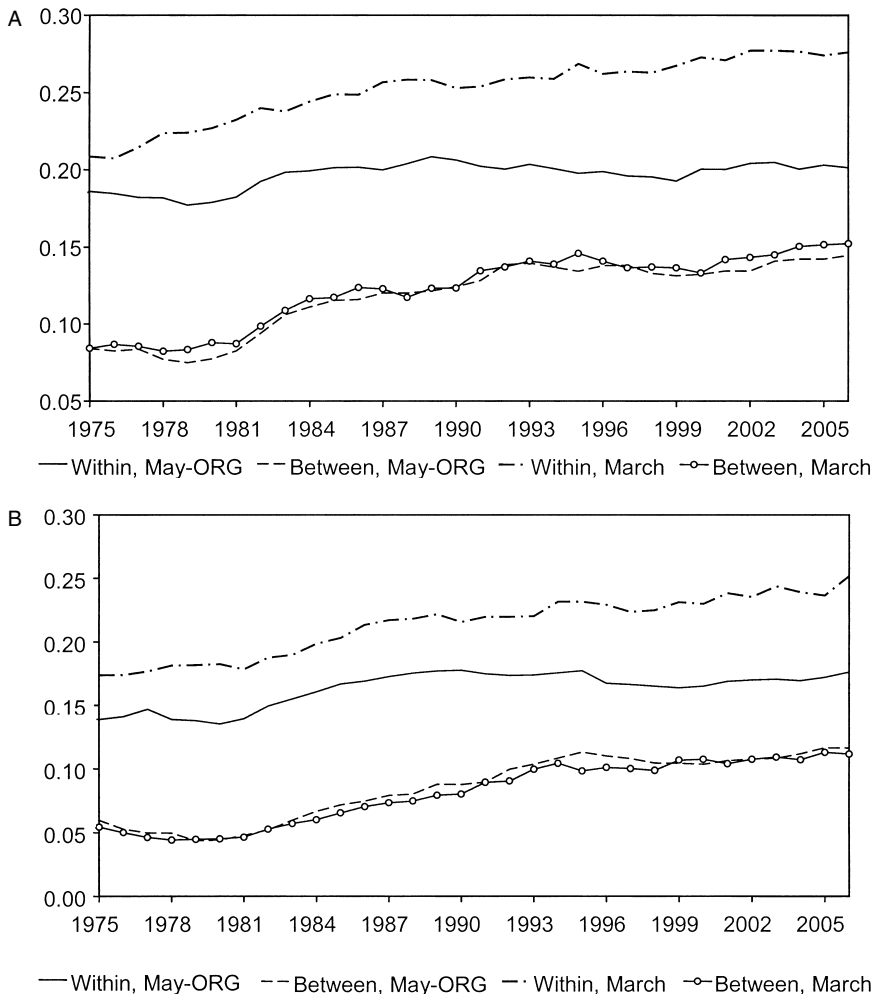
A

B

Fig. 1.12   Variance in May-ORG versus March CPS, controlling for composition effects: *A,* **Men;** *B,* **Women**

hour has increased over time (Hamermesh 2002), thereby magnifying the mismeasurement problems in the March CPS. As pointed out by Autor, Katz, and Kearney (2005), however, this could also bias down the trend in the within-group variance if there is less measurement error in wages of hourly than nonhourly paid workers in the May ORG CPS. In any case, Lemieux (2006a) also shows, using an error-components analysis, that changes in the fraction of hourly rated workers cannot account for most of the discrepancy between the two series. The evidence rather points out to measurement error (conditional on the hourly pay status) increasing over time

in the March CPS, though no explicit story is provided for why this may be the case.

More important, however, both data sources show that the within-group variance grew much faster before than after the late 1980s, once composition effect are controlled for (as they are in figure 1.12). The only difference is that the within-group variance completely stops growing in the May-ORG CPS, while it grows at a much slower rate in the March CPS. So while the two series yield slightly different patterns of growth, in both cases the secular trends in the within-group variance closely mirror those observed for the between-group variance, which also grows much more slowly after than before the late 1980s. The only case where the two data series clearly differ is men in the 1970s. Consistent with Juhn, Murphy, and Pierce (1993), the within-group variance for men in the March CPS increases in the 1970s, while the between-group component remains more or less stable during this period. This particular series stands out as the only exception where inequality grows in the 1970s. By contrast, the within-group variance for women, the within-group variance for men in the May-ORG CPS, and the between-group variance all remain stable in the 1970s. This suggests that, at a minimum, the growth in the within-group variance for men in the March CPS should be interpreted with caution as it is not a very robust finding.

Leaving the 1970s aside, an important point is that, for men, most of the growth in the variance of wages between 1980 and 2006 comes from the between- as opposed to the within-group component. This is similar to what was documented earlier using the May-ORG CPS data. This highlights, once again, the importance of the growing education premia for college educated workers in the overall growth in wage inequality. For women, the growth in the overall variance during this period is, more or less, evenly split between the within- and between-group components. On balance, the earlier results that most of the growth in inequality is due to the between-group component are thus reasonably robust to the choice of data set.

