Earnings Volatility Across Groups and Time

John Sabelhaus[a,b]
Jae Song[c]

March 2009

Abstract

Inferences about earnings volatility across groups and time depend on underlying models of earnings dynamics, data sources, earnings concepts, and sampling strategies. In this paper we evaluate a model of earnings dynamics in which the permanence of shocks varies by age and education. This specification is consistent with observed earnings changes in administrative panel data, and also with the variance of earnings levels in multiple cross-section (synthetic panel) data. However, expanding the earnings concept to include self-employment and changing sampling strategy to include observations with minimal labor force attachment has first-order effects, and may help explain why some studies conclude that earnings volatility is rising.

JEL Codes: H22, J31

[a]Department of Economics, University of Maryland, College Park, Maryland 20742. Phone: (301) 367-3149. Email: sabelhaus@econ.umd.edu. [b]The Investment Company Institute, 1401 H Street NW, Washington, DC 20005. [c]U.S. Social Security Administration, Washington, DC. Views here do not represent those of the Investment Company Institute or the Social Security Administration.

1. Introduction

Earnings volatility is of fundamental interest to many different groups of economists. Researchers studying tax policy are interested in earnings volatility because of the implications for distributional analysis of tax burdens: in a progressive tax system, the more individual incomes vary over time, the more divergent are conclusions about effective tax rates when comparing annual and multi-year measures.[1] Earnings volatility plays a crucial role in macroeconomics, because of the impact that earnings uncertainty has on consumption behavior.[2] Recently, earnings volatility has become an interesting policy question in its own right, because a growing literature has argued that economic well-being has been adversely affected in recent decades because individual earnings became more volatile.[3] The conclusion that earnings volatility increased is far from universal, however, as other studies have found that variability in earnings growth was flat or even declining since 1980.[4]

Inferences about earnings volatility across groups and time depend on underlying models of earnings dynamics, data sources, earnings concepts, and sampling strategies. The analysis in this paper suggests that all of these inputs have likely played some role in the divergence of conclusions in previous studies. Our base-case model of earnings dynamics with stable permanent shocks that vary by age and education is consistent with

[1] See, for example, Congressional Budget Office (2005) and Cilke, et, al. (2001). Related tax-oriented research has focused on "mobility," meaning systematic movement of individuals across earnings groups over time—see, for example, Auten and Gee (2007) or Kopczuk, Saez, and Song (2007). Mobility in that context is closely related to the "permanence" of earnings shocks as described in this paper.

[2] Indeed, much of the basic research that led to the earnings decomposition in this and other papers was undertaken for the purpose of analyzing stochastic consumption models. See in particular Carroll (1992) and Deaton and Paxson (1994).

[3] The best known of these is a series of papers by Moffitt and Gottschalk in which they implement formal variance decompositions for earnings shocks. The most recent in this series is Moffitt and Gottschalk (2008), and their other seminal contributions were Gottschalk and Moffitt (1994) and Moffitt and Gottschalk (2002). Dynan, Elmendorf, and Sichel (2007) analyze the same data using a less formal approach and come to basically the same conclusions.

[4] See Congressional Budget Office (2008) and Sabelhaus and Song (2008).

a transitory component that has declined over time. However, altering the earnings concept to include self-employment income and changing the sampling strategy to include observations with minimal labor force attachment has first-order effects on the results. Indeed, those effects may be enough to explain why some studies conclude that earnings volatility is rising.

Three data sets are used in this paper. The first is a one percent longitudinal sample of earners ages 25 to 55 between 1980 and 2006 drawn at random from the Social Security Master Earnings File (MEF). The advantage of longitudinal data is that one can measure and analyze changes in log earnings over various time periods, which is the key to our first approach to modeling earnings volatility. The second is another longitudinal sample from the MEF, but this time for the 1940 to 1960 birth cohorts and linked to Survey of Income and Program Participation (SIPP) data files from the 1990s. For the purposes of this paper, the advantage of the SIPP data is that we know educational attainment, so we can investigate differences in earnings shocks by level of schooling. The third data set is a series of cross-sections from the March Current Population Survey (CPS) between 1970 and 2007. Although one cannot measure changes in earnings using the CPS cross-sections, the data are useful for analyzing the variance of earnings levels over time in a synthetic panel framework, and the addition of labor force attachment variables makes it possible to investigate how sampling may be affecting the results.

The starting point for the analysis here is the observation that the variance of earnings growth rates in the MEF longitudinal earnings data declined significantly over the period between the early to mid 1980s and the early to mid 1990s, and has since remained flat. A decline in the variance of growth rates does necessarily imply that volatility has fallen however, because labor economists have long recognized the importance of distinguishing permanent and transitory earnings shocks. Permanent shocks reflect differential earnings growth within some reference group that is expected to persist—another way to describe economic mobility. Transitory shocks are temporary (though not necessarily gone after one period) and thus associated with volatility per se.

There are a few different ways to use panel data to separate earnings growth variability into permanent and transitory components. We use an approach suggested by Meghir and Pistaferri (2004) for measuring the variance of permanent shocks that is both intuitive and robust to varying specifications for the time series properties of transitory shocks. Using that approach, our longitudinal earnings data suggest that (1) the variance of permanent shocks declines with age, (2) the variance of permanent shocks is higher for the college educated, and (3) the variance of permanent shocks has been constant (within age and education groups) over time.

If the variance of permanent shocks within groups has been relatively stable, and overall growth rate variability within groups is falling, that suggests the variance of transitory shocks (volatility) must have fallen. Indeed, that finding would be consistent with evidence from the literature on the "Great Moderation" in macroeconomics (see Davis and Kahn (2008)). Also, there was a significant secular decline in U.S. unemployment rates (at least through our sample period), and unemployment is the event one would generally associate with transitory shocks. However, the book is not completely closed on the decrease in transitory earnings shocks; Moffitt and Gottschalk (2008) argue it is possible that transitory shocks actually got bigger but became more serially correlated. Our residual approach—starting with total variation and subtracting permanent shock variation—cannot distinguish between changes in the transitory variance and changes (in the opposite direction) in the covariance of transitory shocks.

In addition to looking at the variability of earnings growth within groups and over time, one can characterize earnings dynamics by looking at the variance of earnings levels within an age group or birth cohort over time. The canonical model that underlies the transitory and permanent decomposition implies that the variance of earnings at any given age and point in time will depend on the variance of transitory shocks at that point in time, the initial earnings dispersion for the age or cohort group in question, and the accumulated permanent shocks since that initial time period. If the stochastic earnings process is stable, one should see stable earnings variances in the synthetic cross section.

The synthetic cross section we develop is based on March CPS data. The advantage of this data relative to our administrative records is that we know self-reported labor force attachment, so we can investigate how restricting the sample to those who worked full-time or eliminating diminimus earnings affects the answers. The CPS data confirms our inferences about a stable permanent shock variance (at least after 1980) while the overall variance of earnings (for our preferred measures) was falling, which is consistent with declining volatility. However, we also show that changing the sampling criteria to include observations with modest labor force attachment and very low self-employment earnings has a dramatic first-order effect on the estimated variances. The sampling criteria is very likely a part of the explanation for why researchers have disagreed about trends in earnings volatility.

2. Measuring the Variability of Earnings Growth Rates Over Time

Although the concept of earnings volatility may seem straight-forward in principle, measuring it in practice requires a number of decisions about data, sampling, and which summary statistic to use. In this section we use administrative panel data from Social Security earnings records to present some basic findings about the standard deviation of one-period earnings growth rates over time. Although the estimated level of variability at any point in time depends on how the sample is chosen, all of the measures we present suggest a general decline in variability between the early 1980s and mid 1990s, and little change thereafter through the end of our sample in 2006.

There are two main longitudinal data sets used in this section and the two sections that follow. Both data sets are ultimately based on Social Security earnings records in the Master Earnings File (MEF).[5] The first sample is a one percent random draw from the MEF, and the second is a draw from the MEF based on linkages to several panels of the Survey of Income and Program Participation (SIPP). The SIPP-linked data is useful because it introduces more demographic information than is available on the Social Security records—for our purposes, the SIPP provides the level of educational attainment used to further subdivide the sample when looking at various types of earnings shocks in subsequent sections. In both samples, data on wages from W2 reporting is used for all years back to 1980, and data on wages plus self-employment earnings is used for years after 1994.[6]

---

[5] The appendix has a complete description of the data sets used in this paper.

[6] Data on wages from W2s actually goes back to 1978, but the sampling in the one percent file was not really representative until 1980. Before 1978, the wage and salary information is reported only up to the Social Security taxable maximum, which limits practical use for volatility studies. The self-employment data (taken from Form SE) was subject to top coding until the ceiling on taxation for Medicare was lifted in the early 1990s.

The decision about which age and/or cohort groups to include in the sample is somewhat dependent on the question being asked. The one percent random sample results presented here are generally based on ages 25 to 55—what we refer to as prime working years.[7] In the SIPP, the main sample is drawn for birth cohorts 1940 to 1960. The cohort restriction is set so the sample is mid-career around the points in time when the SIPP linkages are established (1990 through 1996) while generally assuring that the educational attainments are effectively completed for every observation (that is, the links are established for everyone 25 or older at the time of the survey).

The concept of earnings growth variability used in this section is the standard deviation of the one-period change in log earnings. The standard deviation is a convenient statistic to focus on for a few reasons: the extent of variability is summarized in one number, and it is useful as a starting point for distinguishing permanent from transitory earnings shocks. There are also drawbacks, because the standard deviation may be hiding important information about the symmetry of shocks, and some very large percentage changes may dominate the results even though their meaning is dubious. In particular, a large percentage change can mean large dollar changes near the average earnings, but it can also mean relatively small dollar changes at very low earnings. [8]

Figure 1shows the standard deviation of one period changes in log earnings for the one percent MEF sample for 1981 through 2006. There is a significant decline in earnings growth variability between the early to mid 1980s and the early to mid 1990s. However, there are two important observations about Figure 1 that underscore the need

[7] The conclusions do not change if we expand the sample to include everyone age 21 to 64.

