Residual Wage Inequality: A Re-examination*

Thomas Lemieux
University of British Columbia

Revised, February 2004

Abstract

This paper shows that a large fraction of the increase in residual wage inequality between 1973 and 2003 is a spurious 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, I show that *all* of the growth in residual wage inequality occurred during the 1980s once composition effects are accounted for. These findings are hard to reconcile with the skill-biased technical-change hypothesis. By contrast, changes in the real value of the minimum wage are closely related to changes in residual wage inequality.

I also discuss why the findings of the paper differ from those of earlier studies. One difference is that I control explicitly for composition effects. Another difference is that I use the May and Outgoing Rotation Group (ORG) supplements of the CPS, while most other studies have used the March CPS. I show that the May/ORG provides more reliable measures of hourly wages because it measures directly the wage of workers paid by the hour. Because of this problem, both the level and the growth in residual wage inequality are overstated in the March CPS.

*: I would like to thank SSHRC (INE CRI grant to Team for Advanced Research on Globalization, Education, and Technology) and NICHD (grant no. R01 HD39921-01) for financial support, and seminar participants at the NBER Summer Institute, Columbia, Northwestern, UC Davis and the University of Calgary for useful comments.

1. Introduction

The growth in wage inequality over the last three decades is one of the most extensively researched topics in labor economics. It is well documented that the growth in the college-high school wage premium since the late 1970s is an important factor in the overall growth in wage inequality (Bound and Johnson, 1992, Katz and Murphy, 1992). However, residual or within-group wage inequality –i.e. wage dispersion among workers with the same education and experience– appears to account for most of the growth in overall wage inequality (Juhn, Murphy and Pierce, 1993, JMP hereinafter).[1] According to JMP, residual wage inequality increased steadily since the early 1970s. Recent survey pieces by Acemoglu (2002) and Katz and Autor (1999) confirm those findings and argue that residual inequality kept growing throughout the 1990s, while Card and DiNardo (2002) reach somehow different conclusions.

JMP argue that the growth in residual wage inequality primarily reflects an increase in the price of unobserved skills that are not captured by standard regressors like experience and education. They use this argument to conclude that the growth in residual wage inequality is primarily a consequence of a large and steady growth in the relative demand for skills that started in the early 1970s. Most subsequent studies have argued that skill-biased technical change (SBTC) was the main factor responsible for the steady growth in skill prices. In particular, the computer and information technology revolution has emerged as the leading hypothesis for explaining the growth in the relative demand for skills since the early 1970s (Berman, Bound and Griliches, 1994, Autor, Katz and Krueger, 1998). More generally, JMP's (1993) interpretation of growing residual inequality as an increase in the skill premium has laid the foundations for a large and influential literature on economic growth, technical change, and inequality (see Acemoglu, 2002 and Aghion, 2002 for recent surveys of this literature).

The critical assumption required for interpreting the growth in residual wage inequality as a consequence of growing skill prices is that the distribution of unobserved skills remains constant over time. Otherwise, growing residual wage inequality could

[1] Put differently, dispersion in the residuals from a flexible wage regression grew more than the systematic component of wages predicted by the regression. 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).

simply reflect increasing inequality in the distribution of unobserved skills. There are at least two reasons why this assumption may not hold. First, more recent cohorts may have more unequally distributed skills and abilities than older cohorts for a given level of experience and education. JMP argue, however, that there is little evidence of such "cohort effects" in the distribution of unobserved skills.

Another potential source of change in the distribution of unobserved skills among all workers is changes in the distribution of experience and education. For instance, it is well recognized since Mincer (1974) that wage dispersion increases as a function of experience (past the overtaking point) because of differential investments in on-the-job training (OJT). In other words, inequality in the distribution of unobserved skills (OJT) increases with experience. Unless the experience distribution of the workforce remains constant, observed growth in residual inequality could simply be a spurious consequence of changes in distribution of experience. A similar argument can be made about education since wages are more unequally distributed among more- than less-educated workers (Mincer, 1974).

The main finding of this paper is that a large fraction of the growth in residual wage inequality between 1973 and 2003 is indeed a spurious consequence of changes in the distribution of experience and education in the workforce. Since the early 1980s, the U.S. workforce has grown more educated and experienced, two factors strongly associated with more within-group wage dispersion. Most of the growth in residual wage dispersion is due to the increasing weight put on groups of workers with more dispersed wages, as opposed to growth in wage dispersion within these narrowly defined groups of workers.

This finding suggests that changes in the price of unobserved skills only played a modest role in changes in overall wage inequality since 1973. Furthermore, after controlling for composition effects, most of the growth in residual wage inequality is concentrated in the 1980s, and in particular in the first half of the 1980s (for men). This suggests that the relative demand for skills did not grow steadily over time. Most of the growth in relative demand is rather concentrated in the first part of the 1980s, which is difficult to reconcile with the SBTC hypothesis.

The conclusions of the paper are at odds with most of the previous literature. I show that this discrepancy is due to a combination of several factors. First, I use data on hourly wages from the May and Outgoing Rotation Group (ORG) supplements of the Current Population Survey (CPS), while earlier studies have typically used the March Supplement of the CPS. In Section 5, I explain why the May/ORG CPS is better suited than the March CPS for studying the evolution of within-group wage dispersion. The main problem with the March CPS is that it poorly measures the wages of workers paid by the hour (the majority of the workforce). The fraction of workers paid by the hour grows substantially over time within narrowly defined group of workers, which results in spurious growth in within-group wage dispersion in the March CPS.

Another difference with earlier studies is that much more data are now available for studying secular changes in residual wage inequality. For example, I use wage data for up to 2003 while the last year of wage data available to JMP was 1989. Composition effects play a much bigger role in changes in residual inequality in the 1990s and early 2000s than in the 1970s and 1980s. This may explain why composition effects remained relatively unnoticed in the earlier literature.

One final difference is that most other studies simply do not control for composition effects.[2] Two earlier studies that implicitly control for composition effects in residual inequality are DiNardo, Fortin, and Lemieux (1996) and JMP. Using a re-weighting procedure that I also use here, DiNardo, Fortin, and Lemieux (1996) show that a third of the growth in residual inequality between 1979 and 1988 is due to composition effects. This result is very similar to what I find for the same time period. The decomposition procedure of JMP should also, in principle, adjust for composition effects. JMP only report, however, the joint effect of changes in the composition of the workforce on the between- and within-group components of wage inequality. The separate impact of composition effects on residual wage inequality is not separately reported in their study.

The paper is organized as follows. In Section 2, I discuss in more detail the link between residual wage inequality, unobserved skill prices, and composition effects. I

---

[2] For example, recent surveys by Acemoglu (2002) and Katz and Autor (1999) update Juhn, Murphy, and Pierce's (1993) trends in residual inequality using wages for full-time males in the March CPS up to the mid-1990s. They do not control, however, for composition effects.

present the May/ORG CPS data and show basic trends in within-group inequality for twenty experience-education groups in Section 3. Section 4 presents the main results on the evolution of residual wage inequality once composition effects are adjusted for. Section 5 shows why both the level and the growth in residual wage inequality are overstated in the March CPS. Section 6 concludes by suggesting that the SBTC hypothesis is not very useful for understanding the trends in residual wage inequality since the early 1970s. A more promising explanation is based on changes in the real value of the minimum wage, though other factors need to be introduced to explain the large increase in inequality at the top end of the wage distribution (Piketty and Saez, 2003).

2. Residual inequality, skill prices, and composition effects: conceptual framework

From an economic perspective, residual wage inequality is an important concept because of its potential connection to skill prices. In this Section, I explain why this connection is only valid when composition effects are controlled for. I also discuss the re-weighting approach that I later use to control for composition effects.

To fix ideas, consider a standard Mincer-type wage equation:

$$(1) \qquad \ln w_{it} = x_{it} b_t + \varepsilon_{it},$$

where w_{it} is the hourly wage rate of individual *i* at time *t*; x_{it} is a vector of observed skills (education and labor market experience); b_t is the return (or price) of observed skills; ε_{it} is the standard regression residual. Following JMP, the residual is interpreted as the product of some are unobserved skills, e_{it}, with the return to unobserved skills, p_t:

$$(2) \qquad \varepsilon_{it} = p_t \, e_{it}.$$

Equation (2) has no empirical content unless some assumptions are imposed on the distribution of unobserved skills. Following JMP and Chay and Lee (2000), I assume that the distribution of unobserved skills among workers with the same level of experience and education is stable over time. In other words, the conditional distribution function $F_t(e_{it}|x_{it})$ is time-invariant:

(3) $F_t(e_{it}|x_{it}) = F(e_{it}|x_{it})$ for all time periods t.

As mentioned in the Introduction, JMP discuss some possible objections to this assumption. In particular, more recent cohorts may have more unequally distributed skills and abilities than older cohorts at the same level of experience and education. This could be the result, for instance, of growing inequality in school quality. JMP test this hypothesis by decomposing the growth in within-group wage dispersion into cohort and time effects. They conclude that time effects, as opposed to cohort effects, appear to be driving the growth in within-group inequality. This finding is consistent with the assumption that the conditional distribution of unobserved skills is constant across time (and cohorts).

By contrast, the stronger assumption that the unconditional distribution of unobserved skills, $F_t(e_{it})$, is also stable over time is clearly inconsistent with a large body of research in labor economics. In particular, it is well recognized since Mincer (1974) that wage dispersion increases as a function of experience (past the overtaking point) because of differential investments in on-the-job training (OJT). In other words, inequality in the distribution of unobserved skills (OJT) increases with experience. Similarly, Farber and Gibbons (1996) show that inequality in wages and unobserved skills (as valued by the market) also increase as a function of experience in a simple learning model.[3] In both of these models, the aging of the workforce results in a more dispersed unconditional distribution of unobserved skills as increasingly more weight is put on older workers with more unequally distributed unobserved skills. This can result in significant composition effects in the 1980s, 1990s, and 2000s because of the aging of the baby-boom generation. A similar argument can be made in the case of education.

[3] In Farber and Gibbons (1996) model, wages are equal to the expected value of productivity given the available information about the past productivity of workers. There is little wage inequality among inexperienced workers since the market does not yet know who is productive and who is not. Inequality increases as a function of experience as the market learns who is productive (skilled) and who is not. From the point of view of the econometrician, inequality in unobserved skills (what is valued by the market) thus grows as a function of labor market experience.

For example, Mincer (1997) shows that the within-group variance of wages increases as a function of education in a standard Becker (1967) human capital model with heterogeneity in the marginal costs and benefits of investments in education.

The role of composition effects is easily illustrated by looking at the variance of wages when observed skills, x_{it}, are divided into a finite number of cells *j*. The unconditional variance of unobserved skills, $Var(e_{it})$, is linked to the conditional variance, $Var(e_{it} \mid j)$, by the simple variance decomposition

$$(4) \qquad Var(e_{it}) = \sum_j \theta_{jt} Var(e_{it} \mid j),$$

where θ_{jt} is the share of workers in experience-education group *j* at time *t*. Under the assumption that the conditional variances are stable over time, equation (4) can be rewritten as

$$(5) \qquad Var(e_{it}) = \sum_j \theta_{jt} \sigma_j^2,$$

where $Var(e_{it} \mid j) = \sigma_j^2$ for all t. Since the conditional variance σ_j^2 is different for different skill groups *j*, changes in the skill composition of the workforce (the θ_{jt}'s) will also change the unconditional variance of unobserved skills.

The residual variance of wages, $Var(\varepsilon_{it})$, is obtained by substituting equation (5) into the "skill pricing" equation (2)

$$(6) \qquad Var(\varepsilon_{it}) = Var(p_t e_{it}) = p_t^2 \sum_j \theta_{jt} \sigma_j^2.$$

In this model, an increase in the residual variance can only be interpreted as an increase in skill prices p_t when the skill composition of the workforce (the θ_{jt}'s) is held constant. Fortunately, equation (6) suggests a straightforward way of holding the skill composition of the workforce constant. The residual variance just has to be recomputed at some counterfactual values of the shares, θ_j^*, that remain constant over time.

To see this, rewrite the residual variance as a function of V_{tj}, the variance of wages within each skill group *j*

(7) $\quad \text{Var}(\varepsilon_{it}) = \sum_j \theta_{jt} V_{jt},$

where $V_{jt} = p_t^2 \sigma_j^2$. The counterfactual residual variance, V_t^*, is

(8) $\quad V_t^* = \Sigma_j \, \theta_j^* \, V_{jt}.$

When the number of skill groups is small relative to sample sizes, the within-group variance V_{jt} can be computed for each skill group *j*. It is then straightforward to estimate the counterfactual variance by replacing the year-specific shares, θ_{jt}, by some average or base year shares, θ_j^*.

As it turns out, Mincer (1974) computed such counterfactual variances in his famous 1974 book. After dividing the data in about one hundred experience-education cells, he shows that the variance of wages would have been substantially 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 over the last few decades.[4]

In Section 3, I present some basic trends in residual and within-group inequality by dividing data in a limited number of experience-education cells (twenty). Working with these coarse cells helps illustrate which specific factors are driving the overall changes in residual inequality. To see this, consider the following decomposition of the change in the residual variance between a base period *s* and an end period *t*:

(9)
$$\begin{aligned} V_t - V_s &= \Sigma_j \, (\theta_{jt} V_{jt} - \theta_{js} V_{js}) \\ &= \Sigma_j \, \theta_{js}(V_{jt} - V_{js}) + \Sigma_j \, (\theta_{jt} - \theta_{js}) V_{jt}. \end{aligned}$$

[4] 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 (1974). Mincer shows that the variance of log annual earnings in 1959 would have increased from 0.668 to 0.721 if workers at all experience levels had had the same level of education as younger workers (7-9 years of experience).

Equation (9) shows that the overall change in the residual variance can be decomposed into two terms. The first term on the right hand side of equation (9), $\Sigma_j\ \theta_{js}(V_{jt} - V_{js})$, is a weighted average of changes in the within-group variance for each group *j*. In terms of equation (8), this term represents the change in the counterfactual variance, V_t*, when the counterfactual weights, θ_j*, are set at the base period level ($\theta_j^* = \theta_{js}$).

The second term on the right hand side of equation (9), $\Sigma_j\ (\theta_{jt} - \theta_{js})V_{jt}$, captures the extent of composition effects. Composition effects result in a spurious growth in the residual variance when changes in the weights, $\theta_{jt} - \theta_{js}$, are positively correlated with the within-group variances, V_{st}.

When the number of cells is small enough, equation (9) suggests a simple approach for separating the role of rising skill prices from composition effects. Since $V_{jt} = p_t^2\sigma_j^2$, the most direct evidence on rising skill prices is that the within-group variances, V_{jt}, are also growing over time. This can be readily checked by comparing these variances in a base and an end period. Equation (9) then shows how these changes can be aggregated into a single factor, $\Sigma_j\ \theta_{js}(V_{jt} - V_{js})$.

From an estimation point of view, however, dividing the data in a limited number of coarse experience-education cells may be too restrictive. One alternative is to construct finer cells based on single years of education and experience. Unfortunately, cell sizes based on single years of age and education are too small (and sometimes empty) in most CPS samples (see footnote 7). 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.

Another major advantage of the re-weighting procedure over the variance decomposition in equation (9) is that, in addition to the variance, any other distributional statistics like the Gini coefficient or the 90-10 gap can also be computed in the re-weighted samples (DiNardo, Fortin and Lemieux, 1996). In Section 4, I report estimates of residual wage inequality based on the 90-10 gap in addition to the variance that I use throughout the paper.

In principle, JMP's "residual imputation procedure" could also be used to account for the role of composition effects in changes in residual wage dispersion. JMP essentially suggest replacing the empirical distribution of residuals for given values of experience and education by a counterfactual distribution for the same skill group (e.g. the empirical distribution of residuals for all years pooled together). As discussed in Lemieux (2002), however, it is unclear how this procedure can be implemented in practice unless the data are divided in large enough cells. By contrast, the re-weighting procedure is very easy to implement even in cases where cells are “too small”.

3. Data and Trends in Within-Group Inequality by Skill Groups.

In this Section, I briefly present the May/ORG CPS data and show the basic trends in within-group wage dispersion for twenty experience-education groups. I use equation (9) to illustrate which factors --rising skill prices or composition effects—are driving the growth in the residual variance.

a. Data

Data issues are discussed in more detail in Section 5 that compares the hourly wage measure constructed from the May/ORG and March CPS Supplements. I only briefly discuss here how the May and ORG supplements of the CPS are processed. 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 has direct implications for the hourly price of labor. By contrast, the impact of these factors on weekly or annual earnings also depends on the responsiveness of labor supply to changes in the hourly wage rate.

I use the Dual Jobs Supplement of the May CPS for 1973 to 1978. The May CPS asks questions about wages on the main job held during the survey week to all wage and salary workers. For workers paid by the hour, the May CPS asks workers directly about their hourly rate of pay. This is the hourly wage measure that I use for this group of workers (about sixty percent of the workforce). For the other workers, I compute an hourly wage rate by dividing weekly earnings by weekly hours of work. I use the same procedure for the 1979 to 1993 ORG supplements that ask the same wage questions as the May CPS. The wage questions in the 1994 to 2003 ORG supplements are similar except that workers not paid by the hour can choose the periodicity at which they report earnings. I compute their hourly wage rate by dividing earnings by hours over the corresponding time period.[5] 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.[6]

Unlike in the ORG and March supplements of the CPS, in the May CPS wages were not allocated for workers who refused to answer the wage questions. To be consistent, I only keep workers with non-allocated wages in the 1979-2003 ORG supplement. As a consequence, I have to drop observations for 1994 and the first eight months of 1995 in which the CPS did not flag workers with missing wages. Following most of the literature, I trim extreme values of wages (less than $1 and more than $100 in 1979 $), adjust top-coded earnings by a factor of 1.4, and weight wage observations by hours of work (in addition to the usual CPS weights). I also keep workers age 16 to 64 with positive potential experience.