1.4.4   Topcoding in the March CPS

Figure 1.13 shows what happens to the within-and between-group variances when (a) topcoding is not corrected for (as was done for the May-ORG data), and (b) a more sophisticated imputation procedure is used to allocate earnings for topcoded observations. Unlike the May-ORG CPS, other data sources can be used to make reasonable assumptions about the distribution of annual earnings above the topcode. In particular, because tax data are not topcoded, they provide direct information on the distribution of earnings above the topcode. The usual assumption made is that the upper tail of the earnings distribution follows a Pareto distribution. Piketty and Saez (2003) use the Pareto distribution to smooth their data on top incomes from
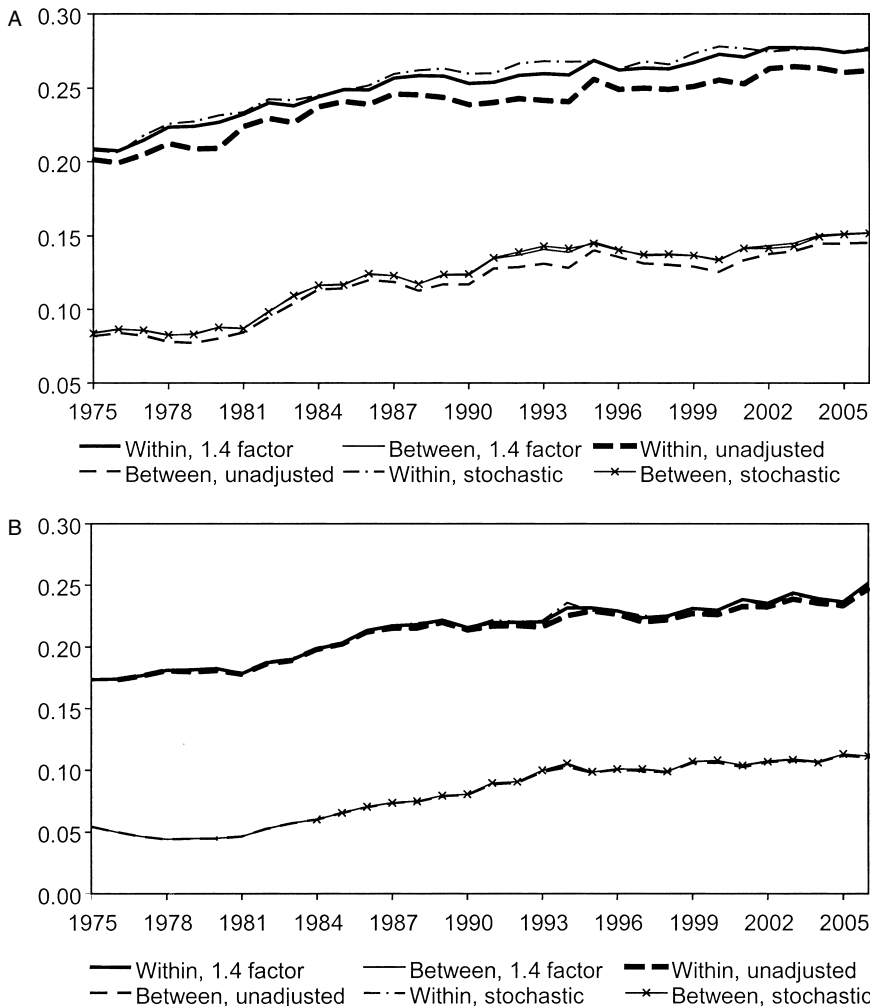
Fig. 1.13   Topcoding adjustments in March CPS, controlling for composition effects: *A,* **Men;** *B,* **Women**

tax data, and I use their implied Pareto parameter for each year to impute earnings above the March CPS topcode.[15]

Another limitation of the standard "fixed" imputation procedure based on a factor of 1.4 or 1.5 is that it does not preserve the distribution of earnings in the upper tail of the distribution. Intuitively, using a fixed as opposed to a stochastic imputation procedure likely has little impact for the between-

15. I use the updated series available at http://elsa.berkeley.edu/~saez/TabFig2005s.xls to compute the Pareto parameter up to 2005 and then use for 2006 the same parameter as in 2005.

group variance, but it could bias down the within-group variance by under-stating wage dispersion among topcoded workers.

Figure 1B.2 shows the topcoding adjustment factors obtained from Pik-etty and Saez's (2003) tax data. One very convenient feature of the Pareto distribution is that the Pareto parameter, α, can be estimated using only the value of a topcode y_{TC}, and the mean value for observations above the topcode, μ_{TC}, using the formula

$$\frac{\mu_{TC}}{y_{TC}} = \frac{\alpha}{(\alpha-1)}.$$

Because we typically work with log wages, it is also convenient to compute the mean log value from a Pareto distribution. Another convenient prop-erty is that when y follows a Pareto distribution with a parameter α, then $\log(y/y_{TC})$ follows an exponential distribution with a parameter α. The mean of $\log(y/y_{TC})$ is thus $1/\alpha$, which implies that

$$E[\log(y)] = \log(y_{TC}) + \frac{1}{\alpha}.$$

With the standard 1.4 imputation factor, it follows that $y = 1.4y_{TC}$. This corresponds to the case of a Pareto distribution with $1/\alpha = \log(1.4) = 0.336$, where the mean value of $\log(y)$ from this distribution is imputed to all observations. Figure 1B.2 shows that the actual value of $1/\alpha$ has been growing over time as top-end inequality expanded, though it decreased a bit following the end of the high-tech bubble around 2000. The figure also shows another adjustment factor from the March CPS. Since 1995, top-coded observations in the March CPS have their earnings replaced by the average earnings for their relevant (broadly defined) demographic group. This is now doable because earnings are no longer topcoded in the CPS (computerized) questionnaire, though they are topcoded in public use files. The implied Pareto parameter is remarkably similar to the one from the tax data, which gives a lot of confidence in the imputation procedure based on the tax data.[16]

In practice, I implement the stochastic imputation procedure by randomly drawing value z from an exponential distribution (with mean 1) and trans-forming them into draws from a Pareto distribution by exploiting the link between the Pareto and the exponential distribution discussed in the pre-ceding:[17]

$$\log(y) = \log(y_{TC}) + \frac{z}{\alpha}.$$

16. I use the number of married white males, by far the largest group of topcoded workers, to compute the Pareto parameter in the CPS.
17. I use a three-year moving average for the adjustment factor from the tax data to smooth for some of the erratic behavior shown in figure 1B.2.