[8] Jensen and Shore (2008) focus on changes in the distribution of volatility over time, and find that the overall average increase in the PSID has been dominated by changes in variability for the most volatile households.

for an alternative sampling strategy in the remainder of the paper. First, the scale for the standard deviation is huge, ranging from about 90 log points in the early 1980s to just over 70 log points today. Simply interpreted, this means that something like one-third of the observations have absolute earnings growth rates of 70 percent or more. The second observation is that the growth rate of the sum of wages and self-employment is actually less variable than wages alone, which is counter-intuitive.

Measuring the change in log earnings for any individual in two subsequent periods imposes a natural sampling restriction: their earnings have to be positive in both years. However, that minimal sampling restriction means that observations with very low earnings in any given year can have disproportionate effects on the estimated standard deviation. In particular, a person whose earnings fall from $50,000 to $25,000 would make the same contribution to the standard deviation as one whose earnings fell from $1,000 to $500, because both are fifty percent declines. The significance of the two declines is obviously very different.

In Figure 2 we impose a sampling restriction based on an arbitrary but useful criterion: observations are included only if the earnings level indicates significant labor force attachment, which we set at the amount needed to qualify for a year of Social Security credits.[9] The qualifying amount varies over time with average wages and was $3,680 in 2005. One way to think about the Social Security coverage threshold is this: a person crossed the coverage threshold if they worked 715 hours at the federal minimum wage, which was $5.15 in 2005. That is either about 14 hours per week for a full year, or

[9] Most other studies of earnings volatility use other restrictions to mitigate the effects of diminimus earnings; for example, Moffitt and Gottschalk (2008) focus on male heads of households. The Social Security earnings data does not allow us to sample based on specific measures of labor force attachment, but we explore that approach with the CPS synthetic panels in the last section of this paper.

18 weeks full time. In the one percent MEF this eliminates just over 10 percent of the sample with positive earnings in any given year, and there is no trend over time.

The striking difference between Figure 1 and Figure 2 is the scale against which the standard deviations are plotted: the scale in Figure 2 is exactly half the scale in Figure 1. This indicates that the bottom ten percent of the sample in terms of earnings levels causes the estimated standard deviation to double. However, looking beyond that effect, the conclusions about trends over time are nearly identical: the standard deviation of earnings growth rates fell significantly between the early to mid 1980s and the early to mid 1990s, and has remained flat ever since. In addition, the variability of total earnings (wages plus self-employment) growth is now above the variability of wage growth, which is certainly more intuitive than the relationship in Figure 1.

Figure 3 shows the same standard deviations of one-period earnings growth rates, but this time the sample is restricted to the 1940 to 1960 birth cohorts. This restricted sample was somewhat younger than the one percent random draw (ages 25 to 55) in beginning of the period, and somewhat older by the end. The oldest person (born in 1940) was only 40 at the beginning of the sample, and the youngest (born in 1960) was 45 at the end. Given the results to follow in subsequent sections, it is not surprising that the decline in earnings variability is larger for this group, because earnings growth variability is higher at younger ages. However, the basic conclusions from Figures 1 and 2 show up clearly: restricting the sample to earnings above the Social Security threshold has a first-order impact on the level, but not on changes over time. Figure 3 also shows that the SIPP-linked file has the same basic patterns as the one percent MEF for the 1940 to 1960 cohorts.

3. Permanent Shocks Over the Life Cycle

The standard deviation of one-period earnings change presented in the last section is a good starting point for the analysis of earnings volatility. One key question that arises is whether any given unpredictable change in earnings growth (or "shock") is permanent or transitory in nature. In this section we use a simple technique suggested by Meghir and Pistaferri (2004) to show that the size of permanent shocks varies by age and education. In particular, the standard deviation of permanent earnings shocks falls significantly as people move from the beginning towards the middle of their working careers, and the variance of permanent shocks is much higher for the college-educated than other groups at any given age.

The usual starting point for decomposing earnings shocks in micro labor economics is the canonical permanent and transitory earnings shock model,

$$y_{it} = \mu_{it} + \varepsilon_{it}$$

$$\mu_{it} = \mu_{it-1} + \eta_{it}$$

where y is log earnings, μ_{it} is the slowly evolving permanent component that changes by η_{it} each period, and ε_{it} is the transitory component. In the simplest versions the transitory and permanent shock terms (ε_{it} and η_{it}) are assumed uncorrelated with zero means and constant variances (σ^2_T and σ^2_P).

In practical applications, the level of y is replaced by the gap between actual y and the predicted value of log earnings based on observable characteristics for each individual—the so-called "earnings differential." In this case the two error terms are more appropriately described as exogenous zero-mean shocks, because the explainable part of earnings growth by age, education, and other observables is removed. Thus, the canonical model basically explains how individual's earnings evolve over time relative to their group average.

Even the simplest implementations of the canonical model also generally acknowledge an ARMA structure for ε_{it}. That is, shocks to earnings are not perfectly transitory or perfectly permanent—some shocks may have a large initial effect but then persist (with declining effects on earnings) for a few years, while others lead to permanent shifts in the earnings differential (at least until another permanent shock comes along). In general, the ARMA specification makes estimation of the canonical model somewhat more complicated, because the distinction between transitory and permanent will depend on exactly how the ARMA is specified and estimated, and in particular, which parameters are allowed to vary and along which dimensions.[10]

A recent paper by Meghir and Pistaferri (1994) offers a key insight into the nature of permanent versus transitory shocks that makes it possible to work around the ARMA specification issue. This insight was also used extensively by Jensen and Shore (2008) in their analysis of growth rate variability in the Panel Survey of Income Dynamics (PSID), and what follows builds directly on that analysis.[11] The suggested Meghir-Pistaferri moment that identifies permanent shocks under very general ARMA structures is given by,

[10] See Moffitt and Gottschalk (2008) for an excellent discussion of the problems inherent in distinguishing between permanent and transitory shocks. The analysis in that paper is the motivation for the approach taken here.

[11] Although it is not the focus of the current paper, it is worth noting that a quick check of the distribution of the Meghir-Pistaferri moment in our data reveals the same skewness that Jensen and Shore (2008) find in the PSID.

$$V_t = \Sigma_i (y_{it} - y_{it-2})\cdot(y_{it+2} - y_{it-4})$$

The intuition underlying this estimator is straight-forward: high frequency (two year) changes in residual earnings growth rates that are correlated with low-enough frequency (six year) changes in residual earnings growth rates around the same period can be characterized as permanent shocks.

Figure 4 shows standard deviations of the Meghir-Pistaferri moment by age for the entire one percent MEF sample with earnings above the Social Security threshold. There are two sets of estimates for wages: the first averages the estimated moments by age for the first half of the sample (1984 to 1993), and the second averages the estimated moments for the second half of the sample (1994 to 2004). The third line shows the average of the estimated moment by age for the sum of wages and self-employment earnings, which covers the period 1994 to 2004.

All three lines on Figure 4 indicate that the standard deviation of permanent shocks declines noticeably with age. This result is intuitive, because earnings distributions widen faster at younger ages as people are being sorted into their ultimate lifetime earnings groups (indeed, this principle is the basis for analyzing the variance of earnings levels in the last section of this paper). The results in Figure 4—both patterns and magnitudes of permanent shocks at each age—are consistent with estimates of permanent shocks generated using a different approach to error decomposition in Sabelhaus and Song (2008).[12] Figure 4 also provides two other key pieces of

[12] The alternative approach, generally traced back to Carroll (1992), relies on the fact that earnings growth measured at different frequencies will have different combinations of transitory and permanent shock

information: the data suggest that permanent shocks to total earnings at any given age are slightly larger than permanent shocks to wages alone, and that there was little if any change to the levels of permanent shocks between the two halves of the sample period.

Figure 5 shows that the variance of permanent shocks depends on educational attainment as well as age. This figure is constructed using the SIPP linked data with the wages-only data for the entire 1984 to 2004 period, so the cohort aging effect described in the last section is a possible confounding factor. However, the fact that the levels of permanent shocks at any given age do not seem to be varying across time (Figure 4) suggests that the overall levels of the Meghir-Pistaferri estimator (at any given age) should be comparable. Figure 5 bears this out, as the average for all education groups in the SIPP-linked data is nearly identical to the averages for the one percent sample (wages only) in Figure 4.

The key contribution of Figure 5 is that permanent shocks vary with both age and education. In particular, the variance of permanent shocks is larger for the college educated at any given age, which suggests more dispersion in their earnings (around the overall average earnings) at any given age. The differences in trajectories within the college-educated population are much larger than in other education groups, and this is reflected in larger (permanent) differences in growth rates, especially at younger ages. The implication is that the variance of earnings levels in the college educated population will be rising faster at any given age, which cross-section data confirms.

components. Sabelhaus and Song (2008) use that logic to separate permanent and transitory shock variances across ages, education, and time. The approach used in Sabelhaus and Song (2008) can be criticized because the ARMA structure is ignored in the estimation; the fact that those results and the results here based on Meghir and Pistaferri's moment estimator are so close suggests that ignoring the ARMA structure is not a problem after all.

4. Permanent Shocks and Transitory Residuals Over Time

In the canonical model each individual's idiosyncratic earnings differential is subject to permanent and transitory shocks in every period. Since the overall variability of earnings growth depends on underlying permanent and transitory shocks, one is tempted to say that transitory shocks (volatility) must have declined the U.S. since the early 1980s. This follows from the observations that overall growth rate variances fell (Section 2) while the variances of permanent shocks have remained stable (Section 3). However, subtracting the variance of permanent shocks from the overall variance in each period leaves a combination of terms that involves transitory variances and covariances, which is best described as a transitory "residual."