All the measures of residual and within-group inequality used in this paper are computed from the residuals of a regression of log wages on an unrestricted set of dummies for age, years of schooling, as well as interactions between nine schooling dummies and a quartic in age.[7] Separate regressions are estimated for both men and women in each year.

[5] 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 in reporting 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. 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.

[6] 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. 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 to 2003 files are almost half as large (instead of a third as large) as the MORG files for these years. The March samples are also slightly larger than the May samples as they include an over sample for Hispanics.

[7] 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-

b. Basic trends in within-group variances

I first divide workers into twenty skill groups based on five education categories (high school dropouts, high school graduates, some college, college graduates, and college post-graduates) and four experience categories (1-10, 11-20, 21-30, and 31 years or more of potential experience). Table 1 and Figure 1 show the within-group variances for each experience-education group at the beginning and end of the sample period. Since the residuals are computed from a very flexible regression, the within-group variance (variance of residuals) for a given group is smaller than the variance of unadjusted wages for the same group. To improve the precision of the estimates, I pool years 1973 to 1975 for the base period, and years 2000 to 2002 for the end period.

Figure 1a (men) and 1b (women) plot the within-group variances in 2000-02 against the 1973-75 variances. In these figures, different symbols are used for different education groups, while the size of the symbol identifies the experience level (smallest symbol for the least experienced group, largest symbol for the most experienced group). If the within-group variances were constant over time, they would line up exactly on the 45 degrees line. Perhaps surprisingly, Figure 1a shows that, for men, most skill groups are scattered around the 45 degrees line. Only four (college graduates with 1-10, 11-20, and 21-30 year of experience, and college post-graduates with 1-10 years of experience) of the twenty groups are clearly above the 45-degree line. For the sixteen other education-experience groups, there is no discernable pattern of increase in the within-group variance.

Several other clear patterns also emerge from Figure 1a. In particular, the within-group variance grows as a function of both experience and education. For example, in both 1973-75 and 2000-02, high school dropouts with 1 to 10 years of experience have the lowest variance (around 0.10) while college post-graduates with 31 years and more of experience have the largest variance (around 0.40). This suggests that composition effects may be quite important since both experience and education increase over time.

---

education dummies only raises the R-square by about half a percentage point relative to the 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.

Another interesting pattern in Figure 1a is the systematic difference in the growth in the within-group variance by education level. For all four experience levels, the variance declines over time for high school dropouts but increases for college graduates. By contrast, there is little systematic change in the variance for the three other education groups that are more or less scattered around the 45 degrees line.

The results for women are qualitatively similar to those for men with the exception of women with some college education. For this education group, the within-group variance systematically increases between 1973-75 and 2000-02. As in the case of men, the within-group variance increases for college graduates, decreases for high school dropouts, and remains relatively unchanged for high school graduates and college post-graduates.

Table 1a (men) and 1b (women) reproduce these within-group variances in columns 1 and 2. The change in the within-group variance is reported in column 3. The groups for which the within-group variance increases by more than 0.04 are highlighted (in bold) in column 3. Columns 4 and 5 show the share of each skill group in the workforce in 1973-75 and 2000-02, respectively, while column 6 shows the change in the shares over time.

Like Figure 1, Table 1 clearly shows that the within-group variance is strongly linked to experience and education. Since the twenty groups are ranked by education and experience in the table, groups at the top of the table (low education and experience) show much less wage dispersion than groups at the bottom of the table (high education and experience).

Table 1 also shows, however, that there is a large and systematic decline in the share of workers in groups with low within-group dispersion. This is most obvious when groups are divided by education. For women, column 6 of Table 1b shows that, for all experience groups, the share of women with a high school degree or less has declined over time. By contrast, the share of women with some college education or more has increased for each and every experience group. With two minor exceptions, the same pattern holds for men in Table 1a. The other clear pattern is that the share of more experienced workers relative to young workers systematically increases for all education groups (except high school dropouts). This reflects the aging of the baby boom generation.

Overall, Table 1a and 1b clearly show a systematic growth in the share of experience-education groups that exhibit large within-group variances. The correlation coefficient between the within-group variance in 2000-02 (V_{jt}) and the change in share (θ_{jt} - θ_{js}) is 0.55 for men and 0.68 for women. This suggests a large and positive composition effect term in equation (9).

Panel B of Table 1a and 1b shows more explicitly the magnitude of composition effects. The first row of the panel shows the weighted average of the within-group variances when the weights used are the actual shares in the corresponding year. The 1973-75 shares are used to weight the 1973-75 variances, and the 2000-02 shares are used to weight the 2000-02. These weighted averages correspond to the unadjusted residual variances for 1973-75 and 2000-02, respectively. Table 1a shows that the residual variance for men increases by 0.041 between these two time periods.

The second row shows that the change in the residual variance is much smaller (0.012) when the shares are held at their 1973-75 level. This means that only about a quarter of the 0.041 change in the residual variance is due to the increase in the within-group variances. Consistent with Figure 1a, this means that, on average, the within-group variance in 2000-02 is only 0.012 above the 45 degrees line. The remaining change in the residual variance, 0.029 (0.041 minus 0.012) is due to composition effects.

The results for women reported in Table 1b are quite similar. Composition effects account for 0.035 of the 0.047 increase in the residual variance. Only a quarter of the total increase (0.012) is due to the changes in within-group group variances, holding the shares constant at their 1973-75 level.

Interestingly, the last row of Panel B shows that the residual variance increases more when shares are held at their 2000-02 levels instead. The intuition for this result is that using the 2000-02 shares instead of the 1973-75 shares puts more weight on college graduates who experience a sharp increase in their within-group variance, and less weight on high school dropouts who experience a decline in their within-group variance.

Figure 2 provides some information on the detailed year-by-year evolution of the within-group variance for each of the five education groups. To control for changes in the experience distribution of the workforce, the variance for each education group is defined as the simple average of the within-group variances over the four experience groups. For example, the within-group variance for college graduates in Figure 2 is the average of the within-group variances for college graduates with 1-10, 11-20, 21-30 and 31 or more years of experience.

I only show the detailed evolution in the within-group variance by education groups for two reasons. First, it would not be practical to show the detailed evolution in the within-group variance for each of the twenty experience-education groups. Second, Table 1 and Figure 1 suggest that, conditional on education, the change in the within-group variance is relatively similar across experience groups. In other words, education (as opposed to experience) accounts for most of the variation in the growth in the within-group variance across the twenty experience-education groups.

The results for both men (Figure 2a) and women (Figure 2b) are different for different time periods. For both men and women, the within-group variances are either stable or declining during the 1970s. The within-group variances then grow substantially for each and every group during the 1980s. In the 1990s, however, there is a divergence in the trends by education groups. For college graduates and post-graduates, the within-group variance increases slightly or remains constant between 1990 and 2000. For all other education groups, however, the within-group variance declines during the 1990s. The decline is particularly pronounced for high-school dropouts. Finally, the within-group variances grow mildly for most groups during the early 2000s.

Taken together, the results in Figure 1 and 2 indicate that, for most groups, there is relatively little change in the within-group variance between 1973 and 2003. The only exception is college graduates and women with some college education. For these particular groups, however, almost all of the growth in the within-group variance is concentrated in the 1980s.

4. Changes in Residual Inequality: Re-weighting Results

Having established the basic patterns of changes in the within-group variance for twenty coarse experience-education cells, I now turn to a re-weighting approach to analyze in more detail the role of composition effects in changes in residual wage inequality. As discussed in Section 2, one main advantage of the re-weighting approach is that it is easily implemented even when the data cannot be divided into fine experience-education cells. Another advantage of the approach is that it can be used to compute counterfactual measures of residual wage dispersion other than the variance.

The counterfactual variances are computed by replacing the sample weight for worker *i*, ω_{it}, by a counterfactual weight ω_{it}*. The actual residual variance is computed from the individual level data as

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

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

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

The counterfactual weights ω_{it}* are chosen in a way that makes the (counterfactual) distribution of skills at time t the same as in an appropriate base year (for example 1973). These weights are obtained by multiplying the sample weights θ_{it} by a re-weighting factor. The re-weighting factor is 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 2003 when 1973 is used as base year are computed by estimating a logit model on data for years 1973 and 2003 pooled together. The dependent variable is a dummy variable for year 2003, while the explanatory variables are the age and education variables. I use the same set of explanatory variables in the logit as in the wage regressions (full set of indicators for age and education plus interactions between education and a quartic in age). The predicted probability that worker *i* is in year 2003, P_i, is then used to compute the counterfactual weight as

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

Once the counterfactual weights have been computed, it is straightforward to compute alternative measures of residual wage dispersion. For example, the actual 90-10 residual gap is defined as the difference between the 90^{th} and the 10^{th} centile of log wages when the usual sample weights ω_{it} are used. The counterfactual 90-10 residual gap is readily obtained by recomputing the 90^{th} and the 10^{th} percentiles using the counterfactual weights ω_{it}* instead.

Figures 3a (men) and 3b (women) compare the actual residual variance from 1973 and 2003 to the counterfactual variances that would have prevailed if the distribution of skills (experience and education) had remained at its 1973 or 2003 level. The composition effects are the difference between the actual and counterfactual variances. Figure 3a shows that the residual variance grows by about 0.04 over the whole sample period. Consistent with Figure 2a, most of the growth is concentrated in the first part of the 1980s. The residual variance remains essentially unchanged in the 1970s and 1990s but grows between 1999 and 2003.

By contrast, the counterfactual variance in the late 1990s / early 2000s is only about 0.01 higher than in the mid-1970s when the distribution of skills is held constant at its 1973 level. Consistent with Table 1a, Figure 3a suggests that about three quarters of the growth in the residual variance is a spurious consequence of composition effects (when the distribution of skills is held at its 1973 level).

In terms of timing, Figure 3a shows that composition effects play a negligible role during the 1970s but become very important during the 1980s and 1990s. It is clear from Appendix Table 1 why composition effects are not important during the 1970s. The table shows that while the workforce got more educated between 1973 and 1980, it also became less experienced with the entry in the labor market of the largest baby boom cohorts (born in the late 1950s). The positive impact of growing educational achievement on the residual variance is offset by the fact that the workforce got younger (lower within-group inequality) during this period. By contrast, Appendix Table 1 shows that both experience and educational achievement increased in the 1980s and 1990s, leading to an unambiguous positive composition bias in the growth of the residual variance.

A closer examination of Figure 3a also shows evidence of a cyclical effect in the residual 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 effects tend to hide the pro-cyclicality of the level of real wages (Barsky et al, 1994). By analogy, Figure 3a suggests that composition effects tend to over-state the counter-cyclical pattern in wage inequality over the business cycle (inequality grows during recessions).

Figure 3a also shows the counterfactual variance that would have prevailed if the distribution of characteristics had remained at its 2003 level. The results are qualitatively similar, though not as dramatic, as those obtained by holding characteristics at their 1973 level. In particular, the counterfactual variance declines less dramatically in the 1990s when characteristics are held at their 2003 instead of 1973 level.[8]

The results for women in Figure 3b are qualitatively similar to those for men. Composition effects explain most of the growth in the within-group variance between 1973 and 2003 when characteristics are held at their 1973 level. Composition effects also play a qualitatively similar, though less dramatic, role when characteristics are held at their 2003 level instead.

The results for both men and women are summarized in Table 2. The table confirms that composition effects account for most of the growth in the residual variance over the 1973-2003 period when the distribution of experience and education are held at their 1973 levels. Once again, the results are less dramatic when the distribution of experience and education is held at its 2003 level instead. Even in this case, however, composition effects still account for about half of the growth in the residual variance for men, and for a third of the growth in the residual variance for women.

Table 2 also compares the growth in the residual variance to the growth in the total variance of wages (both within- and between-group components) over the same periods. Interestingly, over the whole 1973-2003 period, the residual component of the variance accounts for less than half of the growth in the total variance (43 percent for men, 46 percent for women). This finding is at odds with several previous studies that tend to find that most of the growth in wage inequality is due to the residual component. I explain in Section 5 that this earlier finding appears to be an artifact of measurement problems in the March CPS.

[8] The difference stems from the fact 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 2). By contrast, holding characteristics at their 2003 level puts relatively more weight on college graduates who experience a clear increase in their within-group variance.

When the distribution of experience and education is held at its 1973 level, the remaining growth in the residual variance only accounts for 14 percent of the 1973-2003 growth in the total variance of wages for men, and 5 percent for women. These percentages increase to 23 and 31 percent, respectively, when the distribution of experience and education is held at its 2003 level instead. Table 2 also shows that when the skill distribution is held constant, there is more growth in the residual variance between 1979 and 1989 than for the whole 1973-2003 period. This result holds for both men and women when skills are either held at their 1973 or 2003 levels. For example, the residual variance for men increases by 0.034 between 1979 and 1989 when the distribution of skills is held at its 2003 level. This is larger than the 0.025 change over the whole period. This means that for the other sample periods (1973 to 1979 and 1989 to 2003) pooled together, the residual variance *declined* by 0.09.

In light of the discussion in Section 2, these findings suggest that changes in the prices of unobserved skills only play a modest role in the overall growth in wage inequality between 1973 and 2003. For men, changes in the prices of unobserved skills account for no more than a quarter of the growth in overall inequality. For women, changes in the price of unobserved skills account for between 5 and 31 percent of the overall growth inequality.

The results also imply that all of the growth in the price of unobserved skills in concentrated in the 1980s. This finding is very difficult to reconcile with the SBTC hypothesis that typically states that the relative demand for skills also increased during the 1970s and the 1990s. I return to the question of what else may explain the pattern of growth in the residual wage inequality in Section 6.

Finally, the main findings are robust to the choice of alternative measures of wage dispersion. Figure 4 reproduces the results of Figure 3 using the 90-10 residual gap instead of the residual variance. The results are very similar to those for the residual variance. As in the case of the residual variance, almost all the growth in the 90-10 residual gap is concentrated in the 1980s (first half of the 1980s for men in Figure 4a). Furthermore, most of the growth in the 90-10 residual gap appears to be a spurious consequence of composition effects. When the distribution of experience and education is held at its 1973 level (dotted line in the figures), the 90-10 residual gap in the early 2000s is barely higher than in 1973.

5. What is wrong with the March CPS?

As mentioned in the Introduction, the findings of Section 3 and 4 are at odds with most of the previous literature on residual wage inequality. One potential explanation for this difference is that I use data on hourly wages from the May and ORG supplements of the CPS, while earlier studies typically use the March Supplement of the CPS.

In this Section, I show that the key problem with the March CPS is that it poorly measures the wages of workers paid by the hour (the majority of the workforce). To establish this point, however, I first need to process the two different data sources to make them as comparable as possible. I then turn to several pieces of evidence to show that both the level and trends in residual inequality are systematically biased in the March CPS because of the mismeasurement of the wages of workers paid by the hour.

a. Data processing

Both the May/ORG and the March CPS can be used to compute hourly wage rates. The March Supplement of the CPS asks about total earnings during the previous year. An hourly wage rate can then be computed by dividing last year's earnings by total hours worked last year. The latter variable is computed by multiplying two other variables available in the March CPS, usual weekly hours of work last year and weeks worked last year.

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 from May 1973 on. Since one 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.[9]

There are important differences between the way wages are measured in the March and May/ORG CPS. First, while the March CPS asks about retrospective measures of wages and earnings (last year), the May/ORG supplement asks about wages at the time of the survey (Section 3). Second, the May/ORG wage questions are only asked to wage and salary workers. By 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 have both 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

[9] 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 (see JMP), this is one more reason for focusing on the March CPS data starting with the earnings year 1975.

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 do not 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.[10] By contrast, in the May 1973-78 CPS, wages were *not* allocated for workers who failed to answer wage and earnings questions.[11] 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

[10] 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.

[11] 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 (May CPS) 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 7). See Hirsch and Schumacher (2003) for a detailed discussion of how wages are allocated (or not allocated) in the May/ORG CPS.

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 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.

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 (2002), 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).

c. Mismeasurement of the wages of workers paid by the hour in the March CPS

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 particular, around 60 percent of workers in the May/ORG CPS are paid by the hour (see Figure 8 for more detail). These workers report a direct measure of their hourly wage rate in the May/ORG CPS. In the March CPS, however, they have to report their total annual earnings and hours of work that are then used to compute an average hourly wage rate.

In the absence of measurement error, it should not matter whether hourly wages are computed directly from questions about hourly wage rates, or indirectly by dividing earnings by hours of work. Several validation studies show, however, that there is substantial measurement error in the earnings reported in the CPS or similar surveys.[12]

There are good a priori reasons to believe that asking directly hourly-rated workers about their hourly wage rates provides a more accurate wage measure than dividing earnings by hours. If it is easier for workers paid by the hour to report directly their hourly wage rate, this direct measure will likely be less affected by measurement error than the indirect wage measure based on average hourly earnings. In particular, 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 worker would probably have more difficulty reporting total hours and earnings during the year. 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.