The results reported in panel A of figure 1.13 for men show that using either the 1.4 fixed imputation factor or the stochastic imputation procedure yield very similar estimates of the between-group variance. As in the case of the May-ORG, failing to adjust for topcoding slightly biases down the secular growth in the between-group variance, but the overall impact is very small.

Turning to the within-group variance, the fixed and stochastic imputation procedures yield, perhaps surprisingly, fairly similar results. In fact, the change in the within-group variance between 1973 and 2006 is essential identical under the two imputation procedures. One small difference is that the stochastic imputation procedure tends to produce a larger variance from the mid-1980s to the late 1990s, which slightly accentuates the slowdown in the growth in the within-group variance after the late 1980s. Another difference is that the within-group variance obtained using the stochastic imputation procedure does not change when the topcode is raised in 1995 and 2002. By contrast, the within-group variance computed with the fixed 1.4 correction jumps almost as much as the uncorrected series in these two years. This suggests that the stochastic imputation procedure does generally a better job at correcting for topcoding and that year-to-year changes are better measured using this procedure instead of the fixed imputation procedure. In terms of general inequality trends, however, panel A of figure 1.13 suggests that the two topcoding adjustments produce very similar results. Because the topcode has been gradually increased over time, even the unadjusted series more or less capture the correct long-run trends, though year-to-year variations are highly sensitive to changes in the value of the topcode.

1.5   Changes in Wage Inequality: A Summary

The last two sections have shown that there are a number of clear patterns of changes in wage inequality and in the wage structure that are highly robust to measurement issues. The only notable exception has to do with changes in within-group wage inequality for men in the 1970s, where the March and May-ORG supplements of the CPS yield substantially different answers. I will underplay this particular aspect of inequality changes in this section but discuss its implication for the interpretation of inequality changes in the next section.

The main results about "what we really know" about changes in wage inequality can be summarized as follows. Unless otherwise indicated, the results are based on inequality measures adjusted for composition effects:

1. Changes in broad-based measures of inequality, such as the variance, are concentrated in the 1980s. This holds for both the between- and within-group components of wage dispersion.
2. Trends in broad-based measures of inequality hide important

differences at different points of the distribution. At the top end of the distribution, inequality has grown steadily throughout the 1980s, 1990s, and 2000s. At the low end of the distribution, inequality only grew in the 1980s and remained constant or declined during the 1970s, 1990s, and 2000s.

3. The pattern of change in inequality across education groups is highly consistent with the broader changes at the top end and low end of the distribution. Mean wages of college graduates and postgraduates increased steadily relative to high school graduates, though the growth in more marked in the 1980s. Similarly, the within-group variance of these two highly educated groups has grown steadily over time. By contrast, the wage disadvantage of workers without a high school degree relative to those with a high school degree only increased (by a small amount) in the 1980s. The within-group variance for workers with a high school degree or less increased a bit in the 1980s but then remained stable or declined in the 1990s and 2000s.

4. Experience wage differentials go up and down over time but contribute little to long-run changes in wage inequality. These differentials increase more for women, but this likely reflects spurious changes linked to the changing relationship between actual and potential experience.

These main changes are also summarized in table 1.1, which shows the evolution in the different measures of wage inequality over the 1973 to 2006 period. While all the findings reported in the table can also be seen in the various figures, it is clear from the table that the growth in inequality is concentrated in the upper end of the distribution and in the 1980s.

1.6   Possible Explanations: Some Concluding Comments

Looking back at explanations for inequality growth suggested fifteen years ago, the new developments documented in the preceding pose a major challenge to the view based on a general increase in the demand for all dimensions of skill. A first puzzle is that if relative demand for skilled workers kept going up over the last fifteen years, how can one explain the decline in the returns to experience over this period or the stability of the skill premium between high school graduates and less-educated workers?

In the case of the return to experience, a possible answer is that relative supply, as opposed to relative demand, is the key factor behind secular changes in this wage differential. Just like the entry of baby boomers first depressed the wages of young workers in the 1970s and early 1980s (Welch 1979), as this cohort ages, the negative pressure of supply on wages is increasingly moving to the upper end of the experience distribution, which reduces the experience premium. While this hypothesis should be probed in more detail, it is also reasonable to expect that relative demand pressures are less important in the case of experience than education. Indeed, while it is

sensible to think that the computer and information technology revolution is "education-biased," it is far from clear that it is also "experience-biased." This suggests that the growth in both education and experience differentials in the early 1980s may simply be a coincidence linked to demand factors driving the growth in the education premium and supply considerations driving the growth in the experience premium.