The statistical derivation of the transitory residual using our one percent MEF administrative panel data begins with Figure 6. The estimated permanent shock standard deviations using the Meghir-Pistaferri moment calculations are shown for the overall working-age average and three ten-year age groups for each calendar year 1984 to 2004. Although there is some variation in the estimates over time, the scale of Figure 6 makes it clear this variation is modest. There is little or no trend in the overall average or the average within any given age group, which confirms the stability of the permanent shock process over time.[13]

The transitory residual is the difference between the variance of earnings growth rates and the variance of permanent shocks. Figure 7 shows these differences in standard

[13] The previous section showed that one might expect these estimated permanent shocks to vary as the education-composition of each group evolves over time. The education effect is not biasing Figure 6 because changes in educational attainment (at these ages) have been fairly modest since 1980, and the effects of educational differences are small relative to the effects of age (see Figure 5). The estimates in Figure 6 are based on the entire one percent MEF in order to maximize the sample size and to make sure the sample is representative by age in each year. Having said that, the same picture based on the linked SIPP (available from the authors) shows the same basic stability.

deviation terms for the same age groups and time periods in Figure 6. The basic message is that overall earnings growth variability was falling within each age group while permanent shocks variances were constant, and thus the transitory residual was declining. The decline in residual variability was gradual and lasted through the late 1990s, before stabilizing or even beginning to increase around 2000, which is consistent with general observations about macroeconomic volatility.

The derived residual standard deviations in Figure 7 are not estimates of the standard deviation of transitory shocks. As Moffitt and Gottschalk (2008) point out, the only thing we know from the canonical model is that the gap between overall earnings growth variability and permanent shock variances is a combination of three terms,

$$Var(\Delta y_{it}) - Var(\eta_{it}) = Var(\varepsilon_{it}) + Var(\varepsilon_{it-1}) - 2\cdot Cov(\varepsilon_{it},\varepsilon_{it-1})$$

Thus, one cannot say based on Figure 7 that the standard deviation of transitory shocks declined, because it is possible that changes in the covariance of transitory shocks increased. However, this observation about the transitory residual may just underscore the fact that volatility itself is an imprecise concept. When do highly correlated transitory shocks cease to be part of volatility and become part of mobility?

5. The Variance of Earnings by Age and Cohort

The canonical earnings shock model also has testable implications for the variance of earnings levels by age and cohort. For any given age group and at any point in time, the variance of earnings is the sum of the transitory variance, the initial (or age zero) permanent differential variance, and the cumulated permanent shock variances. This characterization suggests a particular trajectory for the variance of earnings levels for any given cohort as they age.[14] In this section we use synthetic panels from CPS cross-sections between 1970 and 2007 to analyze earnings variances across cohorts and time.[15] The results are consistent with our findings from looking at the variance of earnings growths using panel data, and the additional information in the CPS (especially labor force attachment) may be a clue as to why researchers have reached differing conclusions about trends in earnings volatility.

In the canonical earnings shock model the variance of log earnings for a given cohort at a point in time is sum of three terms. Each cohort starts at the beginning of their working period (age zero) with an initial dispersion of log earnings differentials $\mathrm{Var}(\mu_{i0})$. In every year there are permanent shocks (η_{it}) that accumulate over time and transitory shocks (ε_{it}) that disappear after one year. If we check in on this cohort at some time t, the variance of log earnings differentials will be,

$$\mathrm{Var}(y_{it}) = \mathrm{Var}(\mu_{i0}) + \Sigma_t \mathrm{Var}(\eta_{it}) + \mathrm{Var}(\varepsilon_{it})$$

[14] This is a key insight from Deaton and Paxson (1994).
[15] The March CPS data used in this paper was downloaded from the CPS-IPUMS site at the Minnesota Population Center (see King, Ruggles, Alexander, Leicach, and Sobek (2004).

Using this perspective, one can evaluate the stability of the stochastic earnings process by measuring log earnings variance at a particular age over time, or by tracing log earnings variance for various birth cohorts by age.

Figure 8 shows variances of log wages for males age 30 to 39 over time computed a few different ways.[16] The estimates which lead to the highest variances in every year are both based on the entire sample of people with positive earnings. The difference between the top two lines is that the highest variance in each year is for the log of wage levels, while the second highest is the log of the variance of wage differentials. The wage differential has the effect of age and education removed (it is the residual after subtracting the mean). The gap between the top two lines rises over time because the earnings differences across education groups became more pronounced.

The two sets of estimates which lead to lower overall variances are both for earnings differentials, but these are based on restricted samples meant to eliminate the dominating influence of diminimus earnings. The solid line is based on a sample selected the same way as in our earnings growth calculations in the previous sections: an observation is included only if earnings are above the Social Security qualifying threshold. The dashed line which lies close to the solid line is based on excluding observations that had earnings but self-reported that they were not full-time workers. As with our growth rate variance estimates based on the panel data, the effects of the sample restrictions are first order.

From the perspective of the canonical earnings shock model, the key insight of Figure 8 is that all of the measures indicate that the variance of earnings rose between

[16] We are motivated to focus on males 30 to 39 because this is the group Moffitt and Gottschalk (2008) use to illustrate why they think earnings volatility is rising. We find—as they do—that the conclusions do not depend on exactly which age group is used to make the point.

1970 and the early 1980s, but has since fallen or remained stable. The increase in log earnings variance throughout the 1970s is consistent with an increase in earnings growth variability, as first reported by Gottschalk and Moffitt (1992). The decline/stability after 1980 is consistent with the findings presented in this paper.

One clue about why other researchers looking at recent data might be coming to different conclusions is shown in Figure 9. The only difference between Figure 8 and Figure 9 is the measure of earnings: Figure 8 is wages only, while Figure 9 includes wages and self-employment earnings. The scales of Figures 8 and 9 are the same, so the first observation is that most of the variance measures—the exception is the Social Security threshold restricted concept—are noticeably higher. More importantly, there is an increase after 1980 in the variance for the unrestricted samples and in the full-time only sample, which suggests that the combination of including self-employment and diminimus earnings is driving the results.

The same basic conclusion emerges when one considers how the variance of earnings evolves as cohorts age. Figures 10, 11, and 12 all depict log earnings variances at ages 27 to 57 for five year male birth cohorts born between 1940 and 1970. Any upward shift in these lines at a given age indicates increased variance, which indicates an increase in the variance of either permanent or transitory shocks. The most restrictive earnings measure and sample is shown in Figure 10 (wages only, above Social Security threshold) and the most expansive measure and sample is shown in Figure 11 (wages plus self-employment, greater than zero).

Focusing first on Figure 10, and setting aside the oldest cohorts, one sees patterns that are generally consistent with the version of the canonical model presented here; the steeper slope at young ages indicates a larger variance for the permanent shocks, for example. Also, the convergence of variances at any given age for the cohorts who reached working age after 1980 confirms the stability in permanent shocks we found using the earnings growth data.

Figure 11 shows the same upward pattern in variances by age, but does not indicate the same sort of stability across cohorts, because the measure includes self-employment earnings and there are no threshold restrictions. However, there are also no systematic differences by cohort at any given age. The take away message from Figure 11 is that expanding the earnings concept and setting the threshold at zero really just adds a lot of noise. Figure 12, which uses the same total earnings measure but restricts the sample to full-time only, is not much better. The upshot of both figures is that allowing diminimus earnings to enter variance calculations has a first-order effect on estimated variances, and it is easy to see how estimators based on sample variances could lead to spurious conclusions.

The sensitivity of estimated variance levels to diminimus earnings is strongly suggestive about the most promising directions for future research. First, focusing on just wages, it seems important to restrict the sampling to include only full-time earners in the earnings variance estimates. This does not mean that exogenous events such as health-induced labor force withdrawal should be ignored, it just means they don't fit into the logic of the simple canonical model. Second, it seems important to learn more about self-employed earnings—perhaps the decision to become self-employed says something about risk taking, and separate analysis of volatility is suggested.

6. Conclusions

The canonical earnings shock model standard in micro labor economics research has recently played an important role in the debate over whether earnings volatility in the U.S. is rising or falling. Different groups of researchers have started with the same basic framework and the same objective—decomposing earnings shocks into permanent and transitory components—but reached very different conclusions about what is happening to volatility. The results here suggest that the estimates are very sensitive to a combination of sampling restrictions and earnings measures. In particular, excluding observations with diminimus earnings—especially those with self-employment income—has a first-order effect on sample variances and thus on inferences about earnings dynamics which rely on those variances.

The fact that inferences about earnings dynamics are very sensitive to inclusion of very low earnings suggests at least two possible directions for future research. The first, suggested by Jensen and Shore (2008), is to move away from trying to characterize volatility for the whole population over time and to focus on how volatility varies across the population. The second direction involves separating out the sorts of shocks that affect realized earnings but do not fit nicely into the canonical model. In particular, the canonical model does not explicitly incorporate the sorts of health or demographic shocks that might cause large swings in earnings. Some of those shocks are arguably a component of volatility, but in any event isolating them will provide a clearer view of how the underlying permanent and transitory shocks that individuals face in the labor market have evolved over time.

References

Auten, Gerald E., and Geoffrey Gee. “Income Mobility in the U.S.: Evidence from Income Tax Returns for 1987 and 1996.” OTA Paper 99. Washington, DC: U.S. Treasury Department, May, 2007.

Carroll, Christopher D. “The Buffer Stock Theory of Saving: Some Macroeconomic Evidence,” *Brookings Papers on Economic Activity.* (Fall, 1992):61–135.

Cilke, James, Julie-Anne M. Cronin, Janet McCubbin, James R. Nunns, and Paul Smith. “Distributional Analysis: A Longer Term Perspective.” *Proceedings of the Ninety-Third Annual Conference on Taxation*. Washington, DC: National Tax Association, 2001.

