The Extended Semiparametric (ESP) Model of Permanent-Transitory Decompositions of Earnings Variances

May 2018

Robert Moffitt
Johns Hopkins University

Sisi Zhang
Jinan University

This paper is an excerpt from the Online Appendix to "Income Volatility and the PSID: Past Research and New Results" (AER, May 2018). It strips out the 3rd section which discussed the new model of earnings dynamics and therefore deletes the review of the existing literature and the discussion of the merits of the PSID for the study of earnings dynamics. It begins with Section III of that Appendix, which presents some raw patterns of earnings variances and then presents the formal model.

III. Some New Results on Trends in Male Earnings Volatility

The work examining trends in earnings volatility with the PSID reported in the previous section only used data through 2009. Data through 2014 are now available, so we provide new results through that year. The 2009-2014 period is particularly interesting because it encompasses the Great Recession. For our new results, we focus solely on male earnings, which has been the focus of the majority of the literature to date and which can be analyzed without special attention to selectivity of employment. We provide measures both of gross volatility and estimates of an error components model which allows us to decompose trends in gross volatility into trends in permanent and transitory volatility.

We use the data from interview year 1971 through interview year 2015.[1] Earnings are collected for the previous year, so our data cover the calendar years 1970 to 2014. The PSID skipped interviews every other year starting in interview year 1998, so our last observations are

[1]We do not use earnings reported in 1969 or 1970 since wage and salary earnings, which is what we use, are reported only in bracketed form in those years.

for earnings years 1996, 1998, and so on, every other year through 2014. The sample is restricted to male heads of households. Only heads are included because the PSID earnings questions we use are only asked of heads of household. We take any year in which these male heads were between the ages of 30 and 59, not a student, and had positive annual wage and salary income and positive annual weeks of work. We include men in every year in which they appear in the data and satisfy these requirements. We therefore work with an unbalanced sample because a balanced sample would be greatly reduced in size because of aging into and out of the sample in different years, attrition, and movements in and out of employment. Fitzgerald et al. (1998) have found that attrition in the PSID has had little effect on its cross-sectional representativeness, although less is known about the effect of attrition on autocovariances. We exclude men in all PSID oversamples (SEO, Latino) and we exclude nonsample men. All earnings are put into 1996 CPI-U-RS dollars. The resulting data set has 3,508 men and 36,403 person-year observations, for an average of 10.4 year-observations per person. Means of the key variables are shown in Appendix Table 1.

As is common in the literature, we work with residuals from regressions of log earnings on education, a polynomial in age, and interactions between age and education variables, all estimated separately by calendar year (however, we will show gross volatility trends for log earnings itself as well). We use these residuals to form a variance-autocovariance matrix indexed by year, age, and lag length. A typical element of the matrix consists of the covariance between residual log earnings of men at ages a and a' between years t and t'. Because of sample size limitations, however, we cannot construct such covariances by single years of age. Instead, we group the observations into three age groups--30-39, 40-49, and 50-59--and then construct the variances for each age group in each year, as well as the autocovariances for each group at all possible lags back to 1970 or age 20, whichever comes first. We then compute the covariance between the residual log earnings of the group in the given year and each lagged year, using the individuals who are in common in the two years (when constructing these covariances, we trim the top and bottom one percent of the residuals within age group-year cells to eliminate outliers and top-coded observations[21]). The resulting autocovariance matrix represents every individual variance and covariance between every pair of years only once, and stratifies by age so that life cycle changes in the variances of permanent and transitory earnings can be estimated. The matrix has 1,417 unique elements.

Figure 1 shows the variance of 2-year differences in the residuals from the log earnings regression, the usual measure of gross volatility. Gross volatility rose from the 1970s to the mid-1980s and then exhibited no trend (albeit around significant instability) until around 2000, when it resumed its rise. Our results through 2014 show that gross volatility rose sharply during the Great Recession. As shown by the unemployment rate (also in the figure), volatility is correlated with the unemployment rate but with a slight lag. Our findings are consistent with Dynarski and Gruber (1997), who found rising (on average) gross volatility from 1970 to 1991, and with Shin and Solon (2011)'s results through 2005, although those authors found more of a decline in the middle period than a stable and flat trend. Our results for the early and late periods are similar to those of Dynan et al. (2012) although those authors found a slow rise in the middle period. The large number of extreme fluctuations in the middle period in our data may be responsible for these other authors' finding of a slight decline or rise.

[21]If top-coding were the only motivation for trimming, a preferable procedure would be to top-trim the earnings variable directly rather than the residuals. However, our motivation is more general, to avoid distortion of log variances from outliers. In prior work (Moffitt and Gottschalk, 2002), we tested trimming on the residuals versus trimming on earnings itself, and found no qualitative difference in the results.

Figure 2 shows trends in the percentile points of the distribution of the 2-year change, showing that the increasing volatility reflects a widening out at all percentile points but with the largest widening occurring at the top and bottom of the change distribution. Figure 3 shows the variance of 2-year changes of log earnings itself, not of residuals from a regression. The trend pattern and, in particular, the existence of three approximate periods of rise, then flat trend, then rise, is the same as for the residuals.

To decompose gross volatility into its permanent and transitory components, we adopt an error components model similar to those used in the past literature but with some of the more restrictive features of those models eliminated. Error components models have been criticized for being excessively parametric, so, while we maintain many of the restrictions in past work, we also reduce some of their parametric restrictions in two ways. First, we make a clear, non-arbitrary identification assumption to separate permanent from transitory components and, second, we are nonparametric for the evolution of their variances. Letting y_{iat} be the log earnings residual for individual *i* at age a in year *t*, our model is

$$y_{iat} = \alpha_t \mu_{ia} + \beta_t \upsilon_{ia} \qquad (1)$$

where μ_{ia} is the permanent component for individual *i* at age a, ν_{ia} is the transitory component for individual *i* at age a, and α_t and β_t are calendar time shifters for the two components. We shall maintain the usual assumption in these models that the permanent and transitory components are additive and independently distributed, an assumption that can be partially relaxed. We also adopt the common specification that calendar effects do not vary with age, although this could be relaxed by allowing the calendar time shifts to vary with age (but we will not do that here).

The first question is how permanent and transitory components can be separately identified if both are allowed to be a function of age. We assume the dictionary definition of a permanent component, which is a component which has a literally permanent, lasting, and indefinite effect and does not fade away even partially. The transitory component can then be identified as consisting of any residual component whose impact on y does change over time. To make this definition operational, we will assume that the permanent component at the start of the life cycle is μ_0 and that an individual experiences independently distributed permanent shocks $\omega_1, \omega_2, \ldots, \omega_T$ through the end of life at time *T*. We let the permanent component at age *a* be some function of these shocks: $\mu_{ia} = f(\omega_{i1}, \omega_{i2}, \ldots, \omega_{ia}, \mu_0)$. We define a permanent shock ω_{is} to be one for which $\partial\mu_{ia}/\partial\omega_{ia} = 1$ and we assert that the only function *f* which satisfies this condition is the unit root process

$$\mu_{ia} = \mu_{i0} + \sum_{s=1}^{a} \omega_{is} \tag{2}$$

If we similarly define the transitory component to be a linear function of a series of independently distributed transitory shocks $\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iT}$ but we put no restrictions on the impact of each of these shocks on ν_{ia}, then, as noted previously, the impact of transitory shocks can be identified as all shocks which do not have an impact coefficient of 1 on y.

Beyond this assumption, we attempt to make as few restrictive assumptions as possible. We let the distributions of the permanent and transitory shocks, ω_{ia} and ε_{ia}, respectively, be nonparametric functions of a.[22] We do assume that the transitory component is linear in the

[22] This assumption makes the unit root and heterogeneous growth models equivalent and both embedded in the model. The typical heterogeneous growth model assumes that the permanent component to have a subcomponent equal to age times a heterogeneous growth factor. That model is identified only because of the restrictive assumption that individual growth heterogeneity is linear in age. If the growth factor is allowed to be nonparametric in age, the model is not identified from a unit root model with shocks whose distribution varies freely with age.

transitory shocks (this could be relaxed) but we do not impose any ARMA form on the coefficients. Instead, we specify the transitory component to be

$$\nu_{ia} = \varepsilon_{ia} + \sum_{s=1}^{a-1} \psi_{a,a-s} \varepsilon_{i,a-s} \quad (3)$$

and we allow the impact coefficients of transitory shocks, the T(T+1)/2 – T parameters $\psi_{a,a-s}$ to be unconstrained.[23] This model nests the linear models used in the literature but does not nest those which are nonlinear in the shocks and those which have heterogeneous transitory shock impacts (e.g., which allow the ψ parameters or the distributions of the shocks to be individual-specific).[24] We name our model the Extended Semiparametric (ESP) Model because it is a major extension of the semiparametric model proposed by Moffitt and Gottschalk (2012).

