Composition Effects, Wage Measurement, and the Growth in Within-Group Wage Inequality*

Thomas Lemieux
University of British Columbia

June 2003

Abstract

The "standard view" in the literature on wage inequality is that within-group, or residual, wage inequality started growing in the 1970s and accounts for most of the growth in wage inequality over the last two or three decades. This paper first shows that this conclusion is very sensitive to the choice of data used to measure hourly wages (March vs. May/ORG CPS). I use various pieces of evidence to argue that the May/ORG provides a more reliable measure of within-group inequality because it measures directly the hourly wage of workers paid by the hour. The paper also shows that a large fraction of the 1973-2002 growth in residual wage inequality is a consequence of composition effects. As is well known, the workforce grew older and more educated over the last twenty years. Since within-group inequality is larger for older and more educated workers, these composition effects have led to a spurious increase in residual wage inequality. For both men and women, the bulk of the evidence suggests that *all* of the growth in within-group inequality occured during the 1980s. Also, after adjusting for composition effects, I conclude that residual wage inequality accounts for *at most one quarter* of the total growth in wage inequality between 1973 and 2002.

*: I would like to thank NICHD (grant no. R01 HD39921-01) and SSHRC (grant no. 410-2002-0298) for financial support.

1. Introduction

The growth in wage inequality in the United States over the last three decades is one of the most extensively researched topics in labor economics. An important finding first documented by Juhn, Murphy and Pierce (1993) is that residual, or within-group, inequality accounts for most of the growth in wage inequality. In other words, dispersion in the residuals from a standard Mincer wage regression model appears to have grown more than the systematic component of wages predicted by the model. This is perhaps not surprising since standard regressors such as experience and education account for a relatively small fraction the variance of the wages (R-square is typically in the .2-.3 range). More recent survey pieces by Acemoglu (2002) and Katz and Autor (1999) confirm that residual inequality still account for most of the growth in wage inequality in more recent data from the late 1980s and the 1990s.

Another "stylized fact" about residual wage inequality is that it has been increasing steadily since the 1970s (Juhn, Murphy and Pierce, 1993, Katz and Autor, 1999, and Acemoglu, 2002). By contrast, the college-high school wage premium declined in the 1970s before increasing sharply in the 1980s (Bound and Johnson, 1992, Katz and Murphy, 1992). Juhn, Murphy and Pierce argue that the growth in residual wage inequality and the college-high school premium are two consequences of the same underlying increase in the demand for skills that started in the early 1970s. In the case of the college-high school premium the impact of growing demand for skills, was masked, however, by the steep growth in the relative supply of college workers associated with the entry of the baby-boom generation in the labor market during the 1970s.

There is little debate that residual wage inequality grew over the last two or three decades. However, the magnitude and timing of the growth in residual wage inequality is the subject of some controversy. Like Juhn, Murphy and Pierce (1993), DiNardo, Fortin and Lemieux (1996) document a steep growth in residual wage inequality during the 1980s. Unlike Juhn, Murphy and Pierce, however, DiNardo, Fortin and Lemieux find that within-group inequality was stable in the 1970s. Similarly, Acemoglu (2002) and Katz and Autor (1999) find substantial growth in residual wage inequality during the 1990s while Card and DiNardo (2003) and Lemieux (2002) find that residual wage inequality was stable during this period.

These discrepancies aside, Juhn, Murphy and Pierce's conclusion that residual wage inequality start growing in the 1970s and accounts for most of the growth in overall wage inequality remains the "standard view" about residual inequality. In particular, a substantial literature has used this "standard view" as a building block for models of economic growth and technical change (e.g. Aghion, 2001, and Acemoglu, 2002). The general goal of this paper is to assess how robust this "standard view" is to a variety of measurement issues.

For example, one possible explanation for some of the discrepancies among empirical studies is that they do not all rely on the same wage data. In particular, Juhn, Murphy and Pierce (1993) and most other studies construct wage measures from the Annual Demographic Supplement of the March Current Population Survey (CPS). By contrast, DiNardo, Fortin and Lemieux (1996) use wage measures from the May (from 1973 to 1978) and the Outgoing Rotation Group (ORG, from 1979 on) Supplements of the CPS. Both Katz and Autor (1999) and Card and DiNardo (2003) have systematically compared the trends in wage inequality obtained using these two alternative data sources. Despite this, there is still no consensus on how the timing and extent of the growth in residual wage inequality depends on the wage measure used.

The first specific goal of the paper is to re-examine how both the level and trends in residual wage inequality compare in the March and May/ORG CPS. Relative to previous studies, I focus on three specific aspects of the comparison. First, I compare these two alternative measures of inequality for workers in the outgoing rotation group of the March CPS who simultaneously report information about their wages and earnings in both the March and ORG supplements. Second, I contrast trends and levels in wage inequality for workers paid and not paid by the hour. Finally, I exploit the fact that since 1994, the CPS asked workers about the periodicity of earnings (hourly, weekly, annually, etc.) that they prefer to use when report their earnings. On the basis of these various evidences, I conclude that wages as measured in the May/ORG CPS provide a more reliable measure of residual wage inequality.

The second specific goal of the paper is to assess the role of composition effects in the growth in within-group inequality over the last two or three decades. The overall distribution of wage residuals is a mixture of residuals for different skill groups weighted by the proportion of individual in each group. As the workforce becomes older and more educated (as in the last twenty years), increasingly more weight is put on the residuals of the older and more educated groups. Since residual wage dispersion generally increases in both age and education, these composition effects tend to increase overall residual wage inequality. In other words, residual wage inequality may be increasing over time because of composition effects even if wage dispersion remains constant *within* each skill group. Perhaps surprisingly, I find that a large fraction of the growth in residual wage inequality (as measured using the May/ORG CPS) is indeed a spurious consequence of composition effects.

A third specific goal of the paper is to systematically compare the trends in residual wage inequality for men and women. The comparison between men and women is particularly important in the 1970s since the existing evidence (DiNardo, Fortin and Lemieux, 1996, or Katz and Autor, 1999) shows that residual wage inequality did not increase for women during this period.

The main result of the paper is that the "standard view" about residual wage inequality is very sensitive to the source of wage data used, to composition effects, and to the examination of trends in residual wage inequality for women. On balance, I conclude that within-group wage inequality plays a relatively modest role in the overall growth in wage inequality. I also conclude that, for both men and women, all of the growth in within-group inequality is concentrated in the 1980s.

The paper is organized as follows. In Section 2, I present and contrast the wage measures obtained from the March and May/ORG Supplements of the CPS. I then examine in detail the trends in residual wage inequality from the two data sources for the 1975-2001 period. Section 4 examines the role of composition effects in the growth in residual wage inequality. I conclude in Section 5.

2. March vs. May/ORG Supplements of the CPS

a. Data processing

Following most of the literature, the key wage measure on which I focus in this paper 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 have 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.

There are currently two sets of question in the CPS that 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.

For historical reasons, however, many studies based on March CPS data proxy for hourly wage rates by focusing only on the earnings of full-time (and sometimes full-year) workers. The reason is that prior to 1976, the March CPS only asked about full-time/part-time status last year (instead of usual hours of work last year). Furthermore, the information about weeks worked last year was limited to few intervals (0, 1-13, 14-26, 27-39, 40-47, 48-49, 50-52) in the pre-1976 March CPS. One important drawback of this alternative wage measure, however, is that it is limited to the subset of the workforce that works full-time (and sometimes full-year). It also fails to control for the dispersion in hours of work among workers who work full-time (35 hours and more a week).

Since we now have almost 30 years of data for which hourly wages rates can be directly computed for all workers, I limit the analysis of wages in the March CPS to the period starting with the earnings year 1975 (March 1976 survey). 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 starting in May 1973. Since one key contribution of the paper 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.[1]

[1] Another problem discussed later is that since missing wages were not allocated in the May 1973-78 CPS, allocated wages and earning should be excluded from the March CPS for the sake of comparability. Unfortunately, individual earnings allocation flags are not available in the March CPS prior to the 1976 survey (Lillard, Smith, and Welch, 1986). Though family earnings allocation flags can be used instead (Juhn, Murphy, and Pierce, 1993), this is one more reason for focusing on the March CPS data starting with the earnings year 1975.

The second measure of wages was first collected in the May 1973 Dual Job Holders Supplement of the CPS. The same information was collected each in May CPS until 1978. Starting in January 1979, the regular monthly CPS starting asking the same set of wage questions to all workers in the outgoing rotation group. The merged outgoing rotation group (MORG) files combine this information for all 12 months of the year. One important advantage of the MORG supplement is that it roughly three times as large as the May of March supplements of the CPS.[2]

There are also important differences between the way wages are measured in the March CPS and in the May/ORG supplements of the CPS. 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. In the May 1973-78 and ORG 1979-93 supplements, workers are first asked whether there are paid by the hour. Workers paid by the hour are then asked about their hourly rate of pay. Workers not paid by the hour are asked about their weekly earnings. For these workers, an hourly wage rate can then be computed by dividing weekly earnings by usual hours of work (which is also collected in the survey).

Starting with the 1994 CPS, workers are first asked what is the earnings periodicity (hourly, weekly, bi-weekly, annual, etc.) that they prefer to use to report their earnings on their current job. But once again, all workers paid by the hour are asked for their hourly wage rate. Hourly rated workers are asked this question even is "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.[3]

Few other differences between the two wage measures are also worth mentioning. First, the May/ORG wage questions are only asked to wage and salary workers. By

---

[2] The May 1973-78 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 4 and 8) are asked the questions from the ORG supplement each month. But combining the 12 months of data into a single MORG file yields wage data for 24 rotation groups compared to 8 in the March or May supplements. Note that the size the March Annual Demographic Supplement was substantially increased in the survey year 2001 to get more precise estimates of children health insurance coverage by states. As a consequence, the March 2001 and 2002 files are almost half as large (instead of a third as large) as the MORG files for these years.

[3] 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.

contrast, the March CPS asks separate questions about wage and salary earnings and self-employment earnings. To get comparable wage samples, I limit my analysis of the March data to wage and salary earnings. One problem is that when workers both have wage and salary and self-employment earnings, we do not know how many hours of work pertain to wage and salary jobs vs. 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 ten 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, since 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 1 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 paper, 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, a growing fraction of workers refuse to answer questions about wages and earnings. The Census Bureau allocates a wage or earnings item for these workers using the famous "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 supplement.[4] By contrast, in the May 1973-78 CPS, wages were *not* allocated for workers who failed to answer wage and earnings questions.[5] For the sake of consistency across data sources, all results presented in the paper only rely on observations with non-allocated 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 non-hourly 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). In the last sample years (1998 to 2001), the weekly topcode in the ORG CPS is $2884, which is identical to the implied weekly topcode for full-year workers in the March CPS (annual topcode of $150,000 divided by 52). Following most of the literature, I adjust for topcoding by multiplying topcoded wages by a factor 1.4.

In Appendix A, I discuss in detail how the data are processed to handle topcoding in a consistent fashion over time. One particular problem is that 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

[4] Allocation flags are incorrect in the 1989-93 ORG CPS and fail to identify most workers with missing wages. Fortunately, the 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.

[5] There has been some confusion in the literature because of the lack of good documentation on the allocation of missing wages in the 1973-78 CPS. Several papers assume that, like in the March CPS prior to 1976, wages were allocated but not flagged in the May 1973-78 CPS. For example, Katz and Autor (1999) compare a sample without allocated wages in 1973 to a sample with allocated wages in 1979. This likely overstates the growth in residual wage inequality during the 1970s since residual wage dispersion is generally higher when allocated wages are included than when they are not (see Figure 6). See Hirsch and Schumacher (2003) for a detailed discussion of how wages are allocated (or not allocated) in the May/ORG CPS.

variables. The topcodes were later revised to $150,000 for the main job and $25,000 for other jobs in March 1996. I explain in Appendix A how I "re-topcode" total wage and salary earnings at $99,999 in the March 1989 to March 1995 surveys, and at $150,000 from March 1996 on. I also compare trends in the 90-10 wage gap for the March and May/ORG CPS in Appendix A. The advantage of this alternative measure of wage inequality is that it is less sensitive than the standard deviation to topcoding.

Finally, I also follow the existing literature by trimming very small and very large value of wages to remove potential outliers. Following Card and DiNardo (2003), I remove observations with an hourly wage of less than $1 or more than $100 in 1979 dollars. I also limit the analysis to workers age 16 to 64 with positive potential experience (age-education-6).

b. Basic trends in the March and May/ORG wage data

As an initial check on the quality of the data, I compute average (log) hourly wages (deflated by the CPI-U) for the two data sources over the 1975-2001 period. Figure 1 shows the evolution of mean wages for both men and women over this period. Consistent with Abraham, Spletzer and Steward (1998), the figure shows that hourly wages are systematically larger in the March than in the May/ORG CPS. This is particularly striking in the case of men where the difference ranges from 5 to 10 percent. On the other hand, trends in the two wage series are very similar. Both data series show a steep decline in male real wages during the 1981-83 recession, a slower decline in the early 1990s, and a clear recovery in the late 1990s. The trends in the two wage series are also similar for women.

Figures 2a and 2b show the evolution of the standard deviation of log wages for men and women, respectively. A number of clear patterns emerge from these figures. Consistent with the literature, both wage series for both genders show that wage dispersion is clearly growing over the 1975-2001 period. A second clear pattern is that wage dispersion is substantially higher when hourly wages are computed using the March CPS instead of the May/ORG CPS.

A closer examination of the figures also suggests few other noticeable differences between the two wage series. Consistent with Juhn, Murphy, and Pierce (1993), wage dispersion is growing for men in the March CPS during the 1970s. But consistent with DiNardo, Lemieux, and Fortin (1996), wage dispersion is stable or declining (for women) when the May/ORG wage measure is used instead. More generally, wage dispersion tends to grow more over the whole 1975-2001 period in March CPS than in the May/ORG CPS. So while the overall trends from the two wage series are generally similar, features such as the timing of changes in wage dispersion that appear to be sensitive to the choice of data. This raises the obvious question of which of the two wage series provides the most reliable measure of wage dispersion over the last two or three decades.

c. *Which data series is more "reliable"?*

From the above discussion, it is clear that wages computed using the March and May/ORG CPS could differ for a variety of reasons including the treatment of self-employment earnings, topcoding, etc. Instead of looking systematically at all possible sources of differences between the two data sources, I focus on the fact that earnings are collected on a yearly basis in the March CPS, while workers can report their earnings at different periodicities in the May/ORG CPS. In the absence of measurement error in hours of work and earnings, the periodicity used to report earnings should have no impact on the measured hourly wage rate. Several validation studies clearly show, however, that there is substantial measurement error in the earnings reported in the CPS or similar surveys.[6]

The periodicity at which earnings are reported will matter if individuals can provide more accurate reports at some periodicity than others. For instance, a minimum wage worker will likely know and correctly report the exact value of the hourly wage at which he or she is paid. The same workers may experience more difficulties, however, reporting his or her annual earnings. In fact, the U.S. Census Bureau and other national statistical offices often mention the case of the minimum wage as one reason for asking directly workers paid by the hour about their hourly wage rate. By contrast, many professionals (including professors) tend to have their earnings set on an annual basis.

[6] Mellow and Sider (1983) compare employee and employer responses in the January 1977 Validation Study of the CPS. Bound and Krueger (1991) compare employee responses from the March 1977 and 1978 CPS to employer reported Social Security Earnings.

They may thus provide more accurate reports at the annual than hourly level. If most workers can provide the most accurate earnings reports at the hourly level, then the May/ORG CPS should yield more precise measures of hourly wages than the March CPS, and vice versa.

It is thus useful to know the periodicity at which workers feel most comfortable reporting their earnings. Since 1994 the ORG Supplement of the CPS has been asking workers this very question in an effort to improve the quality of earnings report for workers not paid by the hour. Interestingly, workers paid by the hour are asked this question though they are ultimately asked about their wage rate on an hourly basis.

Table 1 shows the frequency distribution (in percentage) of the different periodicities of earnings in the 1995 ORG CPS. For all workers pooled together (first column), 44.4 percent of workers prefer to report their wages by the hour compared to 21.6 percent who prefer to report their wages on a yearly basis. Workers who prefer to report at other periodicities are more or less equally split between reporting earnings on a weekly basis (17.8 percent) or other remaining periodicities (16.3 percent).

The figures for all workers in column 1 suggest that most workers prefer to report the earnings at the periodicity available in the May/ORG CPS (hourly, weekly, and more choices from 1994 on) than in the March CPS. The next two columns show the preferred periodicity for workers paid by the hour and not paid by the hour, respectively. Not surprisingly, most (72 percent) of the 62 percent of workers paid by the hour prefer to report their wages at an hourly rate. By contrast, only 7.7 percent of workers paid by the hour prefer to report their wages on a yearly basis. This strongly suggests that, for hourly workers, direct reports of hourly wages (as in the May/ORG CPS) are more reliable than the indirect measure of hourly wages computed using annual earnings from the March CPS.

The situation is not as clear for the minority (37.9 percent) of workers not paid by the hour. The proportion of these workers who prefer to report their wages on a yearly basis (43.6 percent) exceeds the proportion of workers who prefer to report wages on a weekly basis (26.8 percent). Recall that until 1993, workers not paid by the hour had to report their earnings on a weekly basis in the May/ORG CPS. For this period, the periodicity in the March CPS (yearly) is thus preferable to the one in the May/ORG (weekly) for a plurality of workers not paid by the hour. From 1994 on, however, workers can chose the periodicity they prefer in the ORG CPS, which is better than in the March CPS where they are forced to report their earnings on a yearly basis.

One clear message from Table 1 is that the measure of hourly wages available in the March CPS may be quite problematic for workers paid by the hour who overwhelmingly prefer to report their wage on an hourly basis. Overall, this problem may be quite serious since most workers are paid the hour. Figure 3 show that the fraction of workers in the May/ORG who report being paid by the hour ranges from 55 to 62 percent over the 1973-2002 period. Workers paid by the hour also tend to be at the lower end of the wage distribution, as shown in the last two columns of Table 1. These two columns report the average value of hourly wages (as measured in the 1995 ORG CPS) and the variance of log wages as a function of earnings periodicity. This suggests that the March CPS measure of hourly wages may be more accurate in the upper end than in the lower end of the wage distribution.

Though the "revealed preferences" of Table 1 are quite suggestive, they do not represent direct evidence that hourly wage rates from the March CPS are particularly inaccurate for workers paid by the hour. To look more directly at this issue, I exploit the fact that since 1979, workers in the outgoing rotation group of the CPS in March are asked the questions about wages and earnings from both the ORG and the Annual Demographic Supplements. The two measures of hourly wages can thus be computed for these workers who also report whether they were paid by the hour (or not) at the time of the survey. I use this subsample of workers to compare the standard deviation of the two wage measures for both hourly and non-hourly workers. I also limit the comparisons to workers with non-missing wages for both wage measures. This results in a sample of workers more "attached" to the labor force than in the more general samples used in the rest of the paper. Note also that the fact that a worker is paid by the hour at the time of the survey does not necessarily mean that he or she was paid by the hour during the previous year. A small fraction of workers classified as "paid by the hour" may thus not have been paid by the hour during the period (previous year) captured by the March CPS measure of hourly wages, and vice versa.

Figure 4a and 4b report the standard deviations of both wage measures for hourly and non-hourly workers, respectively. Given the smaller size of these "matched March-ORG" samples, I smooth the graph by reporting moving averages of the standard deviations (three years window). I also pool men and women together to keep reasonable sized samples. The reported moving averages only start in 1985 since the variables from the ORG supplements are only available in the public use files of the March CPS starting in 1984.[7]

The patterns illustrated in Figures 4a and 4b are quite striking. For workers paid by the hour (Figure 4a), the standard deviation of hourly wages as measured by the March CPS is much larger than the standard deviation of hourly wages as measured in the ORG CPS. The gap in the standard deviations actually grows from about 0.07 in the mid-1980s to about 0.010 by 2001. By contrast, the gap in the standard deviations for non-hourly workers is much smaller (0.02 to 0.03) and stable over time.

Consistent with the suggestive evidence in Table 1, the two wage measures seem to yield relatively similar measures of wage dispersion for workers not paid by the hour. For workers paid by the hour, however, hourly wages appear to be more noisily measured in the March than in the ORG CPS. This is consistent with the view that, for these workers, there is more measurement error in the wage measure from the March than the ORG CPS. In fact, the standard deviations of 0.45 (ORG) and 0.55 (March) in 2001 mean that the variance of March wages is about 50% higher than the variance of ORG wages. Under the assumption that ORG wages are measured without error, this suggests that the variance of measurement error in March wages is 50 percent of the true variance of wages. More disturbingly, the implied variance of measurement error rises from about 33 percent to about 50 percent between 1985 and 2001. By contrast, for non-hourly workers the variance of March wages only exceeds the variance of ORG wages by about 7 percent in both 1985 and 2001.

This set of observations suggests that both the level and the trend difference in the standard deviations of log wages illustrated in Figures 2a and 2b are driven by the fact that hourly wages for workers paid by the hour are less precisely measured in the March

[7] I plan to go back to 1979 in a future version of the paper by matching the monthly files (that contain the ORG variables) to the March files.

than in the May/ORG CPS. The level and trends in hourly wage dispersion from the March CPS should be interpreted with great caution because of this problem.

Note also that under the assumption of classical measurement error, the additional noise in the March CPS measure of wages (for hourly workers) should not affect estimates of the conditional means of wage (by education, age, etc).[8] In other words, this type of measurement error should have no effect on the between-group variance of wages (i.e. the dispersion in conditional means). If hourly wages from the March CPS are just a noisier measure of hourly wages from the May/ORG CPS (for hourly workers), then the two wage measures should yield similar between-group variances of wages. The measurement error should just increase the within-group, or residual, variance of wages. I test this hypothesis in the next section that reports estimates of both the between- and within-group variance of hourly wages.

3. Trends in Residual Wage Inequality

I decompose the variance of wages into a between- and within-group component by running standard Mincer-type human capital regressions.[9] More specifically, I estimate regressions of log wages on an unrestricted set of dummies for age, year of schooling, as well as in interactions between schooling dummies and a quartic in age.[10] One well-known problem with using schooling as a regressor in wage equations is that schooling is not measured in a consistent fashion in the CPS. Prior to 1992, the CPS asked about the highest grade attended, and whether the highest grade was completed. Starting in1992, however, the CPS switches to a question about the highest grade or diploma completed.

[8] The assumption is reasonable since both Mellow and Sider (1983) and Bound and Krueger (1991) find that measurement error in the CPS earnings in the late 1970s is uncorrelated with typical regressors like experience and education.

[9] It is common in the literature to report alternative measures of residual wage dispersion like the 90-10 gap. The drawback of this alternative measure, unlike the variance, is that the total 90-10 gap cannot be decomposed as the exact sum of the between- and within-group 90-10 gaps. I nonetheless show in Appendix C that trends in the 90-10 residual gap are very similar to trends in the residual variance.

[10] While it would be ideal to use an unrestricted set of age-education dummies in the wage regressions, in practice many age-education cells are quite small in the March and May supplements of the CPS. The flexible specification I use fits the data quite well. In the larger ORG samples, using a full set of age-education dummies only raises the R-square by about half a percentage point relative to the more flexible specification used in the paper. Note also that variables like race, marital status and other socio-economic variables are often used in standard wage regressions. I only use years of schooling and years of age (or potential experience) as regressors to focus on arguably "purer" measures of skills.

It is nonetheless possible to construct a relatively consistent variable for years of schooling completed over the whole sample period. The nine categories I use for years of schooling completed are 0-4, 5-8, 9, 10, 11, 12, 13-15, 16, and 17+.

The results of the decomposition are shown in Figures 5 and 6. Figure 5a shows the evolution of the between-group variance for men over the 1975-2001 period for the two measures (March and May/ORG) of hourly wages. In the case of hourly wages computed from the March CPS, I report the between-group variance with and without observations with allocated earnings. The figure shows that including observations with allocated earnings has essentially no impact on the between-group variance. This suggests that the mean of allocated wages by age and education categories are similar to the mean for observation with valid (non-missing) wages.

More importantly, the two wage measures yield very similar between-group variances of log wages. Both the levels and the trends in the two series are very similar. In particular, *all* the growth in the between-group variance is concentrated during the first half of the 1980s. The between-group variance is essentially constant between 1975 and 1980, and after 1985. This finding is very robust to the choice of hourly wage measure.

The results for women in Figure 5b are also robust to the choice of wage measure. The between-group variance obtained from the May/ORG and the March CPS (with and without allocators included) all show the same basic pattern. The between group variance declines in the 1970s, grows sharply in the first half of the 1980s, and grow more slowly thereafter. One natural explanation for the continuing growth in the between-group variance throughout the 1980s and 1990s is that age-earnings profiles are getting steeper during this period because of the increased attachment of women to the labor market.[11]

Since total wage dispersion is larger in the March CPS than in the May/ORG CPS (Figures 2 and 4) while the between-group dispersion is identical (Figure 5), within-group dispersion *must* be larger in the March than in the May/ORG CPS. Figures 6a and

---

[11] See Blau and Kahn (1997) and Fortin and Lemieux (1998). The continuing growth in the between-group variance during the 1980s and 1990s may thus be a spurious consequence of the fact that age (or potential experience) is a poor and changing proxy for underlying actual experience. Wage differences across age groups may thus be growing even if wage differences across groups based on actual experience remain constant.

6b show that this is indeed the case. In the case of men (Figure 6a), the within-group variance of March CPS wages (without allocated earnings) is systematically larger than the within-group variance of May/ORG wages. The gap between the two measures grows from about 0.02 in 1975 to about 0.07 in 2001. In percentage terms (relative to the May/ORG within-group variance), the gap increases from 10-15 percent in the mid 1970s to close to 30-40 percent in the early 2000s. Note also that, unlike the between-group variance, the within-group variance is sensitive to the inclusion of allocated wages. Figure 6a shows that keeping allocated wages in substantially increases the within-group variance of March CPS wages.

The large and growing gap between the within-group variances obtained using the two alternative measures of hourly wages has disturbing consequences for the trends in the within-group variance. When hourly wages are computed using the May/ORG CPS, the within-group variance is stable during the 1970s, then grows rapidly in the early 1980s and remains fairly constant from the mid-1980s to the late 1990s. In fact, the most significant increase in the within-group variance since 1983 happens between 1999 and 2001. It will be interesting to see whether this recent change persists over the next few years.

By contrast, the within-group variance grows steadily from 1975 to 2001 when hourly wages are computed using the March CPS. The steady growth in within-group dispersion over the 1970s and 1980s is consistent with Juhn, Murphy, and Pierce (1993)'s findings for full-time male workers. The continuing growth in the 1990s is consistent with the updated trends reported by Acemoglu (2002).