This view that the 1980s was a "perfect storm" where different factors resulted in an expansion in inequality at different points of the distribution, as opposed to a ubiquitous increase in the demand for skill, can also help shed light on some other puzzles. For instance, the large decline in the real value of the minimum wage during the 1980s helps explain why low-end inequality increased sharply during this period but not in other periods. DiNardo, Fortin, and Lemieux (1996) and Lee (1999) find that most of the growth in the 50-10 gap in the 1980s was due to the minimum wage (all of the change in Lee 1999). After adjusting for this factor, the remaining changes in 1980s were, thus, concentrated at the top end of the distribution.

Once several explanations are allowed to affect inequality at different points of the wage distribution, it becomes simpler to think of possible explanations for the secular growth in top-end inequality, without requiring these explanations to also account for swings in inequality growth at the bottom end. Because the growth in inequality at the top end of the distribution has attracted a lot of attention in recent years, a number of candidate explanations are available in the literature. For instance, Autor, Katz, and Kearney (2006) use Autor, Levy, and Murnane's (2003) model of technological change to explain why the labor market became polarized in the 1990s and 2000s. The model is based on a distinction between skilled and routine tasks where computers are substitutes for the latter. In this model, the introduction of computer technologies depresses the middle of the distribution where workers perform skilled but routine tasks, which results in increasing inequality at the top end but decreasing inequality at the low end. Firpo, Fortin, and Lemieux (2007) show that continuing deunionization yields similar predictions. Another possible explanation for the growing inequality at the top end includes the growth in pay for performance (Lemieux, MacLeod, and Parent 2007). None of these three explanations can account very well, however, for why inequality at the low end increased in the 1980s and decreased later, which highlights the value of combining these explanations with changes in the minimum wage.

Another important finding presented here that does not sit well with explanations suggested fifteen years ago is the fact that within-group inequality does not play as an important role in inequality changes as was thought back then. Note, however, that the basic insight of Juhn, Murphy, and Pierce (1993) that changes in the between- and within-group components are driven by similar factors remains consistent with more recent developments. For example, the fact that both the relative wages and the within-group disper-

sion of highly educated workers grew over time suggests that these developments are closely linked. Lemieux (2006b) shows that this follows naturally in a model with heterogenous returns to skill where the demand for college education increases. The fact that trends in the overall within- and between-group inequality are similar over time also points out to similar factors explaining both phenomena.

The one finding of Juhn, Murphy, and Pierce (1993) that is much less clear now than fifteen years ago is related to changes in within-group inequality in the 1970s. Using March CPS data, Juhn, Murphy, and Pierce (1993) argue that the growth in the within-group inequality in the 1970s provides evidence that relative demand was already increasing in the 1970s. As shown earlier, this last conclusion did not turn out to be very robust. This being said, whether within-group inequality increased in the 1970s does not play a crucial role in understanding why inequality has been changing over time. For both theoretical and empirical reasons, it is more appropriate to try to understand what drives changes in returns to education instead of within-group inequality over time. On the empirical side, the evidence presented in this paper shows that the basic facts about returns to education are very robust to measurement issues, which is not the case for within-group inequality. On the theoretical side, the basic idea of a race between relative supply and demand can be tested in the case of education, while the relative supply of unobserved skills underlying within-group dispersion is a fairly nebulous concept.

In terms of potential for future research, arguably the most important fact documented in this chapter and in related work (Lemieux 2006b; Goldin and Katz 2007) is the dramatic importance of education in changes in wage inequality. Fifteen years ago, most observers would probably not have thought that education could play such an important role in inequality growth. After all, the R-square of wage equations in the CPS are typically in the .3 to .4 range, with only part of the explanation coming from education. Furthermore, Juhn, Murphy, and Pierce (1993) had shown that most of the growth in wage inequality in the March CPS was coming from the within-group component, which was not surprising given the low R-square of wage regressions.

As I have shown in this chapter, however, the dominant source of long-run growth in the between-group component is the growth in relative wages of college-educated workers, while the dominant source of growth in the within-group component is the increase in within-group inequality among the same workers. These two related facts also help explain why inequality has mostly increased in the top end of the distribution where these workers are concentrated, as opposed to the low end of the distribution. In retrospect, this is a fairly unexpected development that deserves further investigation.

Data Appendix A

May-ORG and March CPS Data
This appendix explains in more detail how the March and May-ORG CPS are processed to make the wage samples as comparable as possible. It closely follows the data appendix in Lemieux (2006a). Both the May-ORG and the March CPS can be used to compute hourly wage rates. The March Supplement of the CPS asks about total earnings during the previous year. An hourly wage rate can then be computed by dividing last year's earnings by total hours worked last year. The latter variable is computed by multiplying two other variables available in the March CPS, usual weekly hours of work last year and weeks worked last year.

I limit the analysis of wages in the March CPS to the period starting with the earnings year 1975 (March 1976 survey) because an hourly wage rate cannot be computed in earlier years. Another reason for starting with the wage data for 1975 is that the other wage measure available in the May-ORG CPS is only available from May 1973 on. Because one contribution of the chapter is to compare the two data sources, the gain of using a more precise and comparable measure of hourly wages from the March CPS clearly outweighs the cost of losing two years of data for 1973 and 1974.

There are important differences between the way wages are measured in the March and May-ORG CPS. First, while the March CPS asks about retrospective measures of wages and earnings (last year), the May-ORG supplement asks about wages at the time of the survey. Second, the May-ORG wage questions are only asked to wage and salary workers. To get comparable wage samples, I limit my analysis of the March data to wage and salary earnings. One problem is that when workers have both wage and salary and self-employment earnings, we do not know how many hours of work pertain to wage and salary jobs versus self-employment. To minimize the impact of these considerations, I limit my analysis to wage and salary workers with very limited self-employment earnings (less than 10 percent of wage and salary earnings).