My basic hypothesis is that for hourly-rated workers, the hourly wage rate indirectly computed from the March CPS is a more noisy measure of the true hourly rate of pay than the hourly wage rate collected in the May/ORG CPS. For workers not paid by the hour, the hourly wage rate has to be indirectly computed by dividing earnings by hours in both the May/ORG and the March CPS. Therefore, I do not expect the hourly wage from the March CPS to be a more noisy measure for these workers.

---

[12] 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.

Under the additional assumption of classical measurement error, this hypothesis yields several clear empirical predictions.[13] The most direct prediction is that the variance of March CPS wages should be larger than the variance of May/ORG CPS wages among workers paid by the hour. I test this prediction by comparing the variance of the two wage measures for workers paid by the hour and workers not paid by the hour.

One problem with implementing this test is that the March CPS does not ask individuals whether they are paid by the hour or not. Fortunately, this problem can be resolved by exploiting the rotation group feature of the CPS. Among individuals sampled in the March CPS, roughly one quarter of individuals rotate out of the CPS in each of the four following months, including March. This means that from 1979 on, all individuals in the March CPS should eventually be part of the outgoing rotation groups in March, April, May or June. In principle, their responses to the ORG supplement questions can thus be matched to their March CPS records. As discussed below, however, not all March respondents can be matched because of attrition and other data problems.

Prior to 1979, it is still possible to match the May CPS responses to the March responses for the March respondents who are still in the CPS in May (about half of the March respondents if there is no attrition). My empirical strategy is thus to match the March CPS respondents to either their ORG or May CPS records. From this matched sample, I can then use the information from the ORG of May CPS questions to divide workers into those paid and not paid by the hour.

Working with these matched samples involves a number of data issues that are beyond the scope of this paper. In particular, between five and ten percent of respondents cannot be matched because of attrition and other data problems.[14] Also, while the March and May/ORG wage records are for the same respondent, they are not for the same period (March wage is for last year, May/ORG wage is for the current survey month). This means that some workers coded as "paid by the hour" may not have been paid by the hour in the previous year. Focusing on workers who both report a wage in the month of the survey (the May/ORG wage) and in the previous year also results in a more "stable"

[13] Under classical measurement error, the measurement error in March wages (for workers paid by the hour) is assumed to have mean zero and be independent of all observable variables.
[14] Since I am only matching months close by, the matching rates are much higher than in most applications where records in one year are matched to the record for the same respondent one year later.

sample of workers. Fortunately, Appendix Figures 1a and 1b show that while the level of wage inequality is lower for this matched sample than for the full sample, the trends in inequality are very similar for the two samples.

Despite these data limitations, a striking pattern of results emerges from Figures 5a (men) and 5b (women). These figures contrast the variance of the March and May/ORG wages for the two groups of workers (paid by the hour or not). For both men and women, the variance of March wages is systematically larger than the variance of May/ORG wages for workers paid by the hour. By contrast, there is no systematic difference in the variance of March and May/ORG wages for workers not paid by the hour. Figures 5a and 5b provide convincing evidence that the key difference between the March and May/ORG wages is that the wages of workers paid by the hour are more noisily measured in the March CPS.

The extent of measurement error in March CPS for workers paid by the hour is both quantitatively and statistically significant. For men (Figure 5a), the average difference in the variance is 0.064, which represents about a third of the average variance in the May/ORG CPS (0.198). If the May/ORG wages were measured without error, this would imply a noise to signal ratio of about a third in the March CPS. The results are similar for women. The average difference in variances (0.055) also represents a third of the average variance of wages in the May/ORG CPS (0.167).

Despite these large differences in the level of the variances, the pattern of change over time in the variances is very similar in the March and May/ORG CPS. For both wage measures, the variance of wages for hourly workers is flat in the 1970s, grows sharply in the 1980s, and remains relatively constant thereafter. For workers not paid by the hour, however, the variance of wages keeps increasing steadily during the 1990s. This is consistent with the fact that workers not paid by the hour are much more educated than workers paid by the hour (see below), and that within-group inequality increases for college educated workers (Section 3).

There is a second empirical prediction of the hypothesis that the wages of hourly workers in the March CPS are measured with error that can be tested without resorting to the matched sample. 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).[15] This means that 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 simply a noisier measure of hourly wages than wages in 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.

Figures 6 and 7 confirm this empirical prediction. Figure 6a shows the evolution of the between-group variance for men over the 1975-2002 period for the two measures (March and May/ORG) of hourly wages.[16] 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 6b 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 grows 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

[15] 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.

[16] Figures 6 and 7 report the variance of wages by earnings year (year of the survey in the May/ORG CPS, previous year in the March CPS). I report data for 1975 to 2002 that correspond to the 1976 to 2003 survey years in the March CPS.

getting steeper during this period because of the increased attachment of women to the labor force.[17]

Turning to residual wage dispersion, Figure 7a shows that, for men, the residual variance of March CPS wages (without allocated earnings) is systematically larger than the residual variance of May/ORG wages. The results in Figure 7b for women are very similar. A set of strong conclusions thus emerges from Figures 5, 6 and 7. First, Figure 5 clearly shows that wages are more noisily measured in the March CPS. Consistent with classical measurement error, Figures 6 and 7 show that these measurement problems do not affect between-group wage dispersion but spuriously inflate residual wage dispersion in the March CPS.

The findings in Figures 5, 6 and 7 clearly indicate that, relative to the May/ORG CPS, residual wage inequality is biased up in the March CPS because this wage measure poorly captures the hourly wage rate for workers paid by the hour. The ORG CPS supplement thus provides both a much larger sample of wage workers and more accurately measured wages. In terms of econometric practice, there is thus a strong case for using the May/ORG instead of the March CPS for studying the evolution of wage inequality over time.

c. Spurious Trends in Residual Wage Inequality in the March CPS?

If the bias in residual wage inequality in the March CPS were constant over time, using the May/ORG or the March CPS would have little consequence for the interpretation of the sources of change in residual inequality. Unfortunately, Figure 7 shows that both the level and the secular growth in residual inequality are larger in the March than in the May/ORG CPS. In the case of men (Figure 7a), the residual variance of wages in the May/ORG CPS is stable during the 1970s, grows rapidly in the early 1980s, and remains fairly constant from the mid-1980s to the late 1990s. By contrast, the residual variance grows steadily from 1975 to 2002 when hourly wages are computed

[17] See Blau and Kahn (1997) and Fortin and Lemieux (1998). The continuing growth in the between-group variance during the 1980s and 1990s may 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.

using the March CPS. Among women, there is also more growth in the residual variance of March relative to May/ORG wages, though the difference is not as marked as in the case of men.

One potential explanation for this difference is that the fraction of workers paid by the hour has been growing over time. To see this, consider the case where, for hourly-rated workers, March wages are equal to May/ORG wages plus a measurement error of variance σ^2. It follows that the difference between the March and May/ORG variance for group *j* at time *t* is $H_{jt}\sigma^2$, where H_{jt} is the fraction of workers paid by the hour. The variance of residuals for group *j* in the March (V^M_{jt}) and May/ORG (V^O_{jt}) CPS are linked by the formula:

$$V^M_{jt} = V^O_{jt} + H_{jt}\sigma^2.$$

Under the assumption that the measurement error variance is constant over time, it follows that

$$\Delta V^M_{jt} = \Delta V^O_{jt} + \Delta H_{jt}\sigma^2,$$

where $\Delta H_{jt}\sigma^2$ is the spurious growth in the March CPS variance due to changes in the fraction of workers paid by the hour.

Figure 8 shows that the fraction of workers paid by the hour has indeed increased over time. Since education is by far the most important factor explaining the propensity to be paid by the hour, I only report the fraction of workers paid by the hour by education group in Figure 8.[18] Figure 8a (men) and 8b (women) show that the fraction of workers paid by the hour has increased by up to 15 percentage point (depending on the education group) between the mid-1970s and the late 1980s. Since Figure 5 suggests that σ^2 is of the order of 0.05 to 0.07, the growth in the fraction of workers paid by the hours may have resulted in a spurious increase of up to 0.01 in the variance of March wages. This is very substantial relative to the 0.03 growth in the residual variance in the May/ORG CPS during the same period.

Another related problem with the March CPS is that the large difference in the fraction of workers paid by the hour by education category likely masks the importance of composition effects. Remember that composition effects represent the difference

[18] The fraction of workers paid by the hour declines as a function of experience. Relative to education, however, experience has a smaller impact (in absolute value) on the probability of being paid by the hour.

between the actual residual variance and the counterfactual variance obtained by replacing the skill composition in the end period (say 2003) by the skill composition in the base period (say 1973). As discussed earlier, the counterfactual puts much more weight on less educated workers and less weight on more educated workers. This results in large composition effects in the May/ORG CPS because the within-group variance among highly-educated workers is much larger than among less-educated workers.

The difference in the within-group variance across education should be lower in the March than in the May/ORG CPS because the variance among less-educated workers is inflated by the larger fraction of those workers being paid by the hour. Consider, for example, the case of college post-graduates relative to high school dropout. In the early 2000s, between 80 and 90 percent of high-school dropouts are paid by the hour compared to just more than 10 percent among college postgraduates, a difference of about 70 percentage points (Figure 8). In light of the evidence in Figure 3, this suggests that the within-group variance of high-school dropouts in the March CPS is inflated by about 0.05 relative to college postgraduates.[19] This represents about a third of the 0.15-0.20 difference in the within-group variance between college post-graduates and high-school dropouts during the same period. Consistent with this prediction, a closer examination of the March data indeed indicates that the difference between the variance of these two groups observed is about 0.05 lower in the March than in the May/ORG CPS data.[20]

Because of this problem, the reweighting procedure should yield smaller composition effects when applied to the March CPS instead of the May/ORG CPS. Appendix Figures 2 and 3 indeed show that composition effects in the March CPS are about a third smaller than in the May/ORG CPS (see Appendix B for more discussion of these results).

In summary, the mismeasurement of the wages of workers paid by the hour in the March CPS results in two major problems for analysis of residual wage inequality based on this data set. First, the growth in the fraction of workers paid by the hour in the 1970s

---

[19] In Figure 3, the difference between the March and the May/ORG variance among workers paid by the hour is over 0.07 in the early 2000s. Multiplying this difference by the difference in the fraction of workers paid by the hour (about 0.07) yields about 0.05.

[20] For example, among men in 2000-01, the ORG CPS variance is 0.319 and 0.132 for college postgraduates and high school dropouts, respectively (difference of 0.187). The March CPS variance for the same groups in 2000-01 is 0.374 and 0.226 (difference of 0.146).

and 1980s generates a spurious growth in within-group wage dispersion during this period. Second, the role of composition effects in the growth of residual wage inequality is understated because of the large difference between the fractions of highly- and less-educated workers who are paid by the hour. These findings reinforce the earlier conclusion that the May/ORG CPS should be used instead of the March CPS for analyses of the causes and consequences of the growth in residual wage inequality.

6. Concluding comments: If it is not SBTC, what is it?

The main finding of the paper is that, when the composition of the workforce is held constant over time, there is only a modest growth in residual wage inequality between 1973 and 2003. Furthermore, the growth in residual inequality in the 1980s exceeds the growth over the whole 1973-2003 period, which means that residual inequality actually declined in the other time periods. Finally, highly educated workers are the only group for which residual wage dispersion increased over the whole 1973-2003 period.

Looking at more general dimensions of wage inequality, Card and DiNardo (2002) argued that the timing of the growth in wage inequality was difficult to reconcile with the SBTC hypothesis. The findings of this paper pose another important challenge to the SBTC hypothesis since I find that residual inequality actually *declined* in periods other than the 1980s. Technological change can only explain these changes under the implausible assumption that while technological change was biased in favor of skilled workers during the 1980s, it was biased in favor of unskilled workers during the other periods.

If SBTC cannot explain the observed changes, what else could be going on? Another potential explanation for the changes in residual wage inequality is the minimum wage. In particular, DiNardo, Fortin, and Lemieux (1996) find that the decline in the real value of the minimum wage during the 1980s accounts for about a third of the increase in residual wage inequality. Lee (1999) finds an even larger effect by allowing for spillover effects of the minimum wage. Interestingly, the basic trends in the real value of the minimum wage are closely related to the trends in residual wage inequality documented above. For example, Row C of Table 2 shows that the real value of the minimum wage declined in the 1980s and early 2000s, while residual inequality increased during those two periods. By contrast, residual inequality declined when the real value of the minimum wage increased during the 1970s and 1990s.

Figure 9 illustrates the close connection between the evolution of the minimum wage and the residual variance between 1973 and 2003. The figure compares the residual variance when characteristics are held constant at their 1973 level to the predicted variance from a regression that includes a linear trend and the log real minimum wage as regressors. This simple regression model explains the data quite well. The R-square is 0.81 and 0.88 for men and women, respectively. This is a remarkably good fit since there is almost no time trend in the dependent variable.

For both men (Figure 9a) and women (Figure 9b), the minimum wage has a strong impact on the residual variance. The regression models are reported in the figures and show large t-statistics for the effect of the minimum wage (t-statistic of 9 for men, and 12 for women). Consistent with DiNardo, Fortin and Lemieux (1996), the effect of the minimum wage is also larger for women then men. The fit of the model is most impressive for women. The large increases in the minimum wage in 1973-74, 1989-91, and 1995-97 all closely match corresponding declines in the residual variance. By contrast, the three periods where the minimum wage declined in real terms for failing to be indexed (1981-1989, 1992-95, and 1998-2003) all correspond to clear increases in residual inequality.[21]

The estimated effect of the minimum wage is very similar when the regressions are fit to the residual variance that holds the distribution of characteristics at its 2003 instead of 1973 level. In the latter case, however, the underlying time trend is small and positive, while it is negative and significant when characteristics are held at their 1973 level. These results suggest that there is essentially no growth in residual inequality left once composition effects and the impact of the minimum wage have been accounted for.

While the minimum wage explains very well the time series pattern of the residual variance, it is not a credible explanation for the substantial growth in within-

[21] The three most important increases in the minimum wage are: from \$1.60 to \$2.00 in May 1974, from \$3.35 in March 1990 to \$4.25 in April 1990, and from \$4.25 in September 1996 to \$5.15 in September 1997. The real value of the minimum wage was substantially eroded by inflation as the minimum wage remained fixed at \$3.35 from January 1981 to March 1990, at \$4.25 from April 1991 to September 1996, and at \$5.15 from September 1997 to now.

group inequality for the most highly educated workers. Interestingly, this concentration of inequality growth among high-wage workers is consistent with Piketty and Saez (2003) who document a dramatic increase in inequality in the top-end of the earnings distribution (using tax data) since the 1970s. Piketty and Saez argue that both the timing and the extent of the growth in inequality at the top end are hard to reconcile with SBTC. They rather favor an alternative explanation based on social norms.

Finally, one important message of the paper is that the March CPS does not provide very reliable measures of either the level or the trends in residual wage inequality. The ORG supplement of the CPS provides more accurate measures of hourly wages for much larger samples of workers than the March CPS. Over thirty years of data are now available when the ORG CPS is combined with the 1973-78 May CPS. There is thus a strong case for using the May/ORG CPS, instead of the March CPS, for studying the sources of change in wage inequality in the United States since the early 1970s.

References

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

Aghion, Philippe, "Schumpeterian Growth Theory and the Dynamics of Income Inequality," *Econometrica* 70(3), May 2002, 855-82

Autor, David H., Lawrence F. Katz, and Alan B. Krueger, "Computing Inequality: Have Computers Changed the Labor Market?" *Quarterly Journal of Economics* 113(4), November 1998, 1169-1213

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

Becker, Gary S. *Woytinski Lecture*, University of Michigan, 1967. Reproduced in *Human capital (2nd Edition),* University of Chicago Press, 1975.

Berman, Eli, John Bound, and Zvi Griliches, "Changes in the Demand for Skilled Labor within U.S. Manufacturing: Evidence from the Annual Survey of Manufactures," *Quarterly Journal of Economics*, 109(2), May 1994, 367-97

Blau, Francine D., and Lawrence M. Kahn, "Swimming Upstream: Trends in the Gender Wage Differential in the 1980s," *Journal of Labor Economics* 14(4) October 1996, 1-42

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

Chay, Kenneth and David Lee, “Changes in Relative Wages in the 1980s: Returns to Observed and Unobserved Skills and Black-White Wage Differentials,’ *Journal of Econometrics* 99, 2000, 1-38

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

Farber, H. and R. Gibbons, "Learning and Wage Dynamics," *Quarterly Journal of Economics* 111, 1996, 1007-47

Fortin, Nicole, and Thomas Lemieux, “Rank Regressions, Wage Distributions, and the Gender Gap” *Journal of Human Resources* 33, Summer 1998, 610-643

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

Mincer, Jacob “Changes in Wage Inequality, 1970-1990,” *Research in Labor Economics 16*, 1997, 1-18.

Piketty, Thomas, and Emmanuel Saez, “Income Inequality in the United States, 1913-98,” *Quarterly Journal of Economics* 118(1), February 2003, 1-39

APPENDIX A: Topcoding

Topcoding adjustments in the May/ORG and March CPS

As mentioned in Section 5, 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.

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

Appendix Figures 2a (men) and 2b (women) compare the actual residual variance using the March CPS hourly wage rate to the residual variance that would have prevailed if the distribution of age and education had remained at their 1975 level. The re-weighting methodology used to compute the counterfactual variance is the same as for the May/ORG CPS (Figure 3). The figures show that the impact holding the distribution of characteristics constant is less dramatic in the March CPS than in the May/ORG CPS data. As discussed in Section 5c, an important part of this difference is a consequence of differences (over skill groups) and changes (over time) in the fraction of workers paid by the hour.

