The Extended Semiparametric (ESP) Model *AER*, 2018

by Robert Moffitt and Sisi Zhang

James J. Heckman
University of Chicago

Econ 312, Spring 2019

III. Some New Results on Trends in Male Earnings Volatility

- Instead, we specify the transitory component to be

$$v_{ia} = \varepsilon_{ia} + \sum_{s=1}^{a-1} \psi_{a,a-s} \varepsilon_{i,a-s} \tag{1}$$

- $T(T+1)/2 - T$ parameters $\psi_{a,a-s}$.

- Figures 1 and 2 show the trends in a and β, respectively.

Figure 1: Extended Semiparametric (ESP) Model Estimates of Alpha

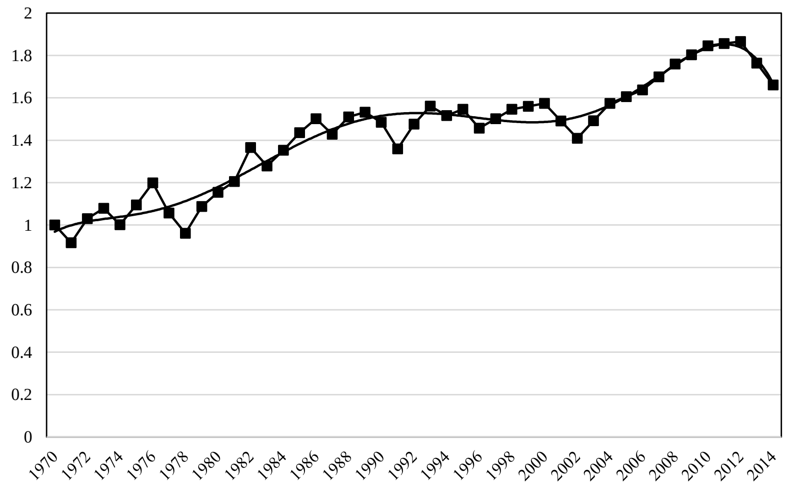
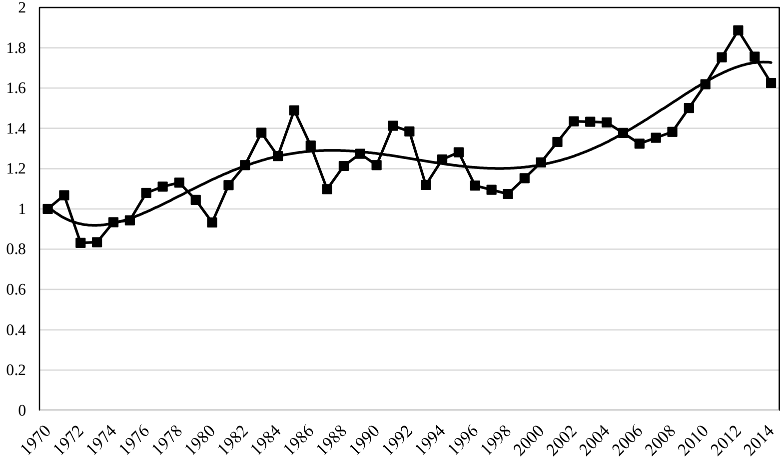
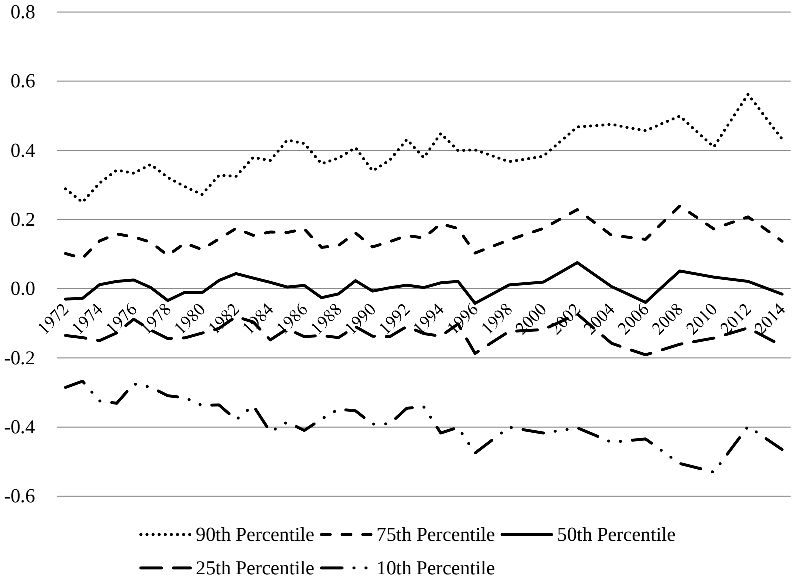


Figure 2: Extended Semiparametric (ESP) Model Estimates of Beta



- Figures 3 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: Percentiles of 2-Year Difference in Male Log Earnings Residuals



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

$$y_{iat} = a_t \mu_{ia} + \beta_t v_{ia} \tag{2}$$

where μ_{ia} is the permanent component for individual i at age a, v_{ia} is the transitory component for individual i at age a, and a_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).