Following the majority of the literature, we restrict our attention to the explaining the second moments of y_{iat} by second moments of the permanent and transitory shocks. We therefore seek to estimate the variances of the permanent and transitory shocks, allowing them to be nonparametric in age. In the Appendix, we show conditions for identification of the parameters. We estimate the parameters with conventional minimum distance. The exact specification of the model and the estimates of the parameters and their standard errors are shown in the Appendix Table 2.

Figures 4 and 5 show the trends in α and β, respectively, which are the calendar time factors in the model. The results show that both permanent and transitory variances trended upward over time and both roughly followed the pattern exhibited by gross volatility, with an initial rise, followed by a middle period when the rise had stopped, and ending with a rising

[23] The coefficient on the contemporary shock, ε_{ia} is not identified and must be set equal to 1.
[24] We also make no attempt to identify measurement error in the model. It can be identified only by untestable parametric assumptions which make such error evolve in a different functional form than the other shocks. For present purposes, which is mainly to identify calendar time trends, measurement error should have no effect unless it has been changing over time.

trend. The turning points—with a necessary caution as to the difficulty of detecting them visually in the facing of considerable instability—are slightly different, however. The transitory variance appears to have stopped rising in the early 1980s whereas the permanent variance continued to rise through the late 1980s. The transitory variance exhibits a slight decline in the middle period whereas the permanent variance is mostly flat. However, both variances turned up toward the end of the period. One reading of the results is that neither variance substantially departed from a process with fluctuations around a stable trend until 2008, when its increase truly started to emerge. This would be consistent with an effect of the Great Recession. The variances also show signs in the last two years of starting to decline from their Recession peaks.

The implications of these trends for the variances of the permanent and transitory components themselves are shown in Figure 6 for those age 40-49 (variances differ by age, with older individuals having higher variances, but the trend is the same at all ages given the model specification). The now-familiar three-phase trend is still apparent. The transitory variance is about two-thirds of the total variance and has risen more than the permanent variance from beginning to end. Thus we find that a larger fraction of the increase in cross-sectional male earnings inequality is accounted for by increases in the transitory component.[25]

We use our estimates to decompose the trend in the variance of 2-year changes of log earnings residuals (see Figure 1) into trends in the 2-year changes in permanent and transitory variances. The variance of 2-year changes involves both the level of the variance at each of the two time points as well as the covariance between them. The results can be found in Appendix Table 4 and show that both the level of the variances and the covariances have trended upward over time, for both the permanent and transitory components. But, on net, the variance of the

[25] The exact numbers for these variables can be found in the Appendix Table 3.

total change is almost entirely the result of increases in the transitory variance. The permanent variance does not have the same volatility as the transitory variance and changes at a slower rate, and the permanent variance is also smaller in magnitude than the transitory variance.

IV. Sensitivity Tests: Imputation and Window Averaging

We conduct two sensitivity tests to our findings. The first estimates the sensitivity of our results to the inclusion of imputed earnings values in the PSID. The second presents estimates of time trends in the transitory variance using the Window Averaging (WA) method, which is a particularly intuitive method of estimating transitory variances that is used in many studies.

Like all survey data sets, a certain fraction of earnings values are imputed in the PSID because of don't know responses and refusals to answer, from implausible values indicating response error, and other reasons. The PSID has conducted imputations for all of these reasons and the exact method of using them has varied somewhat over time, generally with growing sophisticated and complexity. Current imputation procedures for income use a variety of imputation methods, depending on the type of income being imputed and using a different set of variables for each (Duffy, 2011). In our sample of male heads from 1970 to 2014, the percent of wage and salary income observations that are imputed ranges from a low of 0.30 to a high of 4.6, with the high value occurring in 1992, a period when the PSID changed its methodology and interviewing method.

The traditional primary issue with imputation is whether it is ignorable, i.e., whether those observations which are imputed have unobservable differences in earnings from those which are not, and whether the imputation process can adjust for any such differences. The common method of testing for non-ignorability and the accuracy of the process is simply to estimate models with and without imputed observations even though, if non-ignorability holds, both estimates are biased. Figure 7 shows the trend in gross volatility in our sample including and excluding the imputed observations. There is very little difference in the trends in either case, suggesting that the observations being imputed are ignorable or that the imputation process adequately corrects for any non-ignorability.

Moffitt and Gottschalk (2012) dubbed any method of estimating transitory variances based on taking an interval of annual observations and computing transitory components as the deviations from some (possibly trend-adjusted) mean as a Window Averaging (WA) method. This method has been used primarily in the literature on calendar time trends in volatility and was used by the initial paper in that literature, Gottschalk and Moffitt (1994) but has been used in modified form in several subsequent papers (see Table 2 and 3). A traditional ANOVA definition of the transitory variance within a window of T observations is

$$\frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it}-\bar{y}_i)^2 \qquad (4)$$

However, because $y_{it}-\bar{y}_i=\frac{1}{T}\sum_{\tau\neq t}^{T}(y_{it}-y_{i\tau})$, the WA method is based on the variance of pairwise differences between each y and the others within the window. Hence it is closer to an extended version of gross volatility than a true measure of the transitory variance, combining changes in permanent and transitory variances. In addition, if any model like that in equation (1) above holds, the WA method produces some time average of α_t and β_t, weighted by the variances of the pairwise differences.

Figure 8 shows estimates of equation (4) using a 9-year window for our male head data set 1970-2014, plotted against the year in the center of the window. The levels of the estimated variances is quite a bit below those of the transitory variance in Figure 6 (exact numbers in

Appendix Table 3) which is to be expected since the WA method averages over years and hences damps down the year-to-year variances from the ESP model. But the three-phase pattern revealed previously for both gross volatility and the transitory variance continues to hold here, although the turning points are considerably more indistinct than in the ESP model because of the smoothing inherent in the use of a 9-year average.

References

Duffy, Denise. 2011. 2007 PSID income and wage imputation methodology. Technical Series Paper #11-03. Ann Arbor, Michigan. Available from https://psidonline.isr.umich.edu/Publications/Papers/ (Accessed 1/22/18).

Dynan, Karen, Douglas Elmendorf, and Daniel Sichel. 2012. The evolution of household income volatility. *The B.E. Journal of Economic Analysis & Policy* 12 (2): 1-42.

Dynarski, Susan and Jonathan Gruber. 1997. Can families smooth variable earnings? *Brookings Papers on Economic Activity* 1997 (1):229–303.

Fitzgerald, J., Peter Gottschalk, and Robert A. Moffitt. 1998. An analysis of sample attrition in panel data: The Michigan Panel Study of Income Dynamics. *Journal of Human Resources* 33 (2): 251-99.

Gottschalk, Peter, and Robert Moffitt. 1994. The growth of earnings instability in the U. S. Labor Market. *Brookings Papers on Economic Activity* 1994 (2): 217–72.

Moffitt, Robert A, and Peter Gottschalk. 2002. Trends in the transitory variance of earnings in the United States. *Economic Journal* 112 (478): C68–73.

Moffitt, Robert A, and Peter Gottschalk. 2012. Trends in the transitory variance of male earnings: methods and evidence. *Journal of Human Resources* 47(1): 204-36.

Shin, Donggyun, and Gary Solon. 2011. Trends in men's earnings volatility: What does the Panel Study of Income Dynamics show? *Journal of Public Economics* 95 (7–8). Elsevier B.V.:973–82.