Despite these differences, adjusting for composition effects still has a significant impact on the economic interpretation of the trends in the residual variance in the March CPS. In particular, the figures show essentially no growth in the within-group variance after 1987-88 when the distribution of experience and education is held at its 1975 level. For women, the pattern of growth in residual inequality in the March CPS is 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 residual inequality also becomes qualitatively similar to the one in the May/ORG CPS. The only remaining discrepancy is that residual inequality grows rapidly in the March CPS during the 1970s, while it remains stable in the May/ORG data.

Table 1a: Within-group variance of wages by experience-education cell for men, 1973-75 and 2000-02

	Within-group variance			Workforce share		
	1973-75	2000-02	Change	1973-75	2000-02	Change
	(1)	(2)	(3)	(4)	(5)	(6)
A. By education and experience						
Dropout:						
1-10	0.118	0.083	-0.035	0.065	0.035	-0.030
11-20	0.169	0.130	-0.038	0.052	0.026	-0.026
21-30	0.170	0.154	-0.017	0.055	0.025	-0.029
31+	0.180	0.162	-0.019	0.123	0.028	-0.095
High school graduates:						
1-10	0.130	0.130	0.000	0.137	0.082	-0.055
11-20	0.145	0.181	0.035	0.094	0.085	-0.009
21-30	0.162	0.196	0.034	0.069	0.086	0.017
31+	0.188	0.217	0.029	0.074	0.058	-0.016
Some college:						
1-10	0.143	0.152	0.008	0.076	0.077	0.001
11-20	0.173	0.204	0.031	0.036	0.075	0.039
21-30	0.216	0.227	0.012	0.025	0.072	0.048
31+	0.245	0.256	0.011	0.020	0.046	0.026
College graduates:						
1-10	0.161	0.224	**0.064**	0.048	0.061	0.014
11-20	0.204	0.276	**0.072**	0.022	0.063	0.041
21-30	0.220	0.310	**0.091**	0.017	0.051	0.034
31+	0.299	0.332	0.033	0.009	0.024	0.015
Post-graduates:						
1-10	0.217	0.316	**0.099**	0.034	0.023	-0.010
11-20	0.324	0.324	0.000	0.023	0.033	0.009
21-30	0.327	0.302	-0.025	0.015	0.033	0.018
31+	0.420	0.369	-0.051	0.006	0.016	0.010
B. Weighted Average (using alternative shares)						
Actual shares	0.173	0.214	0.041			
1973-75 shares	0.173	0.185	0.012			
2000-02 shares	0.191	0.214	0.023			

Table 1b: Within-group variance of wages by experience-education cell for women, 1973-75 and 2000-02

	Within-group variance			Workforce share		
	1973-75	2000-02	Change	1973-75	2000-02	Change
	(1)	(2)	(3)	(4)	(5)	(6)
A. By education and experience						
Dropout:						
1-10	0.099	0.056	-0.043	0.057	0.026	-0.031
11-20	0.130	0.090	-0.040	0.039	0.015	-0.024
21-30	0.125	0.106	-0.019	0.050	0.018	-0.032
31+	0.139	0.123	-0.017	0.103	0.023	-0.080
High school graduates:						
1-10	0.106	0.108	0.002	0.179	0.070	-0.109
11-20	0.145	0.157	0.011	0.095	0.072	-0.023
21-30	0.144	0.172	0.028	0.092	0.086	-0.006
31+	0.162	0.178	0.016	0.097	0.074	-0.023
Some college:						
1-10	0.118	0.137	0.019	0.077	0.091	0.014
11-20	0.134	0.198	**0.065**	0.025	0.081	0.057
21-30	0.152	0.209	**0.057**	0.020	0.084	0.064
31+	0.160	0.220	**0.060**	0.020	0.054	0.034
College graduates:						
1-10	0.134	0.179	**0.045**	0.055	0.076	0.020
11-20	0.170	0.260	**0.090**	0.015	0.058	0.043
21-30	0.173	0.262	**0.088**	0.014	0.052	0.038
31+	0.195	0.254	**0.059**	0.010	0.021	0.010
College post-graduates						
1-10	0.154	0.239	**0.085**	0.022	0.026	0.004
11-20	0.238	0.259	0.021	0.012	0.027	0.015
21-30	0.204	0.217	0.013	0.011	0.034	0.023
31+	0.280	0.234	-0.046	0.006	0.013	0.007
B. Weighted Average (using alternative shares)						
Actual shares	0.136	0.183	0.047			
1973-75 shares	0.136	0.148	0.012			
2000-02 shares	0.149	0.183	0.034			

Table 2: Composition Effects and Changes in the Residual Variance of Log Hourly Wages, May/ORG CPS

	1973-1979	1979-1989	1989-1999	1999-2003	1973-2003
A. Men					
Residual variance:					
Actual change	-0.003	0.036	-0.003	0.017	0.047 [43]
1973 skills distribution	-0.003	0.027	-0.019	0.011	0.015 [14]
2003 skills distribution	-0.008	0.034	-0.013	0.012	0.025 [23]
Total variance:	-0.002	0.080	0.007	0.014	0.109 [100]
B. Women					
Residual variance:					
Actual change	-0.014	0.047	-0.001	0.013	0.045 [46]
1973 skills distribution	-0.017	0.036	-0.019	0.005	0.005 [5]
2003 skills distribution	-0.012	0.040	-0.006	0.008	0.030 [31]
Total variance:	-0.026	0.092	0.017	0.013	0.098 [100]
C. Real value of the minimum wage (logs)					
	0.103	-0.391	0.135	-0.099	-0.252

Note: Numbers in square brackets represents the percentage of the 1973-2003 change in the total variance of wages (both within- and between-group components) that is attributable to this variance component.

Appendix Table 1: 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

Figure 1a: Change in the Male Within-Group Variance by Experience and Education Levels: 1973-75 to 2000-02

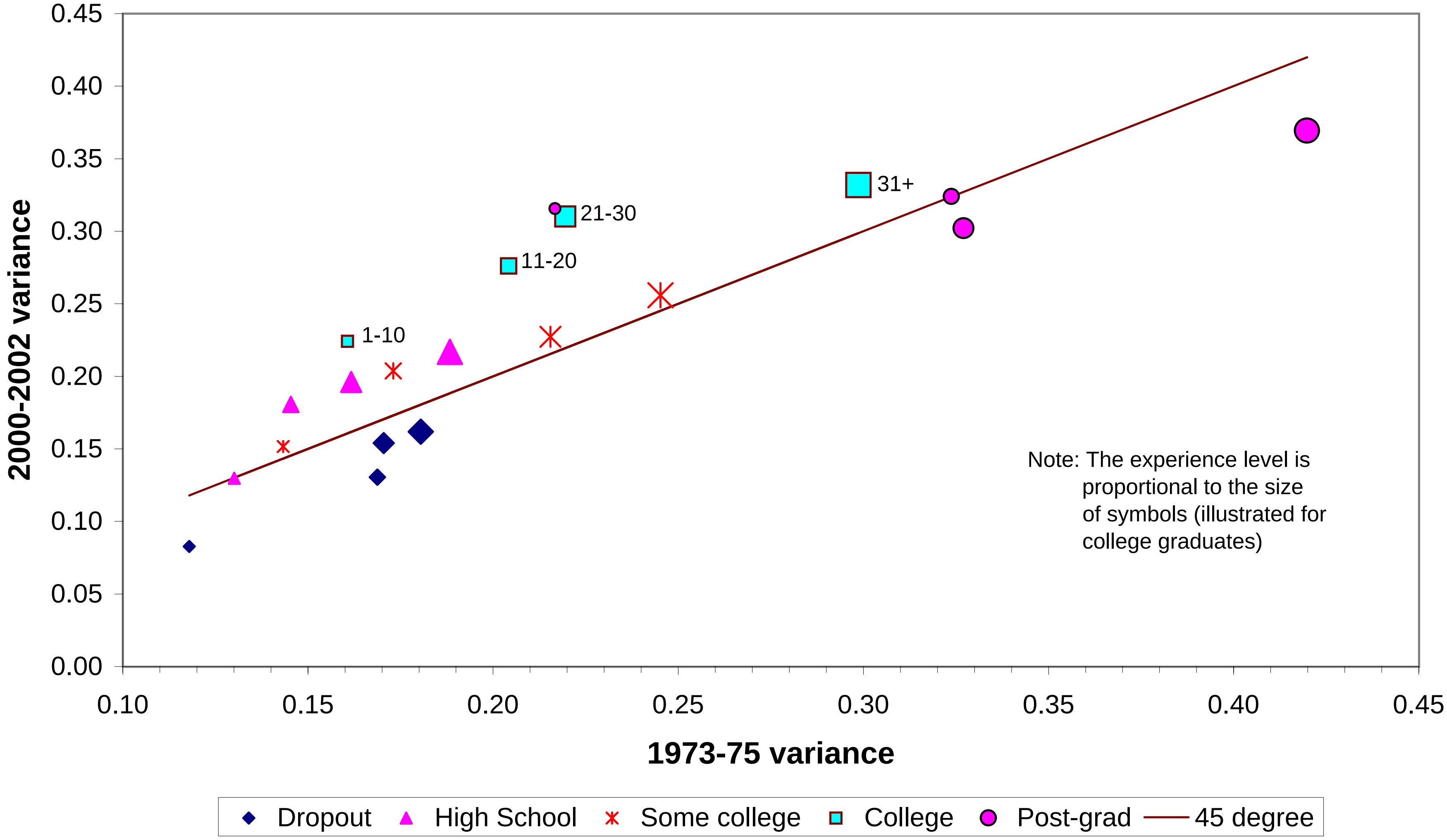


Figure 1b: Change in the Female Within-Group Variance by Experience and Education Levels: 1973-75 to 2000-02

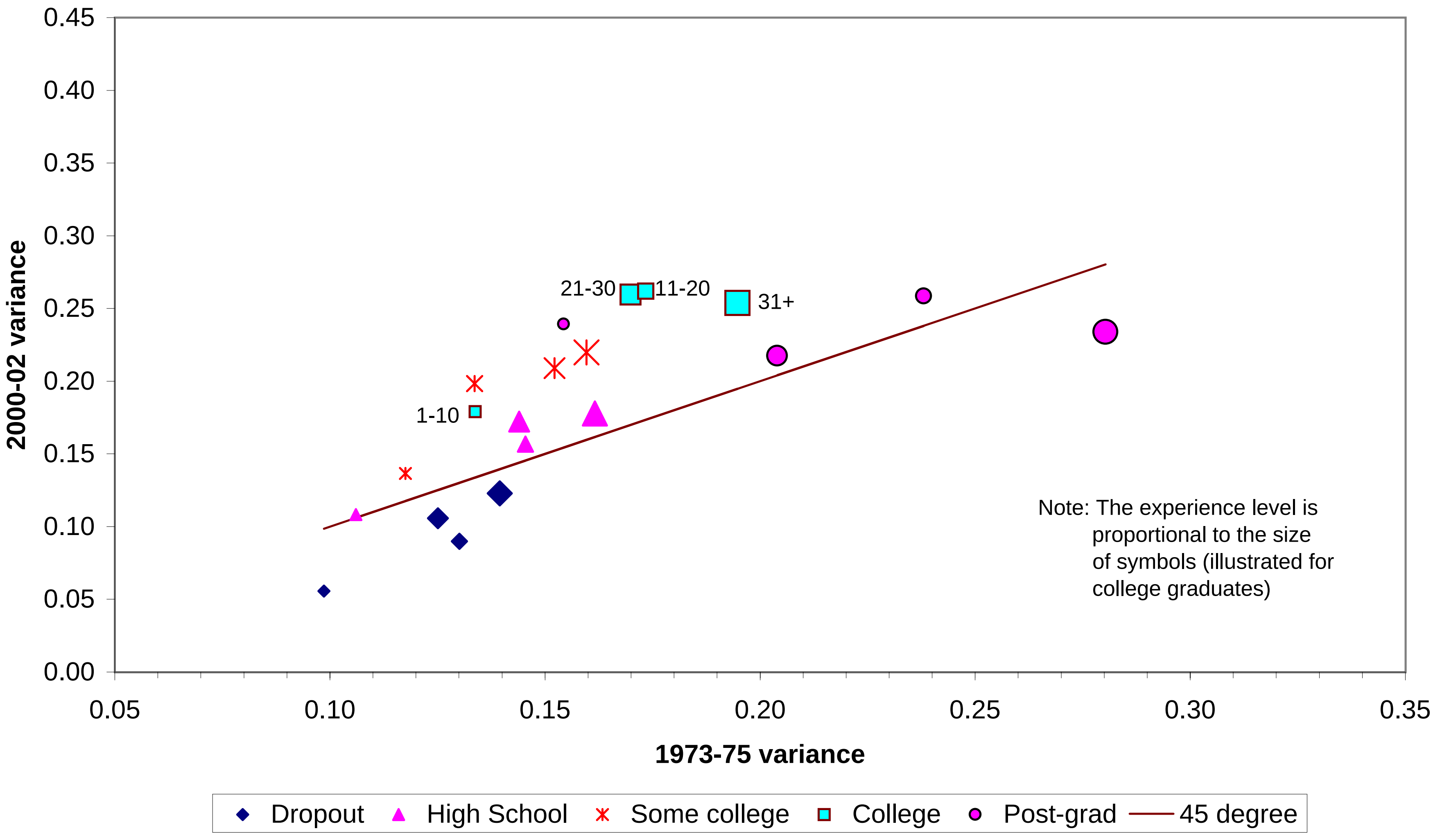


Figure 2a: Within-group variance by education group for men, (average of the four experience groups)

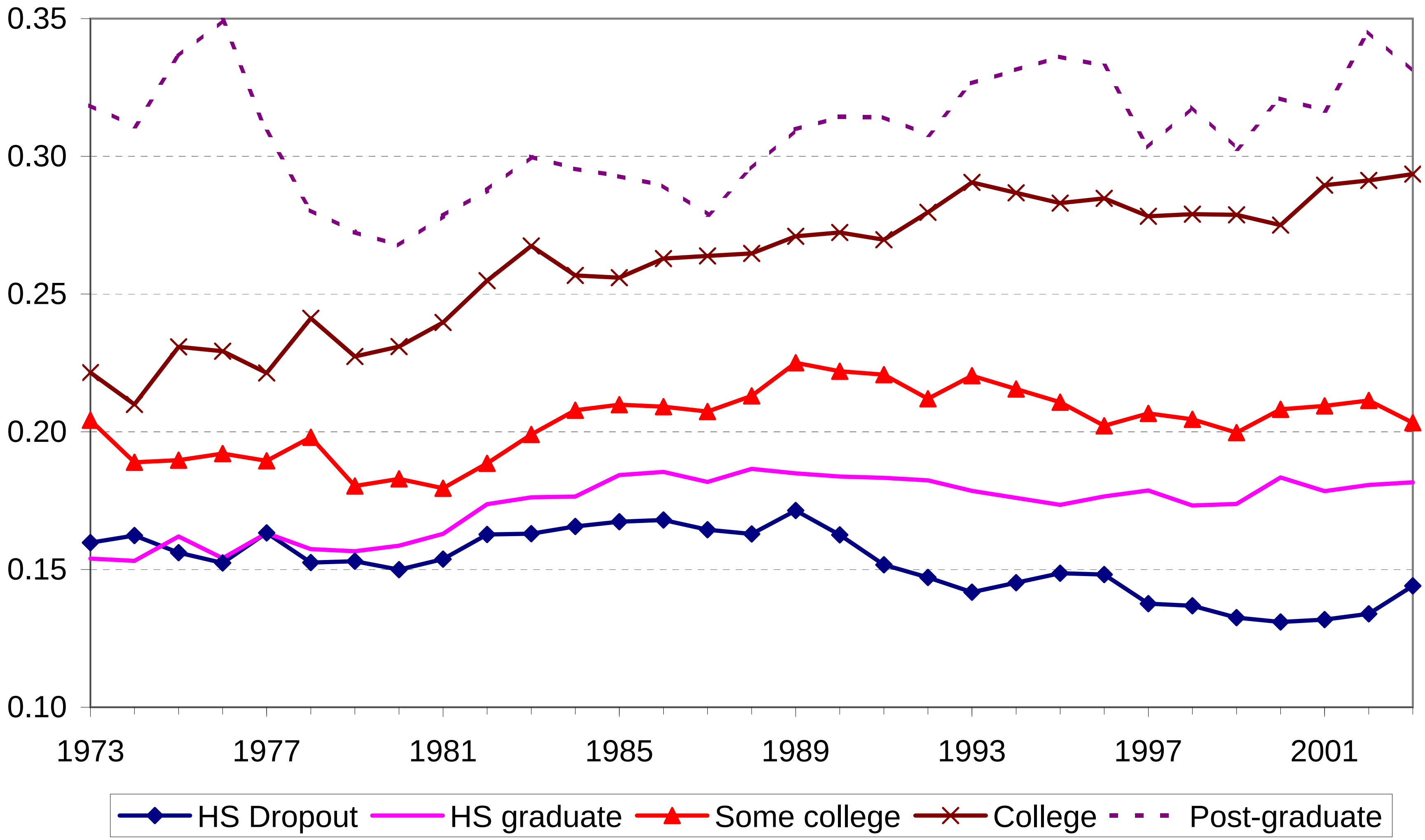


Figure 2b: Within-group variance by education group for women (average of the four experience groups)