Another difference is that the ORG supplement only asks questions about the worker's main job (at a point in time), while the March CPS includes earnings from all jobs, including second jobs for dual job holders. Fortunately, only a small fraction of workers (around 5 percent typically) hold more than one job at the same time. Furthermore, these secondary jobs represent an even smaller fraction of hours worked.

Finally, because the May-ORG CPS is a "point-in-time" survey, the probability that an individual's wage is collected depends on the number of weeks worked during a year. By contrast, a wage rate can be constructed from the March wage information irrespective of how many weeks (provided that it is not zero) are worked during the year. This means that the May-ORG wage

observations are implicitly weighted by the number of weeks worked, while the March wage observations are not.

One related issue is that several papers like DiNardo, Fortin, and Lemieux (1996) also weight the observations by weekly hours of work to get a wage distribution representative over the total number of hours worked in the economy. Weighting by weekly hours can also be viewed as a reasonable compromise between looking at full-time workers only (weight of one for full-time workers, zero for part-time workers) and looking at all workers as "equal" observations irrespective of the number of hours worked. Throughout the chapter, I thus weight the March CPS observations by annual hours of work and weight the May-ORG observations by weekly hours of work.

In both the March and ORG supplements of the CPS, the Census Bureau allocates a wage or earnings item for these workers using an "hot deck" procedure. The CPS also provides flags and related sources of information that can be used to identify workers with allocated wages in all years except in the January 1994 to August 1995 ORG supplements.[18] By contrast, in the May 1973 to 1978 CPS, wages were *not* allocated for workers who failed to answer wage and earnings questions. For the sake of consistency across data sources, all results presented in the chapter only rely on observations with nonallocated wages, unless otherwise indicated.

Wages and earnings measures are topcoded in both the March and May-ORG CPS. Topcoding is not much of an issue for workers paid by the hour in the May-ORG CPS. Throughout the sample period, the topcode remains constant at $99.99, and only a handful of workers have their wage censored at this value. By contrast, a substantial number of workers in the March CPS, and nonhourly workers in the May-ORG CPS, have topcoded wages. When translated on a weekly basis for full-year workers, the value of the topcode for annual wages in the March CPS tends to be comparable to the value of the topcode for weekly wages in the May-ORG CPS. For instance, in the first sample years (1975 to 1980), the weekly topcode in the May-ORG CPS is $999 compared to $962 for full-year workers in the March CPS (annual topcode of $50,000). Toward the end of the sample period (1998 to 2002), the weekly topcode in the ORG CPS is $2,884, which is identical to the implied weekly topcode for full-year workers in the March CPS (annual topcode of $150,000 divided by 52). As discussed in the chapter, I adjust for topcoding in both the May-ORG and the March CPS by multiplying topcoded wages by a factor 1.4.

Several further data adjustments are also performed before applying the 1.4 factor to topcoded wages. In the May-ORG CPS, the topcode on the edited weekly earnings variable for workers not paid by the hour goes from

---

18. Allocation flags are incorrect in the 1989 to 1993 ORG CPS and fail to identify most workers with missing wages. Fortunately, the Bureau of Labor Statistics (BLS) files report both edited (allocated) and unedited (unallocated) measures of wages and earnings. I use this alternative source of information to identify workers with allocated wages in these samples.

$999 in 1973 to 1988, to $1,923 in 1989 to 1997, and to $2,884 in 1998 to 2006. Between 1986 and 1988, however, it is possible to use the unedited weekly earnings variable that is topcoded at $1,999 instead of $999. Though the unedited variable is not computed for workers who fail to respond to the earnings question, this does not matter here because I focus on workers with unallocated wages and earnings. I thus use the unedited earnings variable for the 1986 to 1988 period.

Several adjustments also have to be performed before applying the 1.4 factor to the March CPS data. Until March 1989, wages and salaries were collected in a single variable pertaining to all jobs, with a topcode at $50,000 until 1981 (survey year), $75,000 from 1982 to 1984, and $99,999 from 1985 to 1988. Beginning in 1989, the March CPS started collecting wage and salary information separately for main jobs and other jobs, with topcodes at $99,999 for each of these two variables. The topcodes were later revised to $150,000 for the main job and $25,000 for other jobs in March 1996, and to $200,000 for the main job and $35,000 for other jobs in March 2003.

Prior to March 1996, the earnings variable of workers who are topcoded simply takes the value of the actual topcode. Starting in March 1996, however, the value of earnings for topcoded workers is replaced by the mean earnings among all topcoded workers. Mean earnings are separately computed for different demographic groups. To maintain consistency over time, I first construct a topcoded variable for total wage and salary earnings from March 1989 on. For 1989 to 1995, I simply keep the pre-1989 $99,999 topcode. Because both main job and other job earnings are separately topcoded at $99,999, I simply add these two earnings variables and topcode the sum at $99,999. After various experiments, I decided to use a topcode of $150,000 for total wage and salary earnings from 1996 on. Unfortunately, it is not possible to topcode total wage and salary earnings in a way that is completely consistent with the pre-1996 situation. The problem is with workers who earn less that $125,000 on their main job but have earnings from other jobs topcoded at $25,000. It is not possible to know whether total earnings of these workers are above or below $150,000. After some experiments, I decided to compute total earnings as the sum of main job earnings (censored at $150,000) and earnings on other jobs where I use the actual earnings provided in the CPS (where topcoded observations are imputed the actual mean earnings among topcoded workers).