As in the case of men, the within-group variance for women is systematically larger in the March than in the May/ORG CPS. The gap in the within-group variance ranges from about 0.04 in the mid-1970s to about 0.06 in the early 2000s. Note that differences in the trends in within-group inequality are not as dramatic as for men for most of the sample period. Both wage series show that the within-group variance is relatively stable in the 1970s, but then grows dramatically during the 1980s. The main difference between the two wage series happens in the 1990s. While the within-group variance remains stable in the May/ORG CPS, it keeps growing in the March CPS (at a lower pace than during the 1980s).

The choice of the hourly wage measure has dramatic consequences for understanding the source of growth in within-group inequality and for interpreting the contribution of within-group inequality in the overall growth in wage inequality. For instance, Table 2 shows that in the March CPS, within-group inequality account for 60 percent of the overall growth in the variance of male wages during the 1975-2001 period (last column of panel B). By contrast, the between-group component accounts for most (57 percent) of the growth in wage inequality in the May/ORG data (panel A). As in Juhn, Murphy and Pierce (1993), the within-group component accounts for almost all the growth in male inequality in the 1970s (80 percent) in the March data. By contrast, the within-group component significantly contributes to the *decline* in male wage inequality when the May/ORG data are used instead.

Starting with Juhn, Murphy, and Pierce (1993), the steady growth in within-group inequality since the 1970s has been interpreted as evidence that the relative demand for skills started expanding in the 1970s. Juhn, Murphy, and Pierce argue that the full impact of these changes were somehow masked, however, by the dramatic increase in the relative supply of more educated during the 1970s that depressed the college-high school wage premium. Acemoglu (2002) formalizes this idea using a two-index model. For the sake of the argument, think of schooling or college labor as one skill index, and unobserved skills (school quality, innate cognitive ability, etc.) as the other skill index. Consider an increase in the relative demand for both college labor and unobserved skills due, for instance, to skill-biased technological change. As in Katz and Murphy (1992), the evolution in the return to schooling depends on whether relative demand or relative supply grows fastest. Katz and Murphy argue that the evolution of the college-high school wage gap is consistent with a steady increase in relative demand throughout the 1970s and 1980s. This underlying trend in relative demand is obscured, however, by the fact that relative supply grew much faster in the 1970s (resulting in a decline in the college-high school wage gap) than in the 1980s.

By contrast, under the assumption that the relative supply of unobserved skills is constant over time, within-group inequality should expand steadily over time. Unlike the college-high school premium, underlying trends in within-group inequality induced by increased demand for skill should not be obscured by swings in relative supply.

This prediction of the two-index model depends crucially on the level of substitutability in production between schooling and unobserved skills. The above result that within-group inequality is unrelated to changes in the relative supply of schooling only holds in a CES production function where the elasticity of substitution between all groups of workers (divided on the basis of both schooling and unobserved skills) is the same (Acemoglu, 2002). When unobserved skills and schooling are close substitutes for each other, an increase in the relative supply of college-educated worker should also reduce within-group inequality.[12] In the extreme case where unobserved skills and schooling are perfect substitutes, within-group inequality and the college-high school are two measures of the same wage gap between "skilled" and "un-skilled" workers and should move exactly together over time. This corresponds to the predictions of the single-index model of Acemoglu (2002).

I test these various hypotheses by running simple regressions of the within-group variance on the between-group variance and a time trend. In the single-index model where within- and between-group inequality move perfectly together, the trend should not be significant while the coefficient on the between-group variance should be positive and significant. By contrast, in the version of two-index model typically used in the literature (e.g. Acemoglu, 2002), the trend should be significant while the coefficient on the between-group variance should not be significant. The implicit identification assumption used is that swings in relative supply growth yields variation in the between-group inequality around a smooth trend that captures underlying relative demand changes.

The results reported in column 1 of Table 3 indicate that the single index model cannot be rejected for men when hourly wages from the May/ORG CPS are used. Neither a linear (panel A) nor a quadratic time trend (panel B) is statistically significant. By contrast, the between-group variance is strongly significant in both models. The regression results confirm the graphical evidence that the within and between-group

[12] The CES production function used by Acemoglu (2002) means that the elasticity of substitution between high-ability (high unobserved skills) college graduates and low-ability high school graduates is the same as the elasticity of substitution between high-ability college graduates and low-ability high school graduates. It seems more natural to posit that the latter elasticity of substitution is larger than the former. For example, Card and Lemieux (2001) reject a CES production function for all age-education groups in favor of a nested CES where the elasticity of substitution between college graduates of different age groups is higher than the elasticity of substitution between, say, old college graduates and young high school graduates.

variances follow very similar patterns over the 1975-2001 period (Figures 5a and 6a). They both grow sharply in the first half of the 1980s but remain otherwise stable.

Not surprisingly, the regression results are quite different when hourly wages from the March CPS are used instead (column 2). Both the linear and the quadratic trend terms are now strongly significant. Furthermore, the effect of the between-group variance is much weaker than when the May-ORG data is used. In the model with a quadratic trend, the effect of the between-group variance is very small and not significant. This is consistent with the "standard" two-index model where changes in the relative supply of college-educated labor have no impact on the within-group inequality.

Interestingly, the results for women are less sensitive to the choice of hourly wage rate measures. In particular, the effect of the between-group variance is consistently large and statistically significant. Furthermore, the linear trends (panel A) are not statistically significant under both measures of hourly wages. Looking at Panel A, the results for men from the March CPS stand as a clear outlier since all other models are consistent with the single-index assumption (no significant trend). The results with quadratic trends are more mixed since the quadratic trend terms are now significant for women. But once again, the lack of connection between the within- and between-group components of wage inequality is clearly a peculiarity of the March data for men.

In summary, the results in this Section reinforce those of Section 2 and confirm that hourly wage rates are more accurately measured in the May/ORG than in the March CPS. The "better" May/ORG wage data yield a remarkably simple story about the evolution of wage inequality over the last three decades. For both men and women, both between- and within-group inequality grew sharply in the 1980s but remained otherwise stable in the 1970s and 1990s. These patterns stand in sharp contrast with the "standard view" that within-group wage inequality grew steadily over the last three decades. The "standard view" is based on data for men in the March CPS. It does not hold for women in the March CPS, or in the "better" wage data from the May/ORG CPS.

4. Composition Effects and Residual Wage Inequality

As mentioned in the Introduction, changes in residual, or within-group, wage inequality are potentially sensitive to composition effects. This Section presents some evidence on the importance of composition effects and proposes a method to control for changes in the skill composition of the workforce.

a. Accounting for composition effects

The within-group variances reported in Figure 6 and Table 2 are computed over the set of regression residuals from each sample year. As discussed in Lemieux (2002), it is useful to rewrite the variance of residuals at time t, V_t, as

$$V_t = \Sigma_j \theta_t(j) V_t(j)$$

where $\theta_t(j)$ is the share of the workforce in skill group j, and $V_t(j)$ is the variance of wages within this skill group. Under the assumption that wage residuals are homoskedastic, the within-group variances are the same for all skill groups ($V_t(j) = V_t$ for all j) and the overall residual variance V_t does not depend on the skill composition of the workforce (the $\theta_t(j)$ shares).

It is well known, however, that wage residuals are strongly heteroskedastic. To this date, the most comprehensive study of wage dispersion across skill groups remains the landmark book by Mincer (1974). Consistent with the "overtaking model" of human capital investment, Mincer shows that the variance of wages first declines before increasing steadily as a function of labor market experience. Mincer also documents large differences in the variance of wages as a function of schooling, especially for older workers. Because of these systematic differences in wage dispersion across skill groups, there is significant scope for the skill composition of the workforce to affect the overall residual dispersion in wages. Indeed, Mincer shows that the variance of wages would have been much larger in 1959 if older workers had been as highly educated as younger workers. Since this is basically was happened in the U.S. labor market over the last 40 years, the results in Mincer (1974) suggest that composition effects may indeed be playing an important role in the evolution of residual wage inequality since the mid-1970s.[13]

[13] Card and Lemieux (2001a, 2001b) show that the level of educational attainment of young workers remains relatively constant over the 1975-95 period. As a result, young workers in the early 2000s are not much more educated than older workers. This stands in sharp contrast with the situation that prevailed in the 1959 census data analyzed by Mincer.

Composition effects can be estimated by contrasting the actual residual variance of wages, V_t, to the counterfactual variance, V_t^*, that would have prevailed if the skill group shares had remained constant at some level $\theta^*(j)$:

$$V_t^* = \Sigma_j \theta^*(j) V_t(j).$$

When the number of skill groups is small relative to sample sizes, the variance $V_t(j)$ can be computed for each skill group j, and it is straightforward to estimate the counterfactual variance. As discussed earlier (footnote 10), however, cell sizes based on single years of age and education become too small (and sometimes empty) when the March or May CPS is used. Following Lemieux (2002) and DiNardo, Fortin and Lemieux (1996), I address this problem by estimating a flexible logit model to re-weight the data in a way that keeps the distribution of skills constant over time. In this particular application, I use the same specification as in the wage model (full set of age and education dummies plus interactions between education dummies and a quartic in age) for the logits.

In what follows, I first illustrate the potential importance of composition effects by dividing workers into a limited number of skill groups based on five education groups (high school dropouts, high school graduates, some college, college degree, and post-graduate degree.) and four experience groups (1-10, 11-20, 21-30, and 31 years or more of potential experience). I then use the re-weighting approach to fully account for the role of changes in the distribution of age (experience) and education in changes in residual wage inequality. In light of the results in Section 2 and 3, I focus the analysis on the May/ORG CPS data. This enables me to expand the sample period from 1975-2001 to 1973-2002. Additional results from the March CPS are presented in Appendix B.

b. Results for broader education and experience groups

Table 4 shows the percentage distribution of workers over the different education and experience categories in 1973-74, 1980, 1990 and 2002.[14] As is well known, educational attainment of the workforce steadily grows over the whole sample period. For example, the percent of high school dropouts declines by two-thirds while the percentage of workers with a college degree or some college more than doubles between 1973-74 and 2002.

[14] The May 1973 and 1974 data are pooled together to increase the sample size.

Unlike the distribution of education, the distribution of years of experience does not change uniformly over the sample period. The workforce first becomes younger and less experienced as the largest cohort of the baby boom (those born in the late 1950s) enters the labor market between 1973-74 and 1980. The workforce then becomes older and more experienced as the baby boom cohorts make their way through the experience distribution in the 1980s and 1990s. As a result, by 2002 there are as many "mature" workers with 21 to 30 years of experience as there are "young" workers with 1 to 10 years of experience. By contrast, the fraction of young workers was more than twice as large as the fraction of mature workers in 1980.

Figure 7 plots the within-group variance of wages for each of the 20 experience-education groups for men and women in both 1973-74 and 2002. These variances are computed using the residuals from the wage regressions of Section 3 that implicitly control for finer measures of experience and education within each of the 20 experience-education groups. Consistent with Mincer (1974), the within-group variances tend to grow as a function of labor market experience.[15] The within-group variances also tend to grow as a function of years of schooling. This is most striking for men in 2002 (Figure 7b) where the within-group variance for college-educated men (with or without a post-graduate degree) is about twice as large as the within-group variance of men with a high school degree or less. Interestingly, there is more difference in the within-group variance across education groups in 2002 than in 1973-74. This suggests that the residual variance of wages (for all skill groups combined) is more sensitive to composition effects in educational achievement in 2002 than in 1973. Figures 7c and 7d show that the within-group variance has also become more sensitive to education (and experience) for women in 2002 compared to 1973-74.

Taken together, Table 4 and Figure 7 suggest that skill composition effects have contributed to the growth in the residual variance of wages between 1980 and 2002. Since older and more educated workers tend to have more dispersed wages (Figure 7), the fact that the workforce became older and more educated over this period should have contributed to the growth in residual inequality. The situation is more ambiguous

[15] In the May/ORG CPS, there is little indication that the residual variance first declines before increasing as a function of years of experience, as predicted by a standard overtaking model of investment in on-the-job training.

between 1973-74 and 1980, however, since the workforce became both more educated (higher within-group inequality) and younger (lower within-group inequality).

Figures 8 and 9 show more detailed trends in within-group inequality by education and experience groups, and provide more direct evidence on the role of composition effects. Figure 8a shows the evolution of the within-group variance for each of the five education groups of men. The within-group variances are simply computed using the residuals for men of all experiences groups in each sample year. In terms of the above notation, these within-group variances for each education group s are weighted averages over the four experience groups x:

$$V_t(s) = \Sigma_x [\theta_t(s,x) / \theta_t(s)]V_t(s,x) ,$$

where $\theta_t(s,x)$ is the share of workers with education s and experience x, $\theta_t(s)$ is the share of workers (of all experience groups) with education s, and $V_t(s,x)$ is the variance of residuals of workers with education s and experience x.

Consistent with the trends in within-group inequality for all workers (Figure 6a), Figure 8a shows that the within-group variance grows for all education groups during the first half of the 1980s. The growth in within-group inequality is particularly strong (about 0.05) for college graduates. After 1985, the within-group variance grows mildly for college graduates and post-graduates, but declines for high school dropouts.

Figure 8b shows what would have happened to the within-group variances if the distribution of experience had remained constant over time. I compute these counterfactual variances $V_t(s)^*$ under the assumption that workers are evenly distributed over the four experience groups:

$$V_t(s)^* = \Sigma_x .25\ V_t(s,x).$$

Compared to Figure 8a, the within-group variances are now substantially larger for college workers (some, graduates, or post-graduates) in the 1970s and early 1980s. This is consistent with young (inexperienced) workers, who have lower residual variances, being over-represented among college workers during those early periods. Had college workers been older, their within-group variance would have been substantially higher (as in Figure 8b). In fact, Figure 8b suggests that much of the growth in the within-group variance may just be an artifact of composition effects. Holding experience constant, the within-group variance only grows substantially for college graduates. For other groups, the within group variance in the early 2000s is either slightly above (high school graduates), about the same (college post-graduates or "some college"), or slightly below (high school dropouts), what it was in 1973.

Figures 9a and 9b repeat the same exercise for experience groups. As in the case of education, Figure 9a shows that the within-group variance increases for all experience groups in the first half of the 1980s. Over the whole sample period, however, the within-group variance increases much more for more experienced workers than for young workers with 1 to 10 years of experience. Composition effects (in education) are a natural explanation for this difference. Card and Lemieux (2001a, 2001b) show that the educational attainment of older workers grew steeply over this period, while young workers made little progress in terms of educational attainment. Since more educated workers have more dispersed wages (Figure 7), composition effects may explain the relative trends for younger and older workers in Figure 9a.

This conjecture is explored in Figure 9b that shows the within-group variances that would have prevailed if education had remained constant for each experience group over this period. These counterfactual variances are computed using the average education distribution for all workers in 2002:

$$V_t(x)^* = \Sigma_s \, \theta_{2002}(s) V_t(s,x).$$

Consistent with the conjecture, holding the distribution of education constant has little impact for younger workers but dramatically reduces the growth in inequality for older workers. For example, the within-group variance for the oldest workers in the late 1990s/ early 2000s is about the same as in the 1970s. By contrast, the within-group variance grew by 0.05 between these two time periods when changes in the distribution of education are not controlled for (Figure 9a).

The same trends are reported for women in Figures 8c and 8d (by education) and Figures 9c and 9d (by experience). The results are qualitatively similar to those for men. Consistent with Figure 6b, the within-group variance grows steadily during the 1980s for all experience and education groups. As in the case of men, however, holding either the distribution of experience (Figure 8d) or education (Figure 9d) constant reduces the growth in within-group wage inequality.

c. Re-weighting results

Figure 10 summarizes the impact of composition effects in the evolution of within-group inequality between 1973 and 2002. Figure 10a contrasts the actual within-group variance for men with the within-group variance that would have prevailed if the distribution of education and experience in the workforce had remained as in 1973. The counterfactual variances are computed by replacing the sample weights for each worker i, θ_{it}, by a counterfactual weight θ_{it}^*. The actual within-group variance is computed from the individual level data as

$$V_t = \Sigma_i \theta_{it} r_{it}^2,$$

where r_{it} is the estimated residual for worker i at time t, while the counterfactual variance is

$$V_t^* = \Sigma_i \theta_{it}^* r_{it}^2.$$

The counterfactual weights are such that the sample at time t weighted using the counterfactual weights θ_{it}^* is the same as in an appropriate base year (1973 in Figure 10a). The re-weighting factors are computed using the estimates from a logit model for the probability of being in year t relative to the base year (see Lemieux, 2002, for more detail). For example, the counterfactual weights for 2002 used in Figure 10a are computed by estimating a logit model on data for years 1973 and 2002 pooled together. The dependent variable is a dummy variable for year 2002, while the explanatory variables are the age and education variables (full set of indicators plus interactions between education and a quartic in age). The predicted probability that worker i is in year 2002, P_i, is then used to compute the counterfactual weight as

$$\theta_{it}^* = [(1- P_i)/ P_i]\, \theta_{it} .$$

Note that the re-weighting approach can also be used to compute other counterfactual statistics such as the 90-10 residual gap. Appendix Figures C1 and C2 shows that the basic trends in the 90-10 residual gap and the role of composition effects are very similar to those obtained using the within-group variances instead. I thus focus on the latter measure of residual wage dispersion for the remainder of this section.

Consistent with the evidence in Figures 8 and 9, Figure 10a shows that composition effects play a dramatic role in the growth in the within-group variance between 1973 and 2002. While the actual within-group variance grows by about 0.04

over the whole sample period, the counterfactual variance in the late 1990s / early 2000s is only about 0.01 higher than in the mid-1970s. Consistent with the discussion of Table 4 and Figure 7, composition effects play a steady role in the growth in within-group inequality as the workforce ages and becomes more educated from the early 1980s to the early 2000s. By contrast, composition effects play no role in the 1970s since the effect of growing educational achievement (more within-group inequality) is offset by the fact that the workforce is getting younger (less within-group inequality).

A closer examination of Figure 10a also shows evidence of a cyclical effect in the within-group variance. During the recessions of 1981-83, 1990-92, and 2000-2002, the actual variance grows faster that the counterfactual variance. This is consistent with less-skilled workers, who tend to have a lower within-group variance, being more adversely affected in terms of their employment during recessions. It is well known that composition effect tends to hide the pro-cyclicality of the level of real wages (Barsky et al, 1994). By analogy, Figure 10a suggests that composition effects tend to over-state the counter-cyclical pattern in within-group inequality over the business cycle (inequality grows during recessions).

Figure 10b repeats the same exercise by showing the counterfactual variance that would have prevailed if the distribution of characteristics had always been as in 2002. The results are qualitatively similar, though not as dramatic, as the results in Figure 10a. In particular, the counterfactual variance in Figure 10b remains relatively unchanged in the 1990s, while it steeply declines in Figure 10a. The simplest explanation for this divergence is that holding characteristics at their 1973 level puts relatively more weight on high-school dropouts who experience a clear decline in their within-group variance (Figure 8b). By contrast, holding characteristics at their 2002 level puts relatively more weight on college graduates who experience a clear increase in their within-group variance.

Once again, the results for women in Figures 10c and 10d are qualitatively similar to those for men. Composition effects can explain most of the growth in the within-group variance between 1973 and 2002 when characteristics are held at the 1973 level (Figure 10c). Composition effects also play a qualitatively similar, though less dramatic role, when characteristics are held at the 2002 level instead (Figure 10d).

The results for both men and women are summarized in Table 5. Note that, relative to Table 2 (for the 1975-2001 period), the change in the residual variance over the whole 1973-2002 period is now larger for men but smaller for women (last columns). The main reason for this discrepancy is that the residual variance for women falls sharply between 1973 and 1974. A natural explanation for this drop is that the minimum wage was increased by 25 percent (from \$1.60 to \$2.00) between these two years (DiNardo, Fortin and Lemieux, 1996).

Table 5 confirms that composition effects can account for virtually all the growth in the residual variance over the 1973-2002 period when the distribution of age and education are held at their 1973 levels. Once again, the results are less dramatic when the distribution of age and education is held at their 2002 levels instead. Even in this case, however, composition effects still account for half of the growth in the residual variance for men, and for a third of the growth in the residual variance for women. The remaining growth in the residual variance (0.024 for men, 0.029 for women) is only about half as large as the 0.05 growth in the between-group variance (Table 2). This strengthens the earlier conclusion that, in the May/ORG data, the most important factor in the growth of wage inequality is the between-group variance. Once composition effects are controlled for, the growth in the within-group variance is *at most* half as large as the growth in the between-group variance.

Looking by sub-period, none of the mild growth in residual variance in the 1990s remains when composition effects are controlled for. This results holds when characteristics are either held at their 1973 or 2002 level. Holding the skill composition constant thus strengthens the earlier finding that all the growth in the residual variance is concentrated during the 1980s. In fact, the growth in the residual variance during the 1980s (holding age and education constant) *always* exceeds the growth in the residual variance for the whole 1973-2002 period.

Though the 1973-2002 growth in the residual variance is more or less similar for men and women, there are some differences in the timing of the changes. In particular, the residual variance always declines more for women than for men in the 1970s, while the opposite happens during the 1980s. One natural explanation for this difference in timing is that female wage inequality is more sensitive to changes in the minimum wage than male wage inequality. Indeed, DiNardo, Lemieux, Fortin (1996) find that the increase in the real value of the minimum wage during the 1970s had a larger (negative) effect on female than male residual inequality during the 1970s. They also show that the decrease in the real value of the minimum wage during the 1980s had a larger (positive) effect on female than male residual inequality during the 1970s. DiNardo, Lemieux, Fortin (1996)'s analysis stops in 1992. It is reasonable to assume, however, that the minimum wage had little or no effect on residual inequality during the 1990-2002 period, since the real value of the minimum wage remained relative unchanged over those year. For most of 1990, the minimum wage was at $3.80, which translates to $5.20 in 2002 dollars, about the same as the actual minimum wage ($5.15) in 2002.

Table 6 compares the changes in residual wage inequality (holding characteristics at their 2002 level) to the minimum wage effects estimated by DiNardo, Fortin, and Lemieux (1996). The table shows that about half of the differential male-female growth in the residual variance in the 1970s and 1980s can be explained by the differential impact of the minimum wage.[16] Minimum wage effects also explain all of the small male-female difference in the growth of residual inequality over the whole 1973-2002 period.

5. Conclusion

The "standard view" in the literature on wage inequality is that within-group, or residual, wage inequality started growing in the 1970s and accounts for most of the growth in wage inequality over the last two or three decades. This paper first shows that this conclusion is very sensitive to choice of data (March vs. May/ORG CPS). I use various pieces of evidence to argue that the May/ORG provides a more reliable measure of within-group inequality because it measures directly the hourly rate of pay of workers paid by the hour. For both men and women, the May/ORG data show that residual wage inequality grew substantially in the 1980s but remained relatively stable in the 1970s and 1990s. These trends are very similar to those in the between-group inequality, which

[16] Lee (1999) finds that, if anything, DiNardo, Lemieux, and Fortin (1996) tend to understate the full impact of the minimum wage on the wage distribution because they ignore spillover effects. His results suggest that the minimum wage may in fact explain all of the male-female difference in the differential male-female growth in the residual variance in the 1970s and 1980s.

suggests that changes in both of these dimensions of wage inequality may be driven by the same factors. Also contrary to the “standard view”, between-group inequality accounts for most of the total growth in wage inequality for both men or women.

These findings are reinforced when the changes in residual wage inequality are adjusted for changes in the skill composition of the workforce. In the May/ORG CPS, a large fraction of the 1973-2002 growth in residual wage inequality is explained by the fact that the workforce grew older and more educated over the last twenty years. Since within-group inequality is larger for older and more educated workers, these composition effects have led to a spurious increase in residual wage inequality. In fact, there is very little growth in residual wage inequality when the distribution of age and education is held at its level of 1973. For men, composition effects also account for about half of the growth in residual wage inequality when the distribution of age and education is held at its 2002 level instead. These two bounds indicate that at least half of the growth in the residual variance of wages is a spurious consequence of composition effects. Since the residual variance accounts for less than a half to the growth in the total variance, this means that within-group inequality only accounts for a small fraction of the total growth in the variance of wages over the last three decades.

In summary, there is little support for the “standard view” on within-group inequality once composition effects are controlled for and the “better” hourly wage data from the May/ORG CPS are used. For men, *all* of the growth in within-group inequality happens during the first half of the 1980s. Furthermore, within-group inequality (adjusted for composition effects) accounts for *at most one quarter* of the total growth in wage inequality between 1973 and 2002.

These findings have important implications for understanding the sources of growth in wage inequality over the last three decades. While the demand for skills may have indeed increased steadily since the early 1970s, this factor is not very useful for explaining why both the between- and within-group inequality only increased in the 1980s, an in particular in the early 1980s. Explanations that focus on factors that changed sharply in the first half of the 1980s are more likely to account for the key trends in wage inequality in the United States.

References

Abraham, Katherine J., James R. Spletzer, and Jay C. Steward, "Divergent Trends in Alternative Wages Series, in J. Haltiwanger, M. Manser, and R. Topel (eds.) *Labor Statistics Measurement Issues*, Chicago: University of Chicago Press and NBER

Acemoglu, Daron, "Technical Change, Inequality, and the Labor Market," *Journal of Economic Literature* 40, March 2002, 7-72

Barsky, Robert, Parker, Jonathan a, and Gary Solon, "Measuring the Cyclicality of Real Wages: How Important Is Composition Bias?", *Quarterly Journal of Economics* 109(1), February 1994, 1-25

Bound, John, and George Johnson, "Changes in the Structure of Wages in the 1980s: An Evaluation of Alternative Explanations," *American Economic Review* 82, June 1992, 371-92

Bound, John, and Alan B. Krueger, "The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?" *Journal of Labor Economics 9*, January 1991, 1-24

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

Card, David, and Thomas Lemieux, "Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis," *Quarterly Journal of Economics 116*, May 2001a, 705-46

— , "Dropout and Enrollment Trends in the Post War Period: What Went Wrong in the 1970s?," in Jonathan Gruber (ed.) *An Economic Analysis of Risky Behavior Among Youth*, 2001b, Chicago: University of Chicago Press for NBER

DiNardo, John, Nicole M. Fortin, and Thomas Lemieux (1996) "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach," *Econometrica 65*, September 1996, 1001-46

Hirsch, Barry, and Edward Schumacher, "Match Bias in Wage Gap Estimates Due to Earnings Imputation," forthcoming, *Journal of Labor Economics*, 2003

Juhn, Chinhui, Kevin M. Murphy, and Brooks Pierce (1993) "Wage Inequality and the Rise in Returns to Skill," *Journal of Political Economy* 101, June 1993, 410-42

Katz, Lawrence, and David Autor, "Changes in the Wage Structure and Earnings Inequality," in O. Ashenfelter and D. Card (eds.) *Handbook of Labor Economics*, volume 3A, 1999, Amsterdam: Elsevier Science

Katz, Lawrence, and Kevin Murphy, "Changes in Relative Wages, 1963-1987: Supply and Demand Factors," *Quarterly Journal of Economics 107,* January 1992, 35-78

Lee, David, "Wage Inequality in the United States during the 1980s: Rising Dispersion or Falling Minimum Wage?," *Quarterly Journal of Economics* 114, August 1999, 977-1023

Lemieux, Thomas, "Decomposing Wage Distributions: a Unified Approach" *Canadian Journal of Economics 35*, November 2002, 646-88.