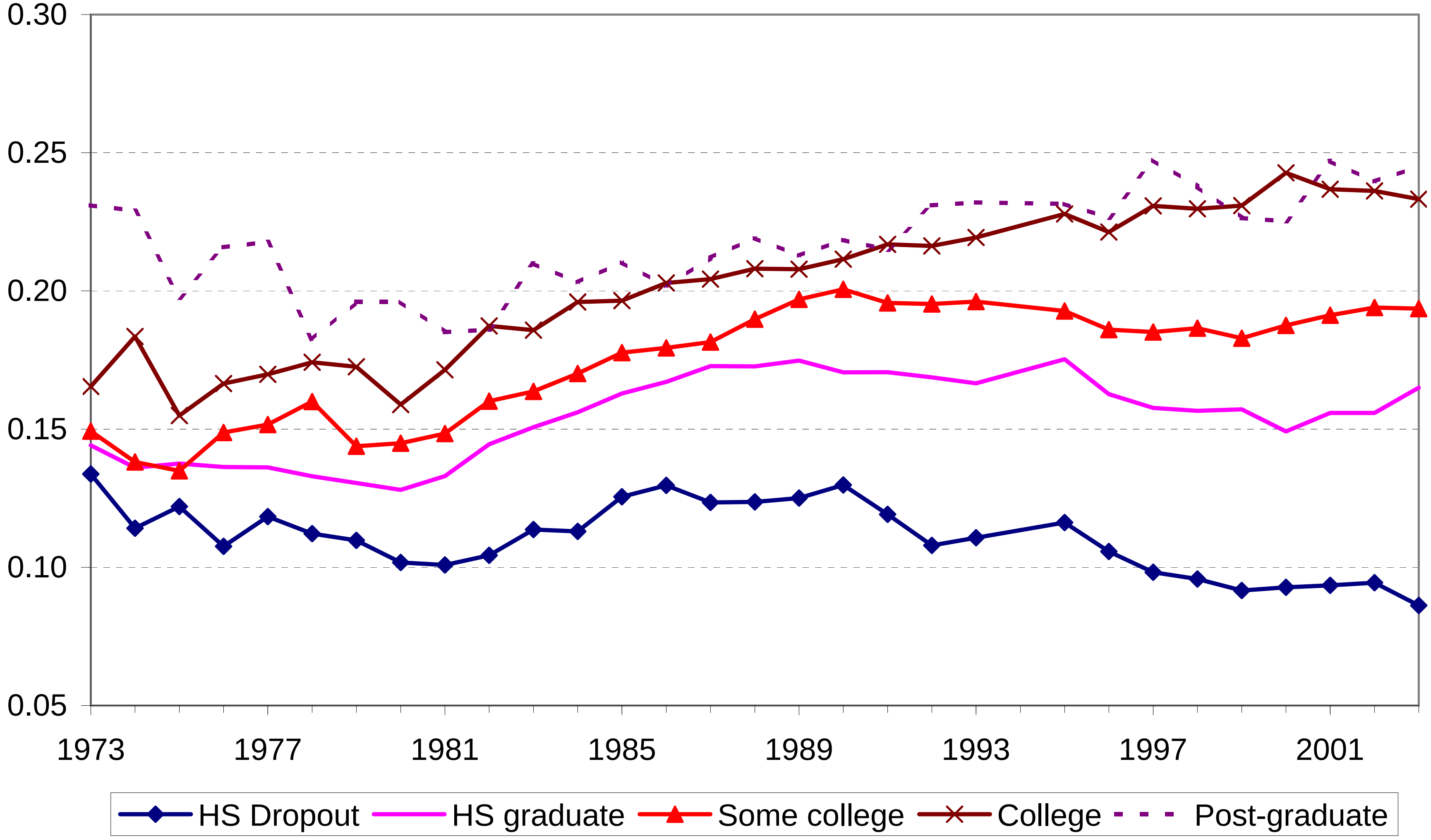


Figure 3a: Actual and counterfactual residual variance of wages for men, 1973 to 2003

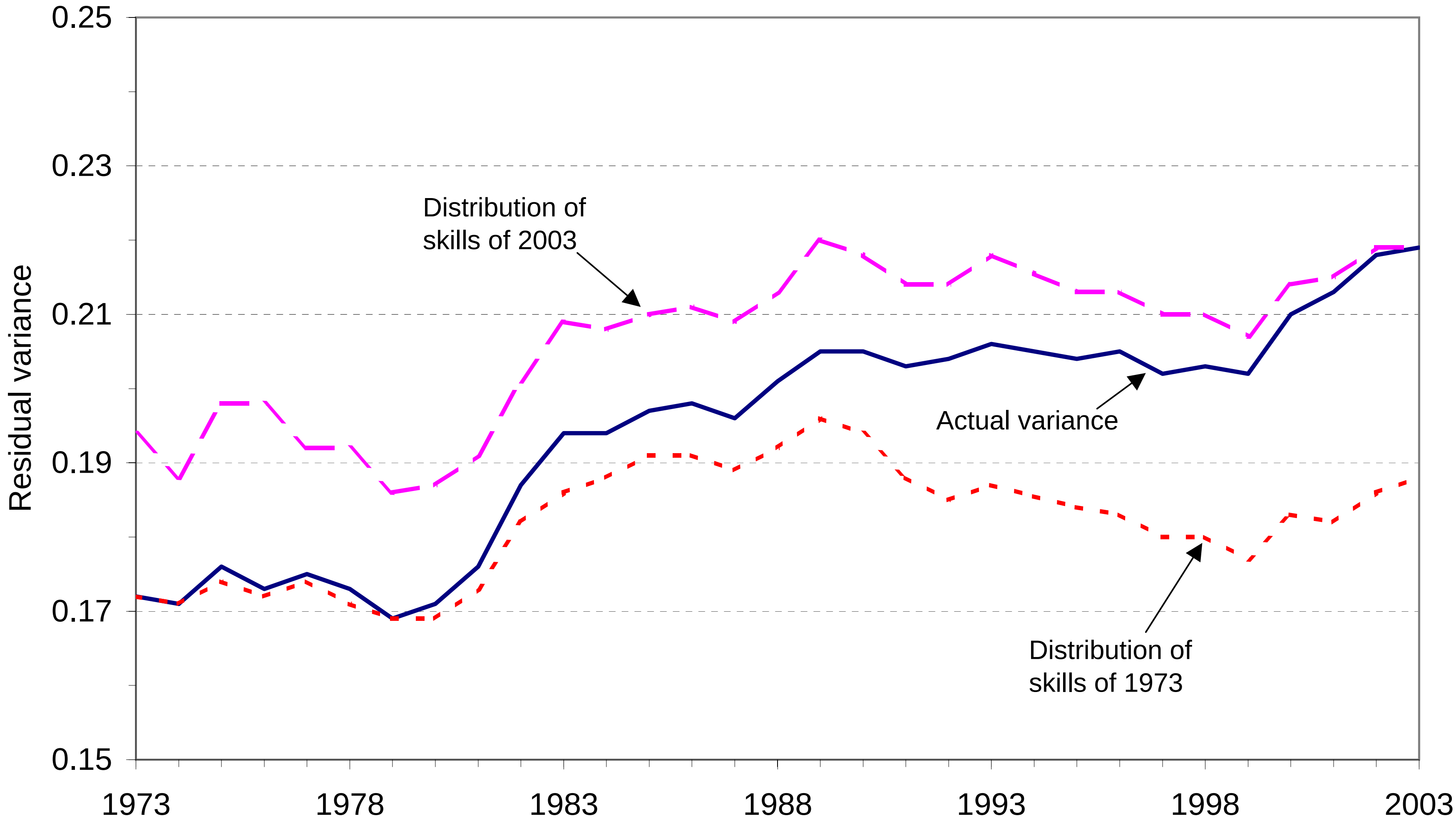


Figure 3b: Actual and counterfactual residual variance of wages for women, 1973 to 2003

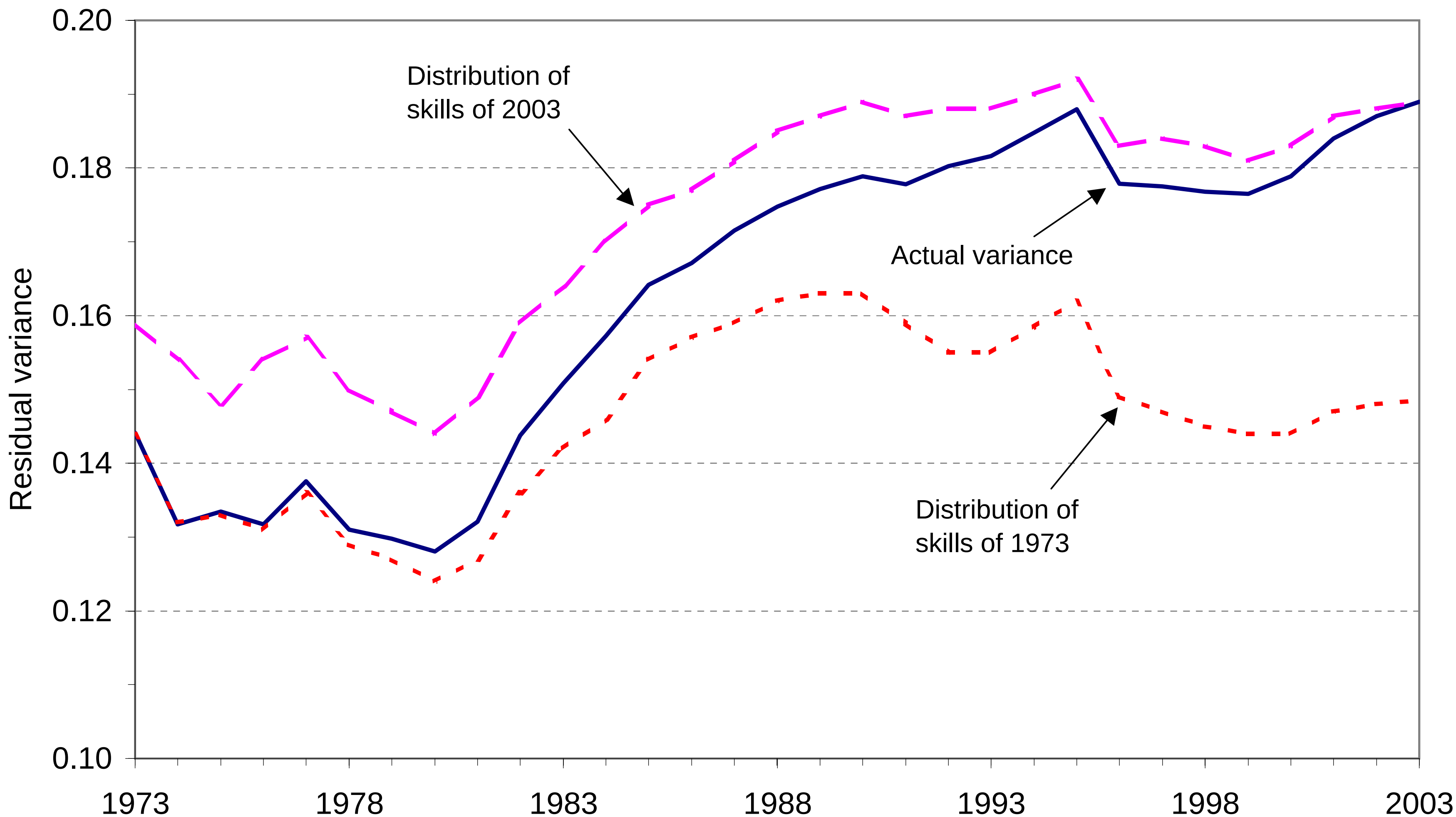


Figure 4a: Residual 90-10 wage gap for men, holding distribution of skills at their 1973 level

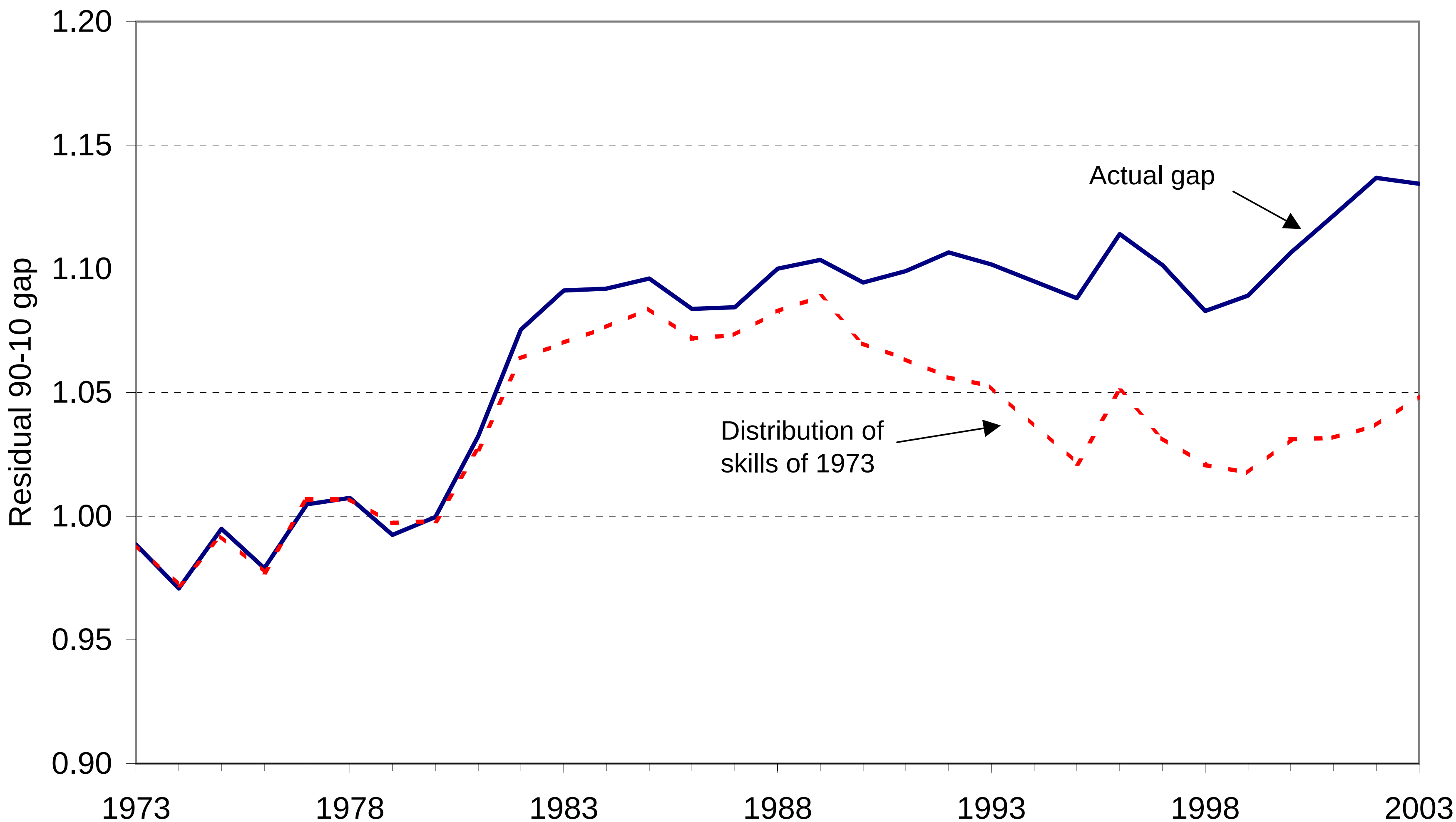


Figure 4b: Residual 90-10 gap for women, holding distribution of skills at their 1973 level