Appendix
The Extended Semiparametric (ESP) Model

Letting y_{iat} be the earnings residual for individual i at age a in year t, the model is:

$$y_{iat} = \alpha_t \mu_{ia} + \beta_t \upsilon_{ia} \tag{1}$$

$$\mu_{ia} = \mu_{i0} + \sum_{s=1}^{a} \omega_{is} \tag{2}$$

$$\upsilon_{ia} = \varepsilon_{ia} + \sum_{s=1}^{a-1} \psi_{a,a-s}\varepsilon_{i,a-s} \quad \text{for } a \geq 2 \tag{3}$$

$$\upsilon_{i1} = \varepsilon_{i1} \text{ for } a = 1 \tag{4}$$

for $a = 1, ..., A$ and $t = 1, ..., T$ and where the shocks ω_{ia} and ε_{ia} are independently distributed from each other and over time. The autocovariances implied by this model, which will be fit to the autocovariances in the data, are:

$$Var(y_{iat}) = \alpha_t^2 Var(\mu_{ia}) + \beta_t^2 Var(\upsilon_{ia}) \tag{5}$$

$$Var(\mu_{ia}) = Var(\mu_{i0}) + \sum_{s=1}^{a} Var(\omega_{is}) \tag{6}$$

$$Var(\upsilon_{ia}) = Var(\varepsilon_{ia}) + \sum_{s=1}^{a-1} \psi_{a,a-s}^2 Var(\varepsilon_{i,a-s}), \text{ for } a \geq 2 \tag{7}$$

$$Var(\upsilon_{i1}) = Var(\varepsilon_{i1}), \text{ for } a = 1 \tag{8}$$

$$Cov(y_{iat}, y_{i,a-\tau,t-\tau}) = \alpha_t \alpha_{t-\tau} Cov(\mu_{ia}, \mu_{i,a-\tau}) + \beta_t \beta_{t-\tau} Cov(\upsilon_{ia}, \upsilon_{i,a-\tau}) \tag{9}$$

$$
\begin{aligned}
Cov(\mu_{ia}, \mu_{i,a-\tau}) &= Var(\mu_{i,a-\tau}) \\
&= Var(\mu_{i0}) + \sum_{s=1}^{a-\tau} Var(\omega_{is})
\end{aligned} \tag{10}
$$

$$
\begin{aligned}
Cov(\upsilon_{ia}, \upsilon_{i,a-\tau}) &= \psi_{a,a-\tau} Var(\varepsilon_{i,a-\tau}) \\
&+ \sum_{s=1}^{a-\tau-1} \psi_{a,a-\tau-s}\psi_{a-\tau,a-\tau-s} Var(\varepsilon_{i,a-\tau-s}), \text{ for } a \geq 3
\end{aligned} \tag{11}
$$

$$
\begin{aligned}
Cov(\upsilon_{ia}, \upsilon_{i,a-\tau}) &= \psi_{a,a-\tau} Var(\varepsilon_{i,a-\tau}) \\
&= \psi_{21} Var(\varepsilon_{i1}), \text{ for } a = 2, \tau = 1
\end{aligned} \tag{12}
$$

We allow the variances of the permanent and transitory shocks to be nonparametric functions of age and we allow the ψ parameters to be nonparametric functions of age and lag length (τ or $\tau + s$).

Identification. Considering first the identification of the parameters of the age-earnings process under the stationary model $\alpha_t = \beta_t = 1$, we note that a data set of age length $a = 1, ..., A$ has an autocovariance matrix of the y_{ia} with $A(A+1)/2$ elements. The unknown parameters in the model are $\sigma^2_{\mu_0}$, the A parameters $\sigma^2_{\omega a}$ ($a = 1, ..., A$), the $A(A-1)/2$ parameters $\psi_{a,a-r}$ ($r = 1, .., a-1$), and the A parameters $\sigma^2_{\varepsilon a}$ ($a = 1, ...A$), for a total of $[A(A+1)/2] + A + 1$ parameters. The stationary model is therefore nonparametrically not identified without $A+1$ restrictions.[1] We allow restrictions by imposing smoothness on the nonparametric functions σ^2_{ω},ψ, and σ^2_{ε} as described below. Our estimation shows that the number of parameters needed to fit the data allow the model to be heavily overidentified.[2] The α_t and β_t parameters are identified, subject to a normalization and conditional on the identification of the parameters of the age-earnings process, from the change in the

[1]Because the equations of the model are nonlinear in the parameters, we also require that the solutions for the parameters exist and are unique if the number of elements of the autocovariance matrix equals the number of unknowns.

[2]We note that the model is identified for a data set of length $A \geq 4$ under homoskedasticity of the permanent and transitory shocks, defined as the model with $\sigma^2_{\omega a} = \sigma^2_{\omega}$, $\sigma^2_{\varepsilon a} = \sigma^2_{\varepsilon}$ for $a = 2, , , A$, and with $\sigma^2_{\varepsilon 1}$ left as a free parameter for initial conditions purposes. We test for, and reject, homoskedasticity of the transitory variances.

autocovariance matrix elements at the same age and lag position but at different points in calendar time, which therefore requires multiple cohorts. Since α_t and β_t constitute two parameters, any two elements of the matrix observed at two calendar time points is sufficient for identification. For example, using the variances at ages a and a' observed at times t and $t+1$, we have

$$Var(y_{iat}) = \alpha_t^2 \sigma_{\mu a}^2 + \beta_t^2 \sigma_{\upsilon a}^2 \tag{13}$$

$$Var(y_{ia't}) = \alpha_t^2 \sigma_{\mu a'}^2 + \beta_t^2 \sigma_{\upsilon a'}^2 \tag{14}$$

$$Var(y_{ia,t+1}) = \alpha_t^2 r_\alpha^2 \sigma_{\mu a}^2 + \beta_t^2 r_\beta^2 \sigma_{\upsilon a}^2 \tag{15}$$

$$Var(y_{ia',t+1}) = \alpha_t^2 r_\alpha^2 \sigma_{\mu a'}^2 + \beta_t^2 r_\beta^2 \sigma_{\upsilon a'}^2 \tag{16}$$

where $r_\alpha = \alpha_{t+1}/\alpha_t$ and $r_\beta = \beta_{t+1}/\beta_t$. We normalize the calendar shifts at $t = 1$ by setting $\alpha_1 = \beta_1 = 1$. Equations (13)-(16) can be solved for α_t and β_t for $t = 2, ..., T$.

Nonparametric Estimation. To estimate the functions $\sigma_{\omega a}^2$, $\sigma_{\varepsilon a}^2$, and ψ, we specify the functions as series expansions in basis functions and use a generalized cross-validation (GCV) statistic, which has a penalty for the number of parameters, to choose the degree of the expansion. Our specific functional forms are:

$$Var(\omega_{ir}) = e^{\Sigma \delta_j (r-25)^j} \tag{17}$$

$$Var(\varepsilon_{ir}) = e^{\Sigma \gamma_j (r-25)^j}, \text{ for } r \geq 2 \tag{18}$$

$$Var(\varepsilon_{i1}) = k e^{\Sigma \gamma_j (1-25)^j}, \text{ for } r = 1 \tag{19}$$

$$\psi_{A,A-b} = [1 - \pi(A-25)][\Sigma w_j e^{-\lambda_j b}] + \Sigma \eta_j D(b=j) \tag{20}$$

The variances use exponential functions of polynomial expansions in age minus 25 (the approximate minimum age), with the initial transitory variance allowed to differ by factor k for an initial conditions adjustment. The ψ parameters are allowed to expand in a weighted sum of exponentials, which force the parameters to asymptote to 0 as the lag length goes to infinity, and with a linear age-function factor in front of that weighted sum. Deviations from the smooth exponential expansions are allowed at each lag length. The unknown parameters in the model are $Var(\mu_{i0}), \delta_j, \gamma_j, k, \pi, \lambda_j, w_j$,and the η_j as well as the α_t and β_t. The parameters are fit to the second-moment matrix of the data using minimum distance.

Appendix Table A-2, column 1, reports the results of the estimation. As is often the case using the PSID, only a small number of basis functions in the expansion improve the parameter-adjusted fit. The initial variance of the permanent component is significant but the variances of the permanent shocks do not vary with age.[3] The transitory variance is also weakly positive in a linear function of age. The initial transitory variance is over twice the size as subsequent transitory shocks (as expected) but the transitory autocovariance curve is only weakly (and negatively) correlated with age and with only a single exponential. The λ parameter confirms that autocovariances decline with lag length and the η parameters indicate that the most recent three lags have a different impact on the current transitory component than the age-adjusted smooth exponential curve indicates. The estimates of the α and β parameters are also shown; the figures in the text are plots of these estimates. The second column in the Table shows the estimates of the parameters if a model stationary in calendar time is estimated (i.e., constraining $\alpha_t = \beta_t = 1$). The parameter estimates are quite different than those estimated when calendar time shifts are allowed.

