HIP, RIP, and the Robustness of Empirical Earnings Processes

by Florian Hoffmann

Quantitative Economics, Volume 10 (2019), Pages 1279–1315

James J. Heckman

THE UNIVERSITY OF CHICAGO

Econ 350, Spring 2022

1. Introduction

- The dispersion of individual returns to experience, often referred to as heterogeneity of income profiles (HIP), is a key parameter in empirical human capital models, in studies of life-cycle income inequality, and in heterogeneous agent models of life-cycle labor market dynamics.
- It is commonly estimated from age variation in the covariance structure of earnings.
- In this study, I show that this approach is invalid and tends to deliver estimates of HIP that are biased upward.

- The reason is that any age variation in covariance structures can be rationalized by age-dependent heteroscedasticity in the distribution of earnings shocks.

- Once one models such age effects flexibly the remaining identifying variation for HIP is the shape of the tails of lag profiles.

- Credible estimation of HIP thus imposes strong demands on the data since one requires many earnings observations per individual and a low rate of sample attrition.

- To investigate empirically whether the bias in estimates of HIP from omitting age effects is quantitatively important, I thus rely on administrative data from Germany on quarterly earnings that follow workers from labor market entry until 27 years into their career.

- To strengthen external validity, I focus my analysis on an education group that displays a covariance structure with qualitatively similar properties like its North American counterpart.

- I find that a HIP model with age effects in transitory, persistent and permanent shocks fits the covariance structure almost perfectly and delivers small and insignificant estimates for the HIP component.

- In sharp contrast, once I estimate a standard HIP model without age-effects the estimated slope heterogeneity increases by a factor of thirteen and becomes highly significant, with a dramatic deterioration of model fit.

- I reach the same conclusions from estimating the two models on a different covariance structure and from conducting a Monte Carlo analysis, suggesting that my quantitative results are not an artifact of one particular sample.

2. Relation to Literature

- As of now, the debate about the importance of HIP does not seem to be settled, possibly because there is little work on credible identification of profile heterogeneity and because of the data limitations discussed in the introduction.

- I consider a considerably larger family of earnings processes, and I explicitly explore the relationship between controlling for age effects flexibly on the one hand and the validity of estimates of the HIP component on the other hand.

- Furthermore, I explore identification by exploiting the equivalence between the common estimation method in the literature, referred to as equally weighted minimum distance estimation, and nonlinear least squares regression.

- This has at least two advantages.

- First, I can address the case with many more moments than parameters, which commonly applies to minimum-distance estimation. In contrast, identification is usually established in the literature by selecting K moments that uniquely solve for K parameters, that is, the exactly identified case.

- Second, conditions for identification in non-linear least squares are well understood and, as it turns out, can be checked quite easily for the family of models considered here. This facilitates the analysis of identification considerably, even though the earnings process considered here features three variance components distinguished by their persistence, all of which feature age-dependent heteroscedasticity and nonparametric time effects.

- While a large literature uses earnings processes primarily to quantify the sources of individual life-cycle earnings variation and their changes over time, I study identification of profile heterogeneity and stress the omitted variable bias coming from omission of age effects in innovation variances.

- The result that age profiles of covariance structures cannot credibly identify slope heterogeneity, and that the bias from not modeling age effects flexibly can be severe is, to the best of my knowledge, new.

- Two areas of research on life-cycle earnings dynamics have received particularly much attention recently.
- The first exploits the joint dynamics of consumption and earnings for parameter identification.
- The central points of my paper that profile heterogeneity imposes strong restrictions on the tails of lag profiles of covariance structures and that omission of age effects leads to an upward bias in the estimates of the variance of these abilities are hard-wired into a HIP process and are thus independent of whether a process for consumption choices is specified or not.
- In practice, one important implication of my findings is that the restrictions on lag profiles imposed by a particular estimate of profile heterogeneity should be tested against the data if the estimation heavily relies on consumption data.

- The second area of active research departs from the conventional approach to earnings dynamics by going beyond autocovariance structures for estimation.
- While these studies paint a richer picture of earnings dynamics than the process considered in my work, the focus is quite different.
- Indeed, if interest is in quantifying the importance of intercept and slope heterogeneity, some parametric restrictions need to be imposed on the earnings process.
- It is in this context that I study identification.
- The focus on credible estimation of the HIP component is therefore one of the central features of my study that distinguishes it from these works.