Congressional Budget Office. *Effective Tax Rates: Comparing Annual and Multiyear Measures*. Washington, DC: U.S. Congress, January, 2005.

Congressional Budget Office. *Recent Trends in the Variability of Individual Earnings and Family Income*. Washington, DC: U.S. Congress, June, 2008.

Davis, Steven J., and James A. Kahn. “Interpreting the Great Moderation: Changes in the Volatility of Economic Activity at the Macro and Micro Levels.” NBER Working Paper No. 14048. Cambridge, MA: National Bureau of Economic Research, May, 2008.

Deaton, Angus, and Christina Paxson. "Intertemporal Choice and Inequality." *Journal of Political Economy* 102 (June, 1994): 437-67.

Dynan, Karen E., Douglas W. Elmendorf, and Daniel E. Sichel. “The Evolution of Household Income Volatility.” Finance and Economics Discussion Paper Series Working Paper 2007-62. Washington, DC: Federal Reserve Board, October, 2007.

Gottschalk, Peter, and Robert Moffitt. “The Growth of Earnings Instability in the U.S. Labor Market.” *Brookings Papers on Economic Activity* 2 (Fall, 1994): 217-272.

Jensen, Shane T., and Stephen H. Shore. “Changes in the Distribution of Income Volatility.” University of Pennsylvania and The Johns Hopkins University (June, 2008).

King, Miriam, Steven Ruggles, Trent Alexander, Donna Leicach, and Matthew Sobek. Integrated Public Use Microdata Series, Current Population Survey: Version 2.0. [Machine-readable database]. Minneapolis, MN: Minnesota Population Center [producer and distributor]. The URL for the IPUMS-CPS site is cps.ipums.org/cps. 2004.

Kopczuk, Wojciech, Emmanuel Saez, and Jae Song. “Uncovering the American Dream: Inequality and Mobility in Social Security Earnings Data Since 1937.” NBER Working Paper No. 13345. Cambridge, MA: National Bureau of Economic Research, August, 2007.

Meghir, Costas, and Luigi Pistaferri. "Income Variance Dynamics and Heterogeneity." *Econometrica* 72 (January, 2004): 1-32.

Moffitt, Robert, and Peter Gottschalk. "Trends in the Transitory Variance of Male Earnings." John Hopkins University and Boston College. (December, 2008).

Moffitt, Robert A., and Peter Gottschalk. "Trends in the Transitory Variance of Earnings in the United States." *The Economic Journal* 112 (March, 2002): 68-73.

Panis, C., R. Euller, C. Grant, M. Bradley, C.E. Peterson, R. Hirscher, P. Stinberg. *SSA Program Data User's Manual*. Baltimore, MD: Social Security Administration, 2000.

Sabelhaus, John, and Jae Song. "The Great Moderation in Micro Labor Earnings." University of Maryland and Social Security Administration, December, 2008.

Data Appendix

This paper uses two different longitudinal micro earnings data extracts from SSA's master earnings file (MEF): SIPP matched extracts and SSA's 1 percent detail earnings extract (Panis, et. al, 2000, Kopczuk, Saez, and Song, 2007). The SSA's MEF includes individuals' annual Social Security covered earnings from 1951 to the present and annual wages directly taken from the W-2 starting from 1978. Other data elements included in the MEF are: the individual's SSN, annual self-employment earnings, type of wage, deferred compensation contribution, and report year. Annual wages reported in the detail segment of the MEF are not top-coded, but earnings from self-employment (SE) are top coded until 1992. The Medicare taxable maximum was completely eliminated in 1993 and this rule change allow SSA to keep SE earnings without top coding starting from 1993. To avoid the complication caused by the rule change, we exclude all SE earnings from our analysis.

Generally, SSA data are limited to information required for program administration and do not include individuals' socio-economic characteristics, labor hours, and family structure. Matched survey data combine the accuracy of SSA administrative data with the variety of demographic and economic characteristics of individuals. SIPP matched samples used in this study are drawn from the 1990, 1991, 1992, 1993, and 1996 panels. The SIPP is a national survey conducted by the U.S. Census Bureau, designed as a continuous series of national panels since 1984, with sample sizes ranging from 14,000 to 36,700 interviewed households. The SIPP is a multistage, stratified sample of the U.S. civilian, noninstitutionalized population. Panel duration ranges from 2½ to 4 years; each panel consists of six to twelve 4-month waves, depending on the panel duration. Each SIPP wave consists of both a core file and a topical module file. The core file contains the demographic, labor force, program participation, and income data designed to measure the economic situations of the individuals in the sample; these data are repeated at each interviewing wave.

The total number of observation interviewed in all core waves is 69,432, 44,373, 62,412, 62,721, and 95,398, respectively for 1990, 1991, 1992, 1993, and 1996 panels. Total number of observations matched with Social Security detail earnings (wages only) data is 53,441, 32,977, 45,621, 44,602, and 75,903 for 1990, 1991, 1992, 1993, and 1996 panels respectively. For our analysis, we used samples born in 1940 through 1960 and those who earn above the amount required for 4 quarters of coverage (QC) each year. The amount needed to get 1 QC in 2000 was $780 (or $3,120 for 4QCs).

In order to get better idea on how much of the standard deviation of logged earnings is accountable for changes in education over the study period, we also repeat the same analysis controlling only for age and sex by using the 1% MEF extract. The 1 percent sample is selected on the basis of certain serial digits of the social security number and generally considered to be random sample. Unlike SIPP matched sample, the 1 percent sample includes institutionalized individuals.

Figure 1. Standard Deviation of Annual Change in Log Earnings (One Percent Master Earnings File, Ages 25 to 55)

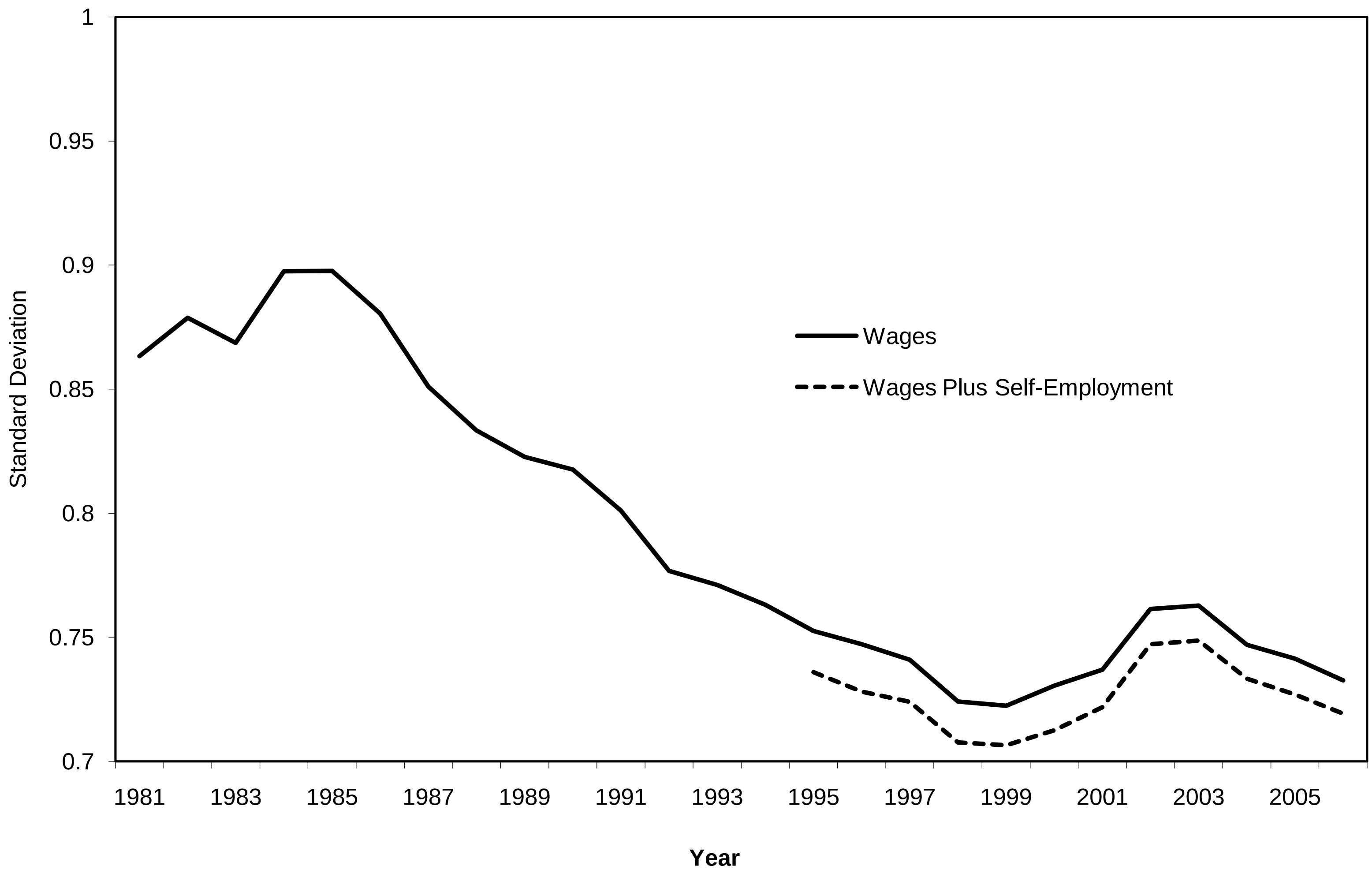

Figure 2. Standard Deviation of Annual Change in Log Earnings Above Social Security Threshold Only (One Percent Master Earnings File, Ages 25 to 55)