Lillard, Lee, James P. Smith, and Finis Welch, "We Really Know about Wages? The Importance of Nonreporting and Census Imputation" *Journal of Political Economy* 94, Part 1, June 1986; 94(3): 489-506.

Mellow, Wesley, and Hal Sider, "Accuracy of Response in Labor Market Surveys: Evidence and Implications" *Journal of Labor Economics 1*, October 1983, 331-44

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

APPENDIX A: Topcoding

Topcoding adjustments in the May/ORG and March CPS

As mentioned in Section 2, adjusting for topcoding is relatively straightforward in the May/ORG CPS. Since the topcode for the hourly wage of workers paid by the hour is quite high ($99.99), topcoding is not an issue for this group of workers. For workers not paid by the hour, the topcode on the edited variable for weekly earnings goes from $999 in 1973-88 to $1923 in 1989-1997 and $2884 in 1998-2002. Between 1986 and 1988, however, it is possible to use the unedited weekly earnings variable which is topcoded at $1999 instead of $999 for the edited variable. Though the unedited variable is not computed for workers who fail to respond to the earnings question, this does not matter here since I only use data for workers with unallocated wages and earnings. I thus use the unedited earnings variable for the 1986-88 period.

The situation is more complicated in the March CPS. As mentioned in Section 2, 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.

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. For example, in the March 2001 CPS, the mean for topcoded main job earnings ranges from $195,699 for white females not working full-time full-year, to $335,115 for full-time full-year white males. The corresponding means for these two groups are $39,320 and $56,879 for wage and salary earnings on other jobs.

To maintain consistency over time, I first construct a topcoded variable for total wage and salary earnings from March 1989 on. For 1989-1995, I simply keep the pre-1989 $99,999 topcode. Since 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 impossible 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 variable provided in the CPS.

For example, consider a full-time full-year white male who earns $90,000 on his main job but has his earnings topcoded at $25,000 for other jobs in the March 2001 CPS. I compute total earnings as the sum of $90,000 and $56,879 (see above), which yields $146,876. Since this is below the $150,000 topcode, I do not compute further adjustments for this worker. By contrast, I would censor at $150,000 the total earnings of the same worker if he earned $100,000 instead of $90,000 on his main job (total of $156,876).

Hopefully, these adjustments have little impact since, in the March 1996-2002 CPS, less than one 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 then multiply the earnings of workers at this consistent topcode by the standard 1.4 factor.

Alternative trends in wage dispersion

Appendix Figures A1 to A4 compare the standard deviation of the wage measure used throughout the paper (where topcoded earnings are multiplied by 1.4) to the 90-10 wage gap and the standard deviation of unadjusted (for topcoding) wages. Unadjusted wages are computed in the same way as usual except that the earnings of workers censored at the topcode are not multiplied by the 1.4 factor. I still process the March CPS data from 1989 on in the way described above to make sure that earnings are topcoded in a consistent fashion over time.

Contrasting the standard deviation of adjusted and unadjusted wages is perhaps the most direct way of assessing the impact of topcoding adjustments. I also report the 90-10 wage gap as an alternative measure of wage dispersion since topcoding adjustments typically only affect workers in the few top percentiles of the wage distribution. I also normalize the 90-10 wage gap by dividing it by the 90-10 wage gap from a standard normal distribution. Under the assumption that wages are distributed lognormal, this normalized 90-10 wage gap should be equal to the standard deviation. This makes it easier to compare the 90-10 wage gap to the standard deviation.

Figure A1 shows that adjusting for topcoding in the May/ORG has a relatively modest impact on the trends in wage dispersion. One exception is in the early 1980s where the unadjusted standard deviation grows substantially less than the adjusted standard deviation until the topcode is increased from $999 to $1999 in 1986. Though the 90-10 gap is not as smooth as the standard deviations, it tends to follow closely the evolution of the adjusted standard deviation.

Consistent with Card and DiNardo (2003), Figure A2 shows that the standard deviation is more sensitive to topcoding adjustments in the March than in the May/ORG CPS. In particular, the unadjusted standard deviation grows by only about 0.01 between 1975 and 1980, compared to about 0.025 for the adjusted standard deviation. But this case aside, the trends in the three measures of wage dispersion are relatively similar over the whole 1975-2000 period.

Figures A3 and A4 show that adjusting for topcoding has very little impact on the standard deviation of wages for women. This is not very surprising since only a small fraction of women have topcoded earnings. For example, between 1 and 4 percent of men earn above the topcode in the March CPS compared to less than one percent of women.

Interestingly, the 90-10 gap is substantially lower than the standard deviation during the 1970s, while the two series closely track each other from the early 1980s on. This is consistent with DiNardo, Lemieux and Fortin (1996)'s finding that the minimum wage strongly compressed the bottom of the female wage distribution during the 1970s, but had much less impact later on.

Finally, following Juhn, Murphy, and Pierce (1993), Figures A5 to A12 show the detailed evolution of all wage percentiles over the 1975-80, 1980-90, 1990-2000, and 1975-2000 periods. This set of figures shows a number of interesting facts. First, Figure A5 confirms that the March CPS wage measure is more sensitive to topcoding than the wage measure from the May/ORG CPS. In the March CPS, there is a strong divergence between adjusted (solid line) and unadjusted wages (dotted line) above the 90th percentile. In fact, the figure suggests that the 1.4 factor may be over-adjusting wages for this period. By contrast, only the very top percentiles appear to be affected by the top-ding adjustment in the May/ORG CPS.

Second, Figures A5 to A8 suggest that most of the differences between the two wage series for men arise in the two tails of the wage distribution. For example, the May/ORG and March CPS yield very similar wage changes between the 25th and 80th percentiles over the whole 1975-2000 period (Figure A8). The difference is that there tends to be more growth in the March than in the May/ORG CPS at the upper end of the distribution, but less growth in the March than in the May/ORG CPS at the lower end of the distribution.

Third, the May/ORG CPS captures much more sharply the impact of the large decline in the real value of the minimum wage for women during the 1980s (Figure A10). This confirms the earlier finding that using March wages is particularly problematic for workers paid by the hour. This also results in important differences in wage changes at the bottom end of the distribution (below the 10th centile) for the whole 1975-2000 period (Figure A12).

APPENDIX B: Accounting for Composition Effects in the March CPS Wage Data

Appendix Figures B1 to B4 compares the actual within-group variance using the March CPS hourly wage rate to the within-group variance that would have prevailed if the distribution of age and education had remained constant over time. The re-weighting methodology used to compute the counterfactual variance is the same as for the

May/ORG CPS (Figure 10) except that I hold the distribution of age and education at their 1975 and 2001 levels (instead of 1973 and 2002 in the May/ORG CPS).

The figures show that the impact holding the distribution of characteristics constant is much less dramatic in the March CPS than in the May/ORG CPS data for two reasons. First, since the overall growth in within-group inequality is more pronounced in the March CPS (Table 4 and Figure 6), the same composition effects will only explain a smaller fraction of the growth in the within-group inequality in the March CPS than in the May/ORG CPS. Second, a comparison of Figures 10a-10d and Appendix Figures B1-B4 shows that composition effects are only about 2/3 as large (in absolute value) in the March CPS as in the May/ORG CPS.

The reason why composition effects are smaller is that there is generally less difference in the within-group variance between high skill and low skill workers in the March CPS than in the May/ORG CPS. For example, when experience is held constant in the May/ORG CPS (as in Figure 8), the average within-group variance for male high school dropouts, high school graduates, and college graduates is 0.15, 0.17, and 0.26, respectively over the whole 1973-2002 period. The comparable within-group variances in the March CPS are 0.23, 0.23, and 0.29, respectively. This amounts to a 0.09 difference between the college and high school variances in the May/ORG, compared to a 0.06 difference in the March/ORG CPS. The gap between the two wages series is even larger when looking at the difference between college graduates and high school dropouts. As a result, changing the distribution of education does not have as large an effect in the March CPS as in the May/ORG CPS data.

The natural explanation for this pattern of results is that high school dropouts and other less skilled workers are much more likely to be paid by the hour than more educated workers as illustrated in the following table:

Appendix Table B1: Percentage of workers paid by the hour by level of education, 1999 ORG CPS

	HS Dropouts	HS Graduates	Some College	College grad	Post-graduate
Men	84%	72%	61%	23%	11%
Women	89%	76%	68%	36%	16%

Recall from Figure 4 that the March CPS appears to largely overstate the variance of log wages for workers paid by the hour, while the two wages series yield much more comparable results for workers not paid by the hour. Combining these pieces of evidence together suggests that the variance of wages for less-educated workers, who are much more likely to be paid by the hour, is artificially high in the March CPS because of the measurement problems for workers paid by the a hour documented in Section 2. Since this biases down significantly the importance of composition effects, I focus on the May/ORG CPS for looking at the importance of composition effects in Section 4.

Despite these shortcomings, adjusting for composition effects still has a significant impact on the economic interpretation of the trends in the within-group variance in the March CPS. In particular, Appendix Figures B1 (men) and B3 (women) show essentially no growth in the within-group variance after 1987-88 when the distribution of age and education is held at its 1975 level. For women, the pattern of growth in within-group inequality in the March CPS is now very similar to the one in the May/ORG CPS (with or without adjustments for composition effects) where all the growth in within-group inequality is concentrated in the first half of the 1980s.

For men, the post-1980 growth in within-group inequality also becomes qualitatively similar to the one in the May/ORG CPS. The only remaining clear discrepancy is that the within-group inequality grows rapidly in the March CPS during the 1970s, while it remains stable in the May/ORG data.

Table 1: Preferred periodicity for reporting earnings, 1995 ORG CPS

Periodicity	Percentage of workers:			Mean wage	Variance of logs
	All	Hourly	Non-hourly		
	(1)	(2)	(3)	(4)	(5)
Hourly	44.4	72.0	0.0	8.31	0.223
Weekly	17.8	12.1	26.8	10.19	0.322
Bi-weekly	6.7	5.0	9.6	11.02	0.304
Twice monthly	2.0	0.9	3.7	11.30	0.325
Monthly	5.9	1.9	12.5	11.64	0.320
Yearly	21.6	7.7	43.6	15.76	0.270
Other	1.7	0.4	3.9	9.66	0.427
Percentage of all workers	100.0	62.1	37.9		

Table 2: Decomposition of Changes in the Variance of Log Hourly Wages

	1975-1980	1980-1990	1990-2001	1975-2001
A. Men, May/ORG CPS				
Total Change	-0.008	0.077	0.018	0.087
Within	-0.005	0.034	0.008	0.037
	[64]	[44]	[47]	[43]
Between	-0.003	0.043	0.010	0.050
	[36]	[56]	[53]	[57]
B. Men, March CPS				
Total	0.027	0.059	0.045	0.130
Within	0.021	0.028	0.029	0.078
	[80]	[48]	[64]	[60]
Between	0.005	0.030	0.016	0.052
	[20]	[52]	[36]	[40]
C. Women, May/ORG CPS				
Total change	-0.016	0.095	0.025	0.103
Within	-0.005	0.051	0.005	0.051
	[33]	[54]	[21]	[49]
Between	-0.011	0.044	0.020	0.053
	[67]	[46]	[79]	[51]
D. Women, March CPS				
Total change	-0.001	0.075	0.053	0.126
Within	0.007	0.039	0.030	0.075
	[-]	[52]	[57]	[60]
Between	-0.008	0.036	0.023	0.051
	[-]	[48]	[43]	[40]

Note: Percentage of total change in square bracket

Table 3: Regression models of within-group variance on between-group variance

	Men		Women	
	May/ORG	March	May/ORG	March
	(1)	(2)	(3)	(4)
A. Models with linear trend				
Between-group variance	0.749	0.403	0.979	0.813
	(0.082)	(0.138)	(0.204)	(0.205)
Year/100	-0.027	0.193	-0.053	0.097
	(0.021)	(0.032)	(0.062)	(0.054)
B. Models with quadratic trend				
Between-group variance	0.706	0.025	0.887	0.865
	(0.108)	(0.151)	(0.115)	(0.150)
Year/100	0.007	0.455	0.233	0.240
	(0.058)	(0.075)	(0.053)	(0.050)
$(\text{Year}/100)^2$	-0.088	-0.704	-0.998	-0.601
	(0.141)	(0.189)	(0.137)	(0.128)

Note: Standard errors in parentheses

Table 4: Percentage distribution of workers by education and experience groups, May/ORG CPS

	Men				Women			
	1973-74	1980	1990	2002	1973-74	1980	1990	2002
A. Education categories								
High School Dropout	30.4	23.0	15.9	11.3	25.7	17.5	11.4	7.8
High School Graduate	37.4	37.9	38.1	31.1	46.3	46.0	41.5	29.7
Some college	15.3	18.1	20.4	27.1	13.7	18.7	23.2	31.1
Bachelors' Degree	9.1	11.6	14.8	20.1	9.3	11.0	14.8	21.1
Post-graduate Degree	7.7	9.4	10.9	10.6	5.0	6.9	9.2	10.3
B. Years of Experience								
0-10	35.8	39.4	31.9	27.0	38.5	41.4	33.8	28.3
11-20	22.7	24.5	32.8	27.8	18.5	22.8	29.5	24.8
21-30	18.2	16.4	19.5	27.1	19.1	16.6	21.0	27.4
31+	23.3	19.7	15.8	18.1	23.9	19.3	15.7	19.4

Table 5: Composition Effects and Changes in the Residual Variance of Log Hourly Wages, May/ORG CPS

	1973-80	1980-90	1990-2002	1973-2002
A. Men				
Actual change	-0.001	0.034	0.013	0.046
Holding education and age as in 1973	-0.003	0.025	-0.008	0.014
Holding education and age as in 2002	-0.008	0.031	0.001	0.024
B. Women				
Actual change	-0.016	0.051	0.008	0.043
Holding education and age as in 1973	-0.020	0.039	-0.015	0.004
Holding education and age as in 2002	-0.015	0.045	-0.001	0.029

Table 6: Role of the Minimum Wage in the Relative Evolution of Residual Wage Inequality for Men and Women

	1973-80	1980-90	1990-2002	1973-2002
A. Changes in the residual variance holding age and education at their 2002 level				
Men	-0.008	0.031	0.001	0.024
Women	-0.015	0.045	-0.001	0.029
Male-Female difference	0.007	-0.014	0.002	-0.005
B. Minimum wage effect from DiNardo, Fortin and Lemieux (1996)				
Men	-0.002	0.007	0.000[a]	0.005
Women	-0.005	0.015	0.000[a]	0.010
Male-Female difference	0.003	-0.008	0.000[a]	-0.005
C. Remaining changes in the residual variance				
Men	-0.006	0.024	0.001	0.019
Women	-0.010	0.030	-0.001	0.019
Male-Female difference	0.004	-0.006	0.002	0.000

a: Since the real value of the minimum wage remained relatively constant over the period 1990-2002 the minimum wage effect is assumed to be zero.

Figure 1: Mean log wages ($2001)

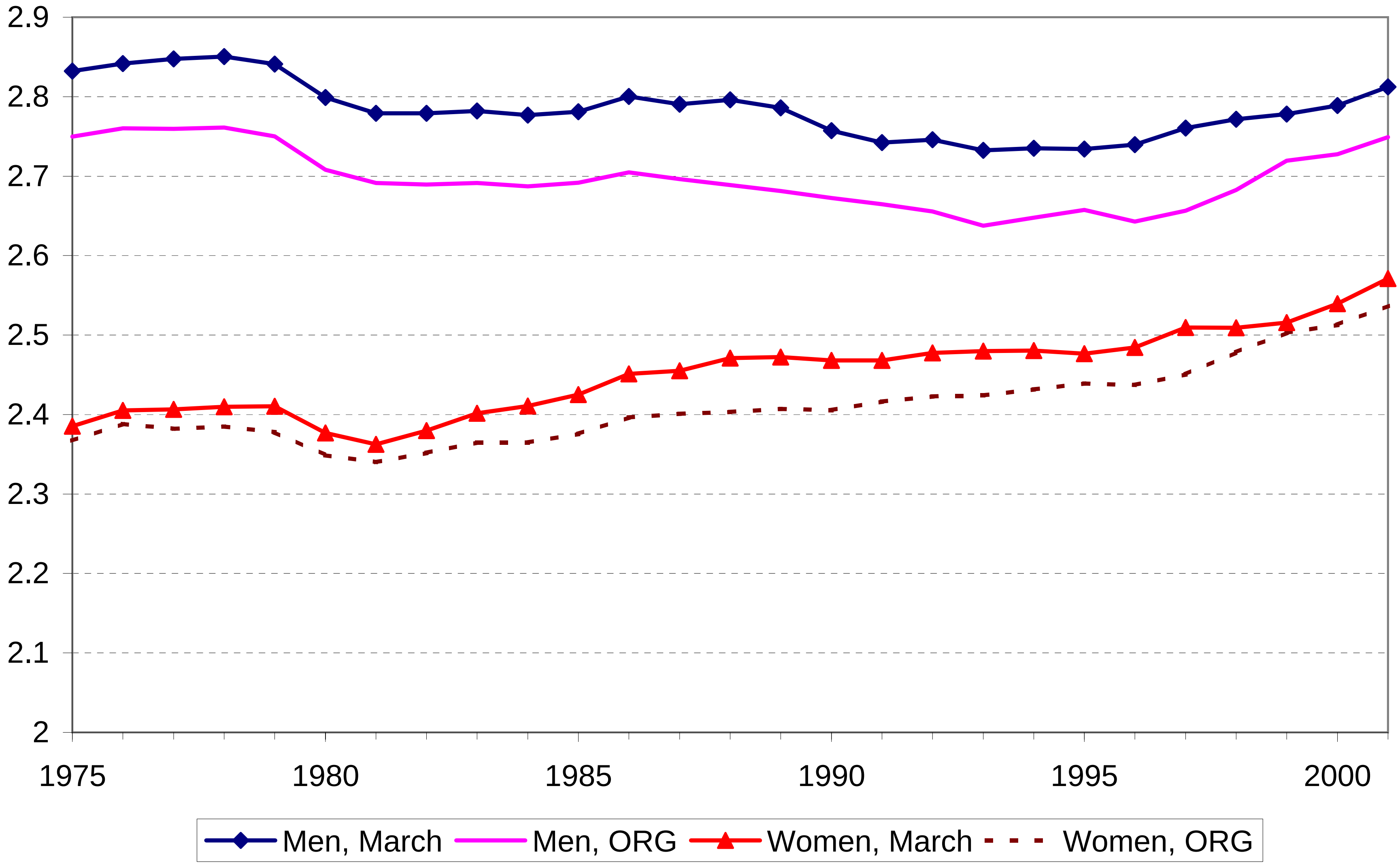


Figure 2a: Standard deviation of log wages, men

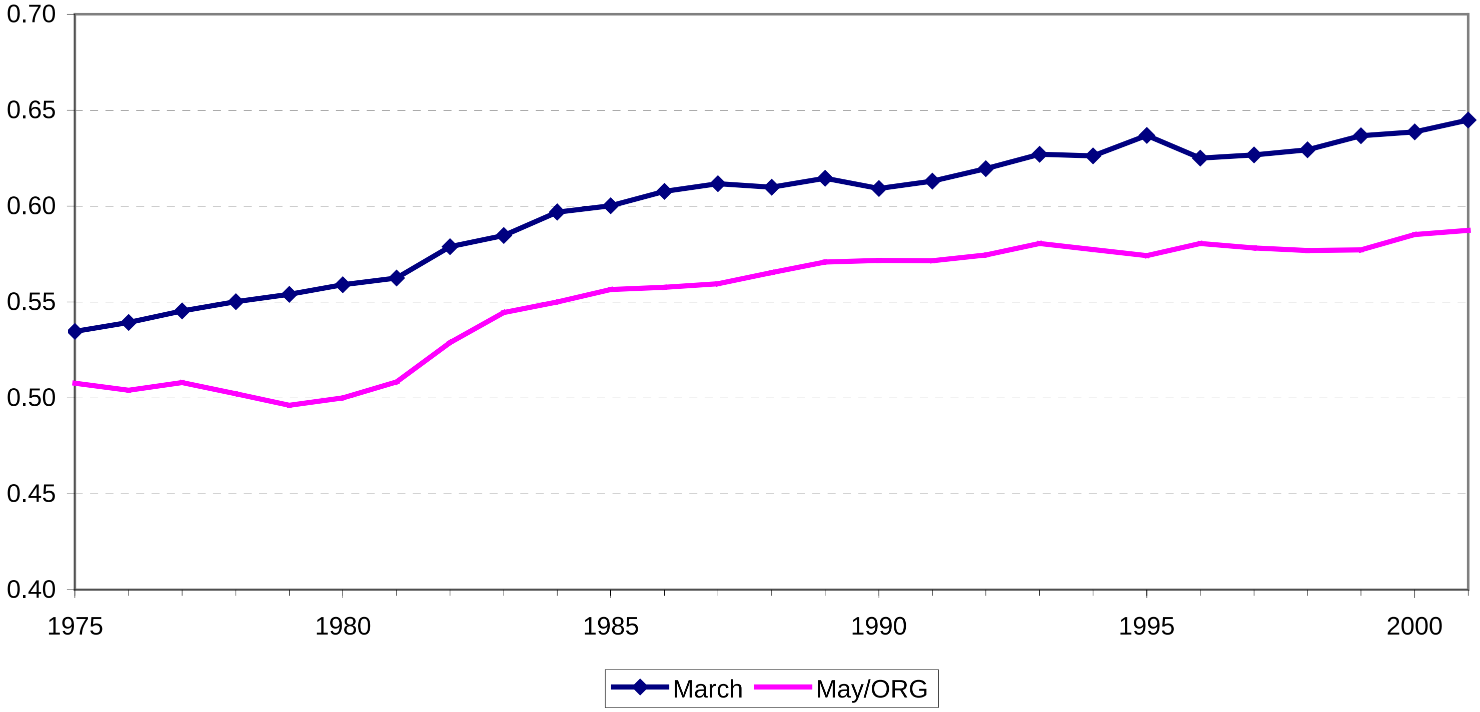


Figure 2b: Standard deviation of log wages, women

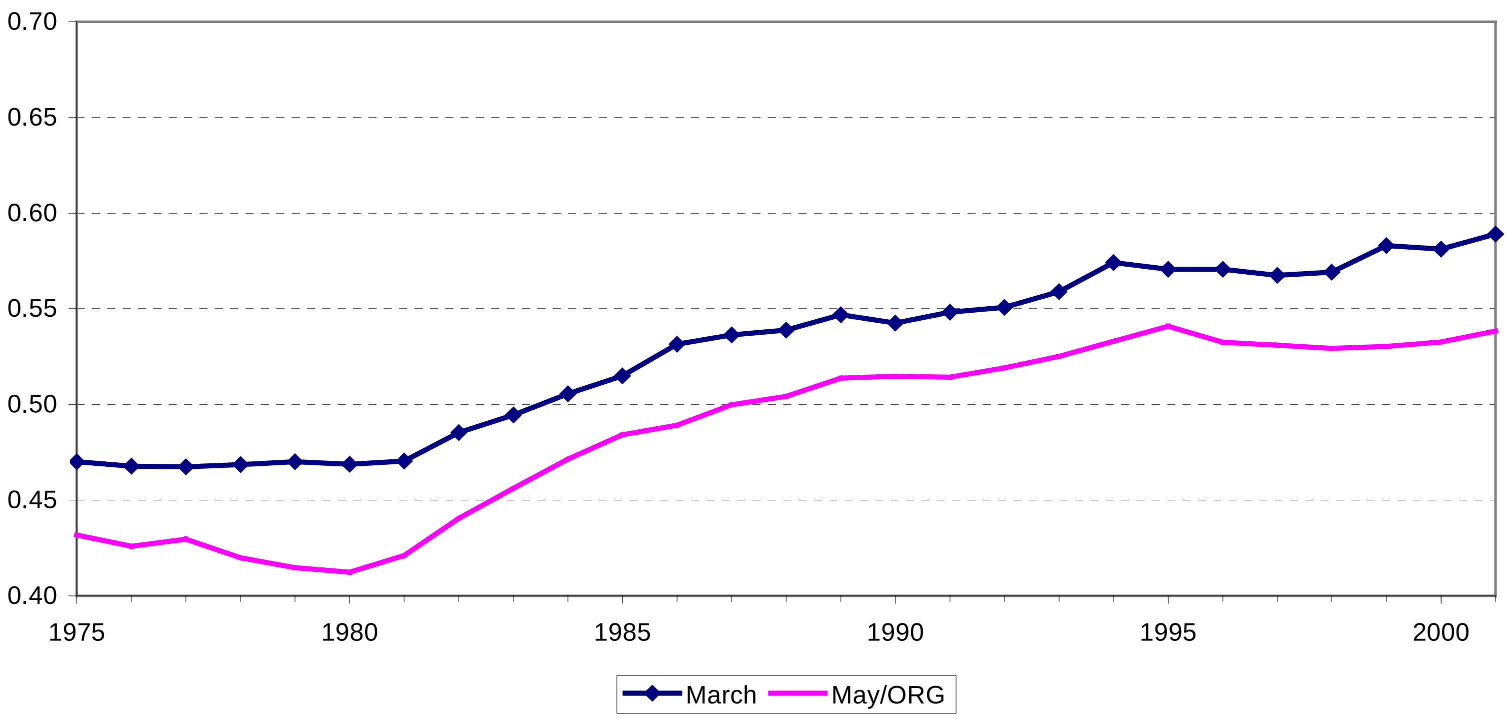


Figure 3: Percentage of workers paid by the hour

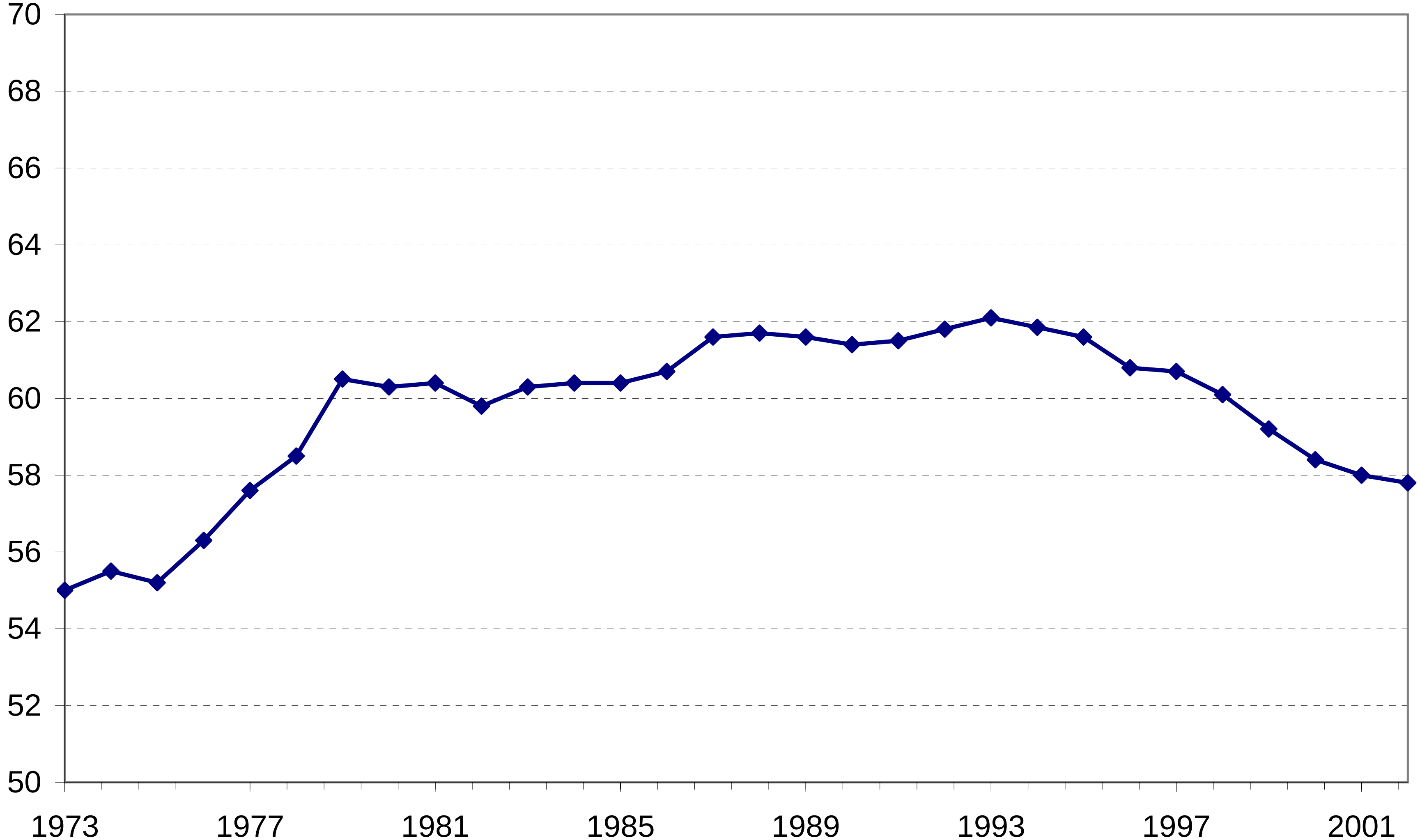


Figure 4a: Standard deviation of log hourly wages for hourly workers with both ORG and March wages