The parameter estimates are inserted into equations (6)-(8) to compute the implied variances of the permanent and transitory components without calendar time effects, and then those estimated components are used in equation (5) to compute the total variance and the two components on the right-hand-side of that equation. The text reports plots of these three variances for those aged 40-49, and Appendix Table 3 reports the exact figures for all three age groups.

[3]The two δ parameters are insignificant but adding the second one lowered the GCV, so we retain both. The total transitory variance is positive and highly significant.

The text reports the implications of the fitted model for the sources of the variance of 2-year changes in y. The 2-year change is

$$\begin{aligned} y_{iat} - y_{i,a-2,t-2} &= (\alpha_t \mu_{ia} + \beta_t \upsilon_{ia}) - (\alpha_{t-2} \mu_{i,a-2} + \beta_{t-2} \upsilon_{i,a-2}) \\ &= \alpha_t \mu_{ia} - \alpha_{t-2} \mu_{i,a-2} + \beta_t \upsilon_{ia} - \beta_{t-2} \upsilon_{i,a-2} \end{aligned} \tag{21}$$

and its variance is

$$\begin{aligned} & Var(y_{iat} - y_{i,a-2,t-2}) \\ &= \alpha_t^2 Var(\mu_{ia}) + \alpha_{t-2}^2 Var(\mu_{i,a-2}) - 2\alpha_t \alpha_{t-2} Cov(\mu_{ia}, \mu_{i,a-2}) \\ &\quad + \beta_t^2 Var(\upsilon_{ia}) + \beta_{t-2}^2 Var(\upsilon_{i,a-2}) - 2\beta_t \beta_{t-2} Cov(\upsilon_{ia}, \upsilon_{i,a-2}) \end{aligned} \tag{22}$$

which contains variances and covariances which have been fitted by the model. Appendix Table 4 shows the exact components by year.