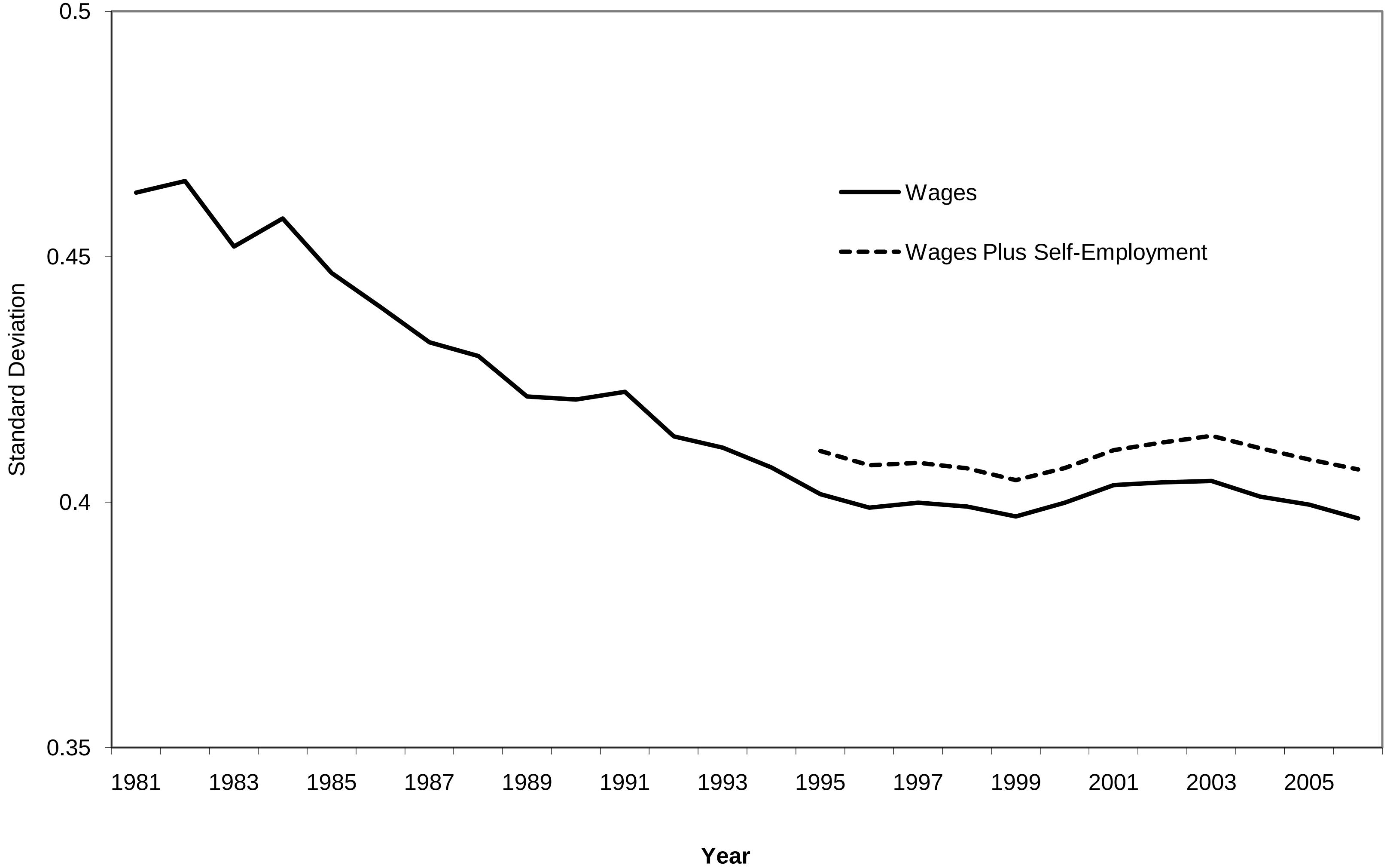

Figure 3. Standard Deviation of Annual Change in Log Earnings (Cohorts 1940 to 1960)

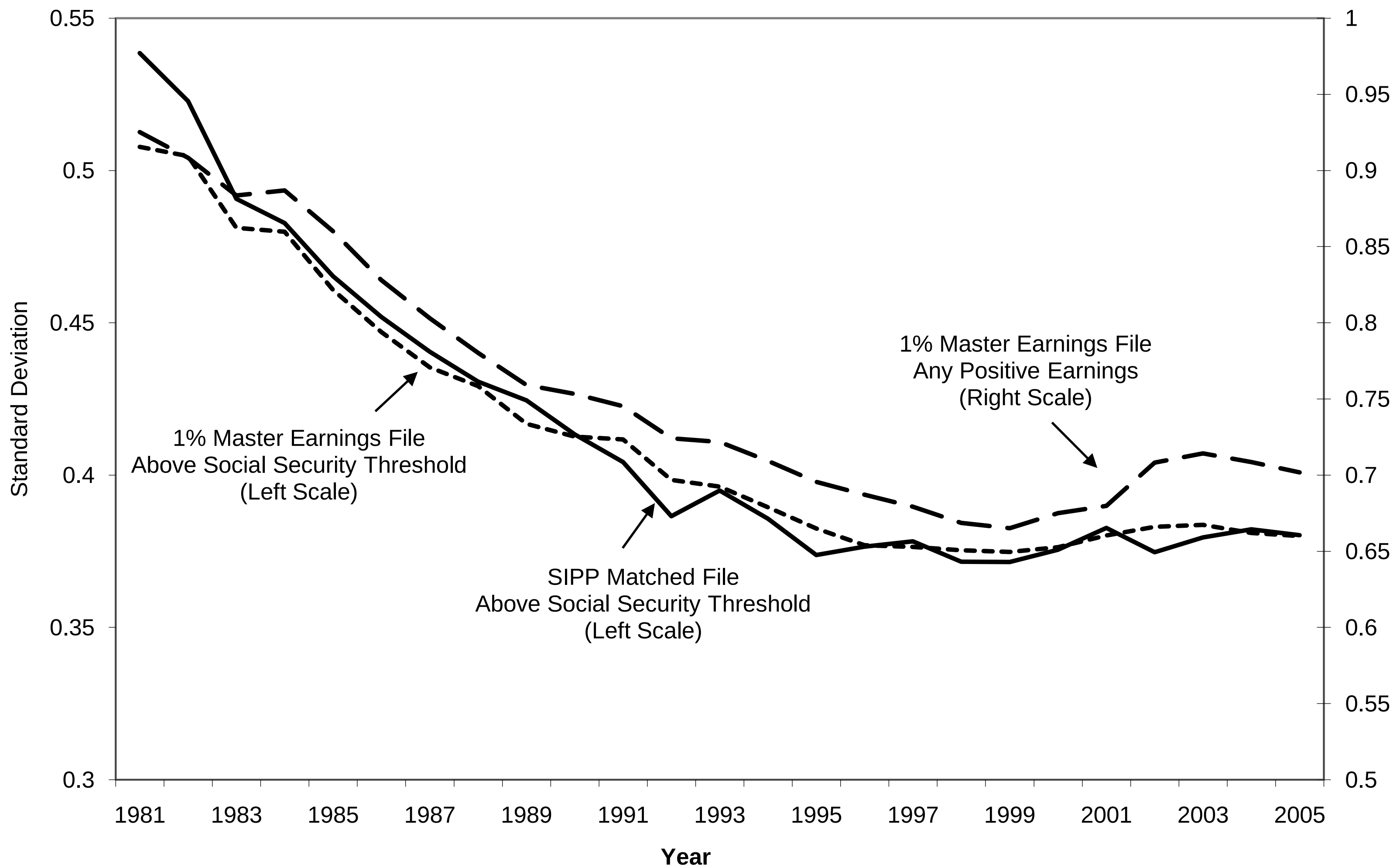

Figure 4. Meghir-Pistaferri Permanent Shock Standard Deviation by Age (Wages and/or Earnings Above Social Security Earnings Threshold)