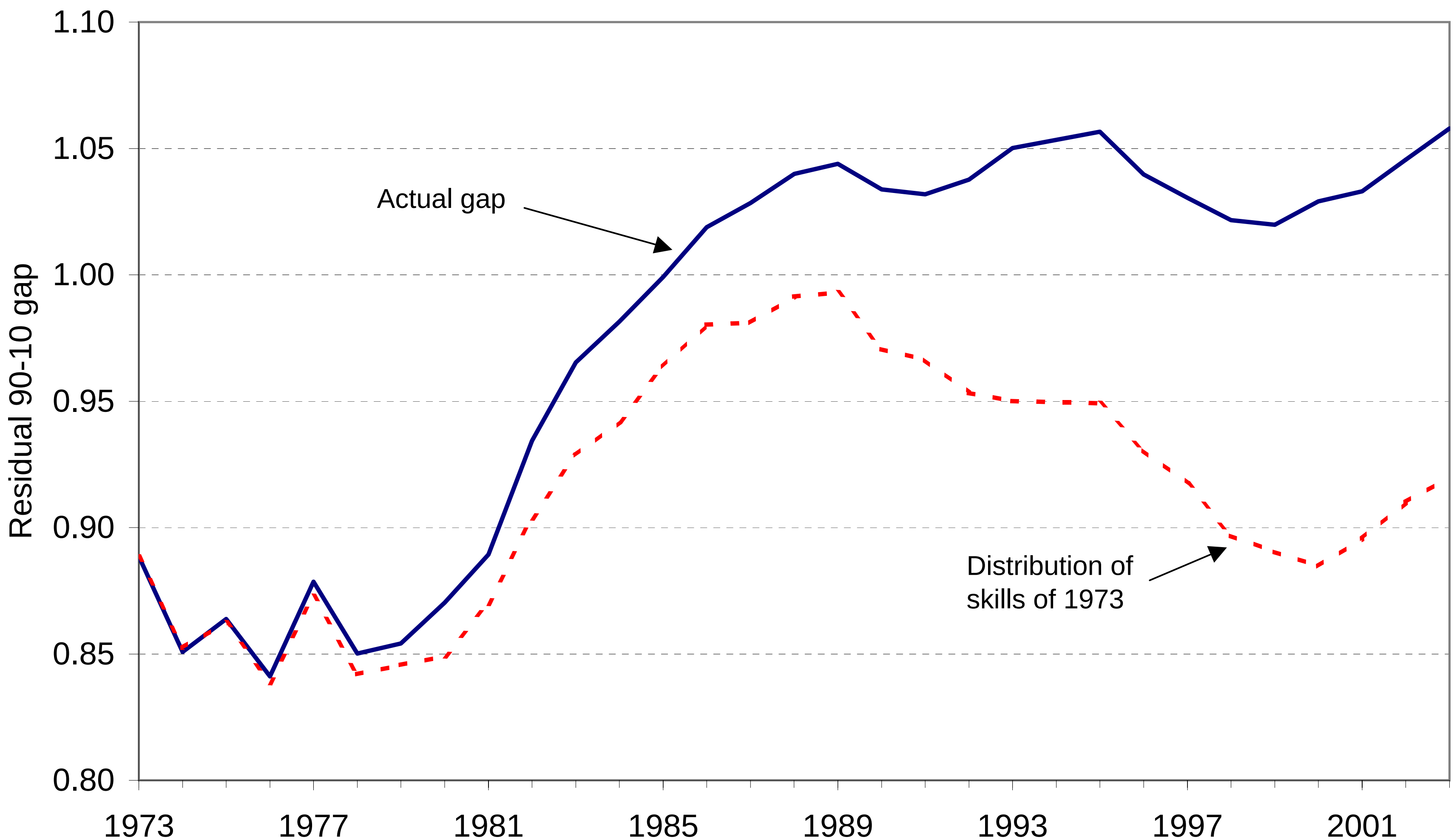


Figure 5a: Variance of log hourly wages of men with both May/ORG and March wages (matched sample)

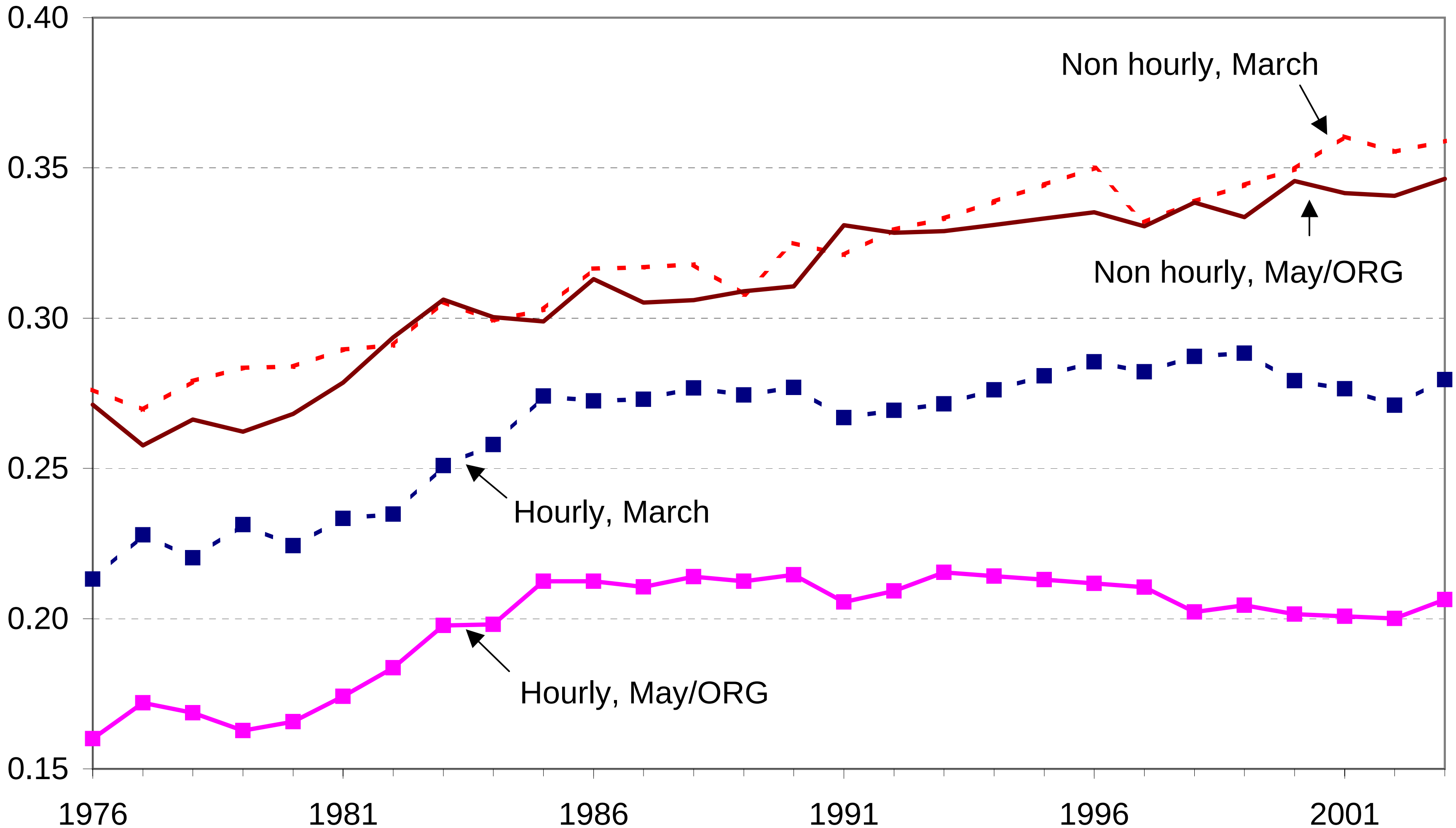


Figure 5b: Variance of log wages of women with both May/ORG and March wages (matched sample)