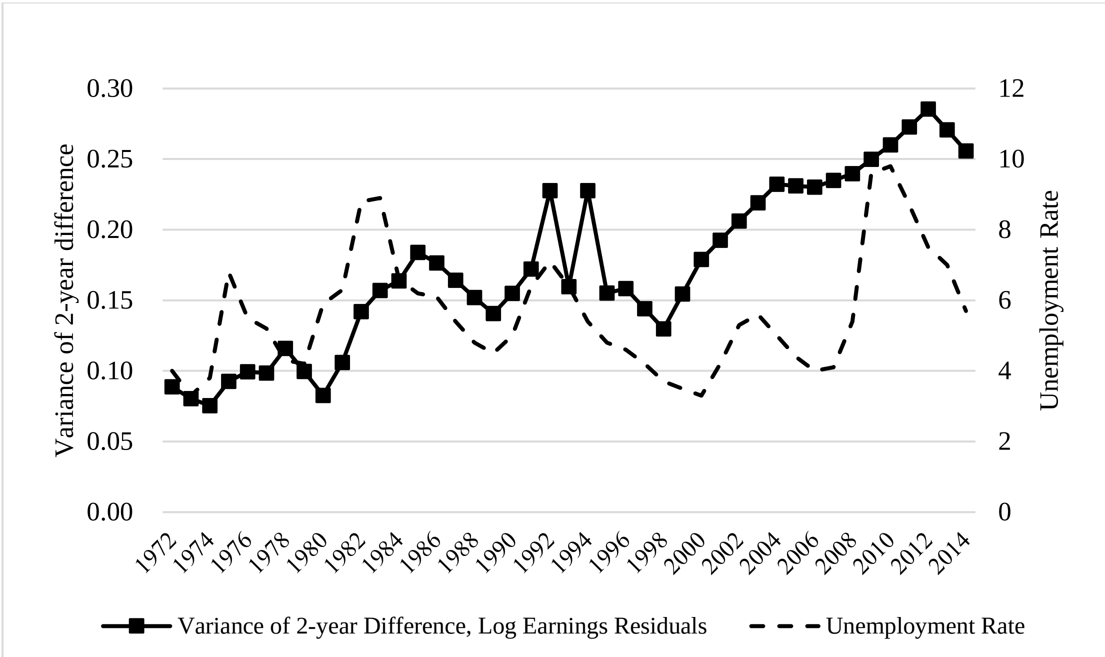

Figure 1
Variance of 2-Year Difference in Male Log Earnings Residuals

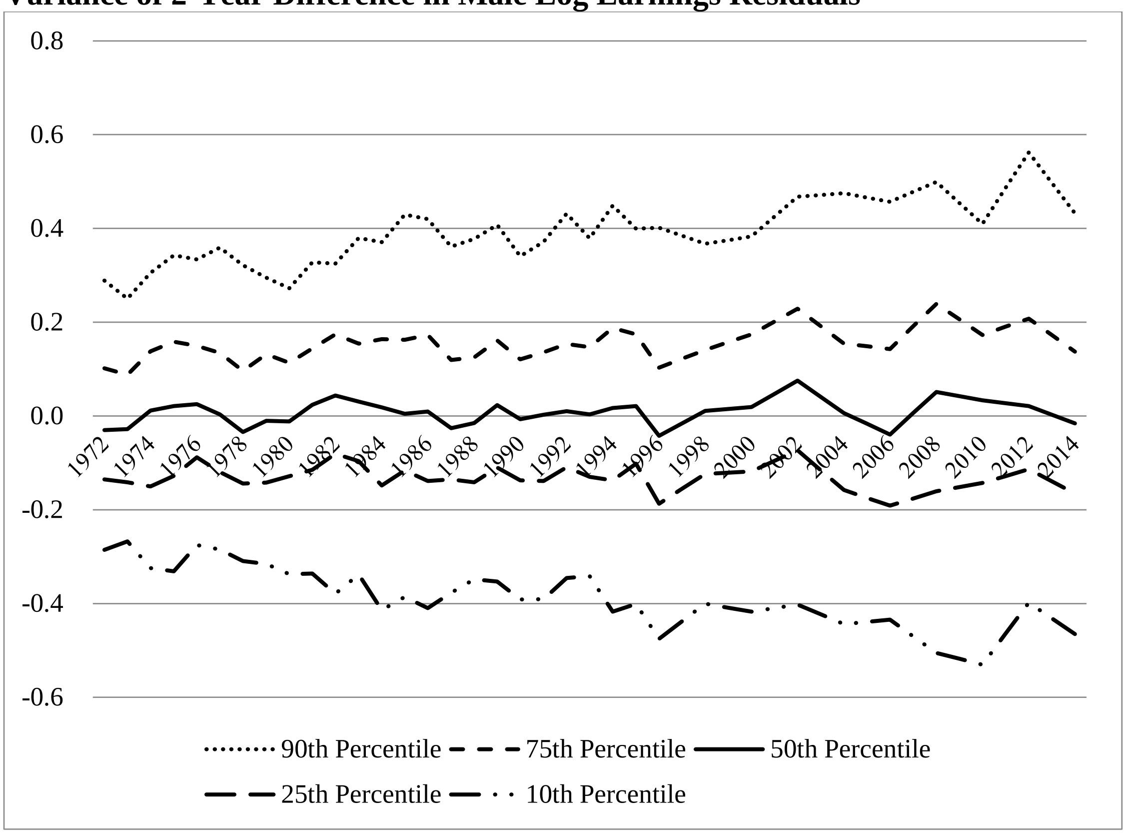

Figure 2
Percentiles of 2-Year Difference in Male Log Earnings Residuals

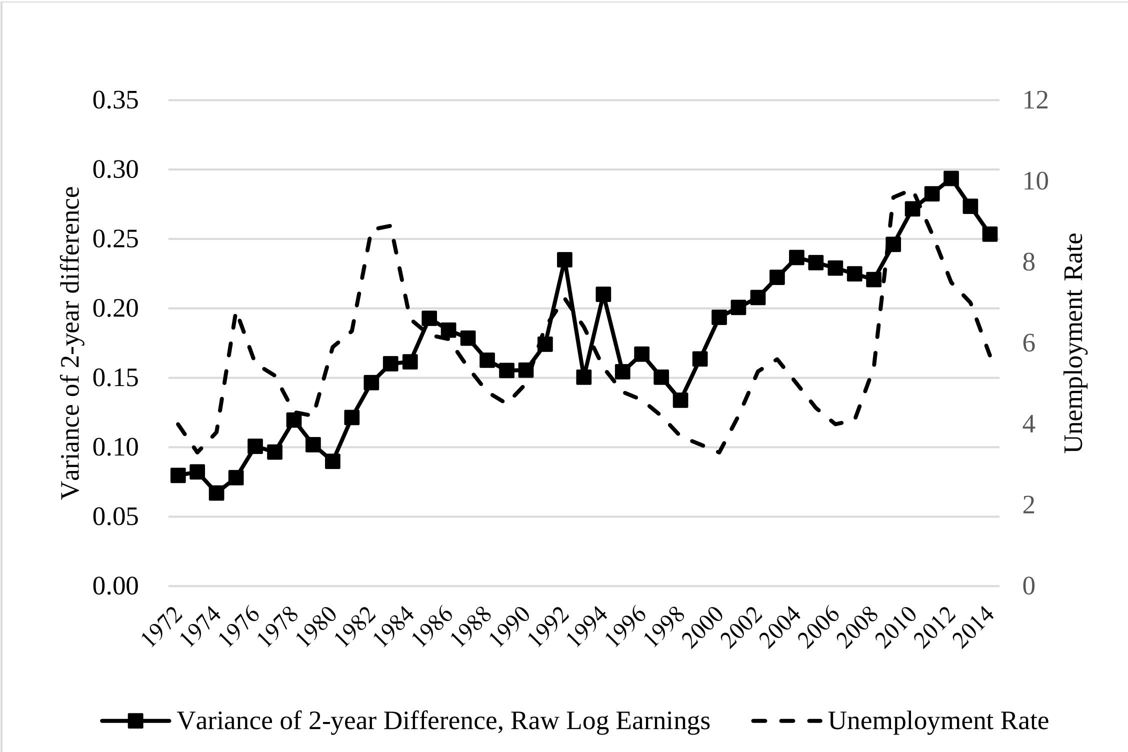

Figure 3
Variance of 2-Year Difference in Raw Male Log Earnings

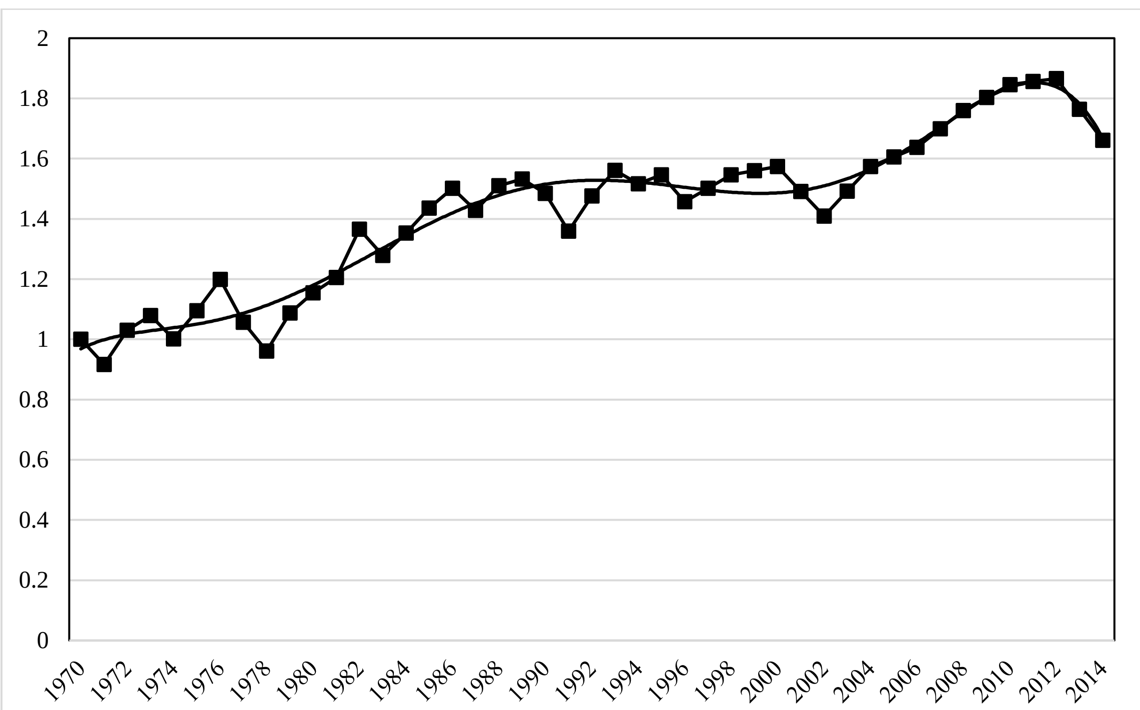

Figure 4
Extended Semiparametric (ESP) Model Estimates of Alpha

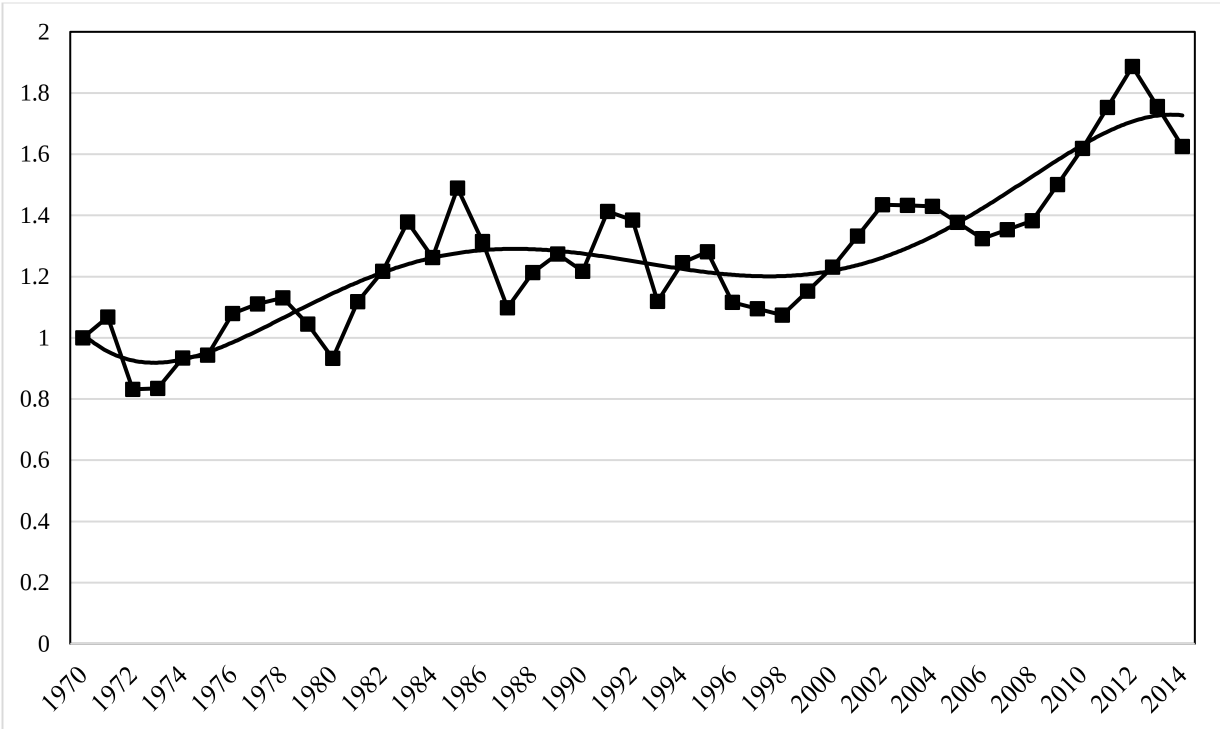

Figure 5
Extended Semiparametric (ESP) Model Estimates of Beta

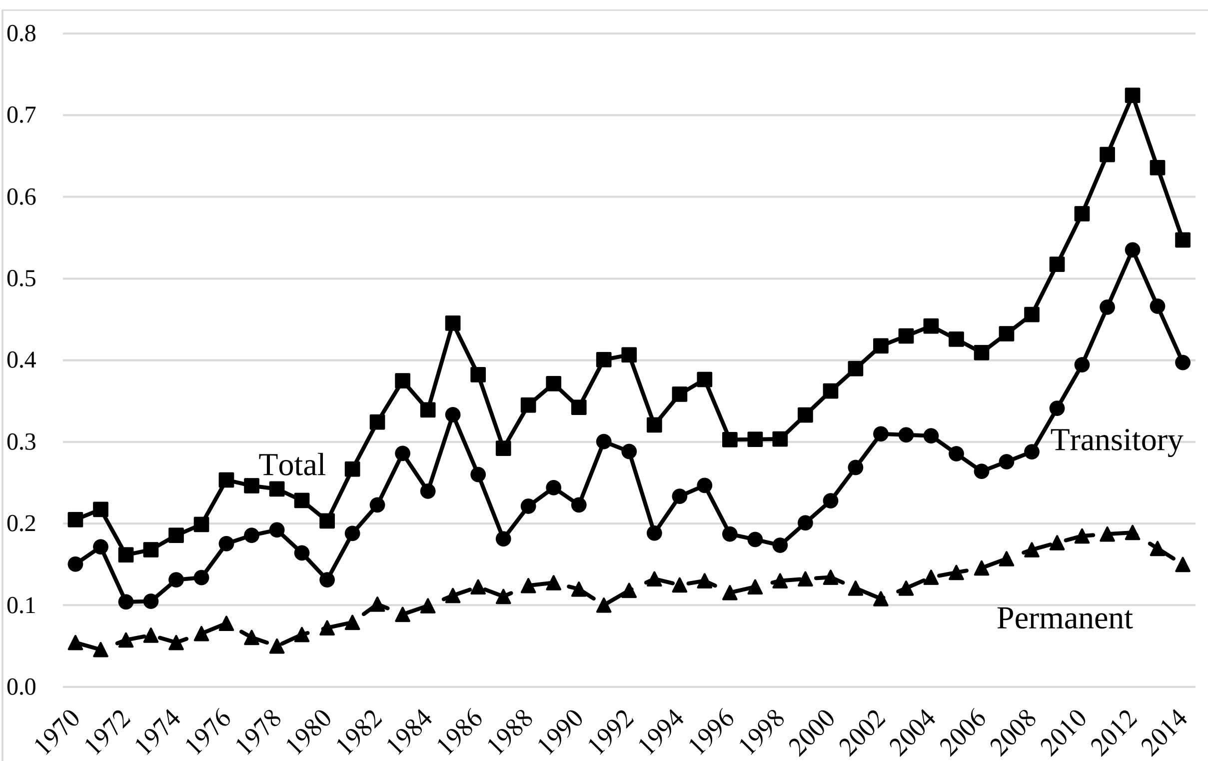

Figure 6