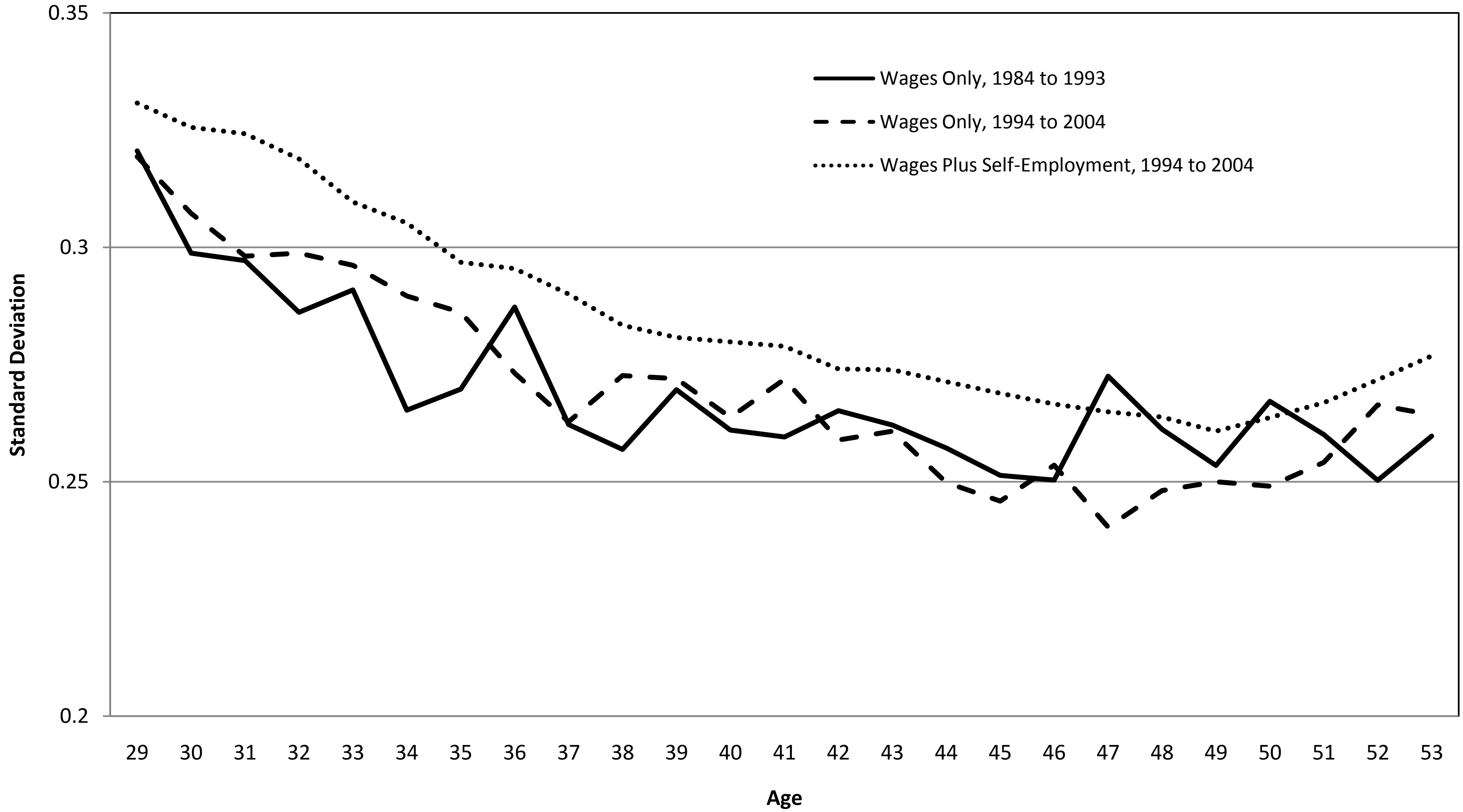

Figure 5. Meghir-Pistaferri Permanent Shock Standard Deviation by Age (Wages Only, 1984-2004, Above Social Security Earnings Threshold)

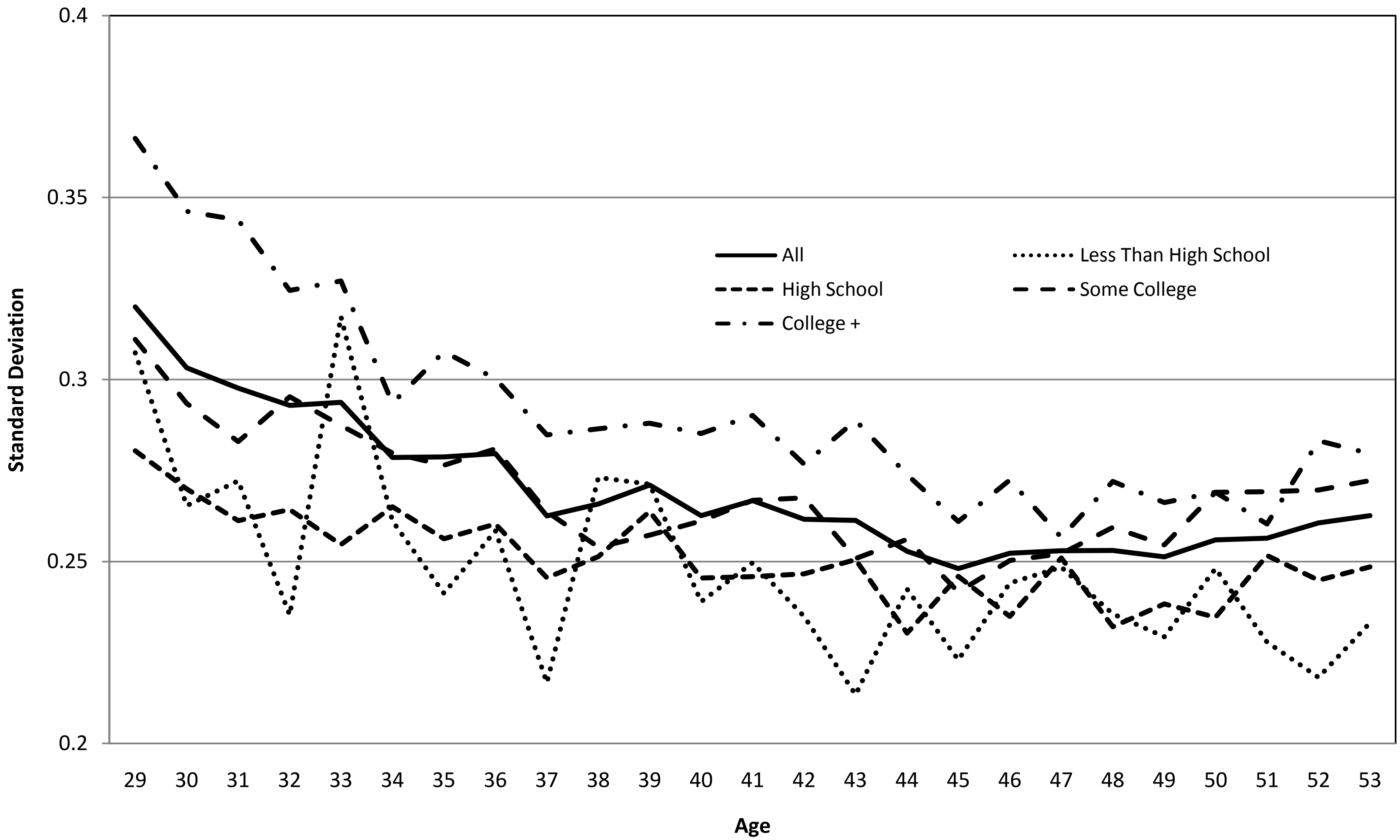

Figure 6. Standard Deviation of Meghir-Pistaferri Permanent Shock (Wages Above Social Security Threshold)

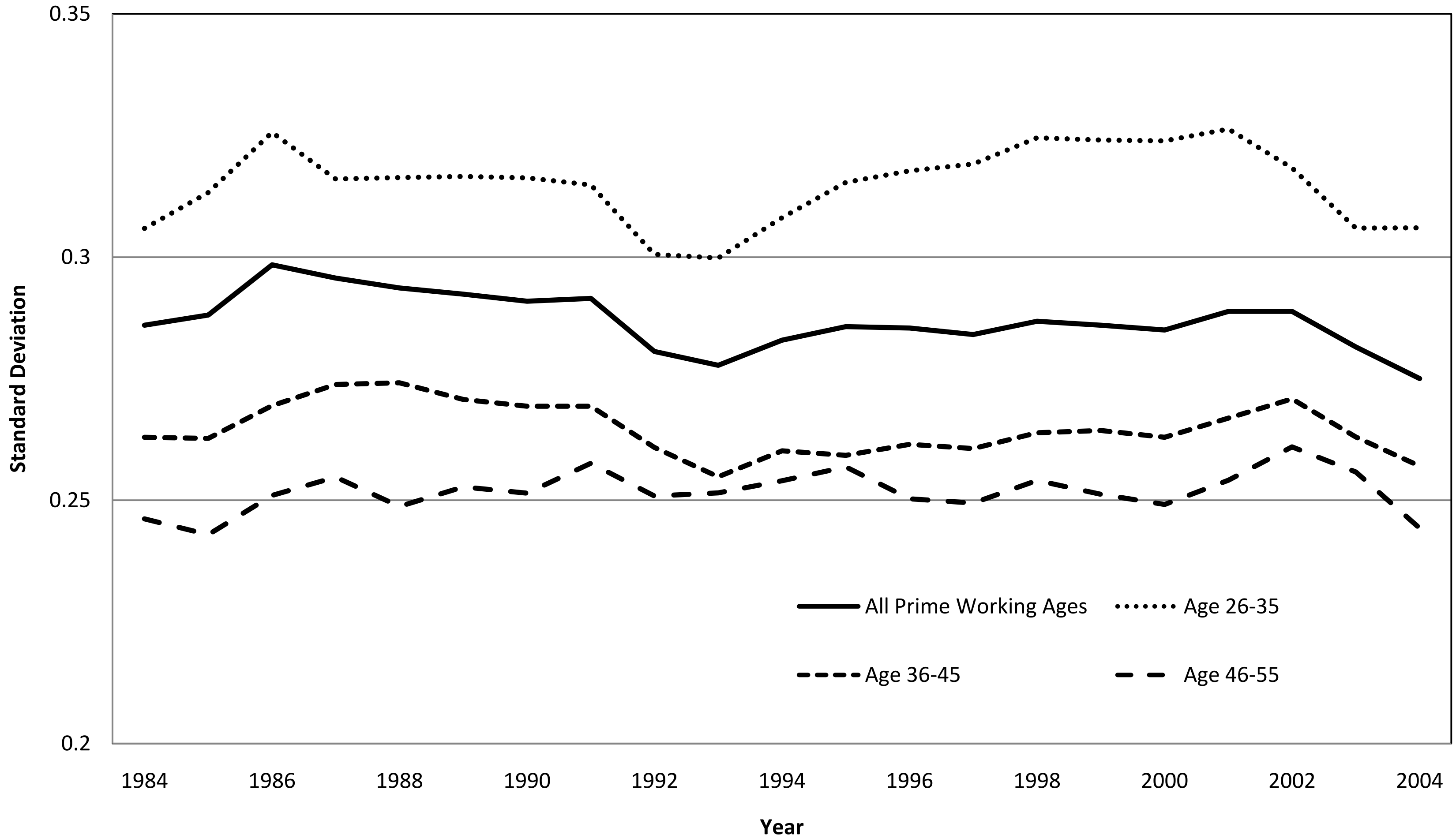

Figure 7. Standard Deviation of Transitory Residual (Wages Above Social Security Threshold)