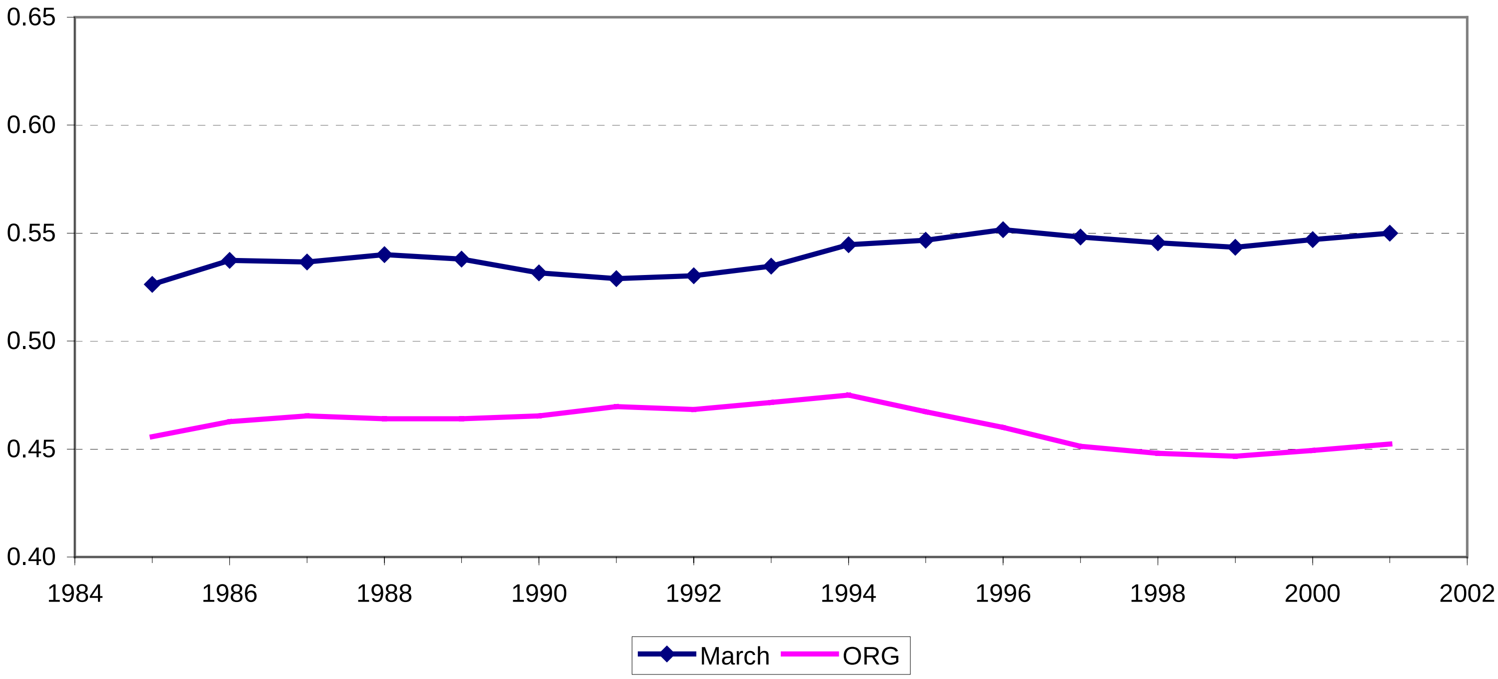


Figure 4b: Standard deviation of log hourly wages for weekly workers with both ORG and March wages