- 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

$$\mu_{ia} = \mu_{i0} + \sum_{s=1}^{a} \omega_{is}. \tag{3}$$

- Let y_{iat} be the earnings residual from a Mincer equation for individual i at age a in year t, the model is:

$$y_{iat} = \alpha_t \mu_{ia} + \beta_t \upsilon_{ia} \tag{4}$$

$$\mu_{ia} = \mu_{i0} + \sum_{s=1}^{a} \omega_{is} \tag{5}$$

$$\upsilon_{ia} = \varepsilon_{ia} + \sum_{s=1}^{a-1} \psi_{a,a-s} \varepsilon_{i,a-s} \quad \text{for } a \geq 2 \tag{6}$$

$$v_{i1} = \varepsilon_{i1} \text{ for } a = 1 \tag{7}$$

  $a = 1, ..., A$ and $t = 1, ..., T$.

- Shocks ω_{ia} and ε_{ia} are independently distributed from each other and over time.

- The autocovariances implied by this model,
- Fit to the autocovariances in the data:

$$Var(y_{iat}) = \alpha_t^2 Var(\mu_{ia}) + \beta_t^2 Var(\upsilon_{ia}) \tag{8}$$

$$Var(\mu_{ia}) = Var(\mu_{i0}) + \sum_{s=1}^{a} Var(\omega_{is}) \tag{9}$$

$$Var(\upsilon_{ia}) = Var(\varepsilon_{ia}) + \sum_{s=1}^{a-1} \psi_{a,a-s}^2 Var(\varepsilon_{i,a-s}), \text{ for } a \geq 2 \tag{10}$$

$$Var(\upsilon_{i1}) = Var(\varepsilon_{i1}), \text{ for } a = 1 \tag{11}$$

$$Cov(y_{iat}, y_{i,a-\tau,t-\tau}) = \alpha_t \alpha_{t-\tau} Cov(\mu_{ia}, \mu_{i,a-\tau}) \tag{12}$$

$$+\beta_t \beta_{t-\tau} Cov(\upsilon_{ia}, \upsilon_{i,a-\tau})$$

$$\ldots$$

$$
\begin{aligned}
Cov(\mu_{ia}, \mu_{i,a-\tau}) &= Var(\mu_{i,a-\tau}) \\
&= Var(\mu_{i0}) + \sum_{s=1}^{a-\tau} Var(\omega_{is})
\end{aligned} \tag{13}
$$

$$
\begin{aligned}
Cov(\upsilon_{ia}, \upsilon_{i,a-\tau}) &= \psi_{a,a-\tau} Var(\varepsilon_{i,a-\tau}) \\
&\quad + \sum_{s=1}^{a-\tau-1} \psi_{a,a-\tau-s}\psi_{a-\tau,a-\tau-s} Var(\varepsilon_{i,a-\tau-s}), \text{ for } a \geq 3
\end{aligned} \tag{14}
$$

$$
\begin{aligned}
Cov(\upsilon_{ia}, \upsilon_{i,a-\tau}) &= \psi_{a,a-\tau} Var(\varepsilon_{i,a-\tau}) \\
&= \psi_{21} Var(\varepsilon_{i1}), \text{ for } a = 2, \tau = 1
\end{aligned} \tag{15}
$$

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

- Identification. Consider first the identification of the parameters of the age-earnings process under the stationary model $\alpha_t = \beta_t = 1$,
- 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.
- **Stationary model nonparametrically not identified without $A+1$ restrictions.**

THE UNIVERSITY OF CHICAGO

- Allow restrictions by imposing smoothness on the nonparametric functions σ_ω^2,ψ, and σ_ε^2 as described below.
- Our estimation shows that the number of parameters needed to fit the data allow the model to be heavily overidentified.
- 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 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

$$Var(y_{iat}) = \alpha_t^2 \sigma_{\mu a}^2 + \beta_t^2 \sigma_{\upsilon a}^2 \tag{16}$$

$$Var(y_{ia't}) = \alpha_t^2 \sigma_{\mu a'}^2 + \beta_t^2 \sigma_{\upsilon a'}^2 \tag{17}$$

$$Var(y_{ia,t+1}) = \alpha_t^2 r_\alpha^2 \sigma_{\mu a}^2 + \beta_t^2 r_\beta^2 \sigma_{\upsilon a}^2 \tag{18}$$

$$Var(y_{ia',t+1}) = \alpha_t^2 r_\alpha^2 \sigma_{\mu a'}^2 + \beta_t^2 r_\beta^2 \sigma_{\upsilon a'}^2 \tag{19}$$