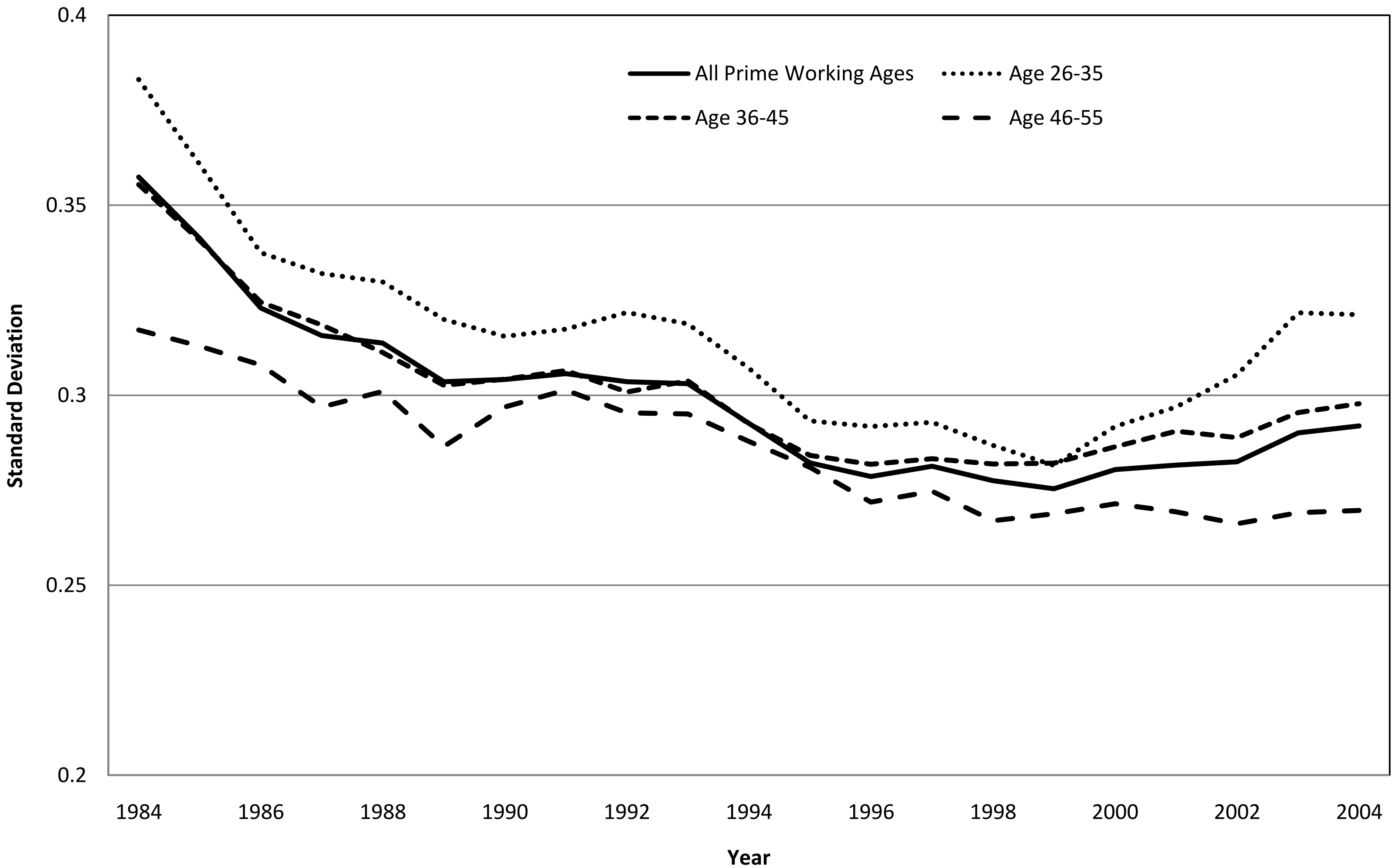

Figure 8. Variance of Male Log Annual Wages, Age 30 to 39 (March Current Population Survey)

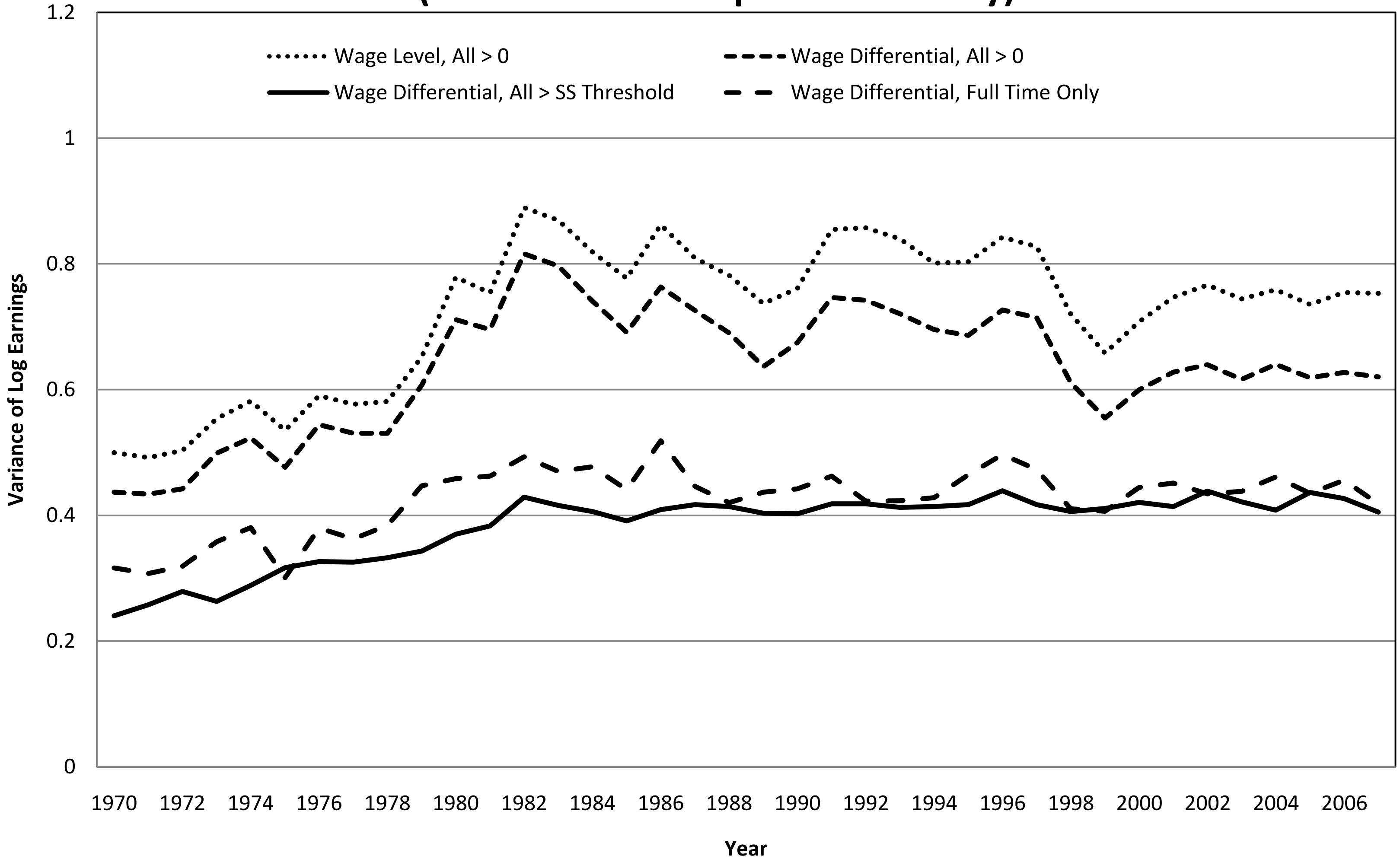

Figure 9. Variance of Male Log Annual Earnings, Age 30 to 39
(March Current Population Survey)