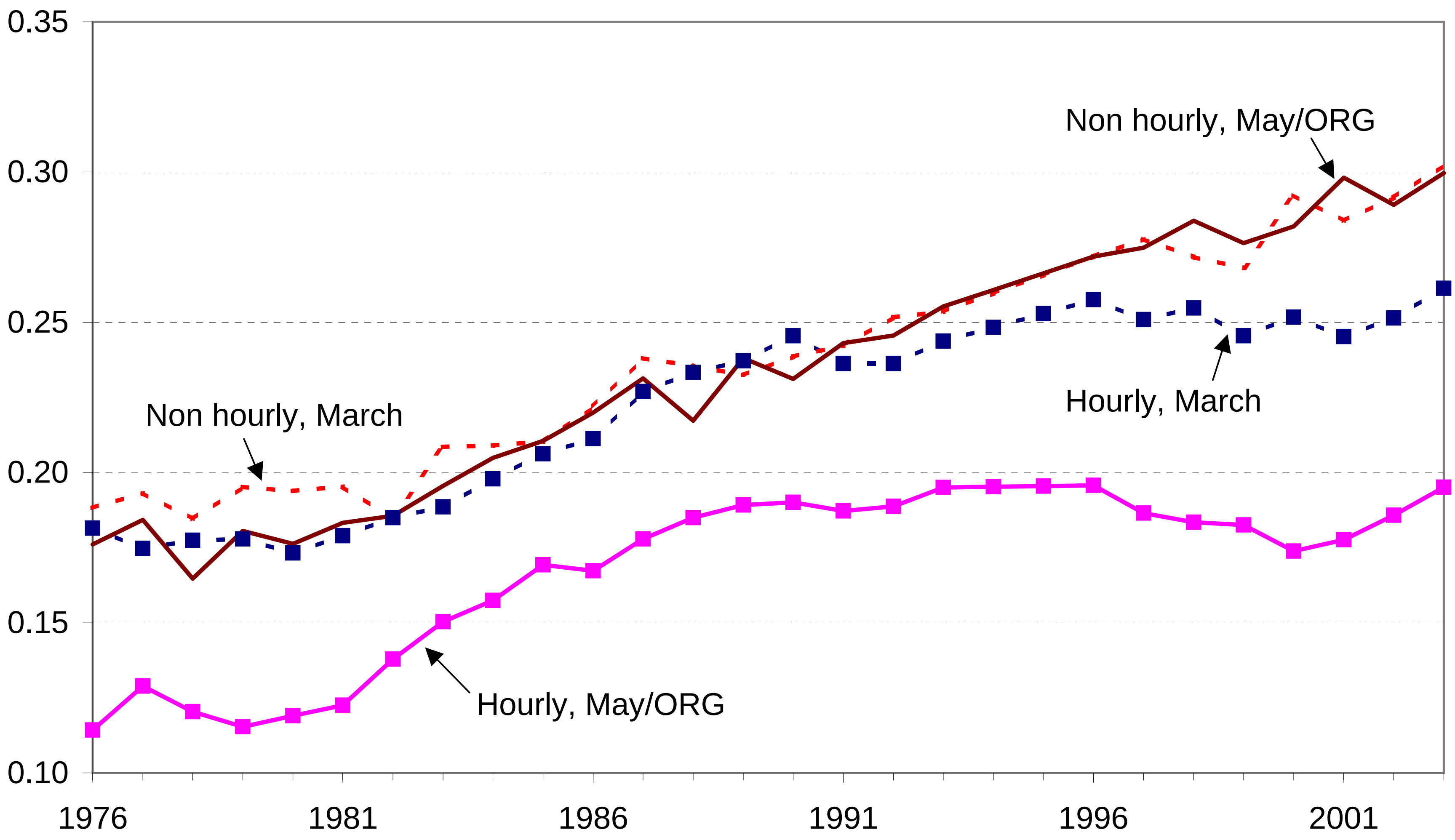


Figure 6a: Between-group variance of wages, men

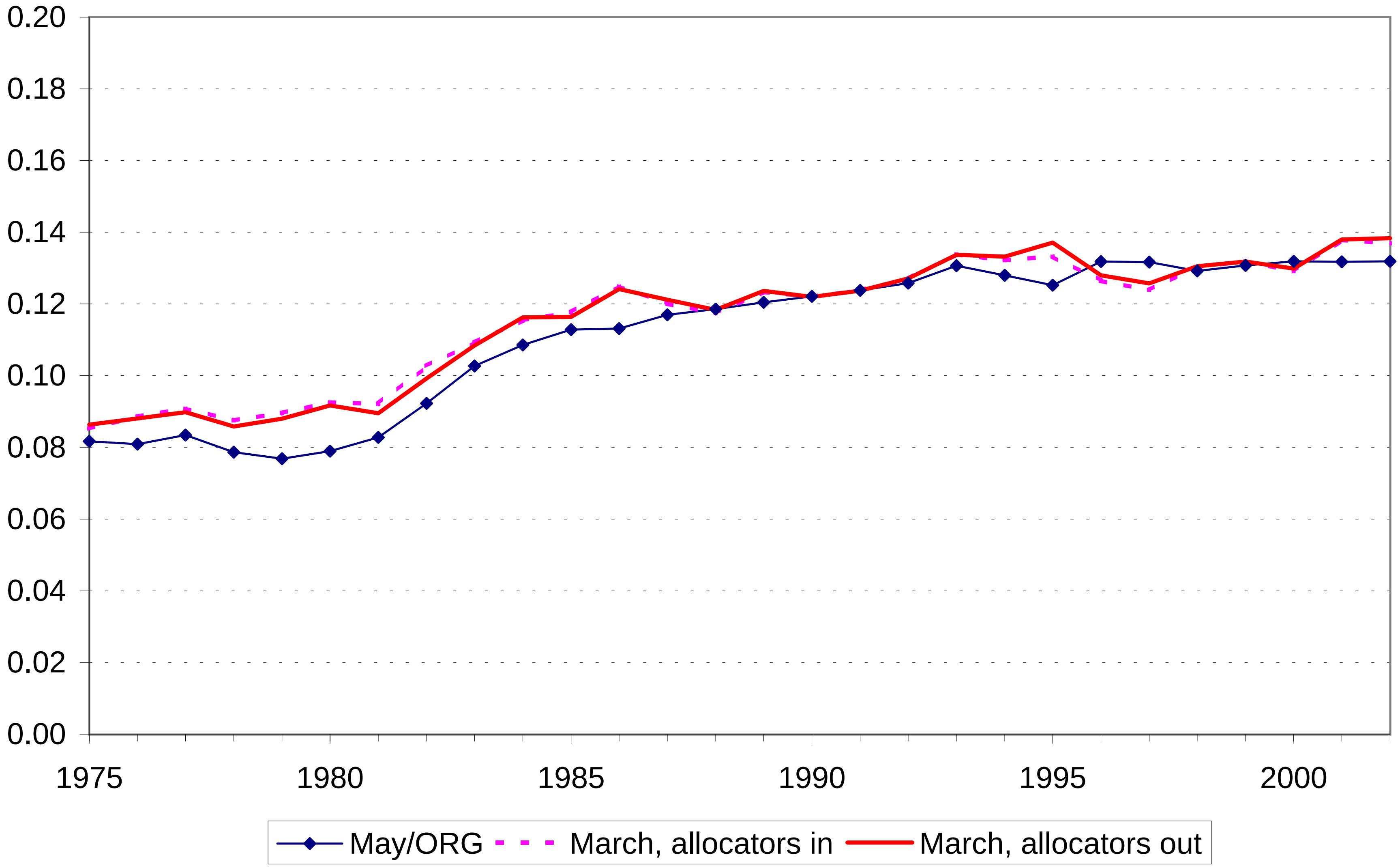


Figure 6b: Between-group variance, women

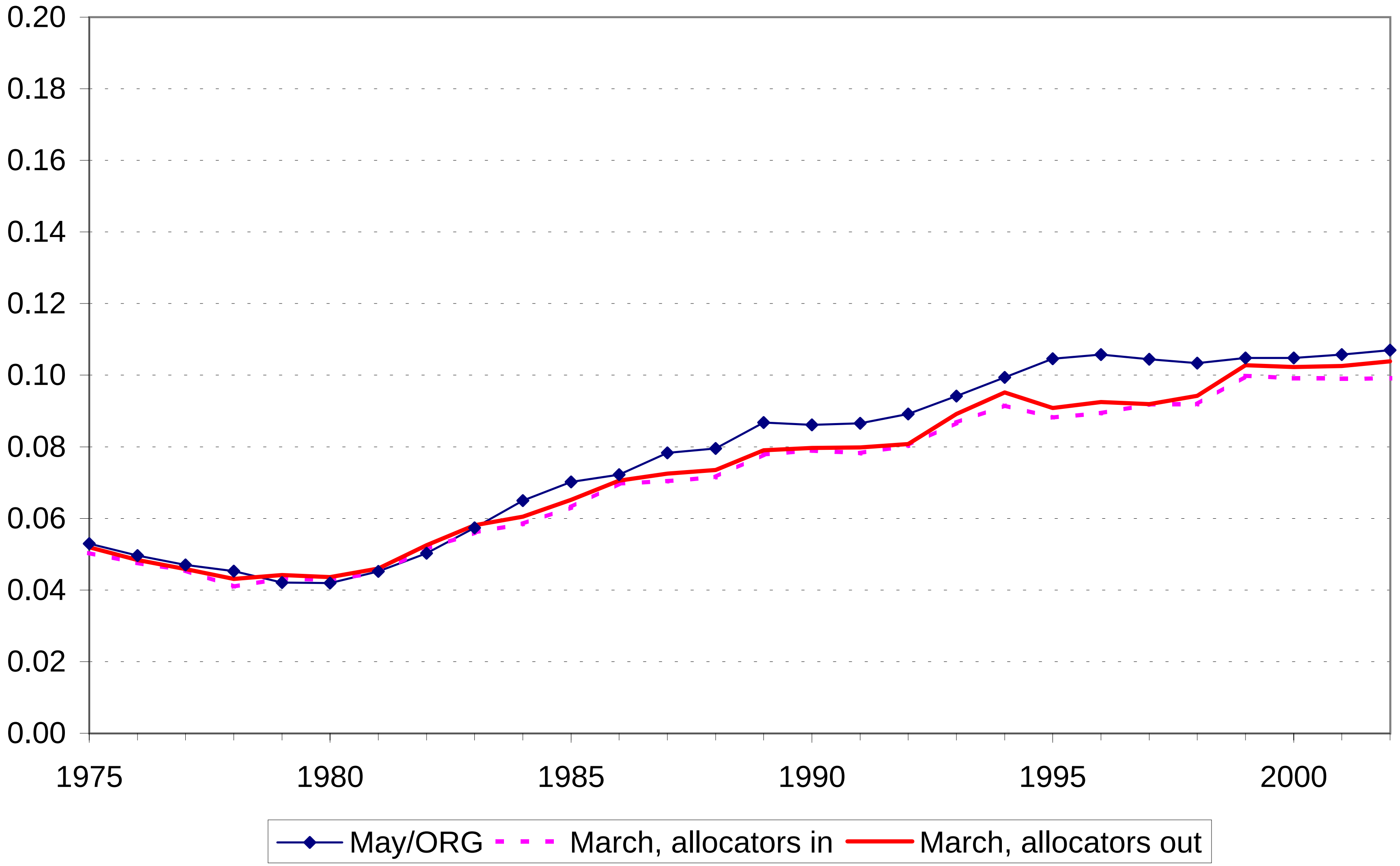


Figure 7a: Residual variance of wages, men

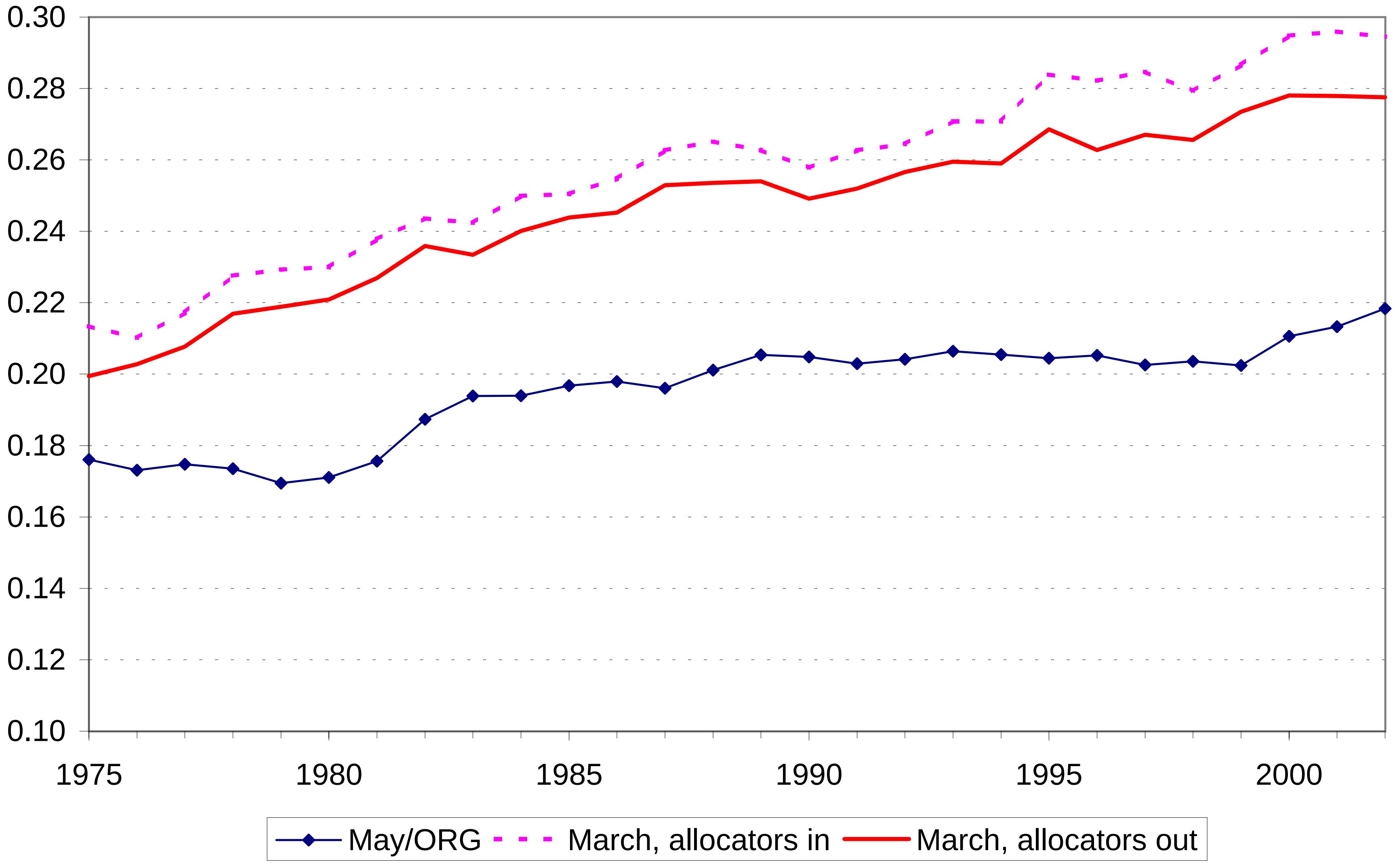


Figure 7b: Residual variance of wages, women

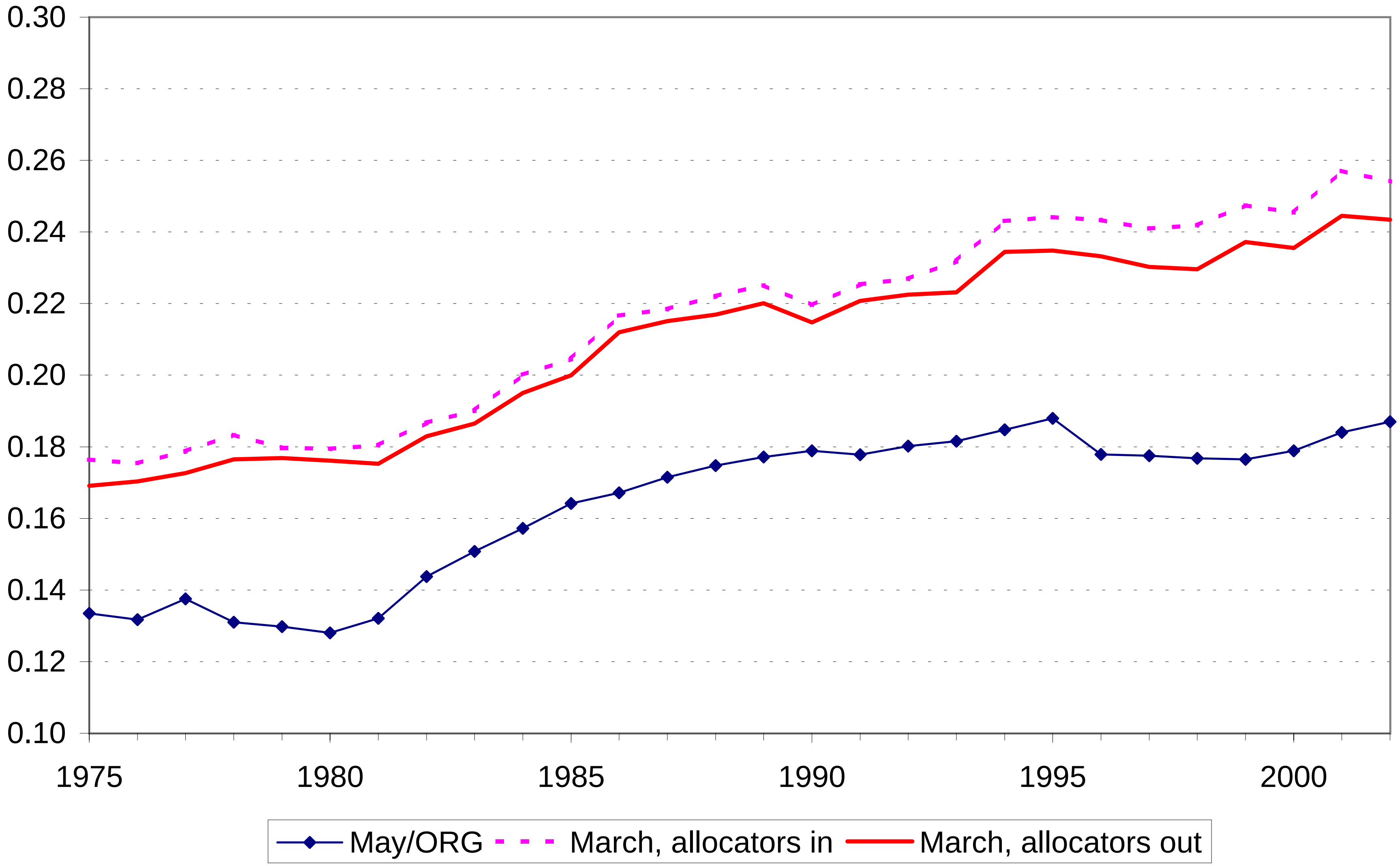


Figure 8a: Fraction of men paid by the hour, by education category

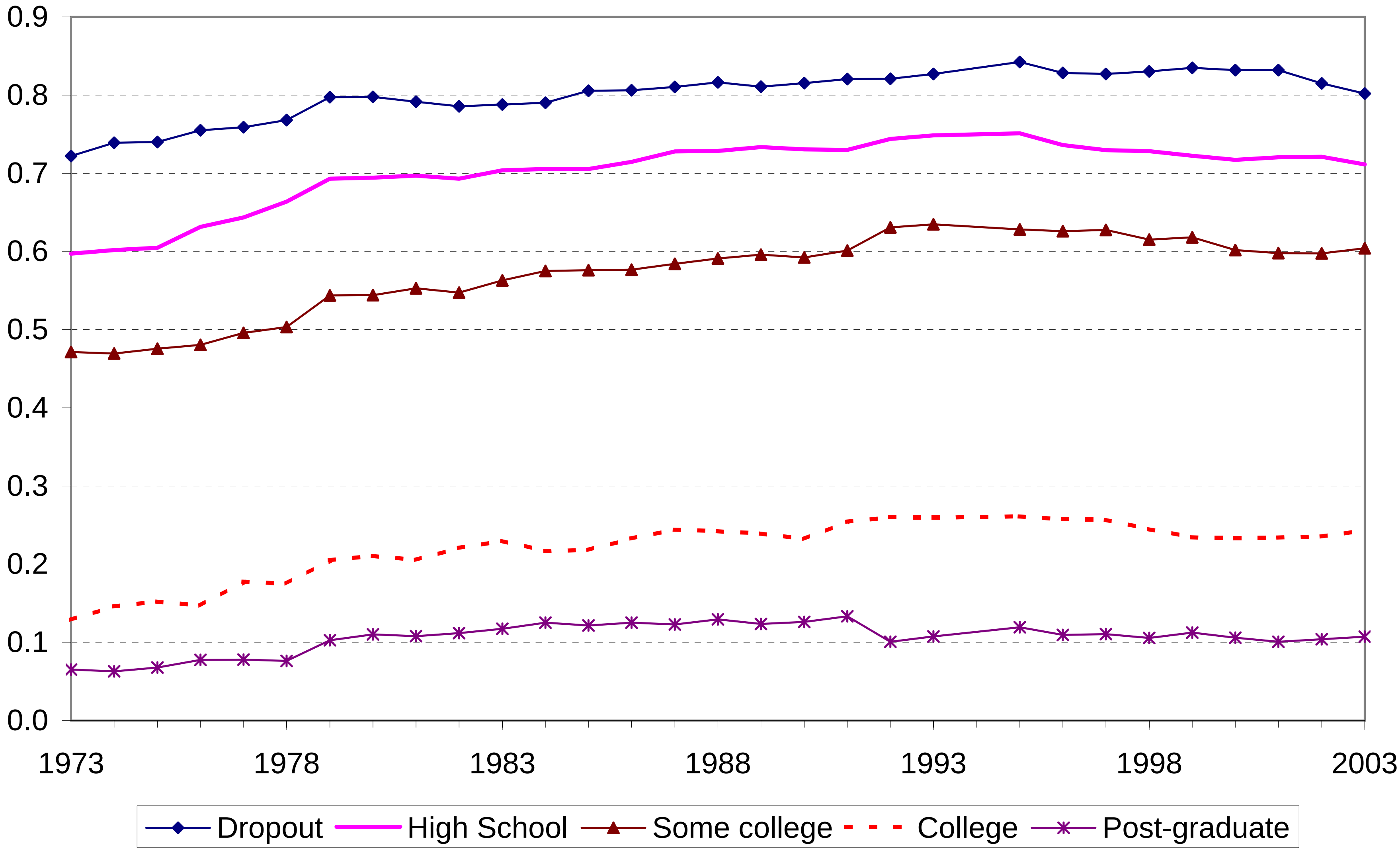


Figure 8b: Fraction of women paid by the hour, by education category

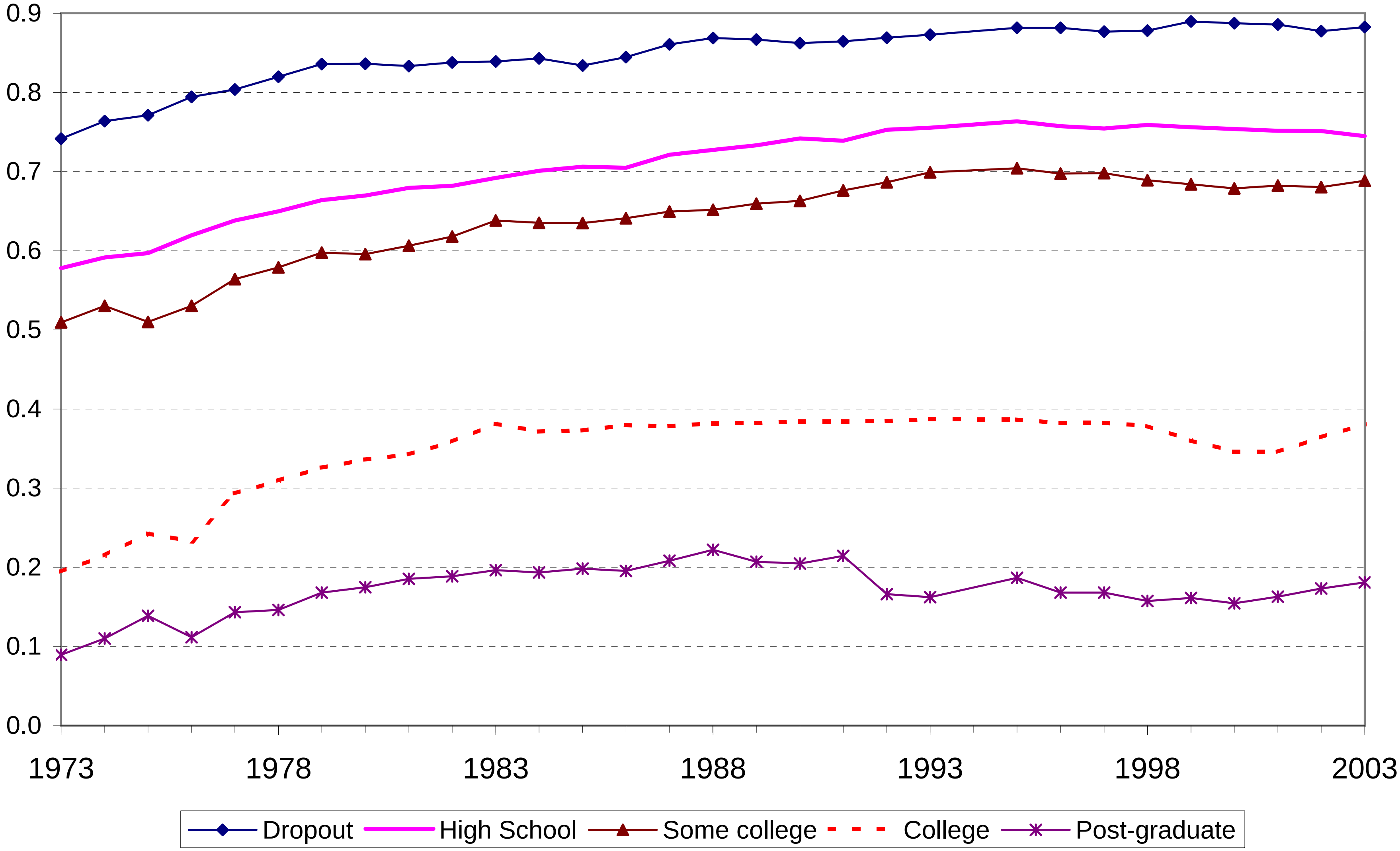


Figure 9a: Male residual variance predicted using the minimum wage (holding characteristics at their 1973 level)