These adjustments likely have little impact because, in the March 1996 to 2007 CPS, less than 1 percent of workers have main job earnings below $125,000 and are topcoded on their other jobs earnings. Finally, once total wage and salary earnings have been censored in a consistent fashion, I multiply the earnings of workers at this consistent topcode by the standard 1.4 factor. I also follow the existing literature by trimming very small and very large value of wages to remove potential outliers (less than $1 or more than

$100 in 1979 dollars) and limit the analysis to workers aged sixteen to sixty-four with positive potential experience (age-education-6).

One last point about the ORG CPS is that, starting in 1994, workers are first asked what is the earnings periodicity (hourly, weekly, biweekly, annual, etc.) that they prefer to use in reporting their earnings on their current job. But as before, all workers paid by the hour are asked for their hourly wage rate. Hourly rated workers are asked this question even if "hourly" is not their preferred periodicity in the first question. Workers not paid by the hour are then asked to report their earnings for the periodicity of their choice. An hourly wage rate can again be computed by dividing earnings by usual hours of work over the relevant period. In 1994, the CPS also introduced "variables hours" as a possible answer for usual hours of work. I impute hours of work for these workers using a procedure suggested by Anne Polivka of the BLS.
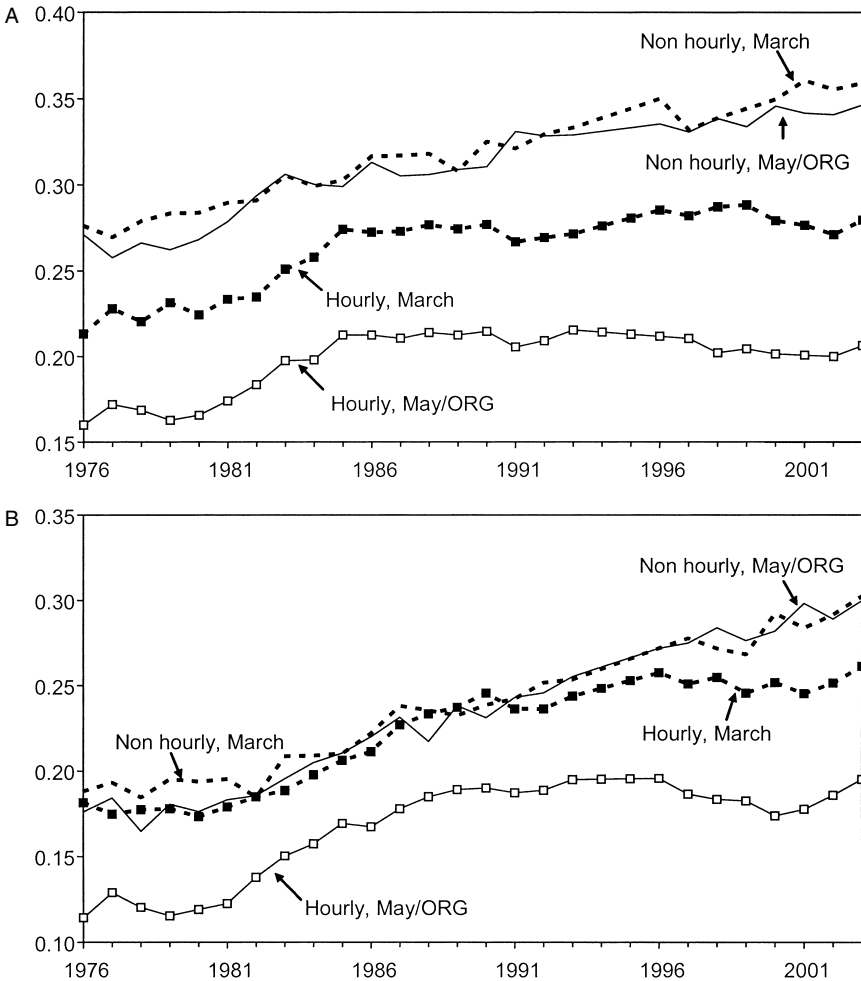
Appendix B

Fig. 1B.1 **Variance of log hourly wages with both May-ORG and March wages (matched sample):** *A,* **Men;** *B,* **Women**