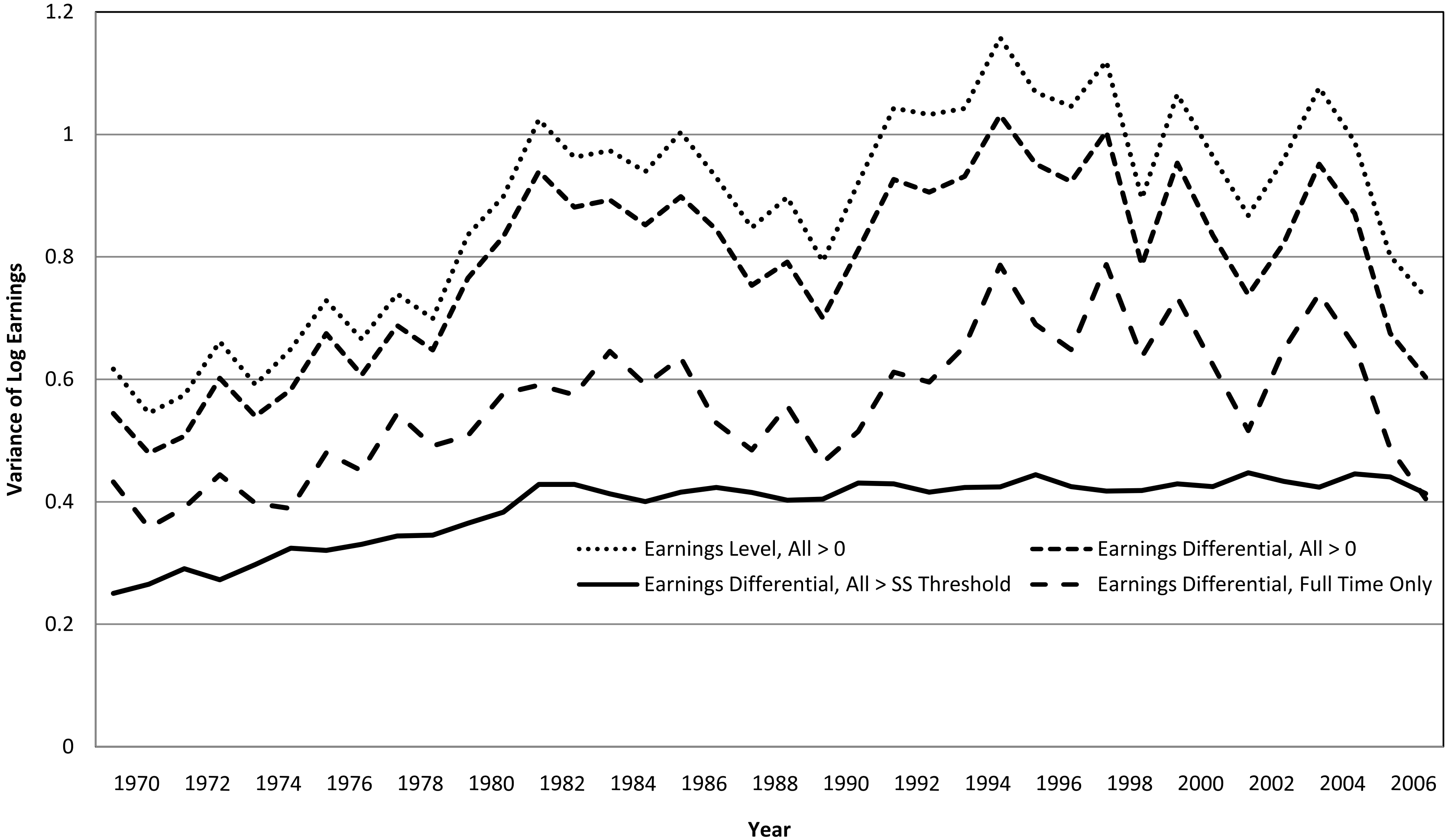

Figure 10. Variance of Male Log Annual Earnings by Birth Cohort (March CPS; Earnings Above Social Security Threshold)

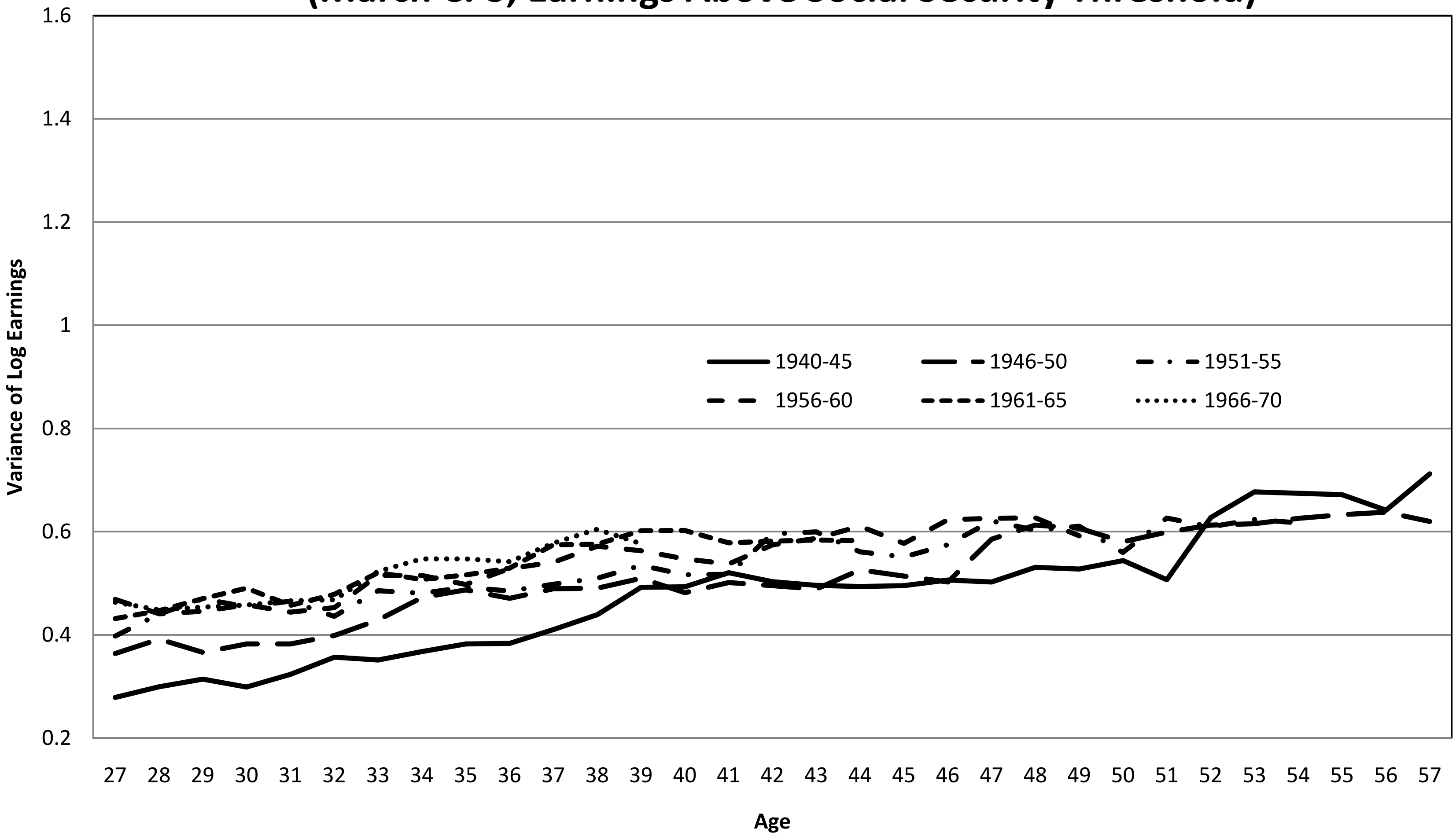

Figure 11. Variance of Male Log Annual Earnings by Birth Cohort (March CPS; Earnings Greater Than Zero)

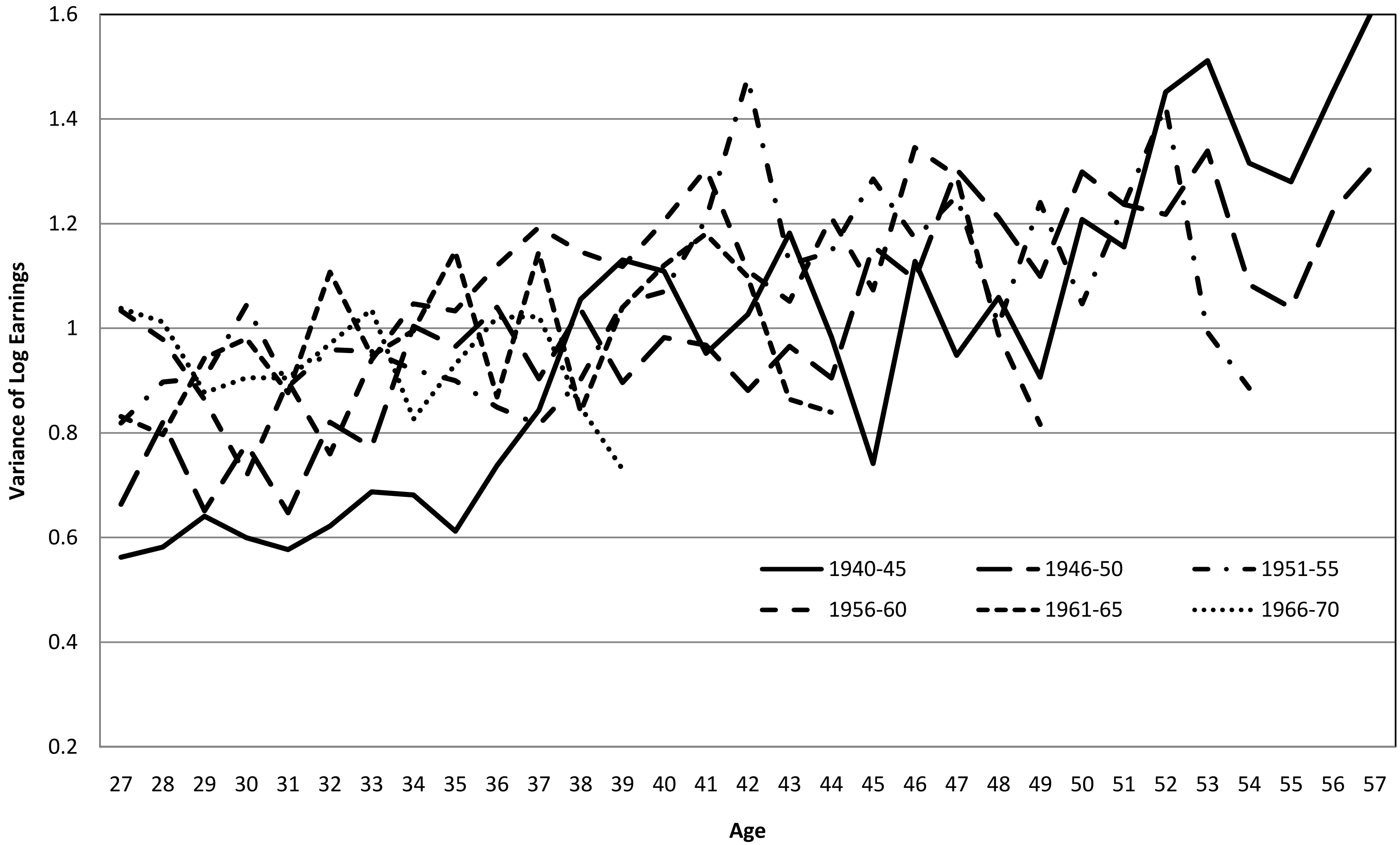

Figure 12. Variance of Male Log Annual Earnings by Birth Cohort (March CPS; Earnings Greater Than Zero, Full Time Only)