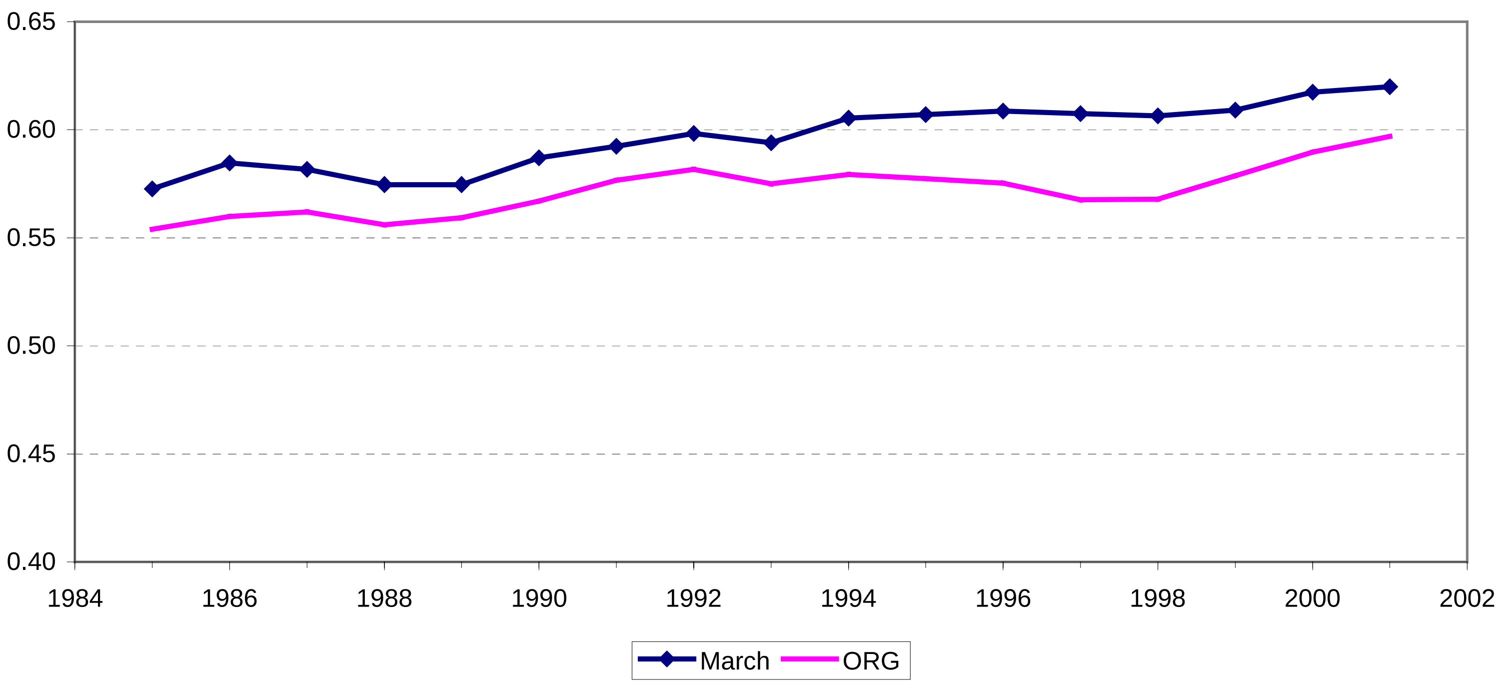


Figure 5a: Between-group variance, men

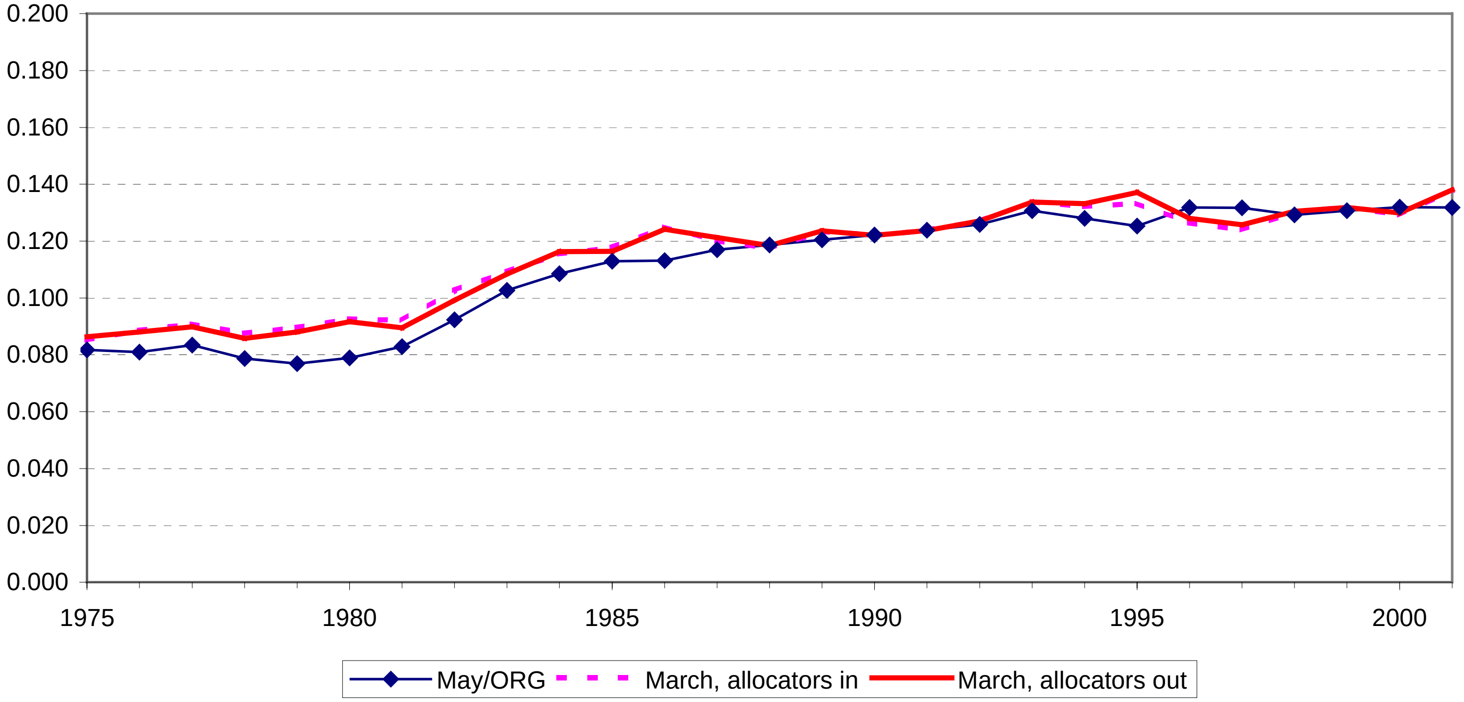


Figure 5b: Between-group variance, women

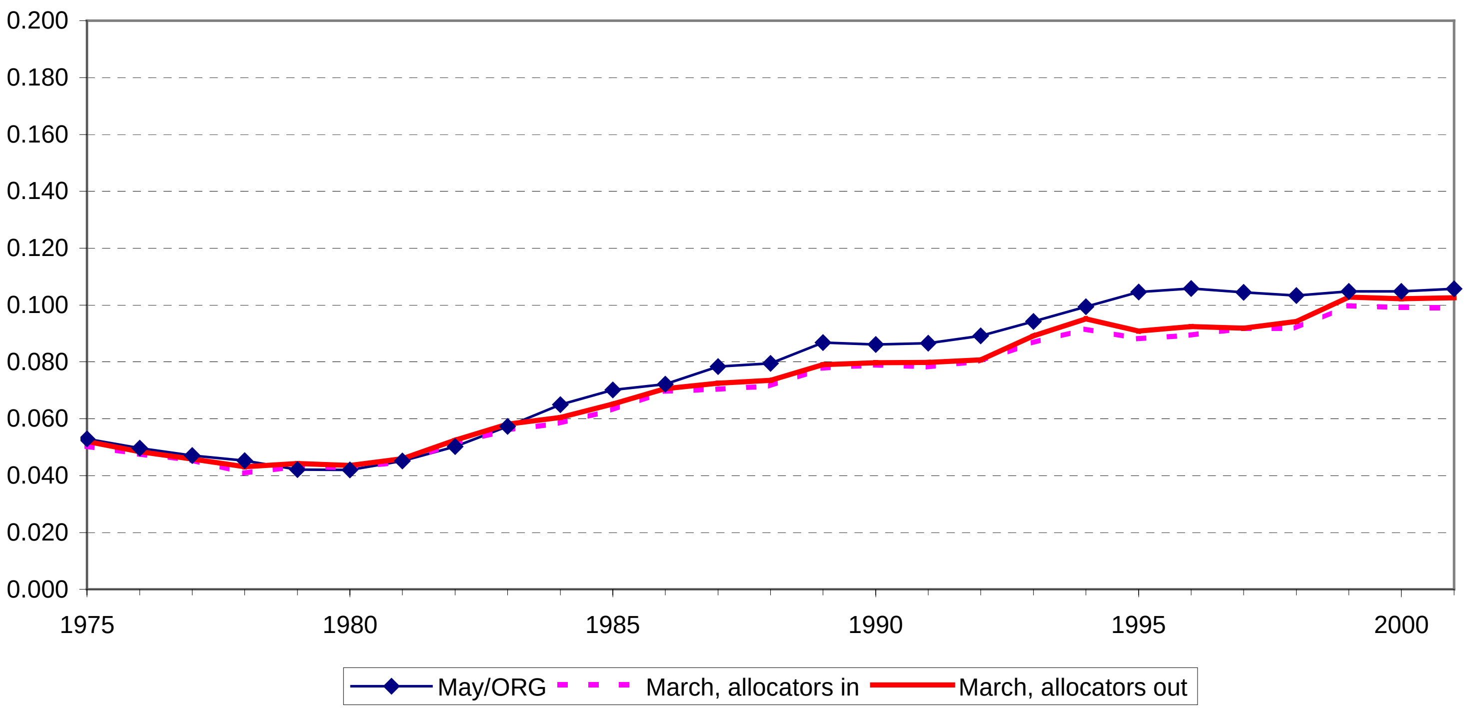


Figure 6a: Within-group variance, men

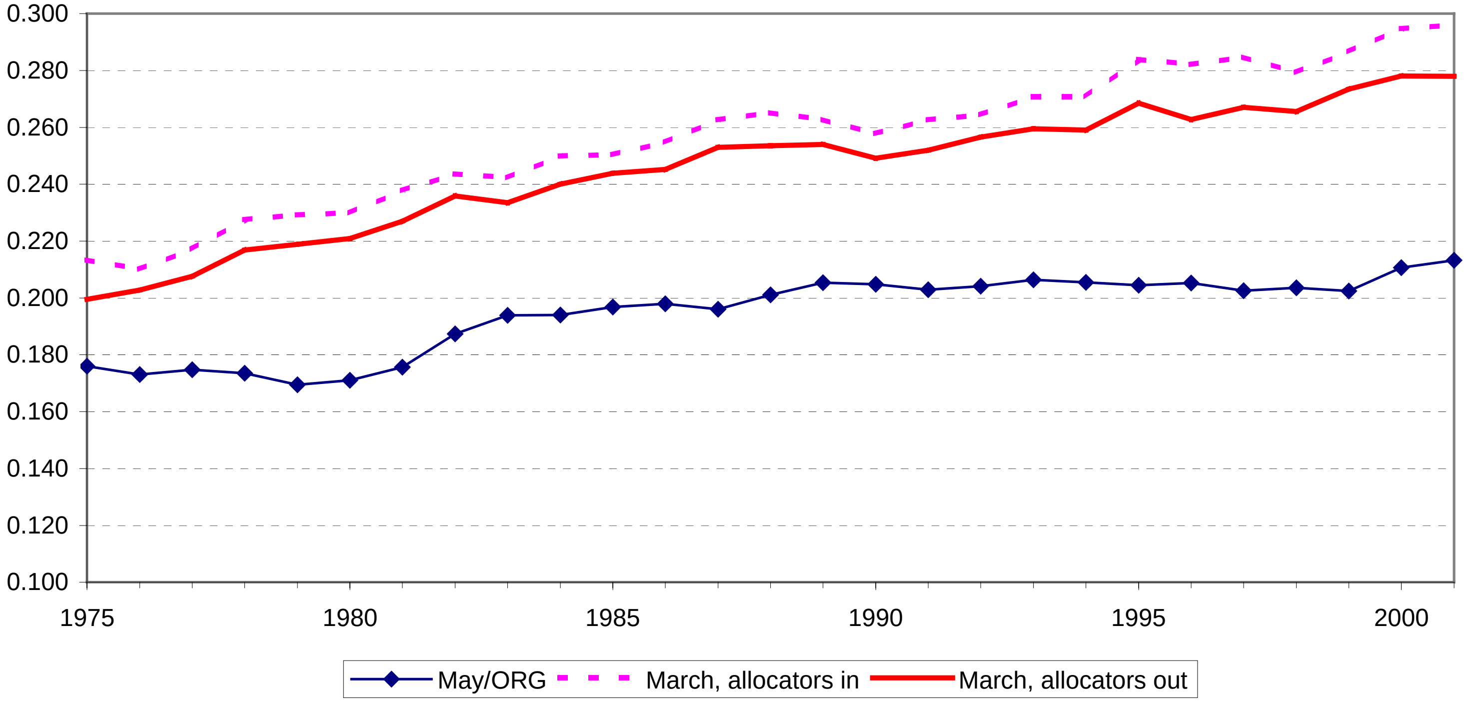


Figure 6b: Within-group variance, women

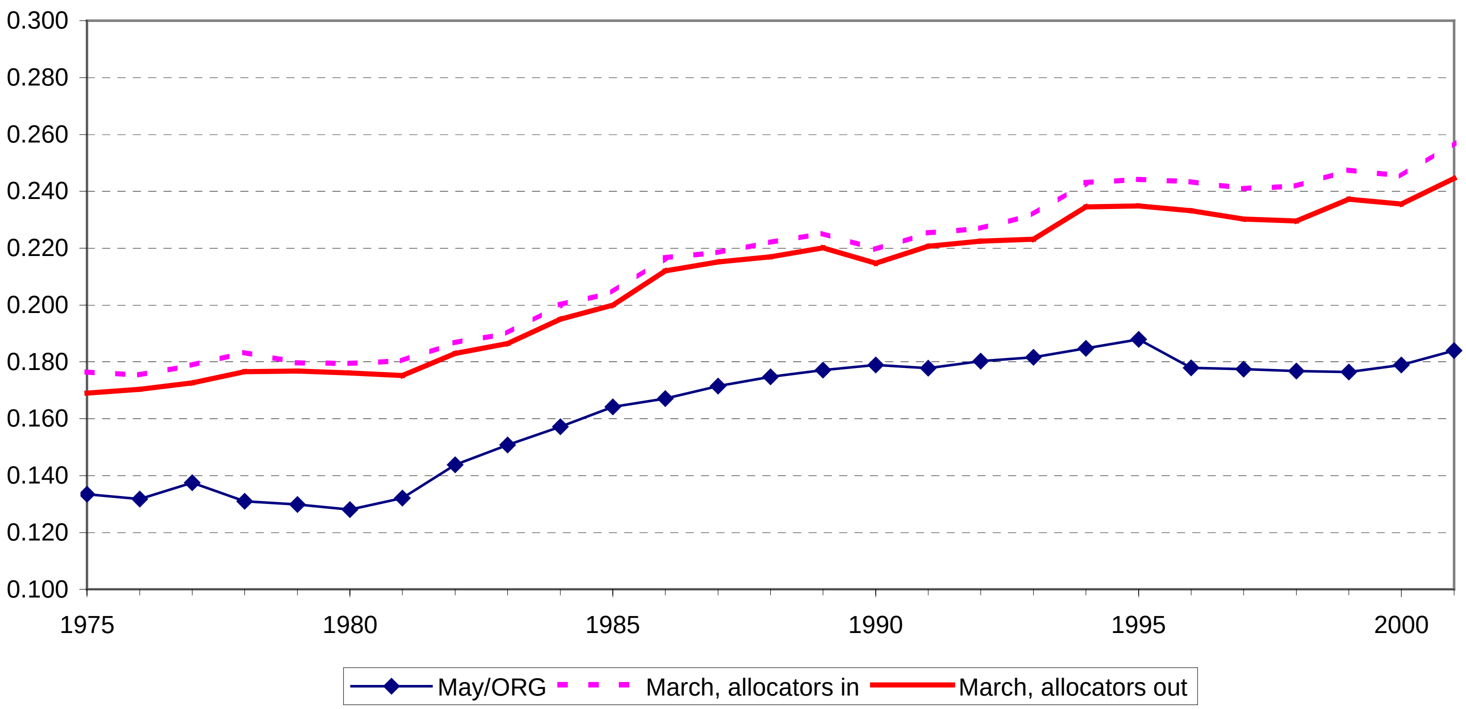


Figure 7a: Within-group variance in 1973-74 May CPS, men

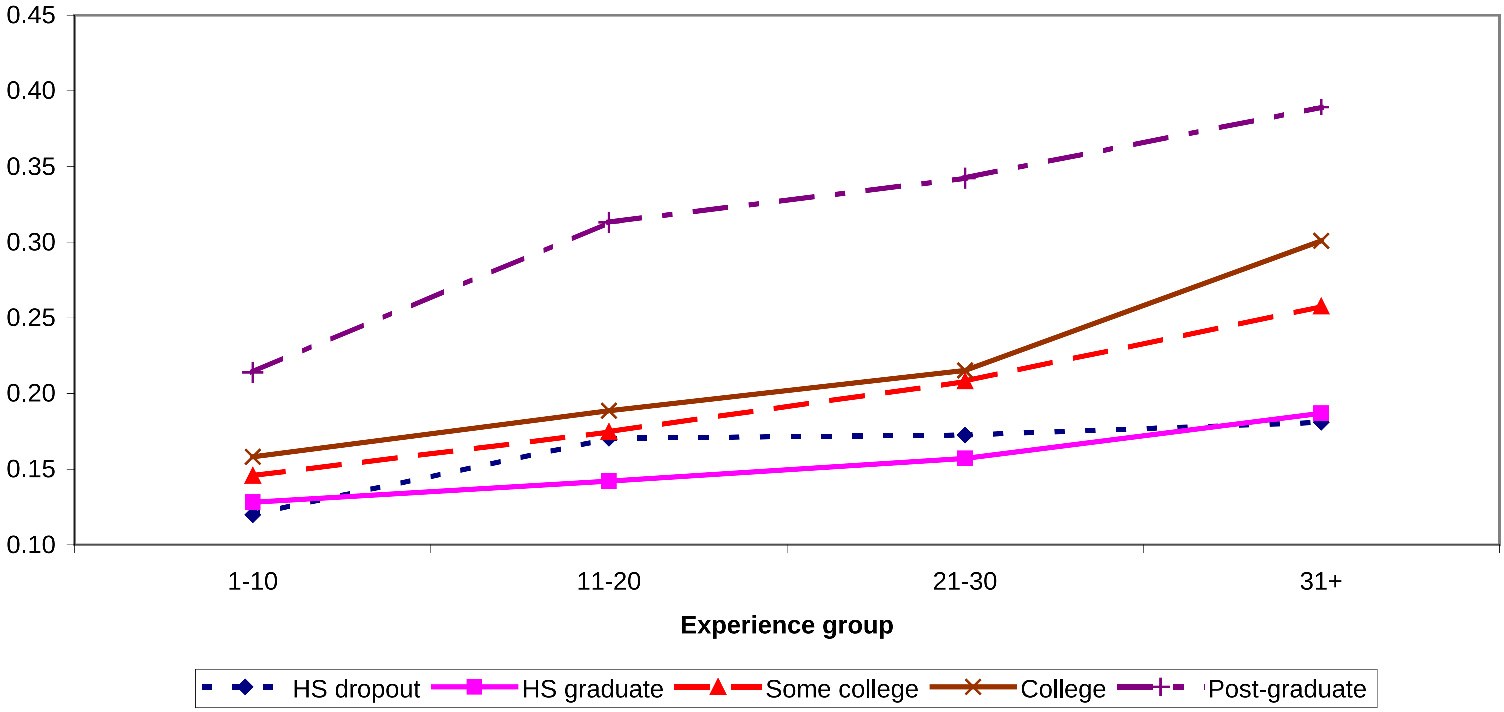


Figure 7b: Within-group variance in 2002 ORG CPS, men

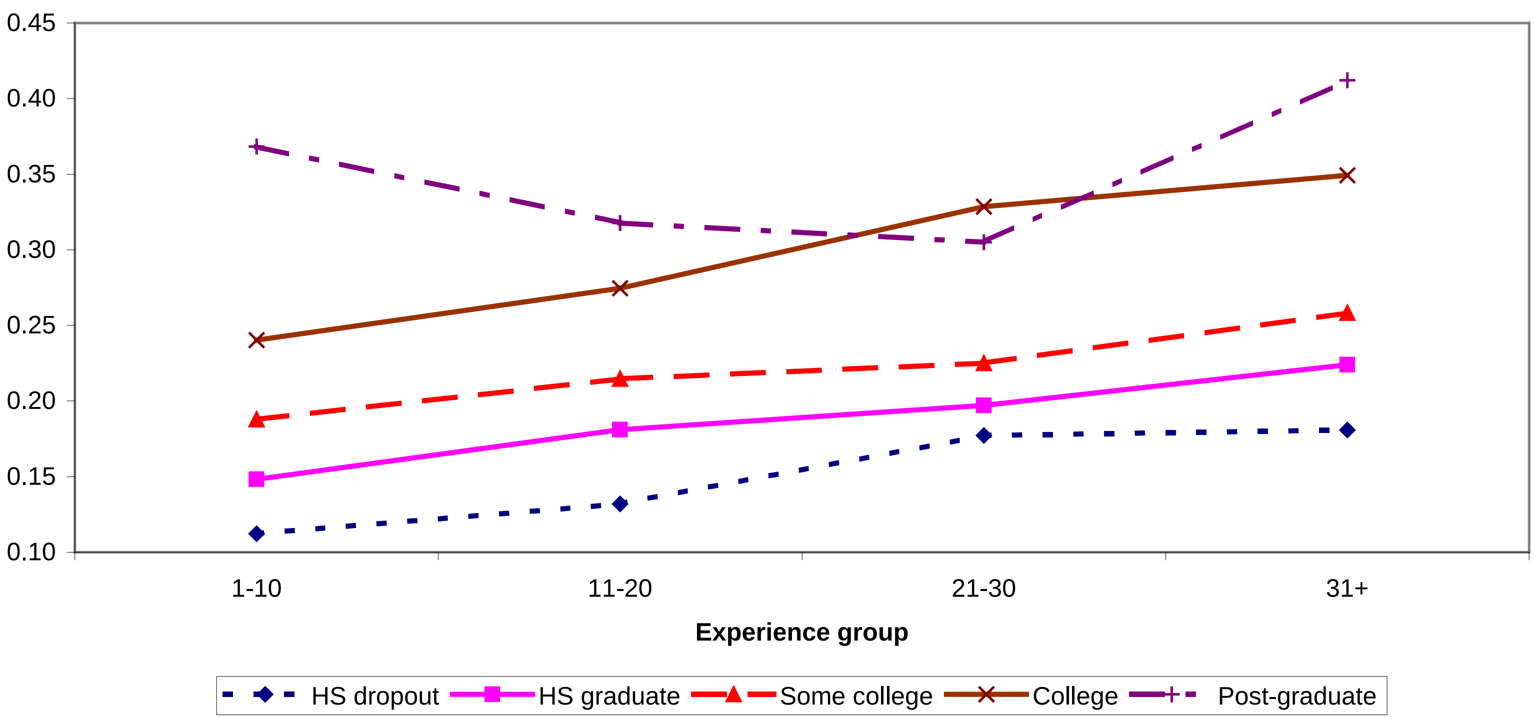


Figure 7c: Within-group variance in 1973-74 May CPS, women

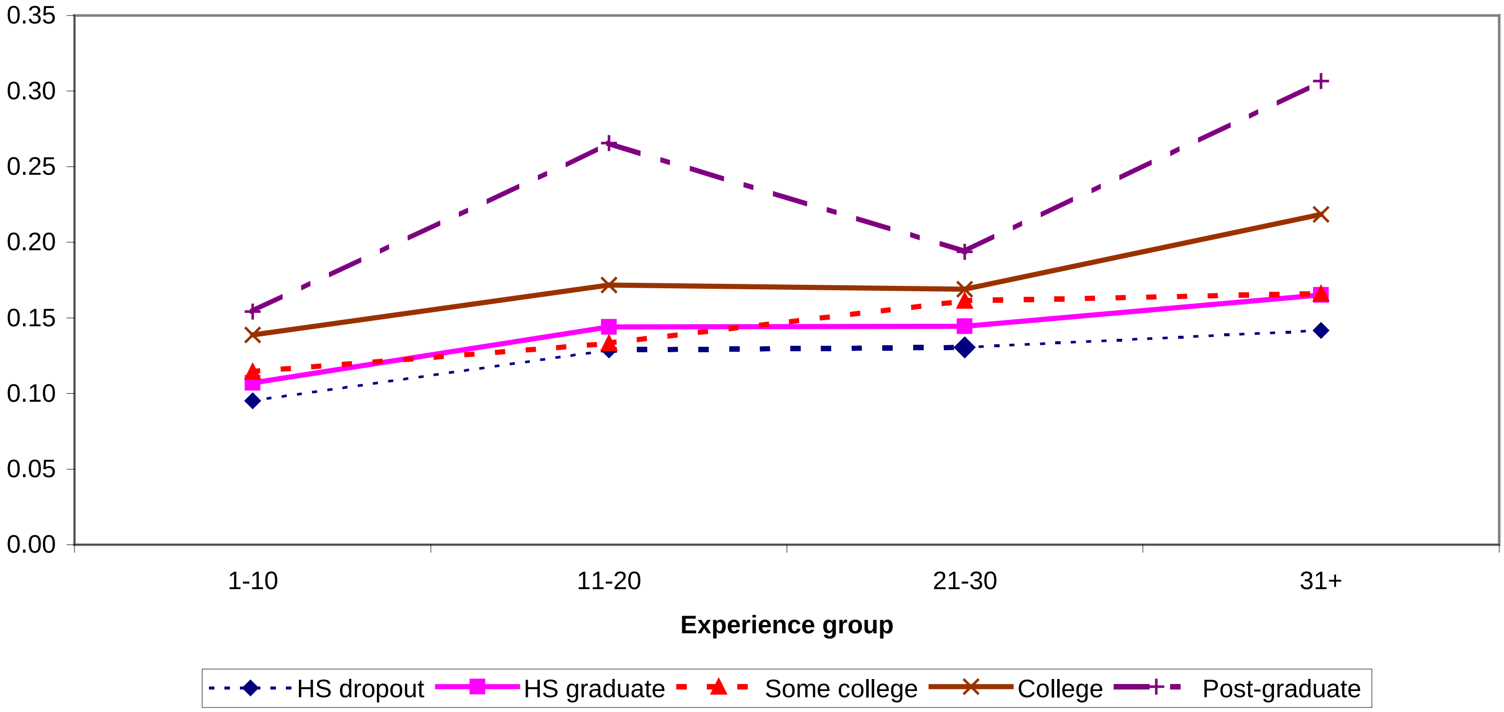


Figure 7d: Within-group variance in 2002 ORG CPS, women

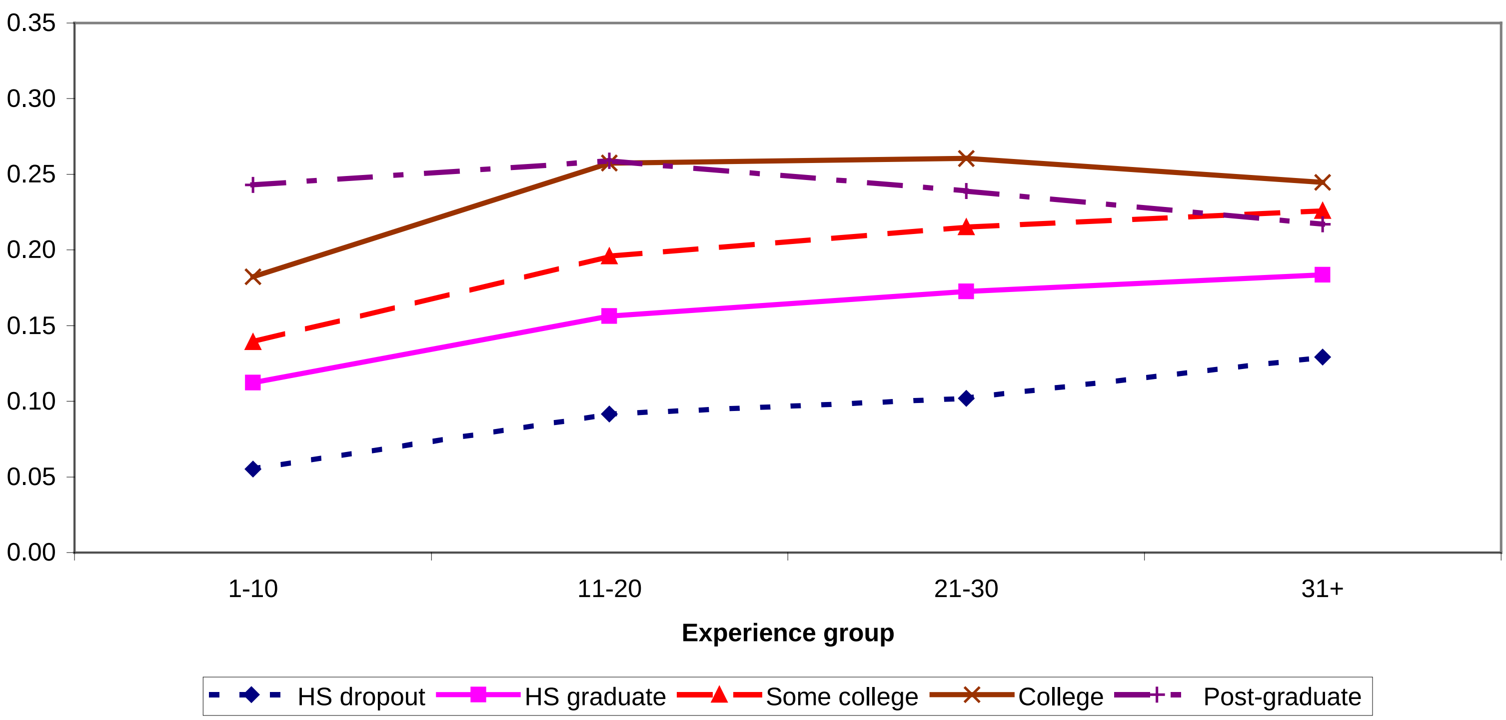


Figure 8a: Within-group variance by education group for men, May/ORG CPS

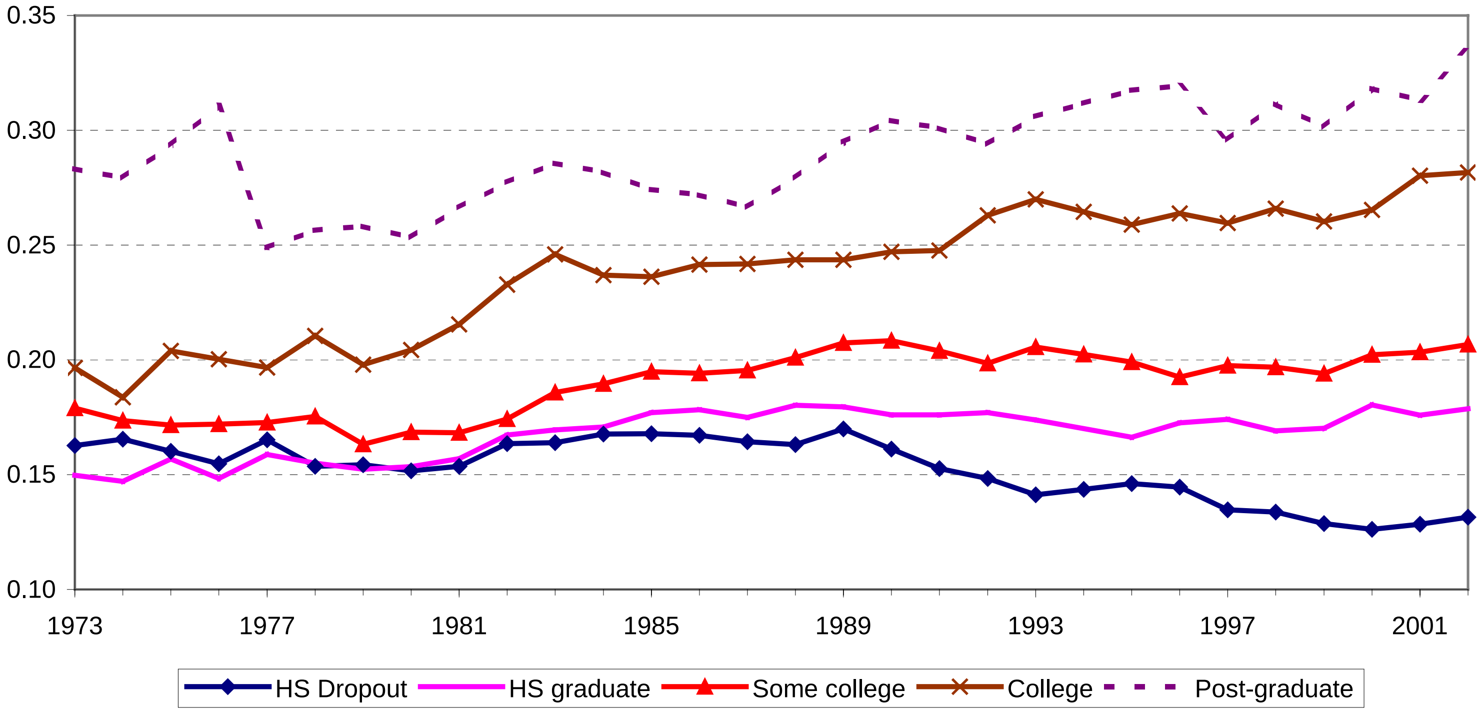


Figure 8b: Within-group variance by education group for men, holding experience distribution constant

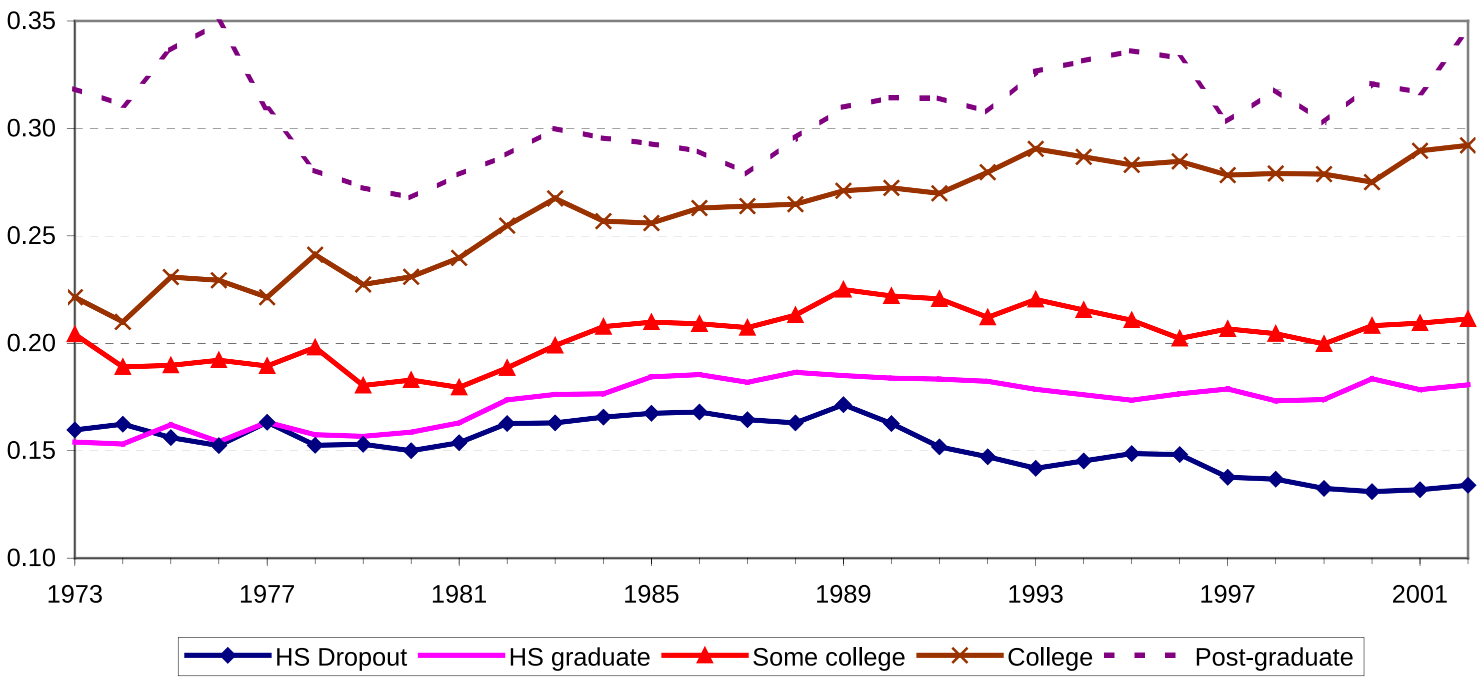


Figure 8c: Within-group variance by education group for women, May/ORG CPS

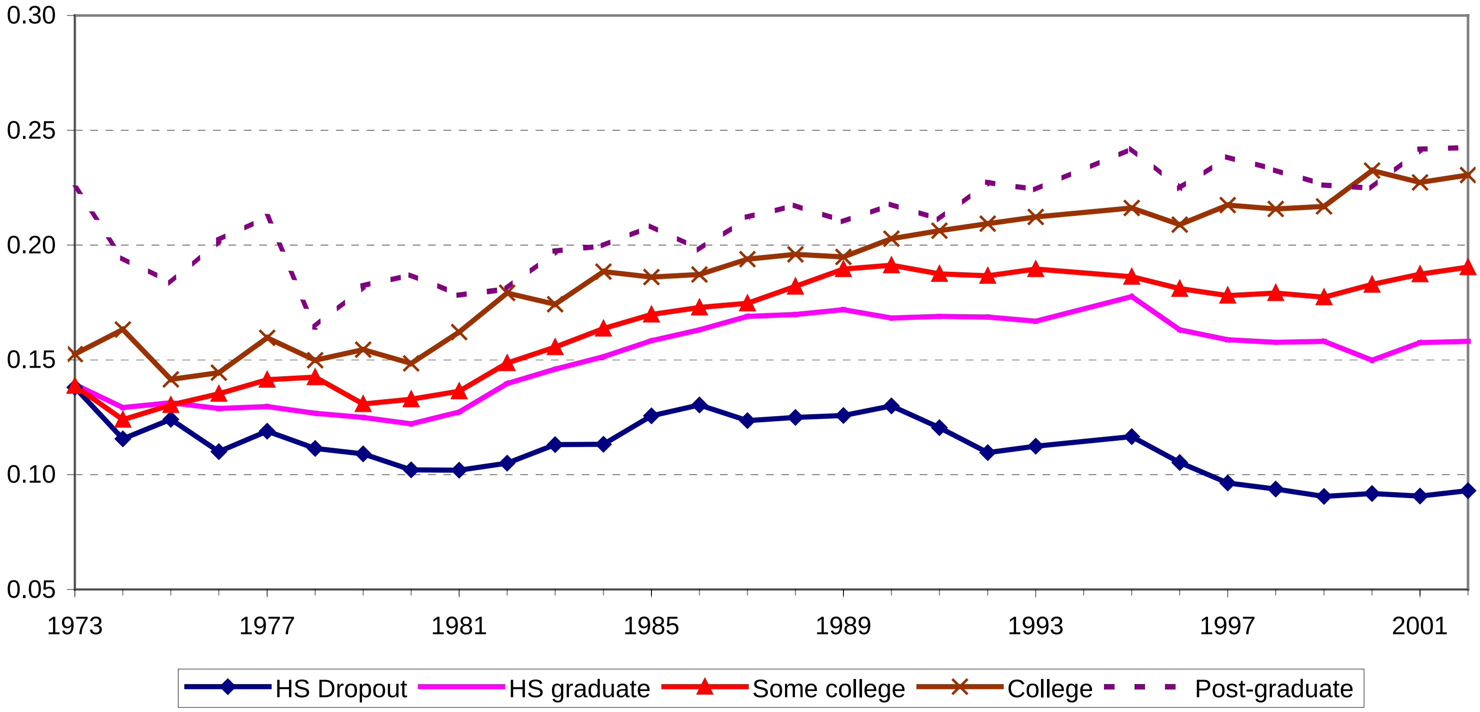


Figure 8d: Within-group variance by education group for women, holding experience distrbution constant

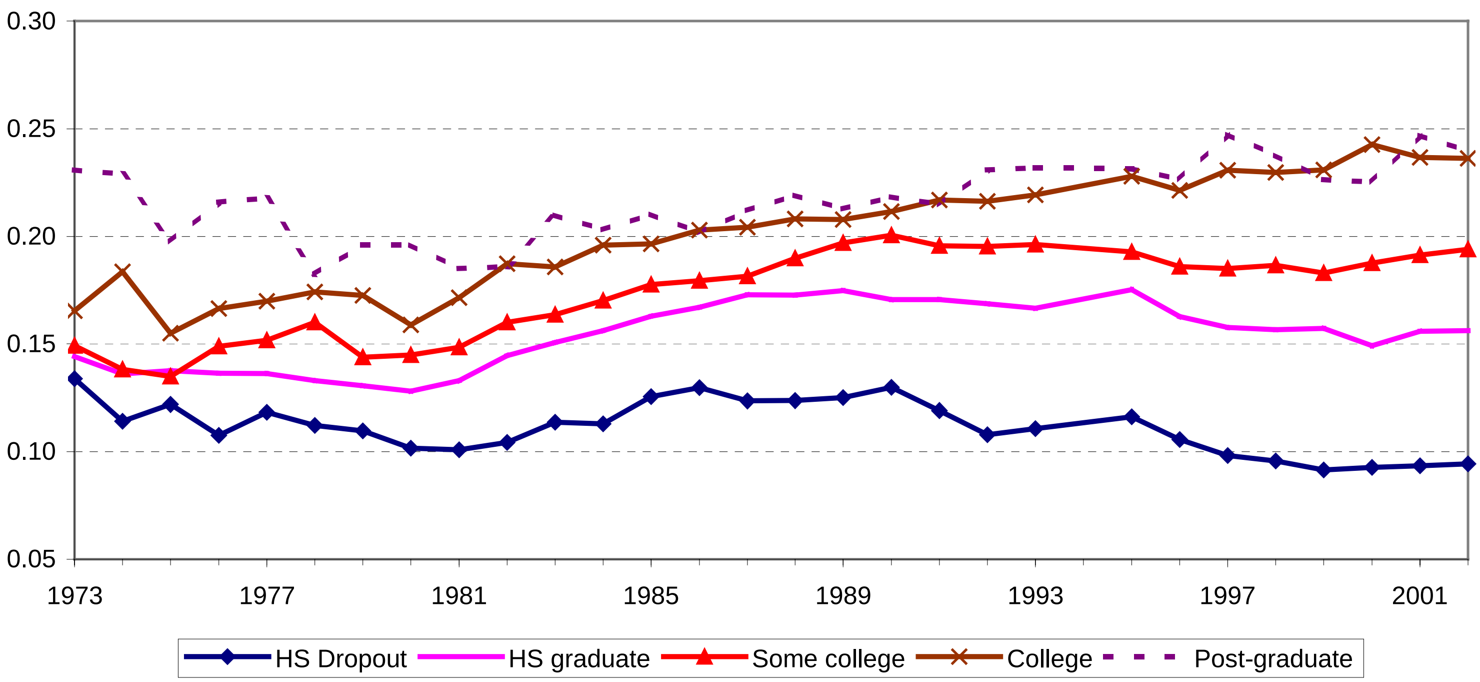


Figure 9a: Within-group variance by experience group for men, May/ORG CPS

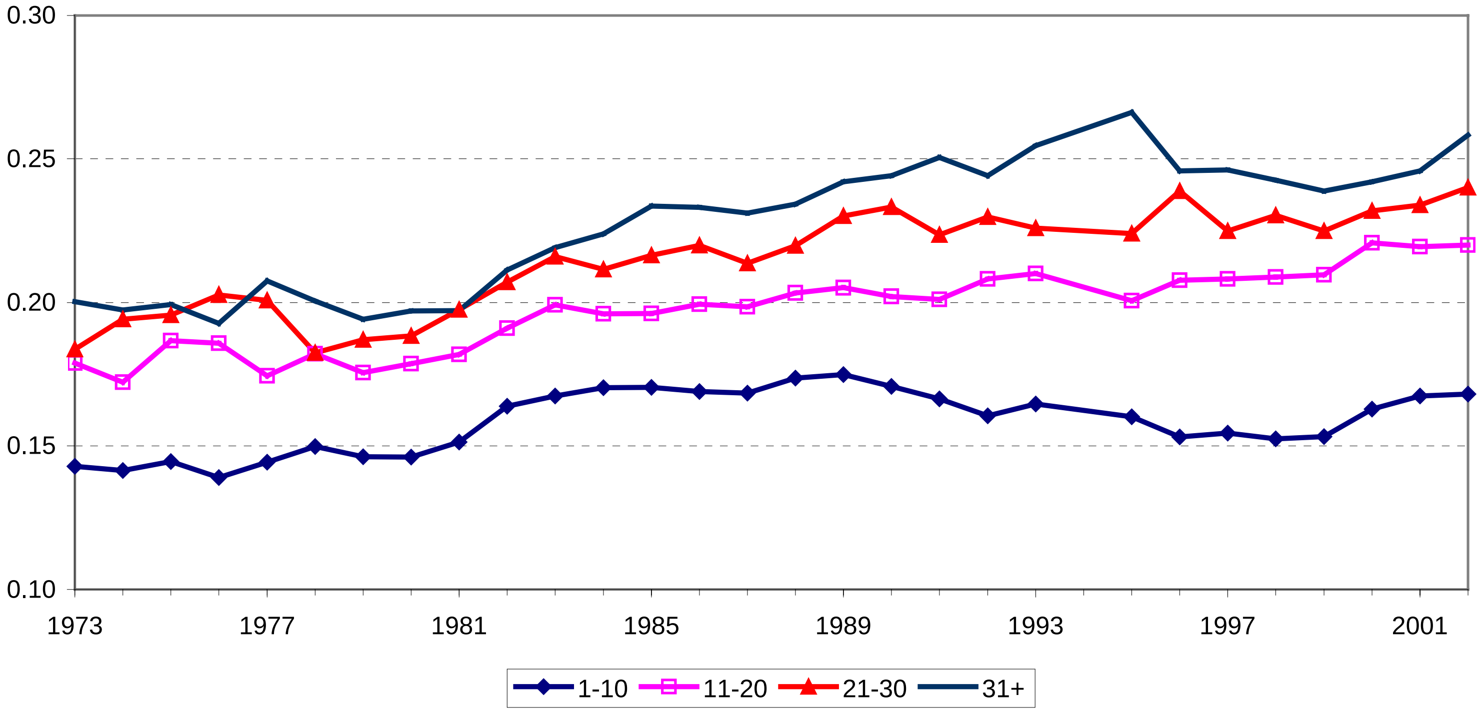


Figure 9b: Within-group variance by experience group for men, holding education distribution constant over time