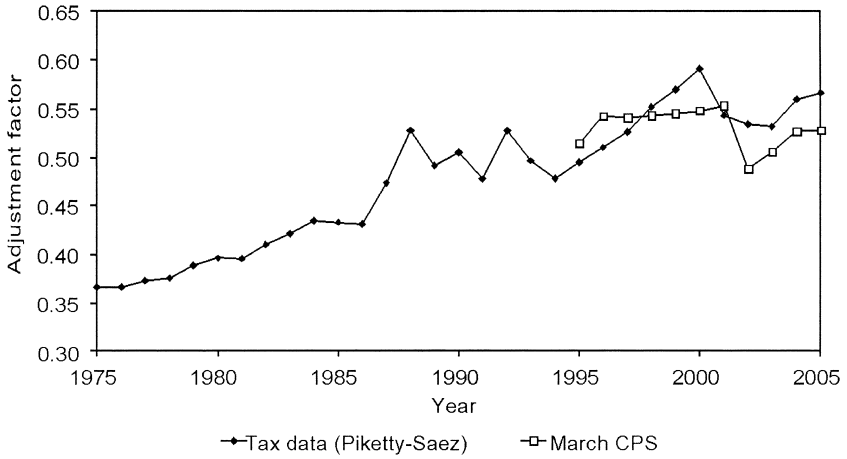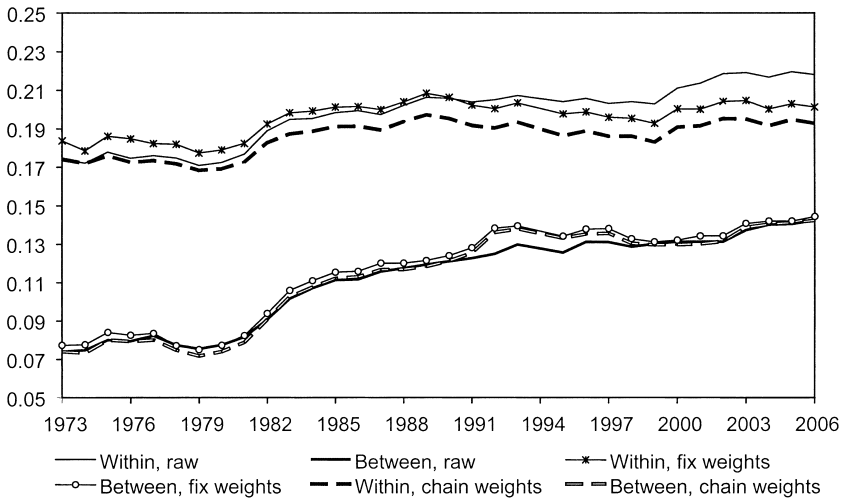**Fig. 1B.2 Topcoding adjustment factor (1/α) log annual earnings**

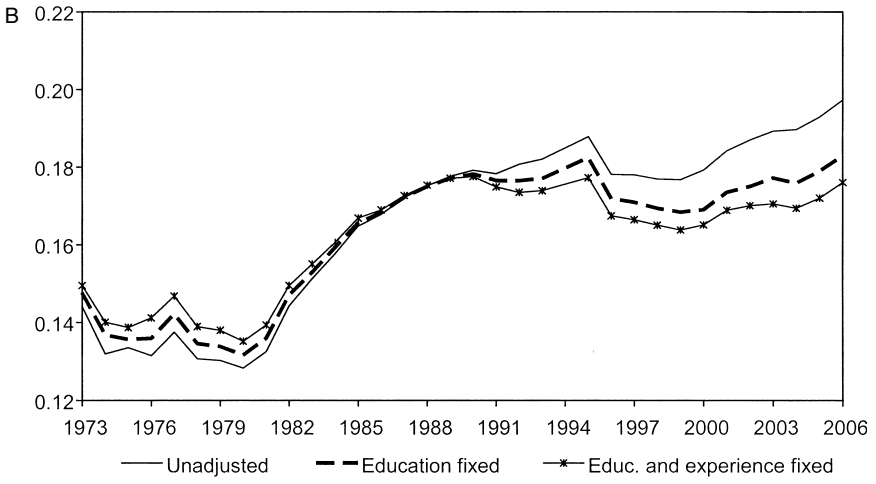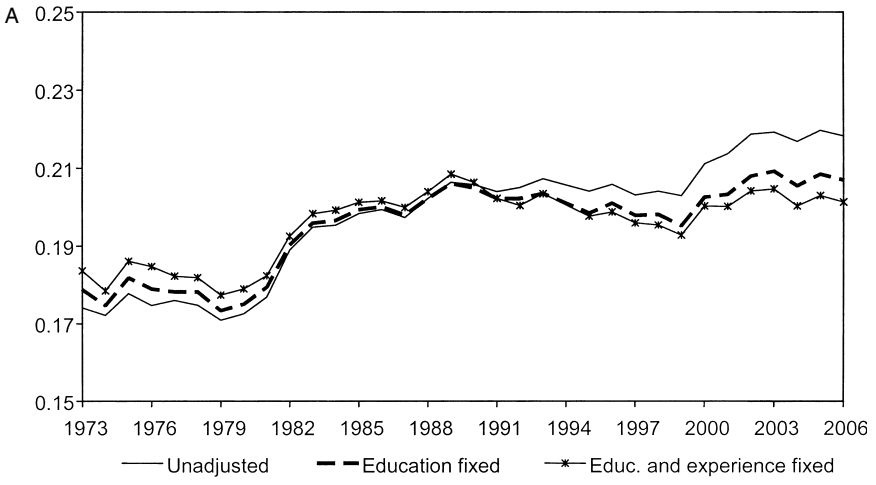**Fig. 1B.3 Comparing chain-weighted and fix-weighted composition-adjusted variances, Men**

Fig. 1B.4   Role of education and experience in composition effects in the within-group variance: *A,* **Men;** *B,* **Women**

References

Abraham, K. G., J. R. Speltzer, and J. C. Steward. 1998. Divergent trends in alternative wage series. In *Labor statistics measurement issues,* ed. J. Haltiwanger, M. E. Manser, and R. Topel, 293–324. Studies in Income and Wealth, vol. 60. Chicago: University of Chicago Press.

Acemoglu, D. 2002. Technical change, inequality, and the labor market. *Journal of Economic Literature* 40 (1): 7–72.

Autor, D. H., L. F. Katz, and M. S. Kearney. 2005. Rising wage inequality: The role of composition and prices. NBER Working Paper no. 11628. Cambridge, MA: National Bureau of Economic Research.

———. 2006. The polarization of the U.S. labor market. *American Economic Review: Papers and Proceedings* 96 (2): 189–94.