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 $\sigma^2_{\omega a}$, $\sigma^2_{\varepsilon a}$, and ψ, 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.
- Specific functional forms are:

$$Var(\omega_{ir}) = e^{\Sigma \delta_j (r-25)^j} \tag{20}$$

$$Var(\varepsilon_{ir}) = e^{\Sigma \gamma_j (r-25)^j}, \text{ for } r \geq 2 \tag{21}$$

$$Var(\varepsilon_{i1}) = ke^{\Sigma \gamma_j (1-25)^j}, \text{ for } r = 1 \tag{22}$$

$$\psi_{A,A-b} = [1 - \pi(A-25)][\Sigma w_j e^{-\lambda_j b}] + \Sigma \eta_j D(b=j) \tag{23}$$

THE UNIVERSITY OF CHICAGO

- 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.

- 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.
- 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 UNIVERSITY OF CHICAGO

- 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.

Table 1: 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

Table 1: Estimated Permanent Variance, Transitory Variance, and Total Variance by Age Group, ESP Model, Cont'd

	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.

- 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

$$\begin{aligned} y_{iat} - y_{i,a-2,t-2} &= (\alpha_t \mu_{ia} + \beta_t \upsilon_{ia}) - (\alpha_{t-2} \mu_{i,a-2} + \beta_{t-2} \upsilon_{i,a-2}) \\ &= \alpha_t \mu_{ia} - \alpha_{t-2} \mu_{i,a-2} + \beta_t \upsilon_{ia} - \beta_{t-2} \upsilon_{i,a-2} \end{aligned} \tag{24}$$

and its variance is

$$\begin{aligned} & Var(y_{iat} - y_{i,a-2,t-2}) \\ = {} & \alpha_t^2 Var(\mu_{ia}) + \alpha_{t-2}^2 Var(\mu_{i,a-2}) - 2\alpha_t \alpha_{t-2} Cov(\mu_{ia}, \mu_{i,a-2}) \\ & + \beta_t^2 Var(\upsilon_{ia}) + \beta_{t-2}^2 Var(\upsilon_{i,a-2}) - 2\beta_t \beta_{t-2} Cov(\upsilon_{ia}, \upsilon_{i,a-2}) \end{aligned} \tag{25}$$

which contains variances and covariances which have been fitted by the model.

- Table 4 shows the exact components by year.

Table 2: 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

Table 2: Decomposition of the Variance of Two-year Changes in Log Earnings Residuals, Age 40-49, ESP Model, Cont'd

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.

Figure 4: Variance of 2-Year Difference in Male Log Earnings Residuals

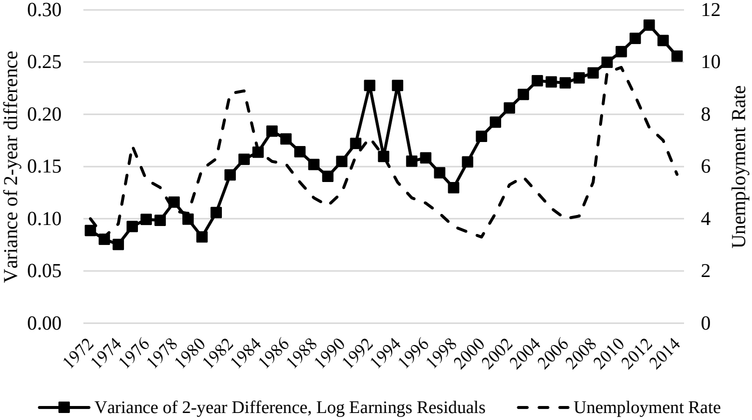


Figure 5: Variance of 2-Year Difference in Raw Male Log Earnings

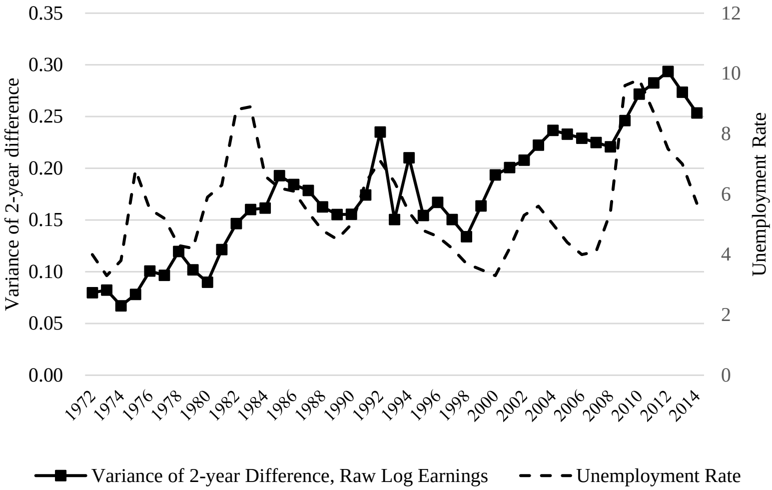


Figure 6: Fitted Permanent, Transitory, and Total Variance of Log Earnings Residuals, Age 40-49, ESP Model