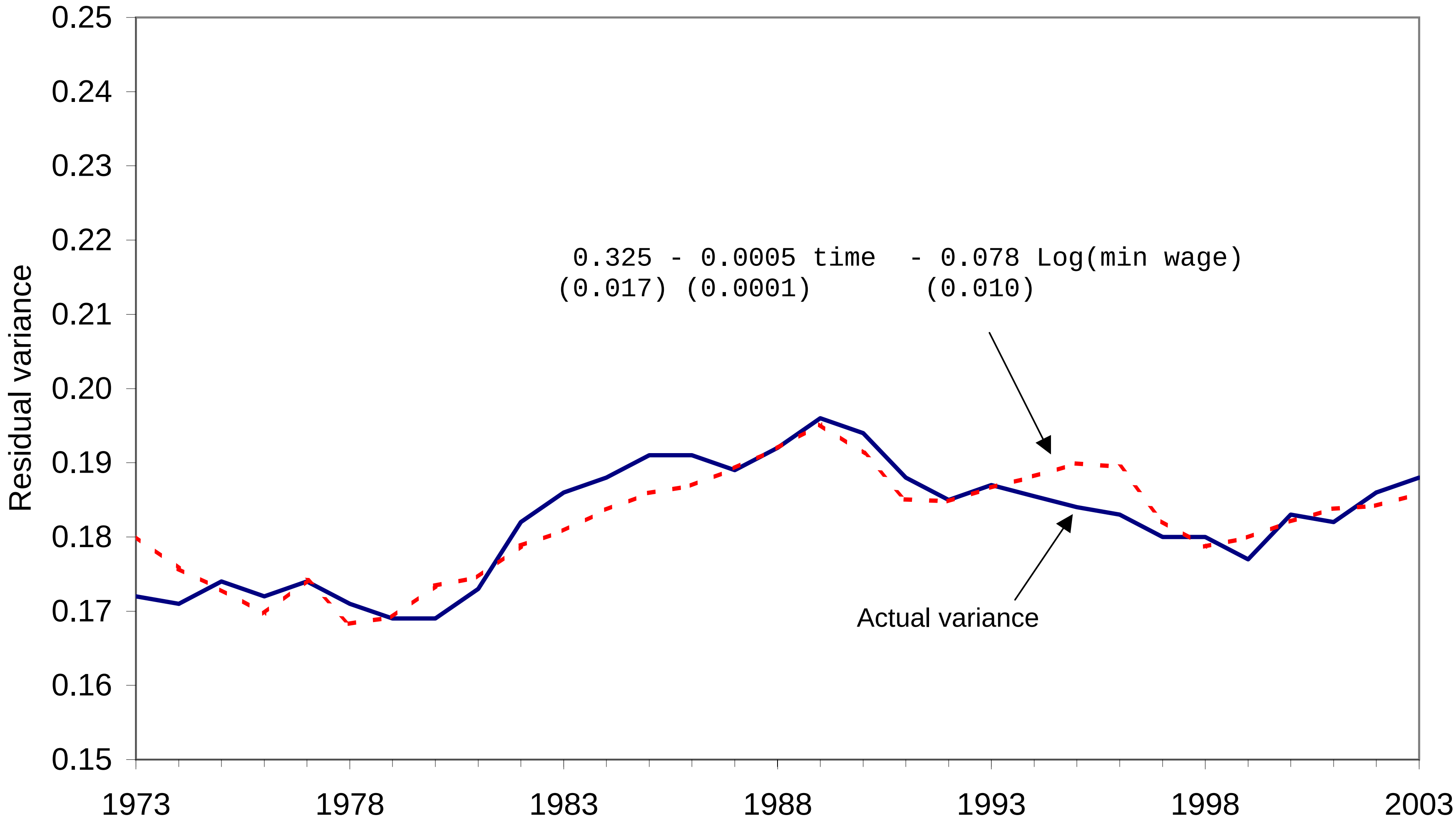
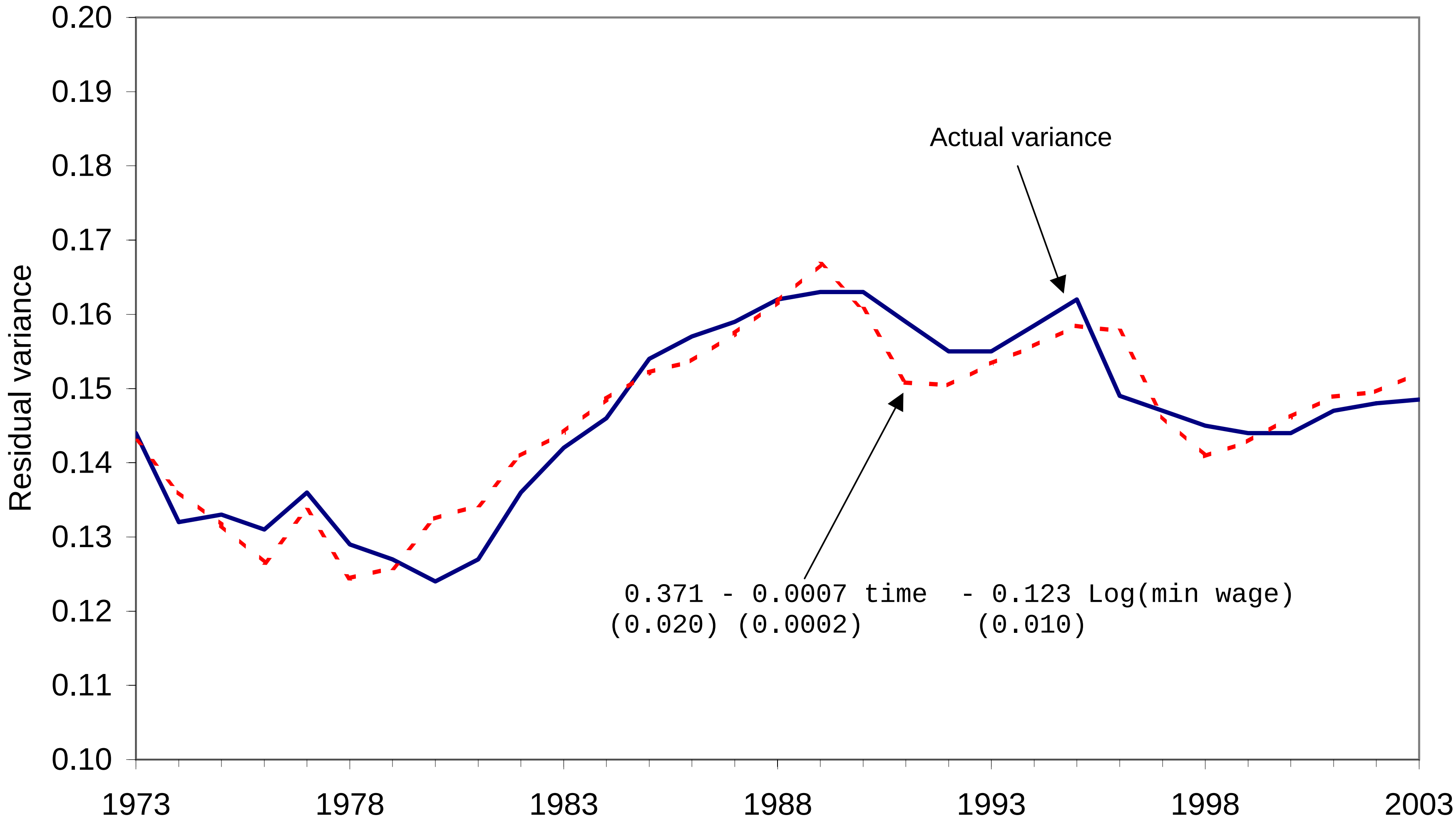
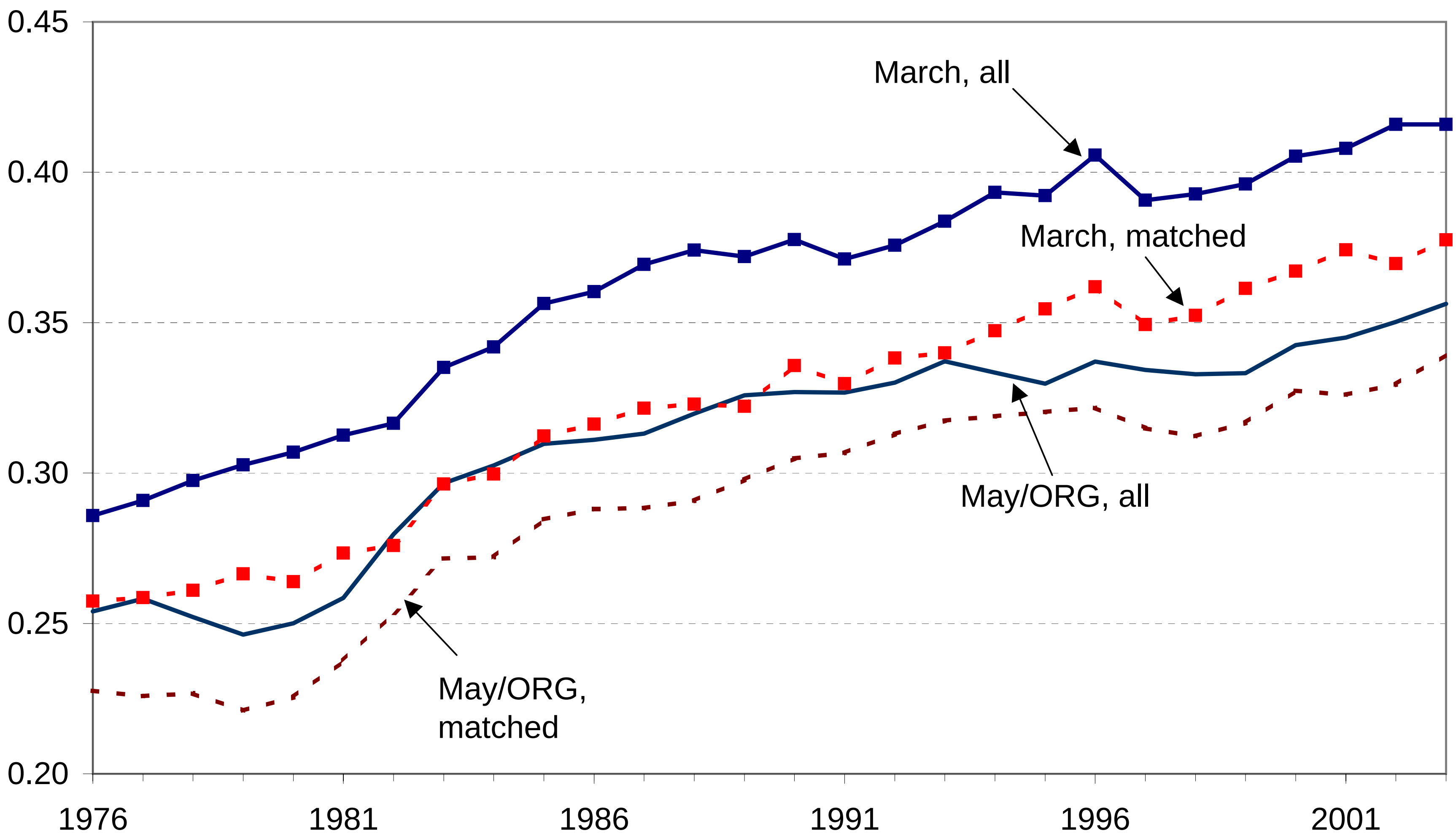


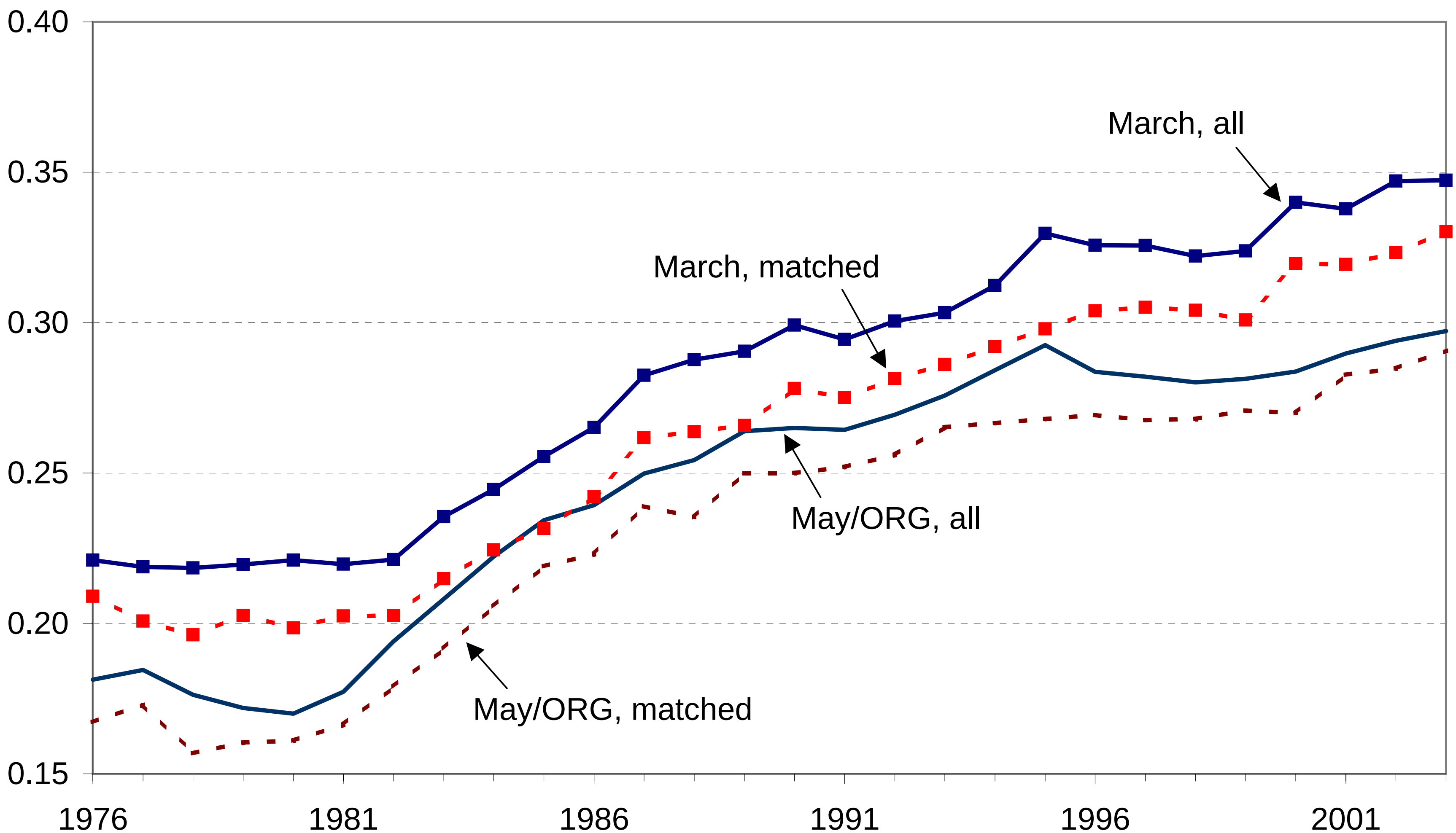
Figure 9b: Female residual variance predicted using the minimum wage (holding characteristics at their 1973 level)



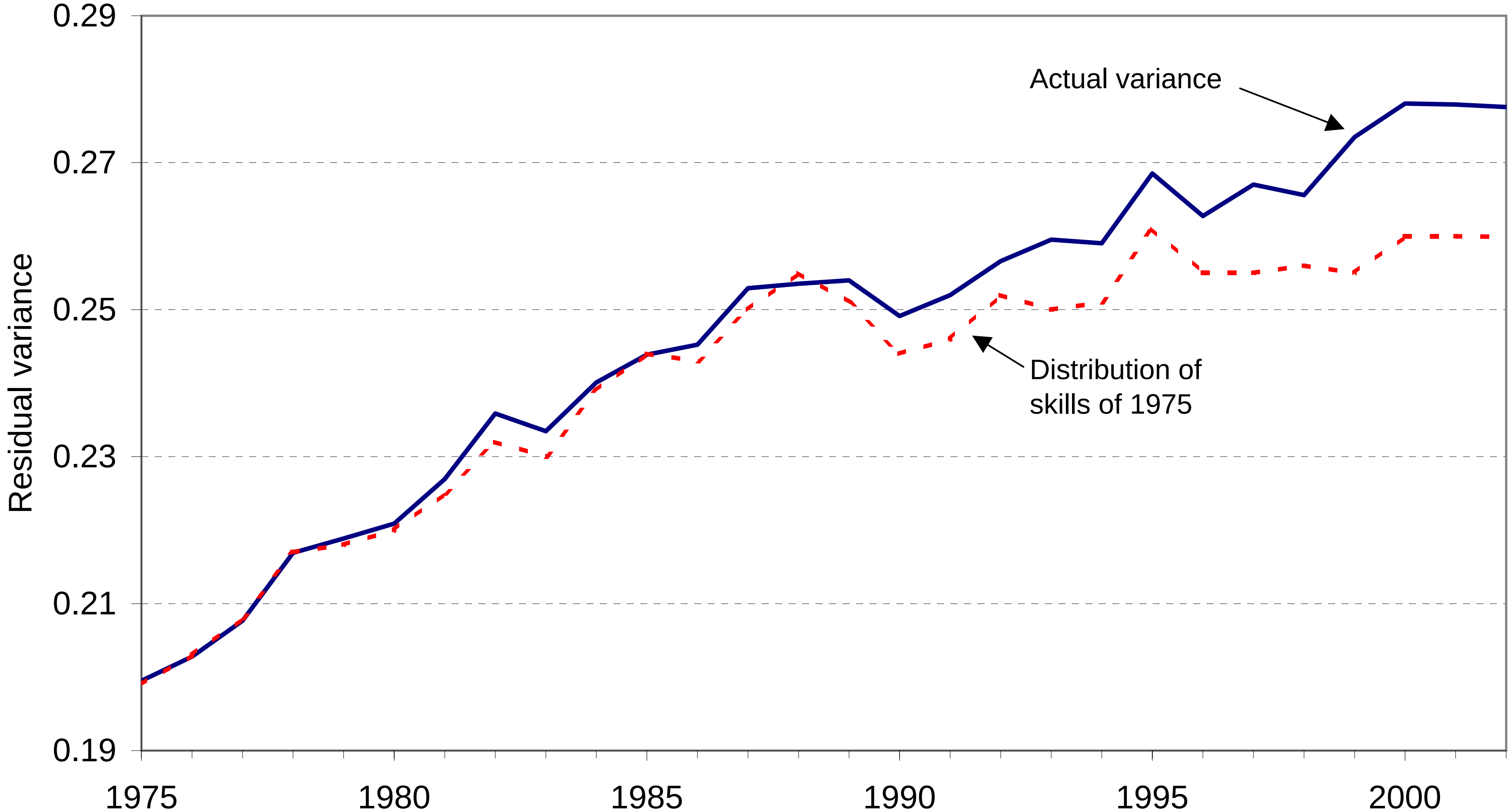
Appendix Figure 1a: Variance of log hourly wages for all workers and matched workers only, men



Appendix Figure 1b: Variance of log hourly wages for all workers and matched workers only, women



Appendix Figure 2a: Residual variance for men in the March CPS, holding distribution of skills at their 1975 level



Appendix Figure 2b: Residual variance for women in the March CPS, holding distribution of skills at their 1975 level