———. 2008. Trends in U.S. wage inequality: Re-assessing the revisionists. *Review of Economics and Statistics* 90 (2): 300–323.

Autor, D. H., L. F. Katz, and A. B. Krueger. 1998. Computing inequality: Have computers changed the labor market? *Quarterly Journal of Economics* 113 (4): 1169–1213.

Autor, D. H., F. Levy, and R. J. Murnane. 2003. The skill content of recent technological change: An empirical investigation. *Quarterly Journal of Economics* 118 (4): 1279–1333.

Berman, E., J. Bound, and Z. Griliches. 1994. Changes in the demand for skilled labor within U.S. manufacturing: Evidence from the Annual Survey of Manufactures. *Quarterly Journal of Economics* 109 (2): 367–97.

Blau, F. D., and L. M. Kahn. 1996. Swimming upstream: Trends in the gender wage differential in the 1980s. *Journal of Labor Economics* 14 (4): 1–42.

Bound, J., and G. Johnson. 1992. Changes in the structure of wages in the 1980s: An evaluation of alternative explanations. *American Economic Review* 82 (3): 371–92.

Card, D., and J. DiNardo. 2002. Skill biased technological change and rising wage inequality: Some problems and puzzles. *Journal of Labor Economics* 20 (4): 733–83.

DiNardo, J., N. M. Fortin, and T. Lemieux. 1996. Labor market institutions and the distribution of wages, 1973–1992: A semiparametric approach. *Econometrica* 64 (5): 1001–46.

Firpo, S., N. M. Fortin, and T. Lemieux. 2007. Decomposing wage distributions using influence function regressions. University of British Columbia. Mimeograph.

Goldin, C., and L. F. Katz. 2007. Long-run changes in the U.S. wage structure: Narrowing, widening, polarizing. *Brookings Papers on Economic Activity,* Issue no. 2: 135–65. Washington, DC: Brookings Institution.

Hamermesh, D. S. 2002. 12 million salaried workers are missing. *Industrial and Labor Relations Review* 55 (4): 649–66.

Hirsch, B. T., and E. Schumacher. 2004. Match bias in wage gap estimates due to earnings imputation. *Journal of Labor Economics* 22 (4): 689–722.

Juhn, C., K. M. Murphy, and B. Pierce. 1993. Wage inequality and the rise in returns to skill. *Journal of Political Economy* 101 (3): 410–42.

Katz, L. F., and D. H. Autor. 1999. Changes in the wage structure and earnings inequality. In *Handbook of labor economics.* Vol. 3A, ed. O. Ashenfelter and D. Card, 1463–1555. Amsterdam: Elsevier Science.

Katz, L. F., and K. M. Murphy. 1992. Changes in relative wages, 1963–1987: Supply and demand factors. *Quarterly Journal of Economics* 107 (1): 35–78.

Krueger, A. B. 1993. How computers have changed the wage structure: Evidence from microdata, 1984–1989. *Quarterly Journal of Economics* 108 (1): 33–60.

Lee, D. S. 1999. Wage inequality in the United States during the 1980s: Rising dispersion or falling minimum wage. *Quarterly Journal of Economics* 114 (3): 977–1023.

Lemieux, T. 2002. Decomposing wage distributions: A unified approach. *Canadian Journal of Economics* 35 (4): 646–88.

———. 2006a. Increasing residual wage inequality: Composition effects, noisy data, or rising demand for skill? *American Economic Review* 96 (3): 461–98.

———. 2006b. Post-secondary education and increasing wage inequality. *American Economic Review: Papers and Proceedings* 96 (2): 195–99.

Lemieux, T., W. B. MacLeod, and D. Parent. 2007. Performance pay and wage inequality. *Quarterly Journal of Economics* 124 (1): 1–49.

Levy, F., and R. J. Murnane. 1992. U.S. earnings levels and earnings inequality: A review of recent trends and proposed explanations. *Journal of Economic Literature* 30 (3): 1333–81.

Lillard, L., J. P. Smith, and F. Welch. 1986. What do we really know about wages? The importance of nonreporting and census imputation. *Journal of Political Economy* 94 (3, Part 1): 489–506.

Piketty, T., and E. Saez. 2003. Income inequality in the United States, 1913–98. *Quarterly Journal of Economics* 118 (1): 1–39.

Welch, F. 1979. Effects of cohort size on earnings: The baby boom babies' financial bust. *Journal of Political Economy* 87 (5): S65–S98.

Comment   Lawrence F. Katz

Thomas Lemieux has produced a terrific chapter documenting the basic facts about changes in the U.S. hourly wage structure over the past three decades using the May-outgoing rotation group (ORG) and March Current Population Survey (CPS) data. He also does a very nice job of examining how conclusions about wage inequality trends are affected by different choices for handling crucial measurement issues related to the topcoding of high earnings, the treatment of imputed (allocated) earnings observations, adjustments for changes in the education and experience composition of the workforce, and whether to use direct point-in-time wage measures from the May-ORG CPS versus wage measures based on past year annual earnings, weeks worked, and usual hours worked from the March CPS.

Lemieux carefully documents large increases in overall hourly wage inequality (as measured by the variance or 90-10 log wage gap) for both men and women since 1980. He shows that overall wage inequality grew most rapidly in the 1980s when both upper-end (90-50) and lower-end (50-10) wage inequality increased. But lower-end wage inequality stopped increasing (and even decreased after adjusting for education-experience composition

Lawrence F. Katz is the Elisabeth Allison Professor of Economics at Harvard University, and a research associate of the National Bureau of Economic Research.