Fitted Permanent, Transitory, and Total Variance of Log Earnings Residuals, Age 40-49, ESP Model

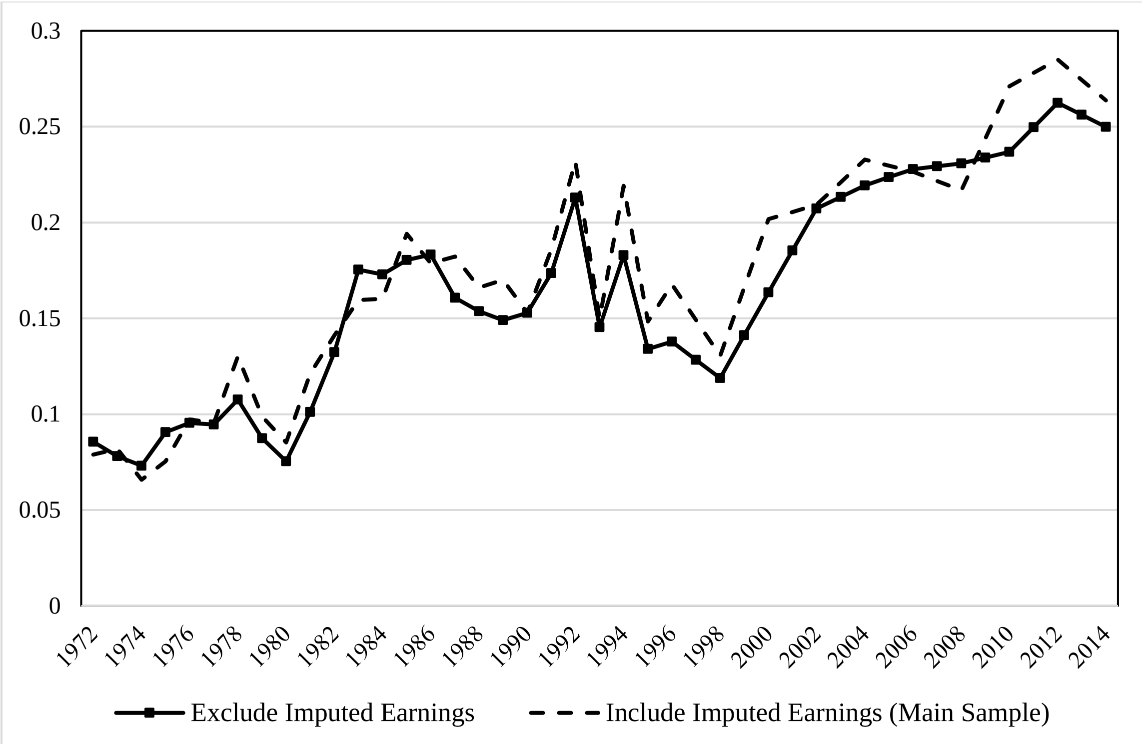

Figure 7
Variance of 2-Year Difference of Log Earnings Residuals, Including and Excluding Imputed Observations

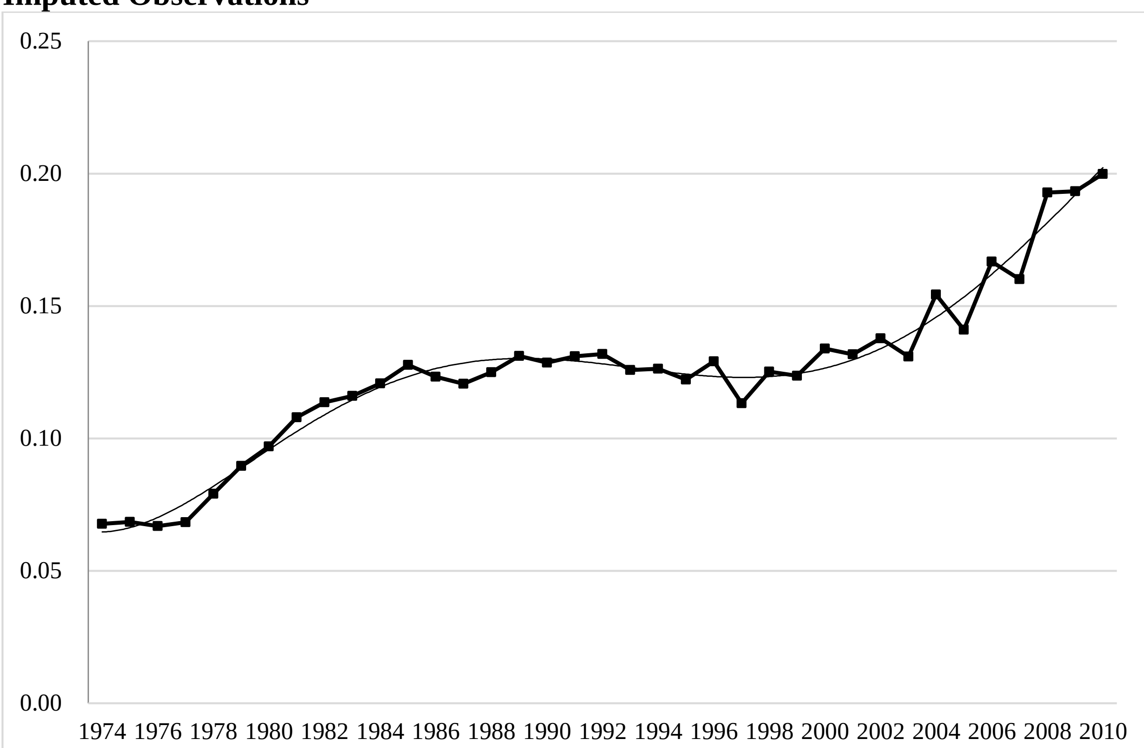

Figure 8
Window Averaging (WA) Estimate of Transitory Variance, 9-year Window

Appendix Table 1
Summary Statistics of Key Variables

Variable	No. of Obs	Mean	Standard Deviation	Minimum	Maximum
Person ID	36,403	1,524,646	826,882	1001	2,930,001
Age	36,403	42.9	8.4	30	59
Income Year	36,403	1989.4	12.4	1970	2014
Log Earnings Residual	36,403	0.020	0.589	-4.716	2.271