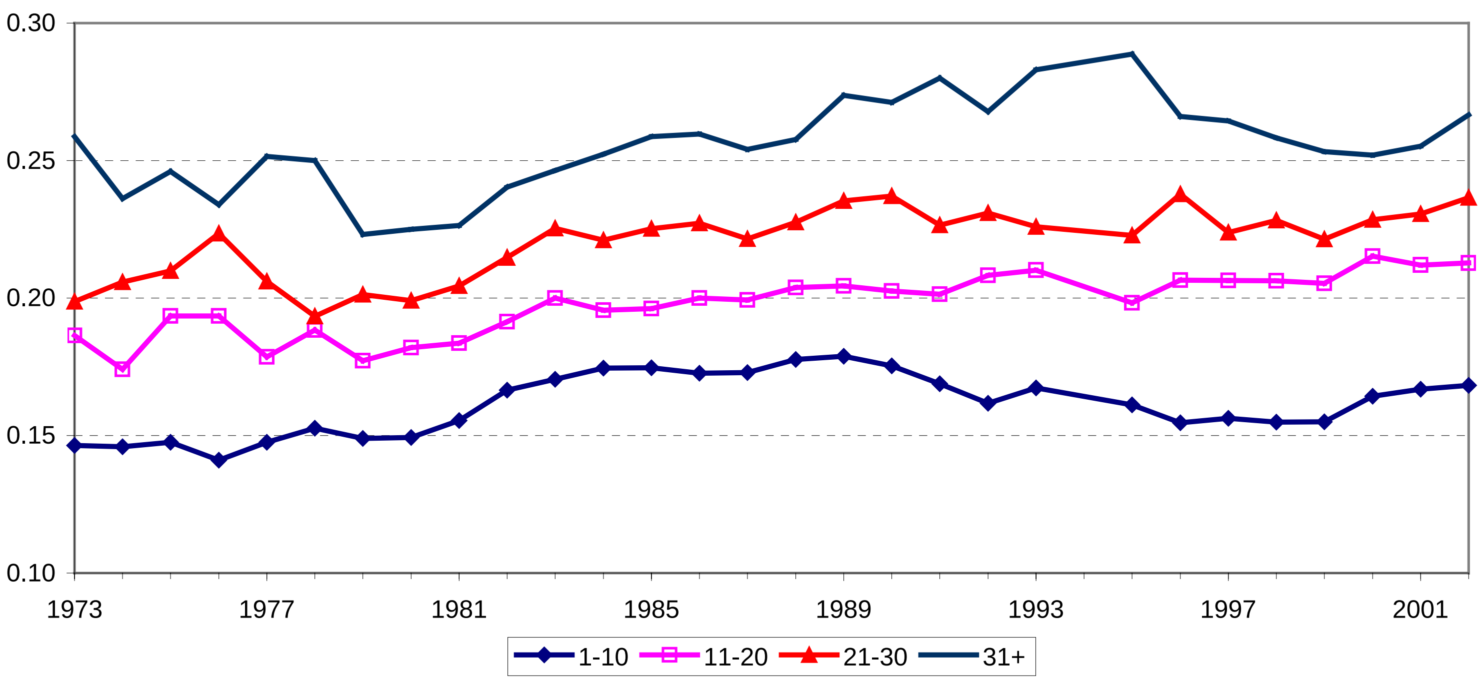


Figure 9c: Within-group variance by experience group for women, May/ORG CPS

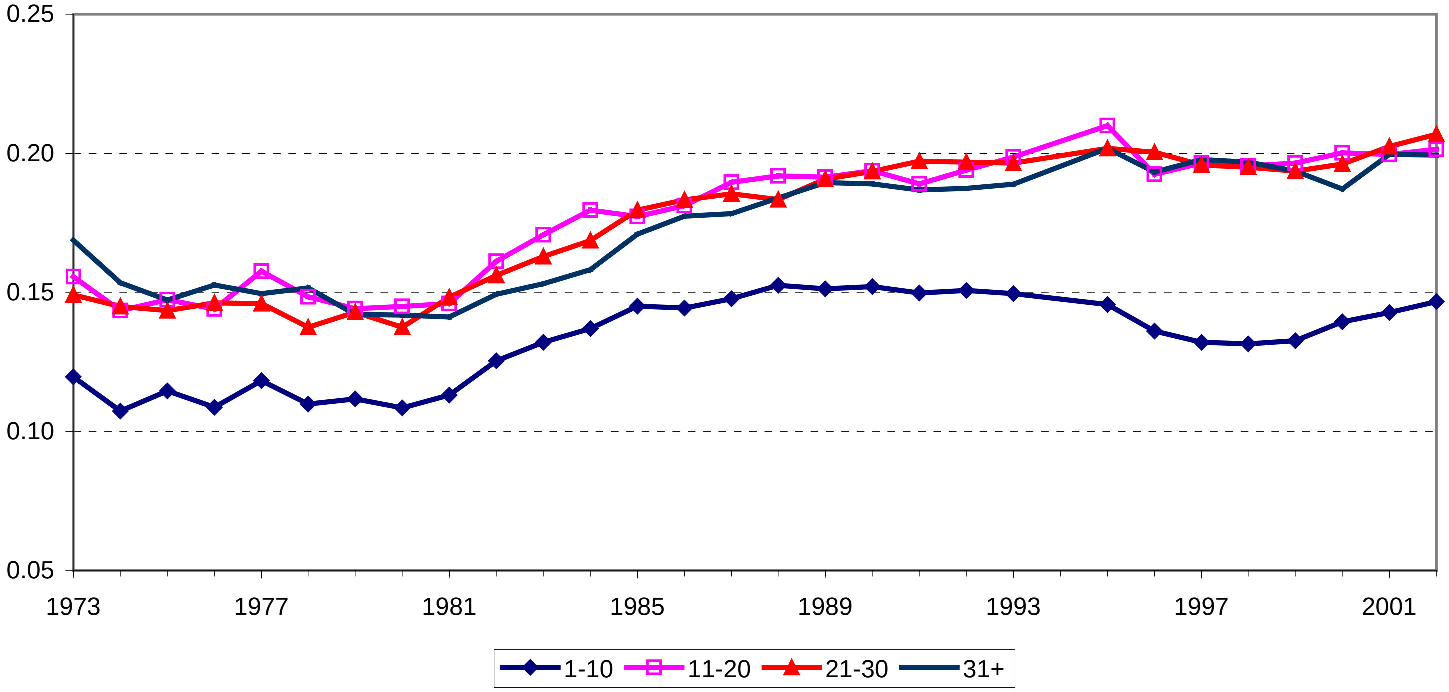


Figure 9d: Within-group variance by experience group for women, holding education distribution constant

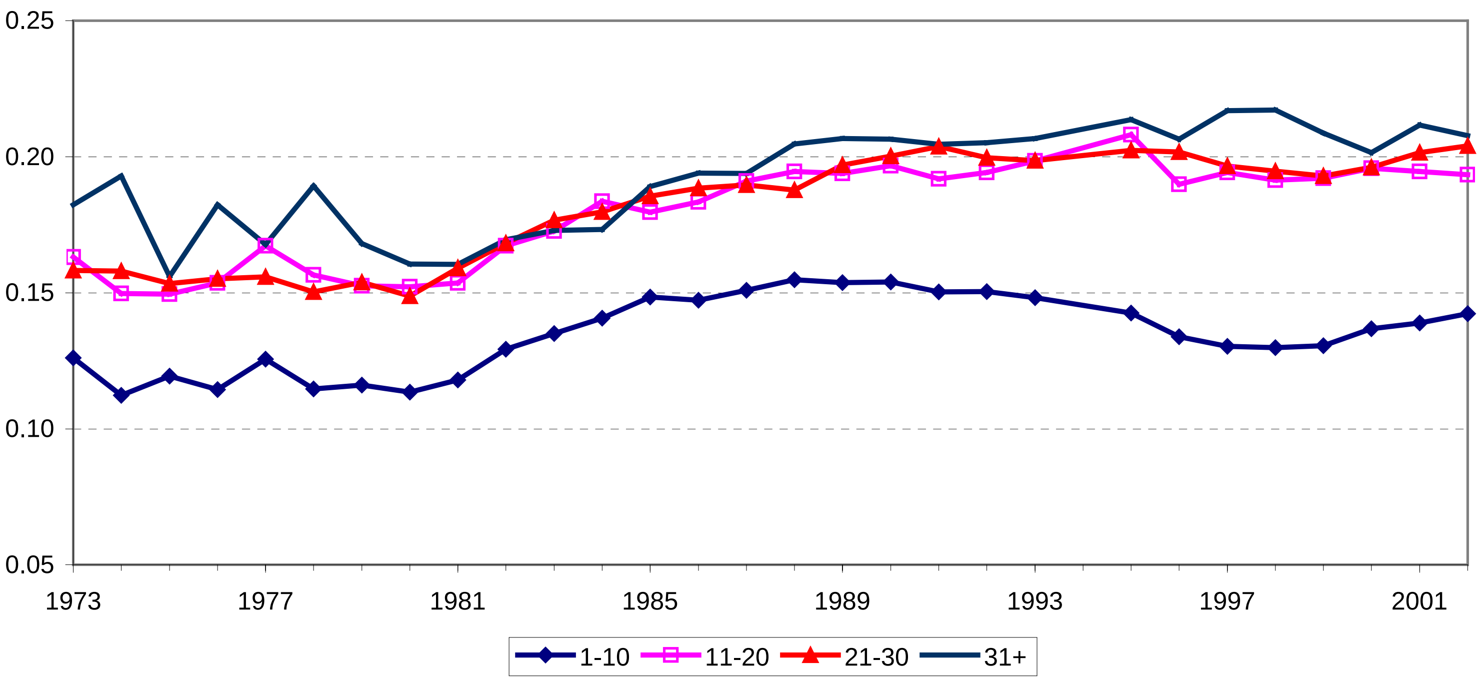


Figure 10a: Within-group variance for men, holding distribution of skills at their 1973 level

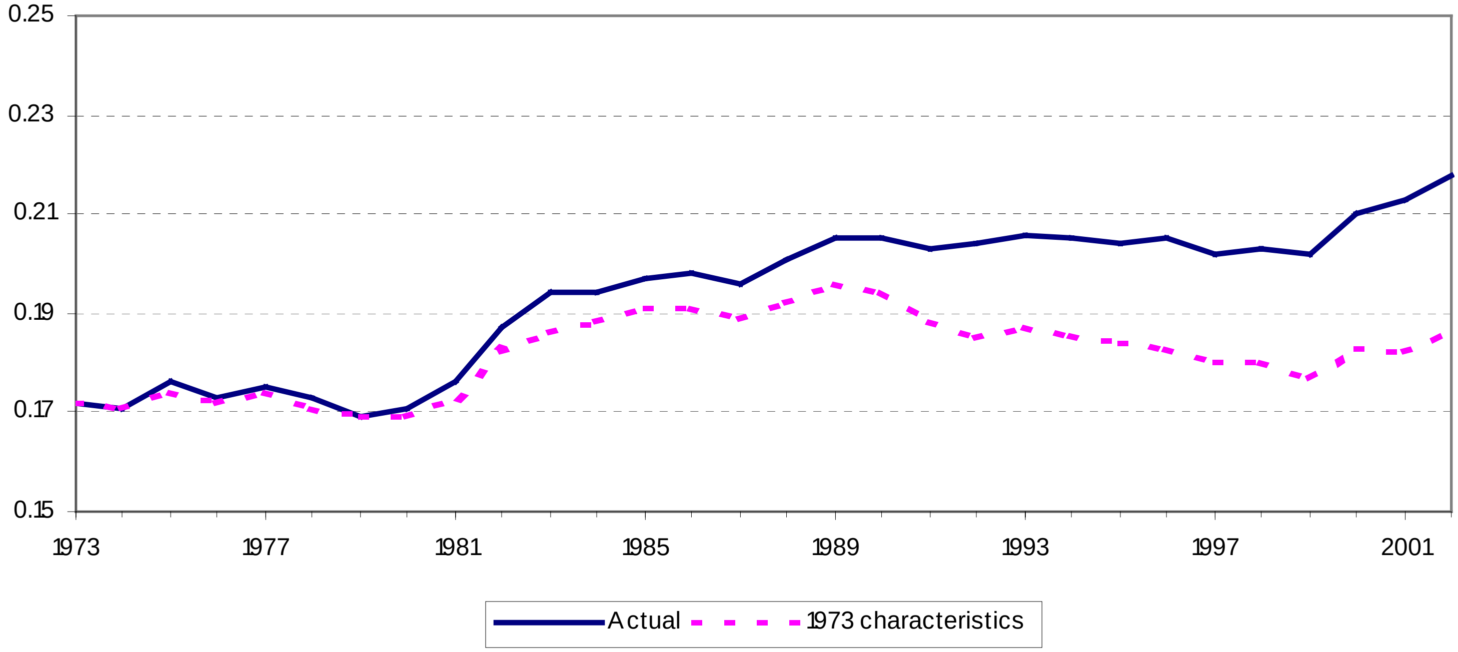


Figure 10b: Within-group variance for men, holding distribution of skills at their 2002 level

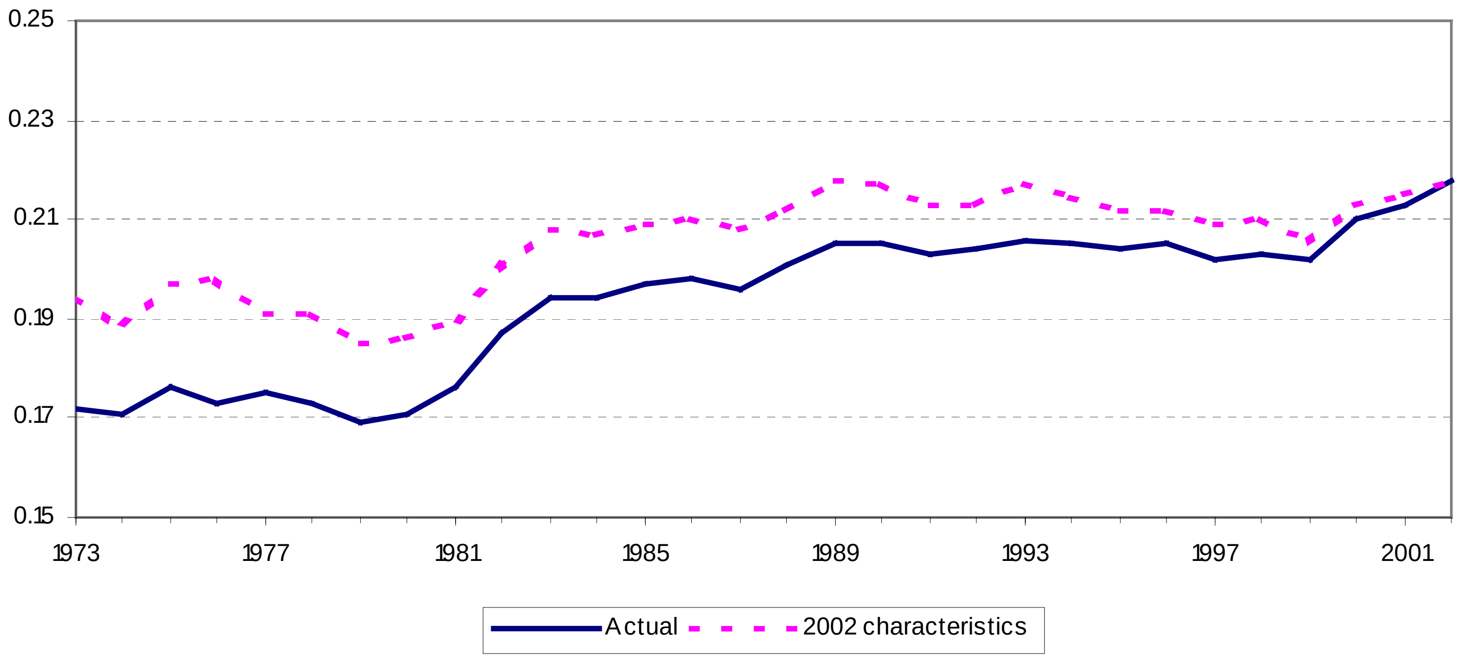


Figure 10c: Within-group variance for women, holding distribution of skills at their 1973 level

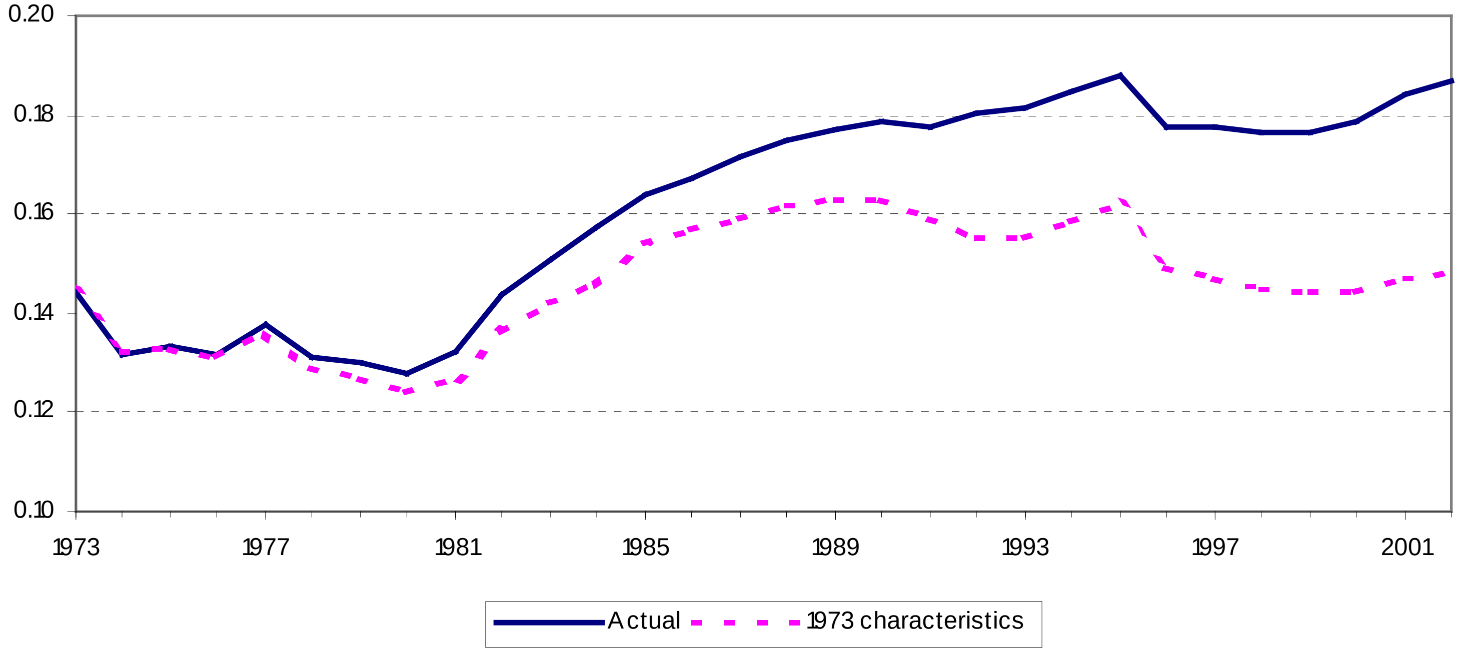
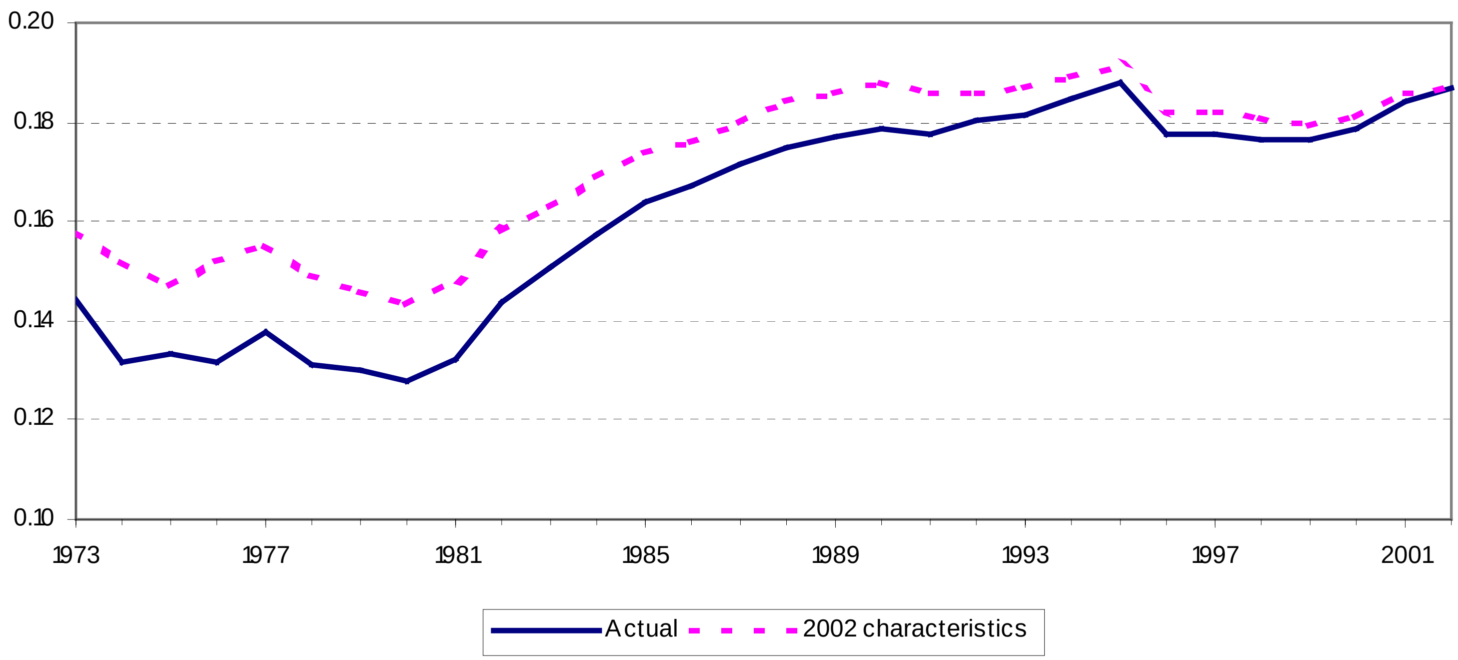
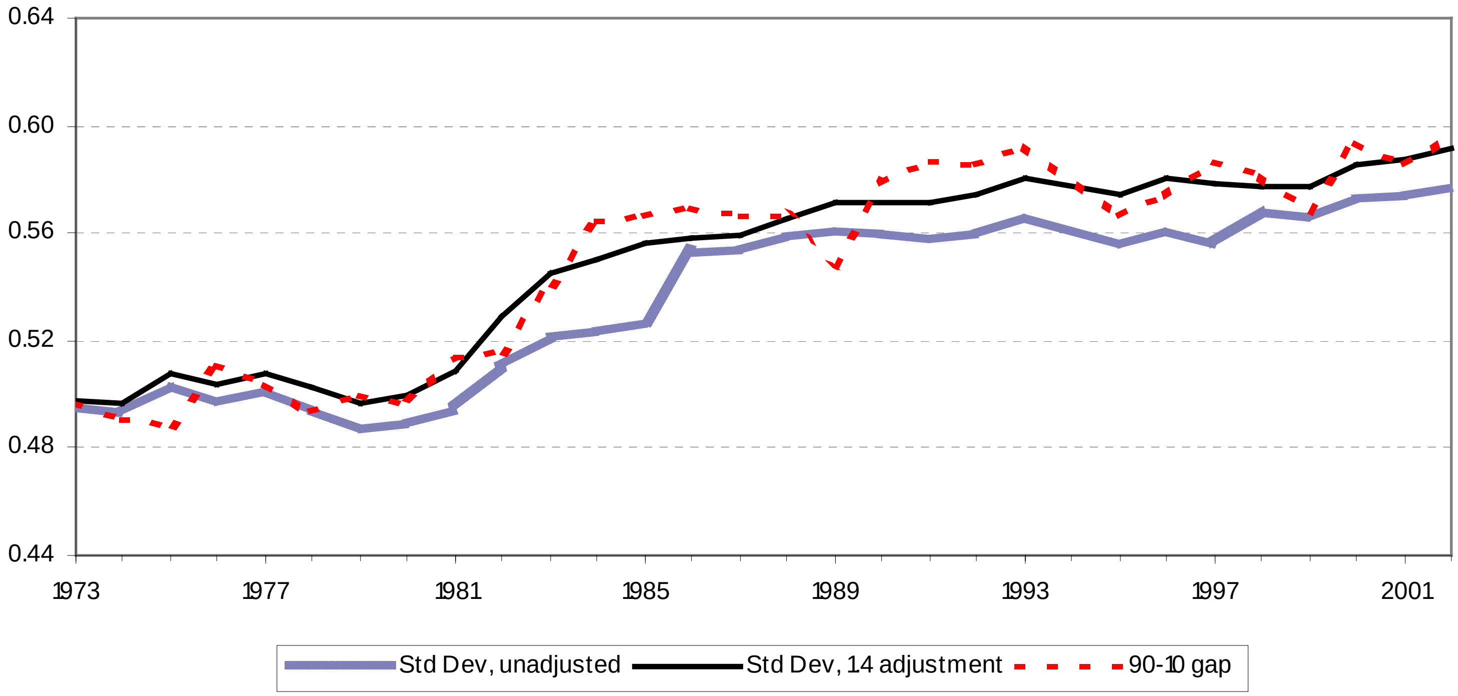