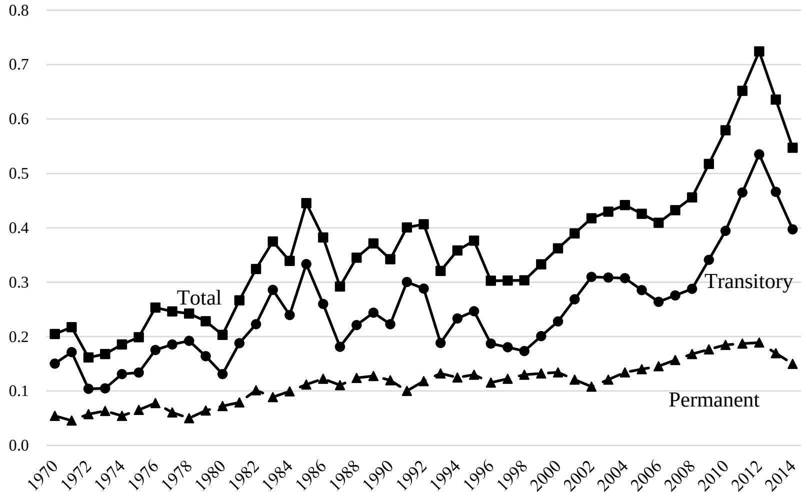


Figure 7: Variance of 2-Year Difference of Log Earnings Residuals, Including and Excluding Imputed Observations

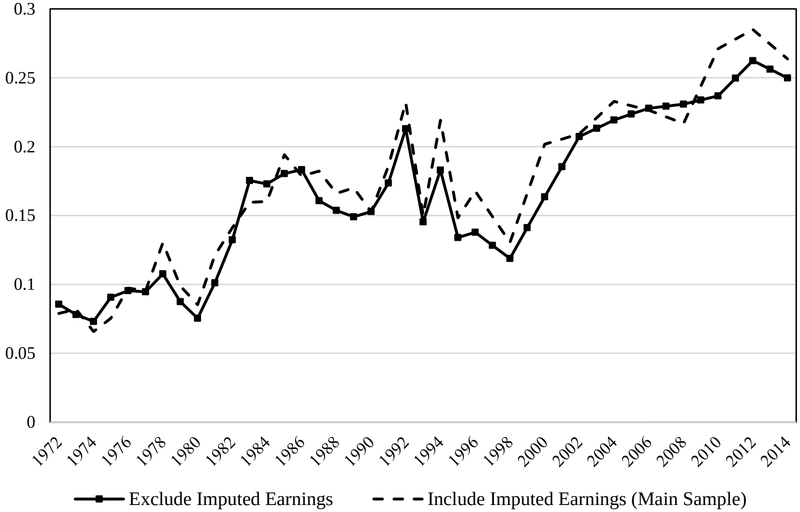


Figure 8: Window Averaging (WA) Estimate of Transitory Variance, 9-year Window

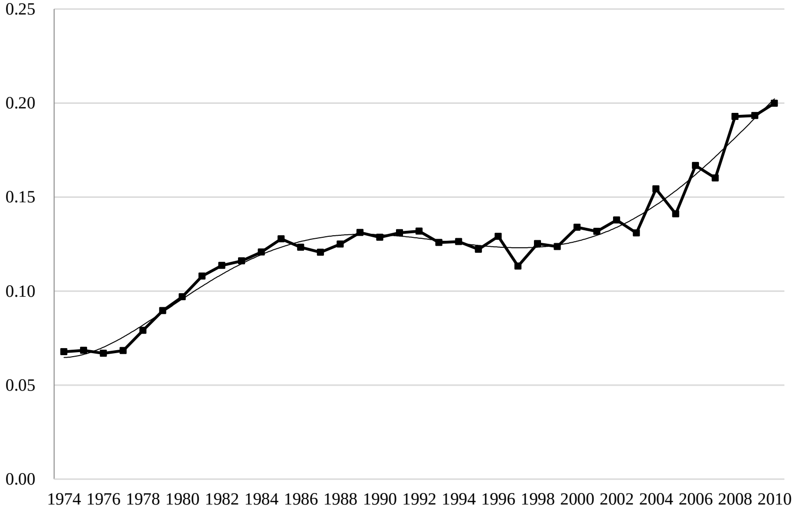


Table 3: Summary Statistics of Key Variables

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