Appendix Table 2
Estimates of the ESP Model Parameters

Parameter	With Calendar Time Trends	Without Calendar Time Trends
Var(μ_{i0})	.054 (.010)	.104 (.006)
δ_0	-12.2 (11.7)	1.41 (.268)
δ_1	.681 (1.25)	-.842 (.104)
γ_0	-4.45 (0.30)	-6.06 (1.36)
γ_1	0.011 (.010)	.115 (.101)
k	2.21 (.38)	.004 (.024)
π	-0.010 (0.007)	3.97 (1.74)
λ_1	.094 (.021)	.070 (.024)
η_1	2.03 (0.41)	-4.97 (.752)
η_2	-.639 (.089)	.010 (.006)
η_3	.208 (.049)	.711 (.195)
α_{1971}	.916 (.096)	--
α_{1972}	1.03 (.103)	--
α_{1973}	1.08 (.101)	--
α_{1974}	1.00 (.106)	--

Appendix Table 2
Estimates of the ESP Model Parameters (continued)

Parameter	With Calendar Time Trends	Without Calendar Time Trends
α_{1975}	1.10 (.125)	--
α_{1976}	1.20 (.138)	--
α_{1977}	1.06 (.125)	--
α_{1978}	.961 (.110)	--
α_{1979}	1.09 (.135)	--
α_{1980}	1.15 (.138)	--
α_{1981}	1.21 (.143)	--
α_{1982}	1.37 (.167)	--
α_{1983}	1.28 (.161)	--
α_{1984}	1.35 (.167)	--
α_{1985}	1.14 (.173)	--
α_{1986}	1.15 (.183)	--
α_{1987}	1.43 (.171)	--
α_{1988}	1.51 (.176)	--
α_{1989}	1.53 (.176)	--
α_{1990}	1.48 (.174)	--
α_{1991}	1.36 (.168)	--
α_{1992}	1.48 (.176)	--
α_{1993}	1.56 (.181)	--
α_{1994}	1.52 (.177)	--
α_{1995}	1.55 (.182)	--

Appendix Table 2
Estimates of the ESP Model Parameters (continued)

Parameter	With Calendar Time Trends	Without Calendar Time Trends
α_{1996}	1.46 (.169)	--
α_{1998}	1.55 (.180)	--
α_{2000}	1.58 (.192)	--
α_{2002}	1.41 (.196)	--
α_{2004}	1.57 (.197)	--
α_{2006}	1.64 (.200)	--
α_{2008}	1.76 (.202)	--
α_{2010}	1.85 (.221)	--
α_{2012}	1.87 (.246)	--
α_{2014}	1.66 (.214)	--
β_{1971}	1.07 (.086)	--
β_{1972}	.832 (.071)	--
β_{1973}	.835 (.074)	--
β_{1974}	.934 (.075)	--
β_{1975}	.943 (.084)	--
β_{1976}	1.08 (.092)	--
β_{1977}	1.11 (.091)	--
β_{1978}	1.13 (.090)	--
β_{1979}	1.05 (.092)	--
β_{1980}	.933 (.092)	--
β_{1981}	1.12 (.104)	--

Appendix Table 2
Estimates of the ESP Model Parameters (continued)

Parameter	With Calendar Time Trends	Without Calendar Time Trends
β_{1982}	1.22 (.120)	--
β_{1983}	1.38 (.123)	--
β_{1984}	1.26 (.120)	--
β_{1985}	1.49 (.132)	--
β_{1986}	1.31 (.126)	--
β_{1987}	1.10 (.108)	--
β_{1988}	1.21 (.112)	--
β_{1989}	1.27 (.117)	--
β_{1990}	1.22 (.112)	--
β_{1991}	1.41 (.131)	--
β_{1992}	1.38 (.124)	--
β_{1993}	1.12 (.110)	--
β_{1994}	1.25 (.117)	--
β_{1995}	1.28 (.120)	--
β_{1996}	1.12 (.098)	--
β_{1998}	1.07 (.101)	--
β_{2000}	1.23 (.113)	--
β_{2002}	.143 (.126)	--
β_{2004}	1.43 (.119)	--
β_{2006}	1.32 (.113)	--

Appendix Table 2
Estimates of the ESP Model Parameters (continued)

Parameter	With Calendar Time Trends	Without Calendar Time Trends
β_{2008}	1.38 (.119)	--
β_{2010}	1.62 (.143)	--
β_{2012}	1.89 (.173)	--
β_{2014}	1.63 (1.43)	--

Notes:
Standard errors in parentheses.
Parameters α and β normalized to 1 in 1970.

Appendix Table 3
Estimated Permanent Variance, Transitory Variance, and Total Variance by Age Group, ESP Model

	Age 30-39			Age 40-49			Age 50-59		
	Permanent Variance	Transitory Variance	Total Variance	Permanent Variance	Transitory Variance	Total Variance	Permanent Variance	Transitory Variance	Total Variance
1970	0.054	0.122	0.176	0.054	0.150	0.205	0.082	0.183	0.266
1971	0.046	0.139	0.185	0.046	0.172	0.217	0.069	0.209	0.278
1972	0.058	0.084	0.142	0.058	0.104	0.162	0.087	0.127	0.214
1973	0.063	0.085	0.148	0.063	0.105	0.168	0.096	0.128	0.223
1974	0.054	0.106	0.161	0.054	0.131	0.186	0.082	0.160	0.242
1975	0.065	0.108	0.173	0.065	0.134	0.199	0.099	0.163	0.262
1976	0.078	0.142	0.220	0.078	0.175	0.253	0.118	0.214	0.332
1977	0.061	0.150	0.211	0.061	0.186	0.246	0.092	0.226	0.318
1978	0.050	0.156	0.206	0.050	0.192	0.243	0.076	0.235	0.311
1979	0.064	0.133	0.197	0.064	0.164	0.228	0.097	0.200	0.297
1980	0.072	0.106	0.178	0.072	0.131	0.203	0.109	0.160	0.269
1981	0.079	0.152	0.231	0.079	0.188	0.267	0.119	0.229	0.348
1982	0.101	0.181	0.282	0.101	0.223	0.324	0.153	0.272	0.425
1983	0.089	0.232	0.320	0.089	0.286	0.375	0.134	0.349	0.483
1984	0.099	0.194	0.293	0.099	0.240	0.339	0.150	0.292	0.443
1985	0.112	0.270	0.382	0.112	0.333	0.445	0.169	0.407	0.576
1986	0.122	0.211	0.333	0.122	0.260	0.382	0.185	0.317	0.502
1987	0.111	0.147	0.258	0.111	0.181	0.292	0.168	0.221	0.389
1988	0.124	0.179	0.303	0.124	0.221	0.345	0.187	0.270	0.457
1989	0.127	0.198	0.325	0.128	0.244	0.371	0.193	0.297	0.490
1990	0.120	0.180	0.300	0.120	0.223	0.342	0.181	0.272	0.453
1991	0.100	0.243	0.344	0.100	0.300	0.401	0.152	0.366	0.518
1992	0.118	0.234	0.352	0.118	0.288	0.407	0.179	0.352	0.531
1993	0.132	0.153	0.285	0.132	0.188	0.321	0.200	0.230	0.430
1994	0.125	0.189	0.314	0.125	0.233	0.358	0.189	0.285	0.474
1995	0.130	0.200	0.330	0.130	0.247	0.377	0.196	0.301	0.497