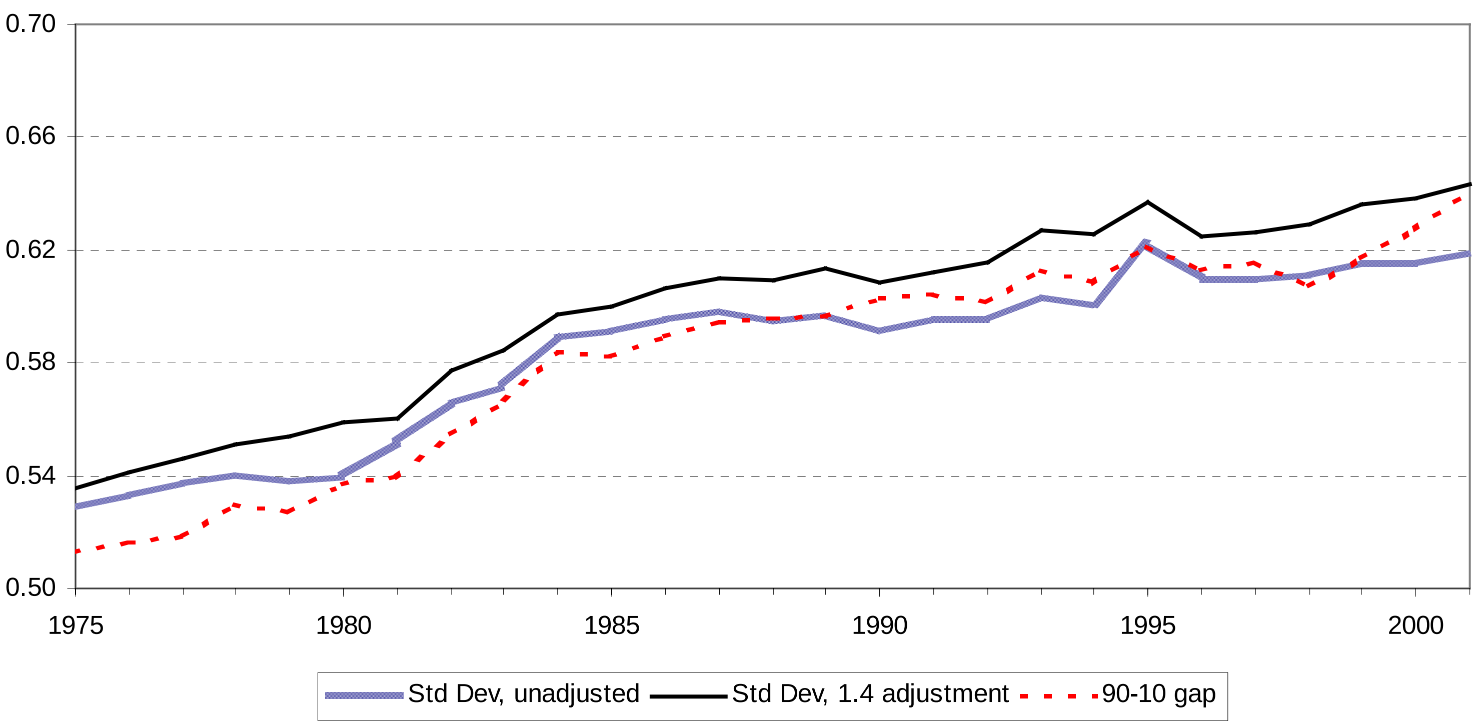
Figure 10d: Within-group variance for women, holding distribution of skills at their 2002 level



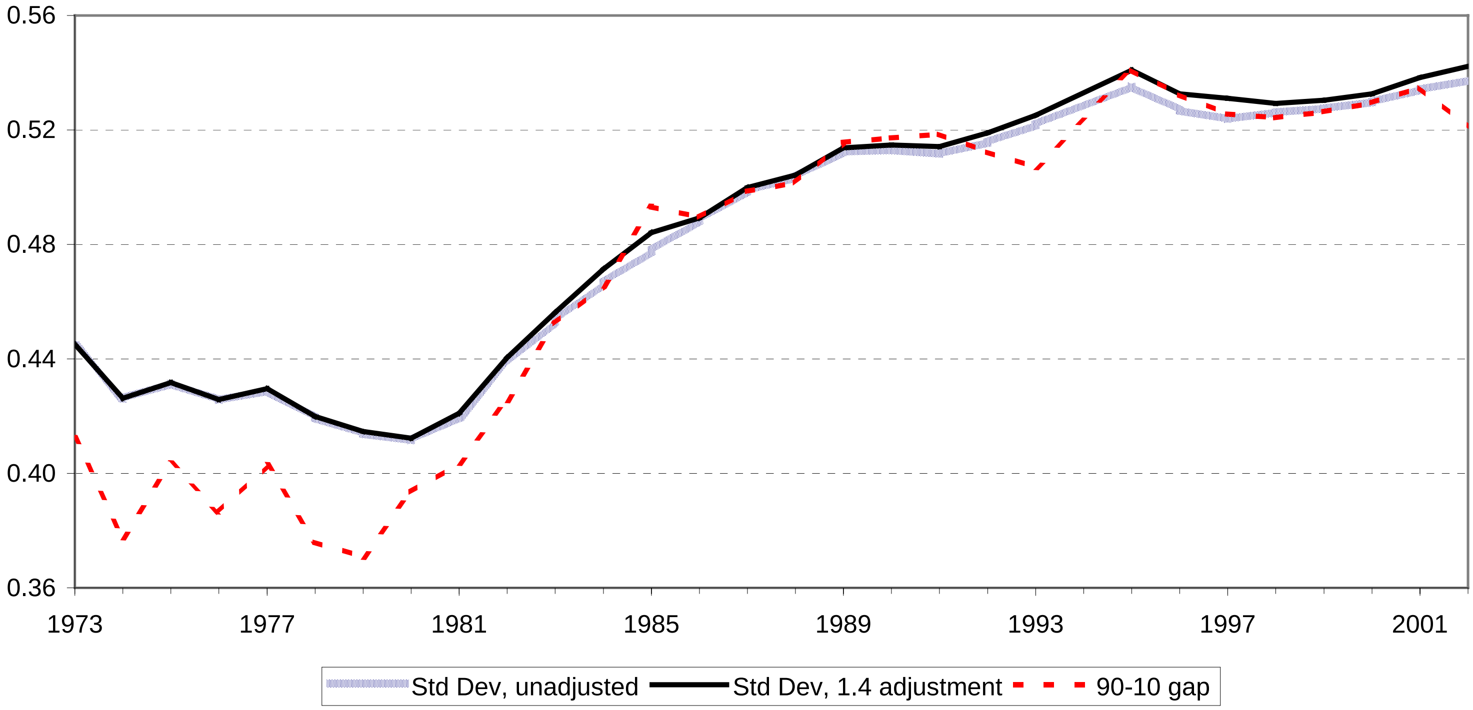
Appendix Figure A1: Top-Coding and Male Wage Dispersion, May/ORG CPS



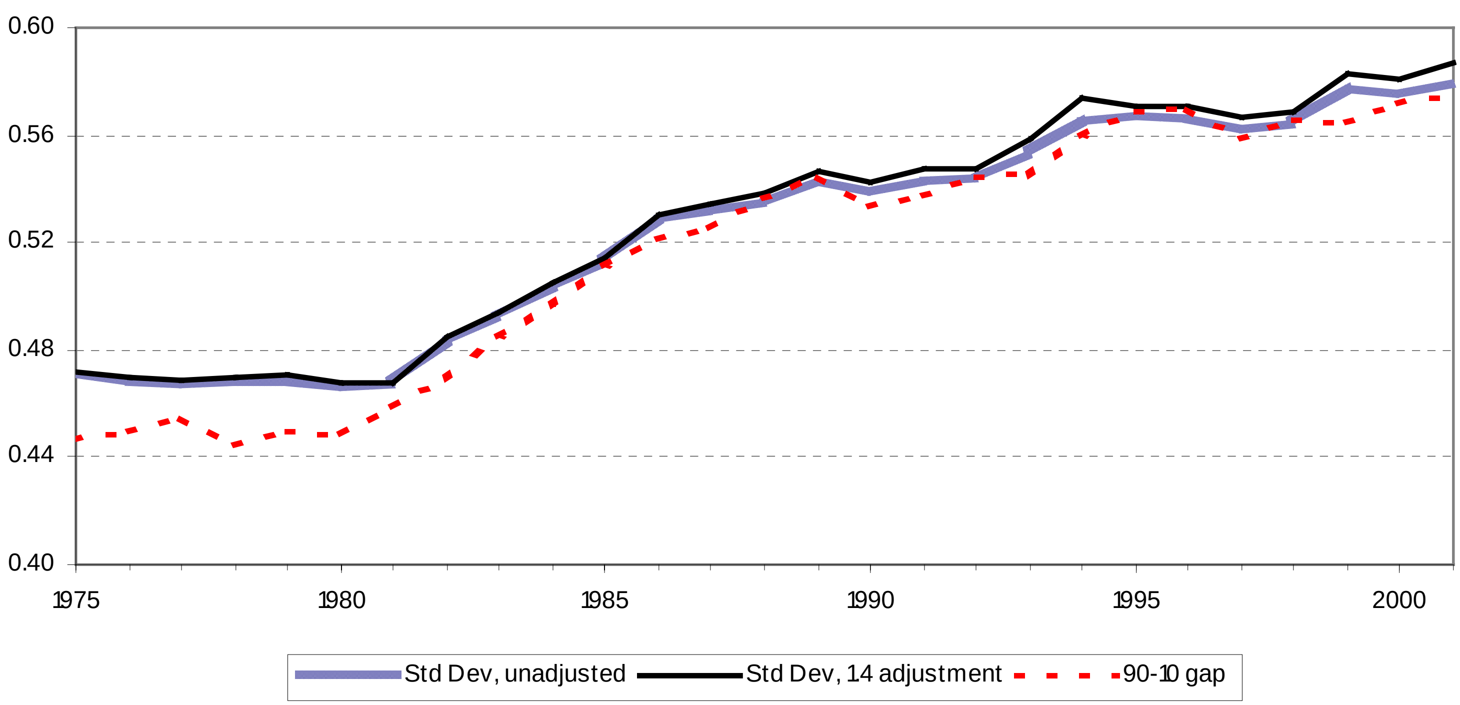
Appendix Figure A2: Top-Coding and Male Wage Dispersion, March CPS



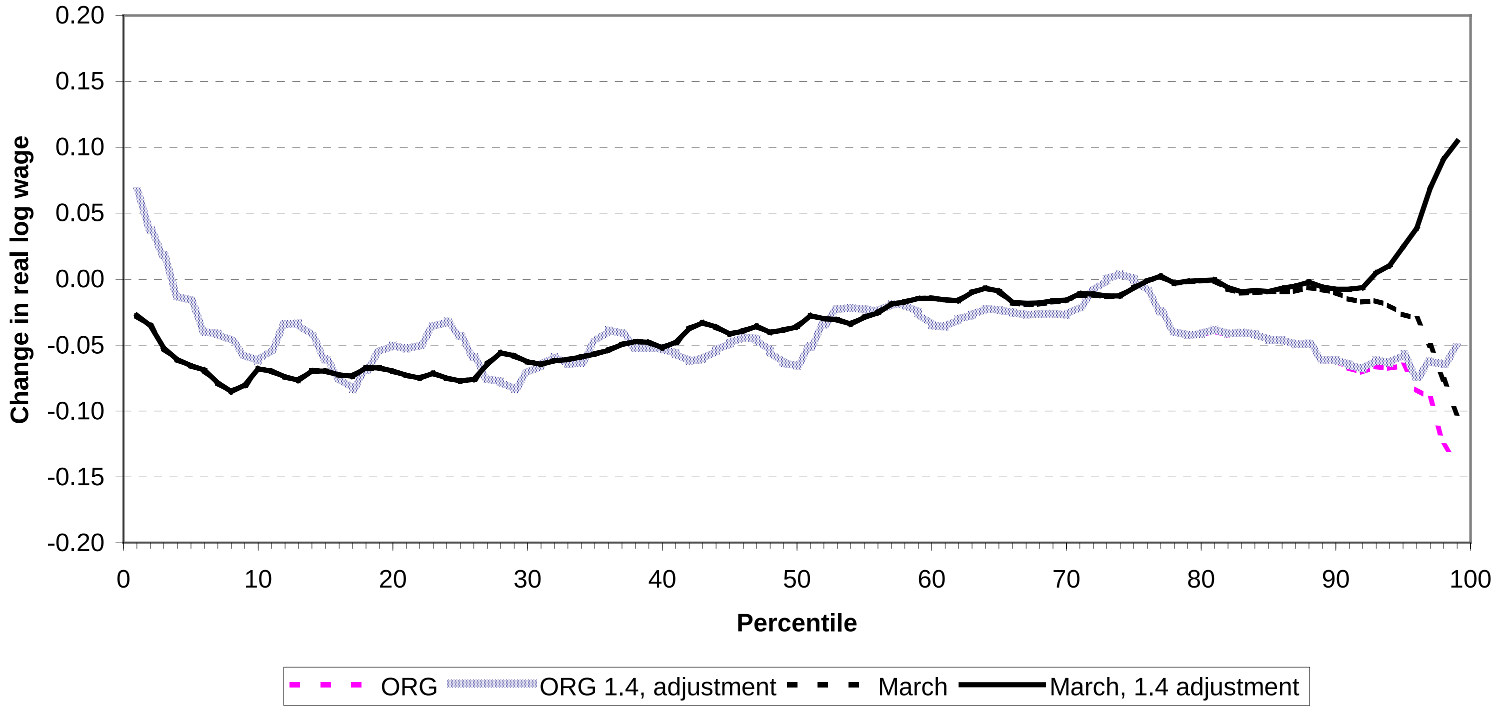
Appendix Figure A3: Top-Coding and Female Wage Dispersion, May/ORG CPS



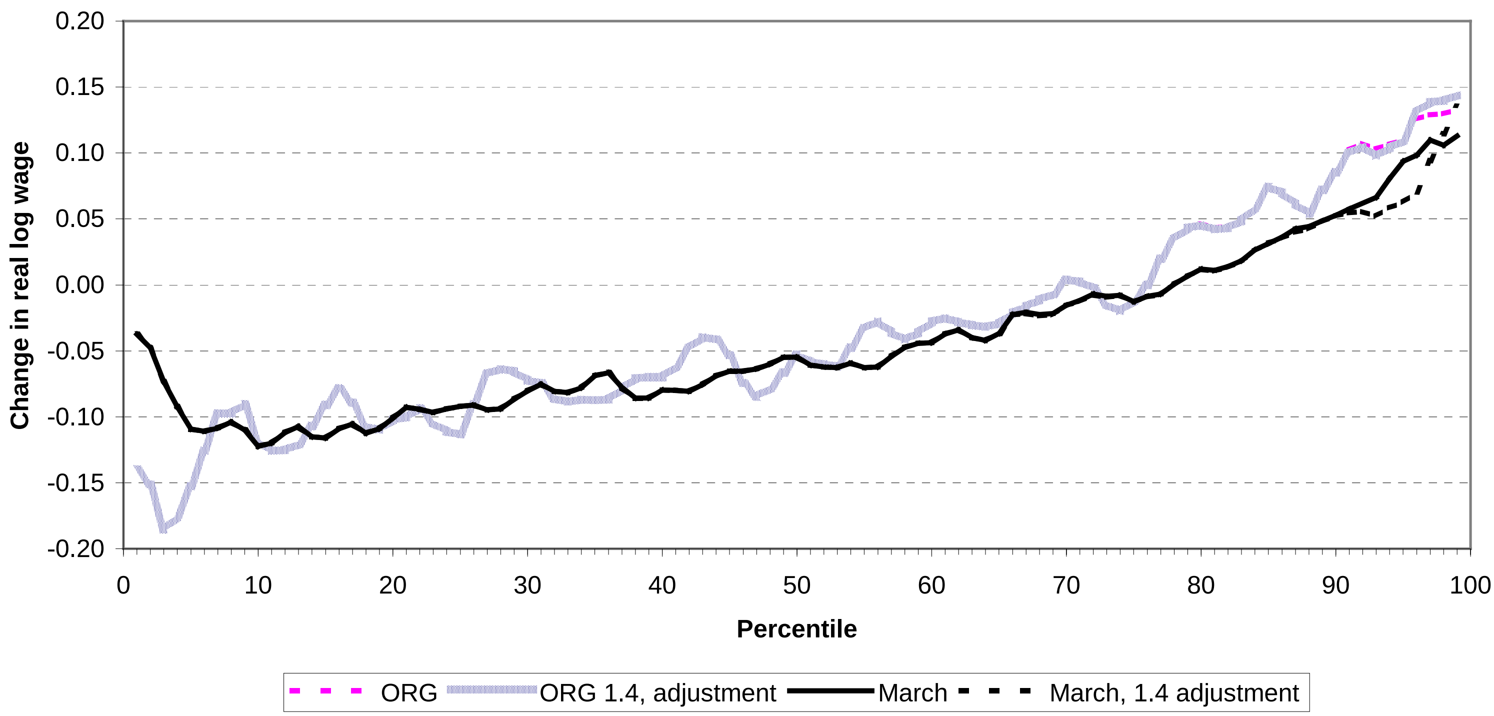
Appendix Figure A4: Top-Coding and Female Wage Dispersion, March CPS



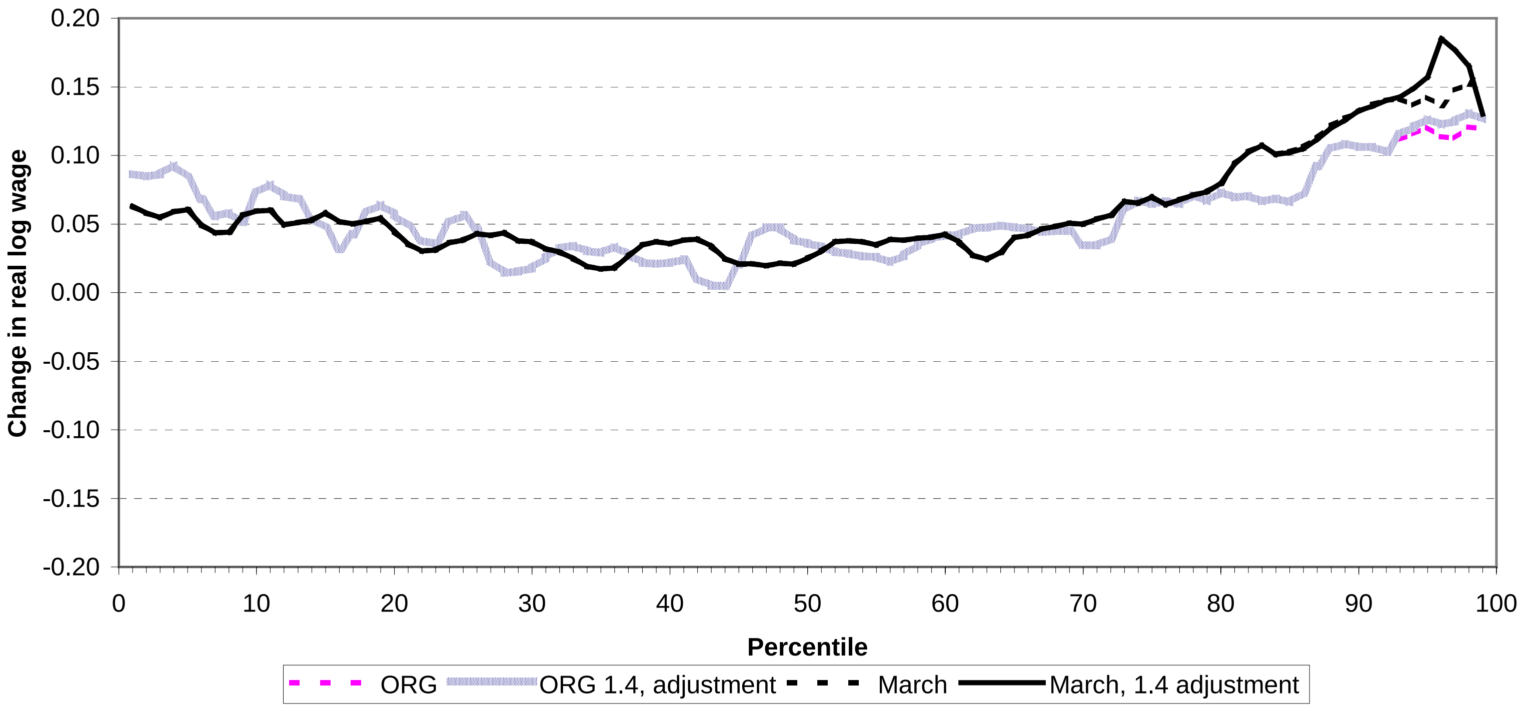
Appendix Figure A5: Change in Male Wages by Percentile, 1975-1980



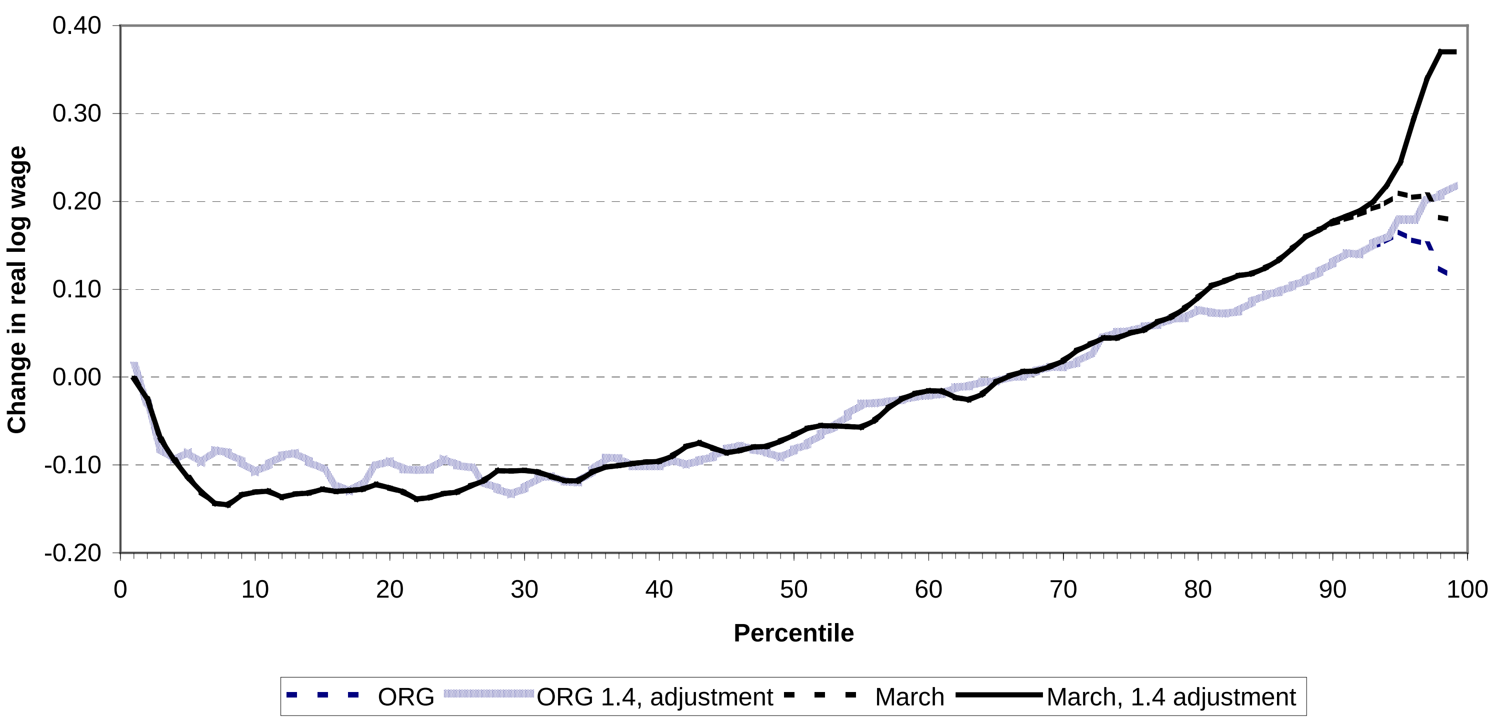
Appendix Figure A6: Change in Male Wages by Percentile, 1980-1990



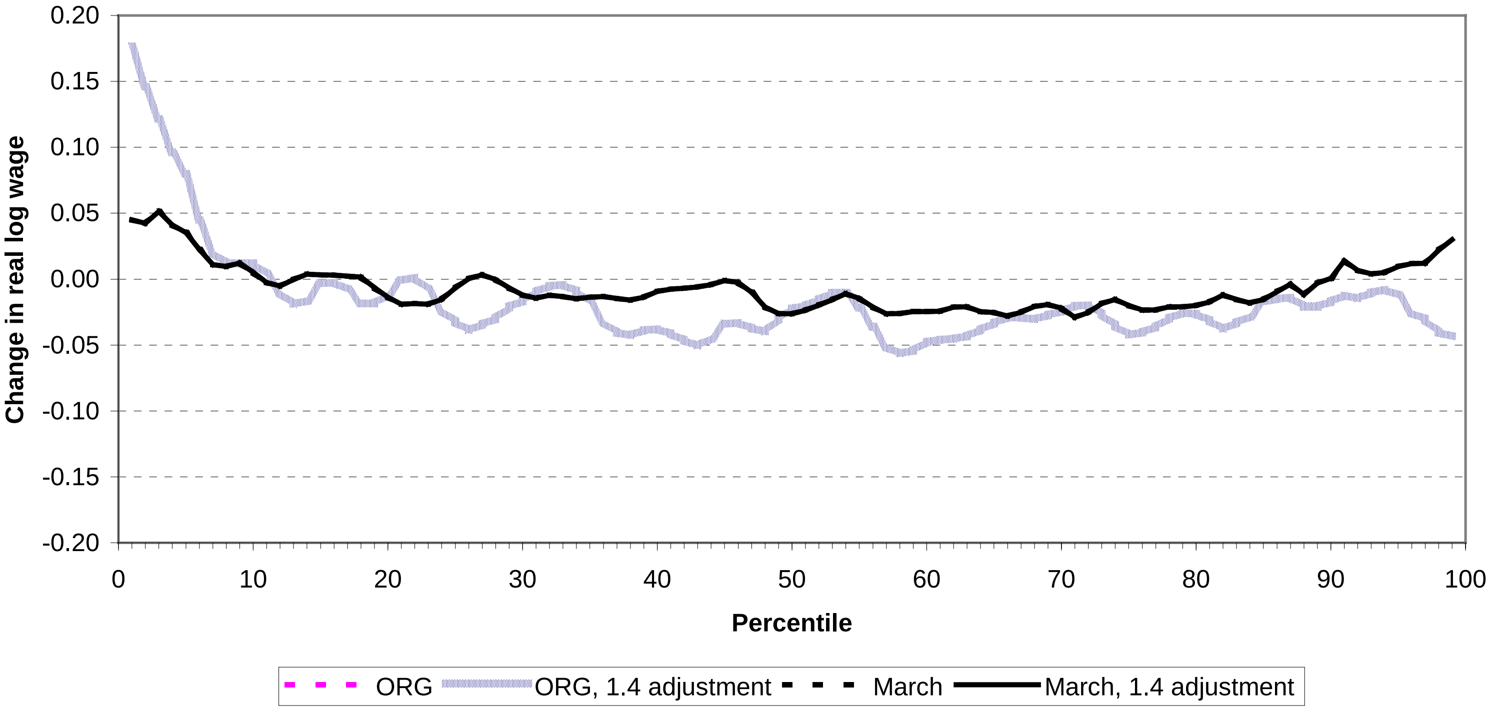
Appendix Figure A7: Change in Male Wages by Percentile, 1990-2000



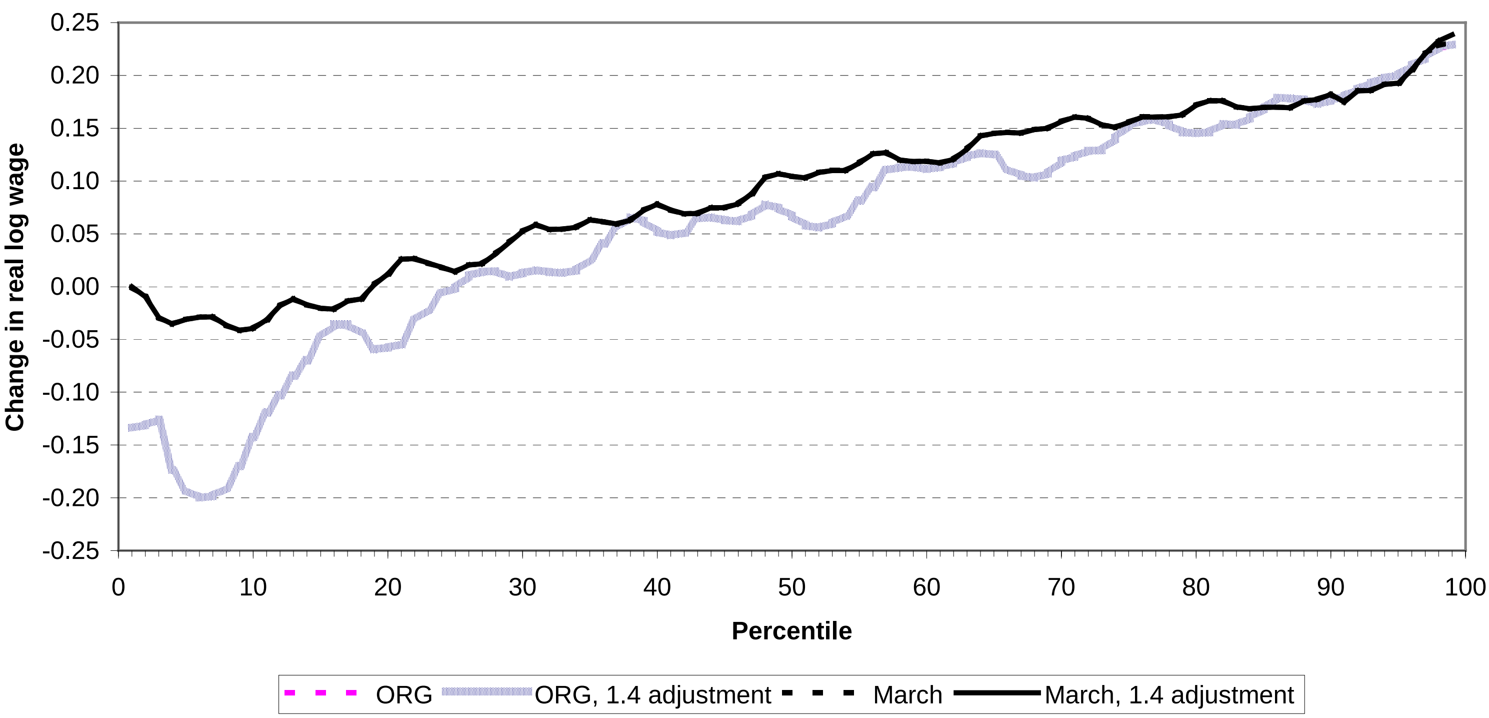
Appendix Figure A8: Change in Male Wages by Percentile, 1975-2000



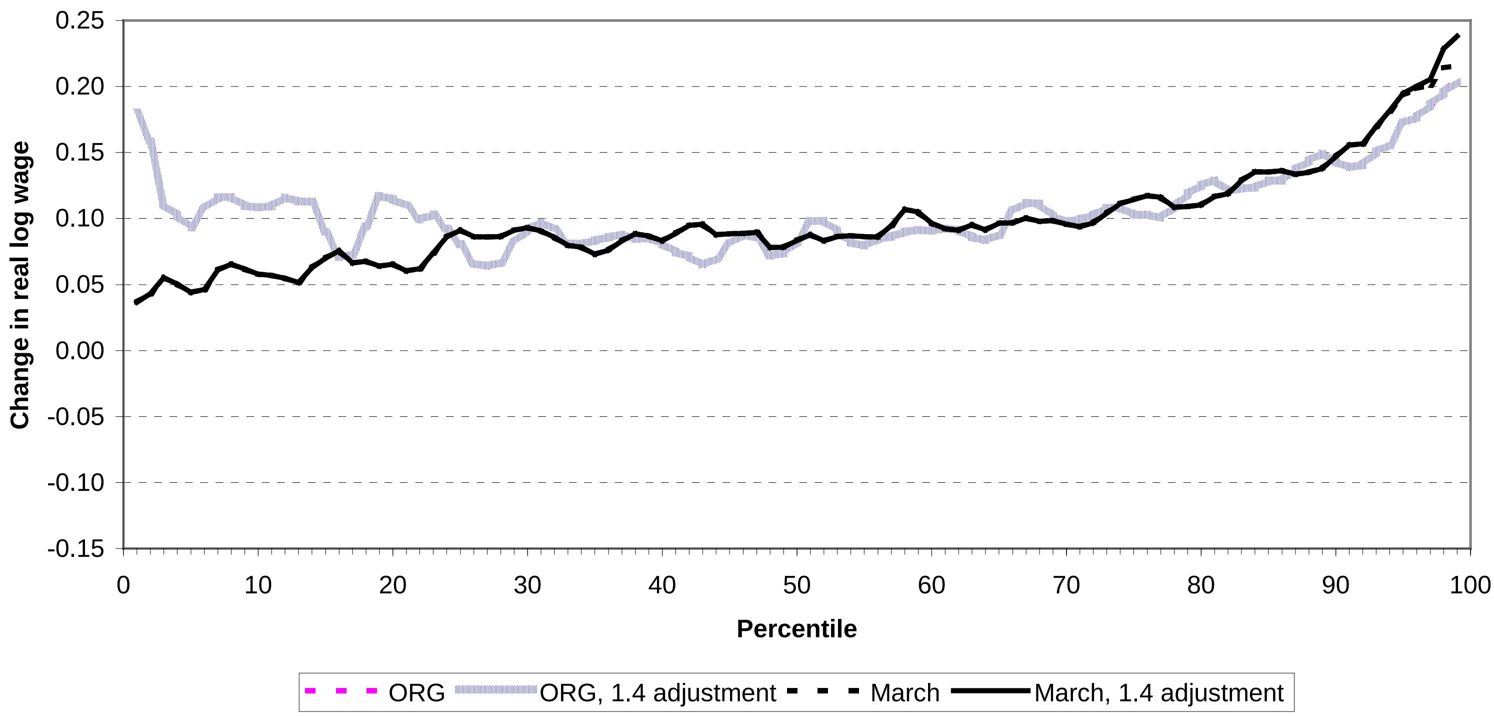
Appendix Figure A9: Change in Female Wages by Percentile, 1975-1980



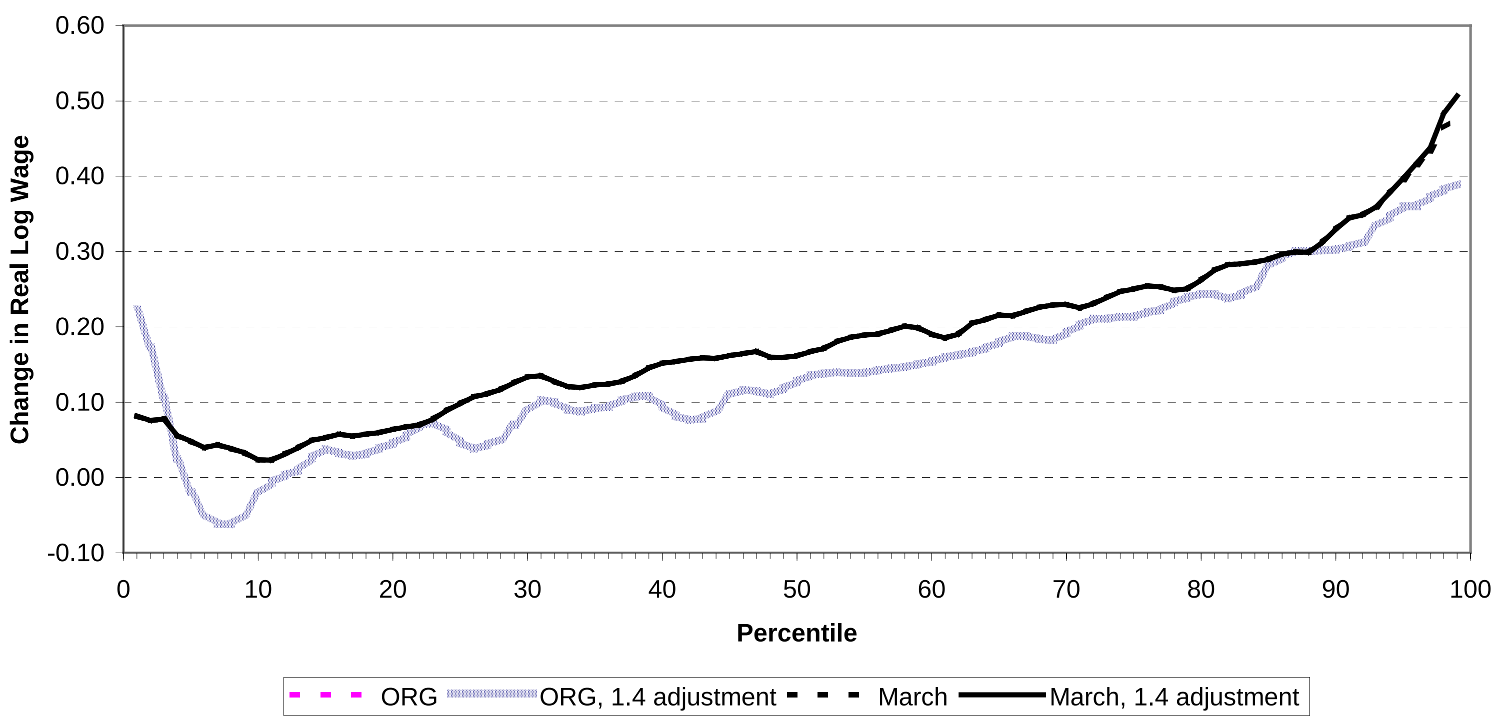
Appendix Figure A10: Change in Female Wages by Percentile, 1980-1990



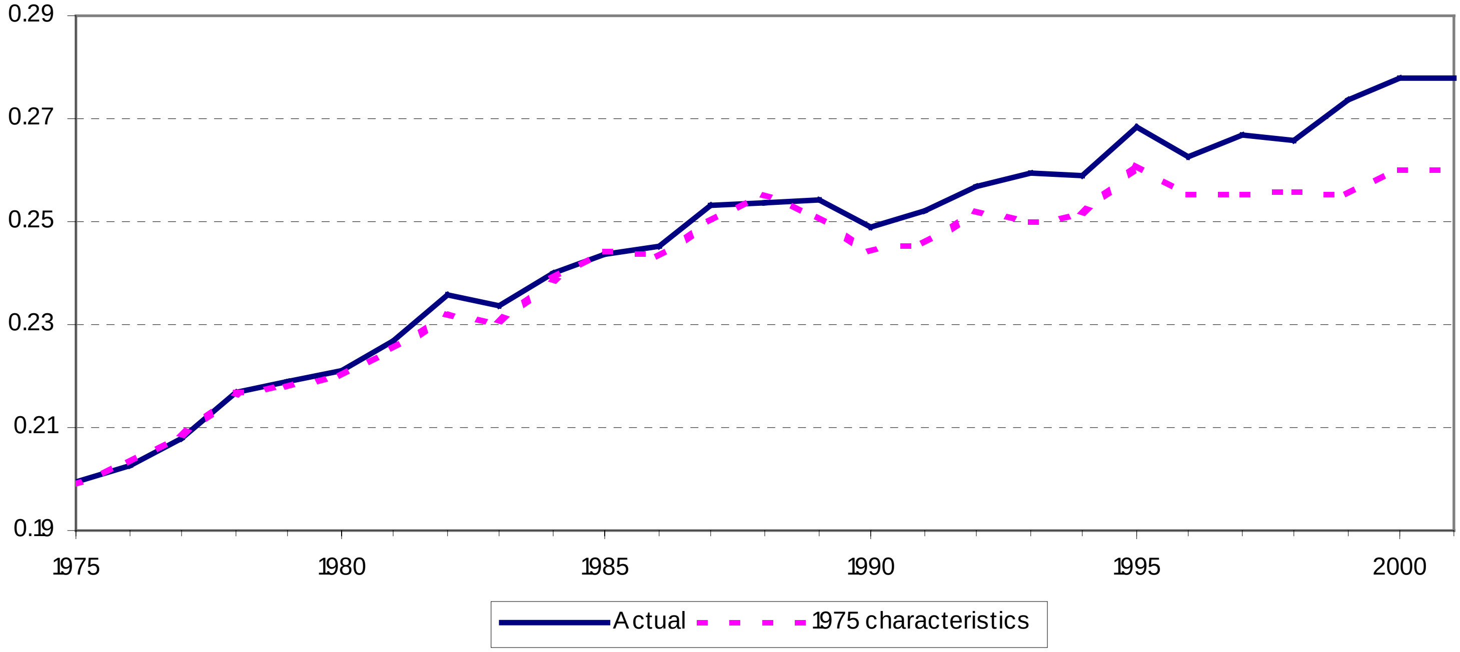
Appendix Figure A11: Change in Female Wages by Percentile, 1990-2000



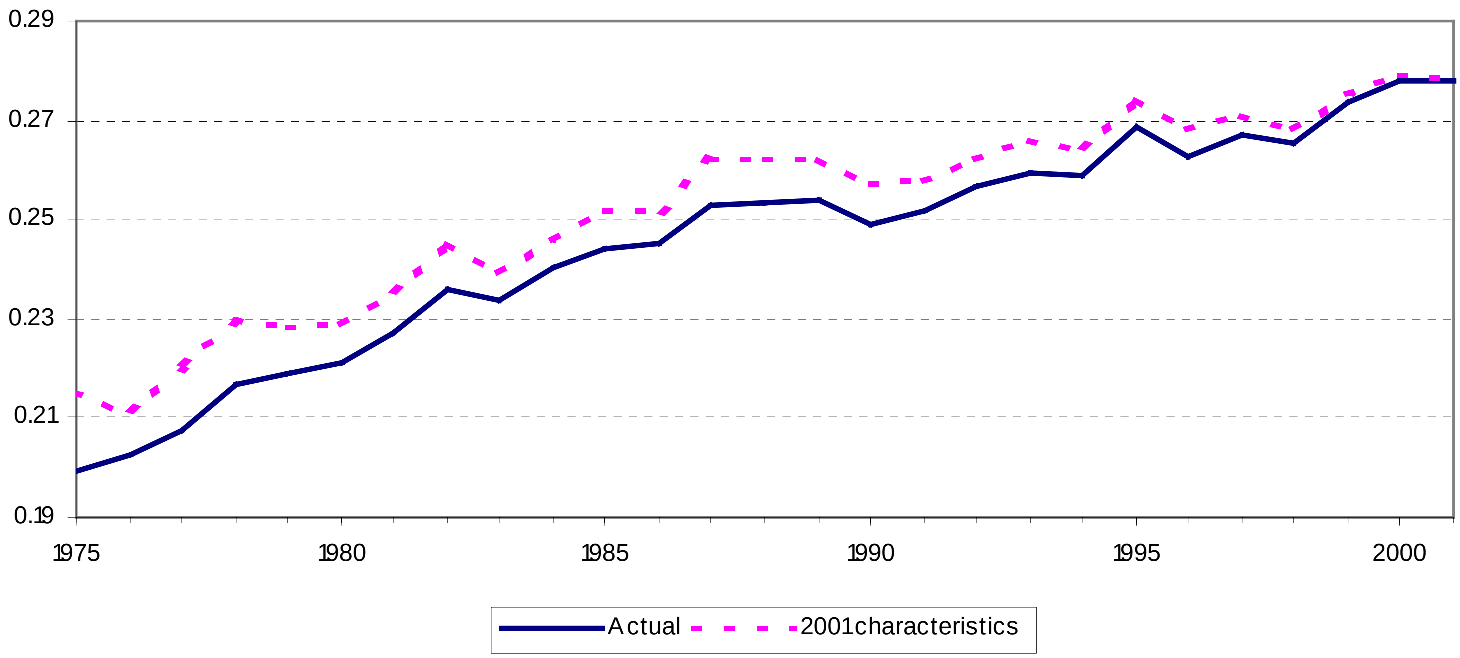
Appendix Figure A12: Change in Female Wages by Percentile, 1975-2000



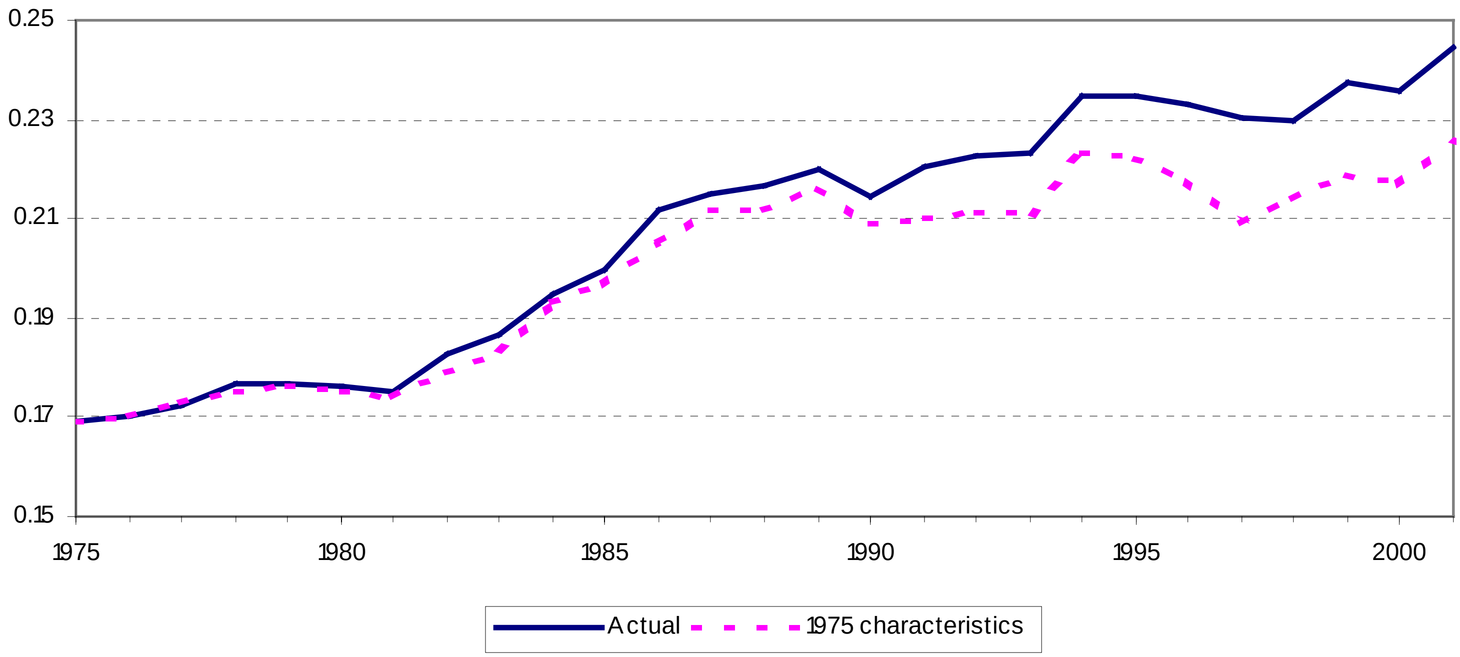
Appendix Figure B1: Within-group variance for men in the March CPS, holding distribution of skills at their 1975 level



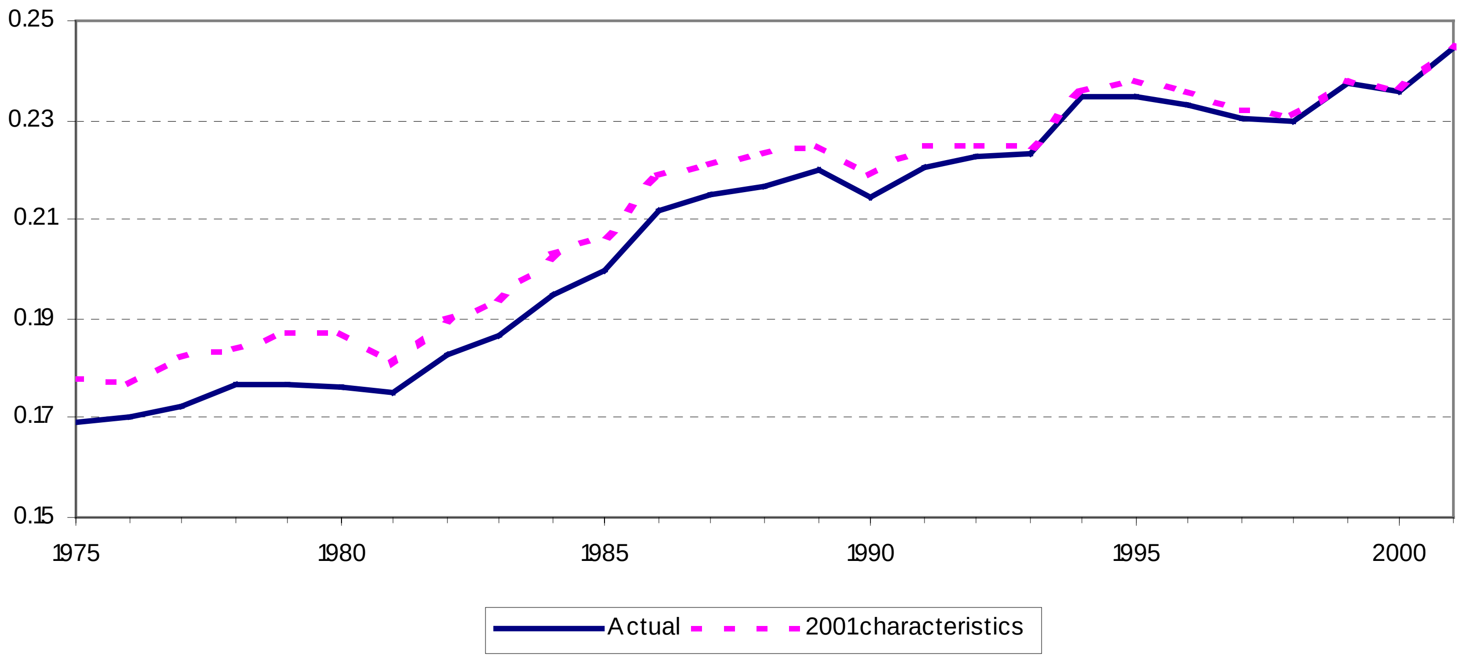
Appendix Figure B2: Within-group variance for men in the March CPS, holding distribution of skills at their 2001 level



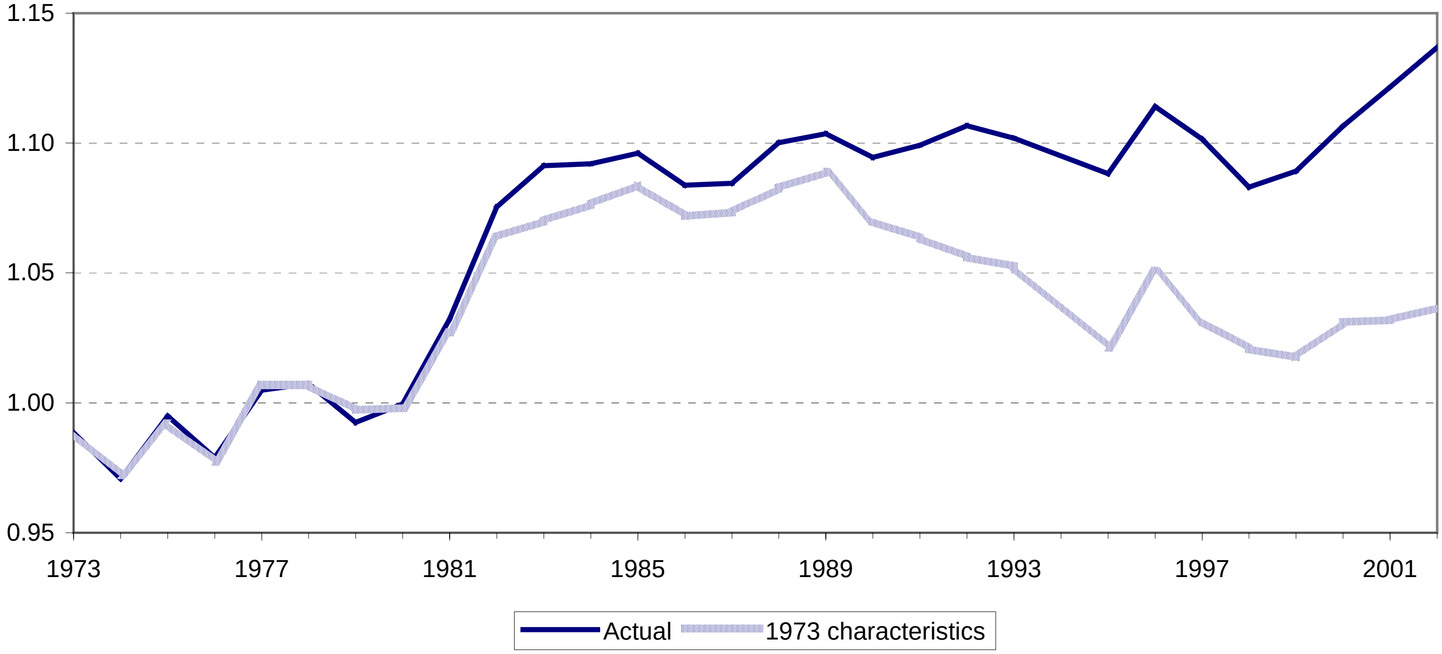
Appendix Figure B3: Within-group variance for women in the March CPS, holding distribution of skills at their 1975 level



Appendix Figure B4: Within-group variance for women in the March CPS, holding distribution of skills at their 2001 level



Appendix Figure C1: Within-group 90-10 dispersion for men, holding distribution of skills at their 1973 level



Appendix Figure C2: Within-group 90-10 dispersion for women, holding distribution of skills at their 1973 level