Appendix Table 3
Estimated Permanent Variance, Transitory Variance, and Total Variance by Age Group, ESP Model (continued)

	Age 30-39			Age 40-49			Age 50-59		
	Permanent Variance	Transitory Variance	Total Variance	Permanent Variance	Transitory Variance	Total Variance	Permanent Variance	Transitory Variance	Total Variance
1997	0.123	0.146	0.269	0.123	0.180	0.303	0.185	0.220	0.406
1998	0.130	0.141	0.270	0.130	0.174	0.303	0.196	0.212	0.408
1999	0.132	0.163	0.295	0.132	0.201	0.333	0.200	0.245	0.445
2000	0.134	0.185	0.319	0.134	0.228	0.362	0.203	0.278	0.481
2001	0.121	0.218	0.339	0.121	0.269	0.390	0.183	0.328	0.511
2002	0.108	0.251	0.359	0.108	0.310	0.417	0.163	0.378	0.541
2003	0.121	0.250	0.371	0.121	0.309	0.430	0.183	0.376	0.560
2004	0.134	0.249	0.384	0.134	0.308	0.442	0.203	0.375	0.579
2005	0.140	0.231	0.371	0.140	0.286	0.426	0.212	0.348	0.560
2006	0.145	0.214	0.359	0.146	0.264	0.409	0.220	0.322	0.542
2007	0.157	0.223	0.380	0.157	0.276	0.433	0.237	0.336	0.574
2008	0.168	0.233	0.401	0.168	0.288	0.456	0.254	0.351	0.605
2009	0.176	0.276	0.453	0.177	0.341	0.518	0.267	0.416	0.683
2010	0.185	0.319	0.504	0.185	0.394	0.579	0.280	0.481	0.761
2011	0.187	0.377	0.563	0.187	0.465	0.652	0.283	0.567	0.850
2012	0.189	0.434	0.622	0.189	0.535	0.724	0.286	0.653	0.939
2013	0.169	0.378	0.547	0.169	0.466	0.636	0.256	0.569	0.825
2014	0.150	0.322	0.472	0.150	0.397	0.547	0.227	0.485	0.711

Note: After income year 1996, we interpolate the variances between two years.

Appendix Table 4
Decomposition of the Variance of Two-year Changes in Log Earnings Residuals, Age 40-49, ESP Model

Second Year	Variance of Change in Permanent Component	Variance of Change in Transitory Component	Variance of Change in Total	$\alpha_t^2 Var(\mu_{ia})$	$\alpha_{t-2}^2 Var(\mu_{i,a-2})$	$-2\alpha_t\alpha_{t-2} * cov(\mu_{ia}, \mu_{i,a-2})$	$\beta_t^2 Var(\upsilon_{ia})$	$\beta_{t-2}^2 Var(\upsilon_{i,a-2})$	$-2\beta_t\beta_{t-2} * cov(\upsilon_{ia}, \upsilon_{i,a-2})$
1972	0.000	0.142	0.142	0.058	0.054	-0.112	0.104	0.144	-0.107
1973	0.001	0.155	0.157	0.063	0.046	-0.107	0.105	0.165	-0.114
1974	0.000	0.131	0.131	0.054	0.058	-0.112	0.131	0.100	-0.100
1975	0.000	0.133	0.133	0.065	0.063	-0.128	0.134	0.101	-0.101
1976	0.002	0.172	0.174	0.078	0.054	-0.130	0.175	0.126	-0.130
1977	0.000	0.179	0.180	0.061	0.065	-0.126	0.186	0.128	-0.135
1978	0.003	0.204	0.207	0.050	0.078	-0.125	0.192	0.168	-0.157
1979	0.000	0.193	0.193	0.064	0.061	-0.125	0.164	0.178	-0.149
1980	0.002	0.180	0.182	0.072	0.050	-0.120	0.131	0.185	-0.136
1981	0.001	0.196	0.196	0.079	0.064	-0.142	0.188	0.158	-0.150
1982	0.002	0.203	0.205	0.101	0.072	-0.171	0.223	0.126	-0.146
1983	0.000	0.268	0.269	0.089	0.079	-0.167	0.286	0.180	-0.198
1984	0.000	0.256	0.256	0.099	0.101	-0.201	0.240	0.214	-0.197
1985	0.001	0.344	0.346	0.112	0.089	-0.199	0.333	0.275	-0.264
1986	0.001	0.277	0.278	0.122	0.099	-0.220	0.260	0.230	-0.213
1987	0.000	0.292	0.292	0.111	0.112	-0.223	0.181	0.320	-0.210
1988	0.000	0.266	0.266	0.124	0.122	-0.246	0.221	0.250	-0.205
1989	0.001	0.238	0.239	0.128	0.111	-0.238	0.244	0.174	-0.180
1990	0.000	0.246	0.246	0.120	0.124	-0.243	0.223	0.212	-0.190
1991	0.002	0.303	0.305	0.100	0.127	-0.226	0.300	0.234	-0.231
1992	0.000	0.286	0.286	0.118	0.120	-0.238	0.288	0.214	-0.216
1993	0.002	0.274	0.276	0.132	0.100	-0.230	0.188	0.288	-0.203
1994	0.000	0.289	0.289	0.125	0.118	-0.243	0.233	0.277	-0.222
1995	0.000	0.244	0.244	0.130	0.132	-0.262	0.247	0.181	-0.184
1996	0.000	0.233	0.233	0.115	0.125	-0.240	0.187	0.224	-0.179
1997	0.000	0.216	0.217	0.123	0.120	-0.242	0.180	0.202	-0.166
1998	0.000	0.199	0.200	0.130	0.115	-0.245	0.174	0.180	-0.154
1999	0.000	0.212	0.212	0.132	0.123	-0.254	0.201	0.173	-0.162

Appendix Table 4

Decomposition of the Variance of Two-year Changes in Log Earnings Residuals, Age 40-49, ESP Model (continued)

Second Year	Variance of Change in Permanent Component	Variance of Change in Transitory Component	Variance of Change in Total	$\alpha_t^2 Var(\mu_{ia})$	$\alpha_{t-2}^2 Var(\mu_{i,a-2})$	$-2\alpha_t\alpha_{t-2} * cov(\mu_{ia}, \mu_{i,a-2})$	$\beta_t^2 Var(\upsilon_{ia})$	$\beta_{t-2}^2 Var(\upsilon_{i,a-2})$	$-2\beta_t\beta_{t-2} * cov(\upsilon_{ia}, \upsilon_{i,a-2})$
2000	0.000	0.225	0.225	0.134	0.130	-0.264	0.228	0.167	-0.170
2001	0.001	0.263	0.264	0.121	0.132	-0.252	0.269	0.193	-0.198
2002	0.002	0.302	0.303	0.108	0.134	-0.241	0.310	0.219	-0.227
2003	0.002	0.322	0.323	0.121	0.121	-0.241	0.309	0.258	-0.245
2004	0.002	0.341	0.343	0.134	0.108	-0.241	0.308	0.297	-0.263
2005	0.001	0.329	0.330	0.140	0.121	-0.260	0.286	0.296	-0.253
2006	0.000	0.316	0.316	0.146	0.134	-0.280	0.264	0.295	-0.243
2007	0.001	0.311	0.311	0.157	0.140	-0.296	0.276	0.274	-0.239
2008	0.001	0.306	0.307	0.168	0.146	-0.313	0.288	0.253	-0.235
2009	0.001	0.344	0.345	0.177	0.157	-0.333	0.341	0.265	-0.261
2010	0.000	0.383	0.383	0.185	0.168	-0.353	0.394	0.276	-0.288
2011	0.000	0.452	0.453	0.187	0.176	-0.363	0.465	0.327	-0.340
2012	0.000	0.522	0.522	0.189	0.185	-0.374	0.535	0.379	-0.392
2013	0.001	0.520	0.521	0.169	0.187	-0.355	0.466	0.446	-0.393
2014	0.002	0.517	0.520	0.150	0.189	-0.336	0.397	0.514	-0.394

Notes: See formula in Appendix.