3. Econometric Framework, Estimation, and Identification

3.1 The econometric model

- Let y_{ibt}^{e} be the log-earnings in period t of individual i born in year b who belongs to education group e.

- Assume that log earnings are described by the equation

$$y_{ibt}^{e} = \mu_{bt}^{e} + \widehat{y}_{ibt}^{e}, \tag{3.1}$$

where μ_{bt}^{e} represents a set of education specific cohort-time fixed effects and $\hat{y}_{ibt}^{e}$ is the error term.

- The focus of this study will be on the life-cycle dynamics of $\hat{y}_{ibt}^{e}$.

- Given that this is a regression error term that needs to be assumed to be conditionally independent from the observed part of (3.1), controlling flexibly for age, cohort, and education effects is important.

- The required flexibility is achieved by using the nonparametric specification in (3.1) for the observed part of the model rather than the more conventional Mincerian approach that estimates parametric linear regressions to obtain the residual of interest, $\hat{y}_{ibt}^{e}$.

- To describe the dynamics of $\hat{y}_{ibt}^{e}$, some additional notation is required.
- To avoid clutter in indexing variables, I suppress the education superscript for the rest of the paper.
- Let $t_0(b)$ be the year a cohort b enters the labor market and define $\underline{t_0} = \min\{t_0(b)\}$, which is the year the oldest cohort enters the data and hence the first sample period.
- Assume that individuals of the same cohort and education group enter the labor market at the same time so that potential experience, interchangeably referred to as age, is given by $h_{bt} = t - t_0(b)$.

- The model of $\hat{y}_{ibt}$ is given by the following set of dynamic equations:

$$\widehat{y}_{ibt} = p_t * [\alpha_i + \beta_i * h_{bt} + u_{ibt}] + z_{ibt} + \varphi_t * \varepsilon_{ibt} \tag{3.2}$$

with

$$u_{ibt} = u_{ib,t-1} + \nu_{ibt} \tag{3.3}$$

$$z_{ibt} = \rho * z_{ib,t-1} + \lambda_t * \xi_{ibt}. \tag{3.4}$$

- This model decomposes the life-cycle dynamics of residual log-earnings into three stochastic processes of different persistences.

- The first term $(\alpha_i + \beta_i * h_{bt} + u_{ibt})$ is a permanent component, updated each period by a permanent shock ν_{ibt}; the second term z_{ibt} is an AR(1)-process with persistence $\rho \in (0,1)$; and the third term ε_{ibt} is a purely transitory component.

- The set of parameters $(p_t, \lambda_t, \phi_t)_{t \geq \underline{t_0}}$ are factor loadings, one for each component.

- They allow the process of $\hat{y}_{ibt}$ to change over time, so that different cohorts are subject to different life-cycle earnings dynamics.

- Let x be some random variable, and assume that experience-dependent heteroscedasticity in its distribution can be described by a polynomial of degree J_x in h.

- All shocks and components of unobserved heterogeneity are assumed to have unconditional mean of zero and the following variance structure:

$$\operatorname{var}(\alpha_i) = \tilde{\sigma}_\alpha^2; \qquad \operatorname{var}(\beta_i) = \sigma_\beta^2; \qquad \operatorname{cov}(\alpha_i, \beta_i) = \sigma_{\alpha\beta}, \tag{3.5}$$

$$\operatorname{var}(\nu_{ibt}) = \sum_{j=0}^{J_\nu} (h_{bt})^j * \delta_j; \qquad \operatorname{var}(u_{it_0(b)}) = \tilde{\sigma}_{u0}^2, \tag{3.6}$$

$$\operatorname{var}(\xi_{ibt}) = \sum_{j=0}^{J_\xi} (h_{bt})^j * \gamma_j; \qquad \operatorname{var}(z_{it_0(b)}) = (\lambda_{t_0(b)})^2 * \sigma_{\xi_0}^2, \tag{3.7}$$

$$\operatorname{var}(\varepsilon_{ibt}) = \sum_{j=0}^{J_\varepsilon} (h_{bt})^j * \phi_j. \tag{3.8}$$

- This specification leaves initial conditions of the three experience-variance profiles unrestricted, which plays an important role in the empirical implementation below.

- No further distributional assumptions are required, but the factor loadings $(p_t, \lambda_t, \phi_t)_{t \geq \underline{t_0}}$ need to be normalized for some periods.

- The following restrictions are sufficient for identification:

$$p_{\underline{t_0}} = p_{(\underline{t_0}+1)} = \lambda_{\underline{t_0}} = \lambda_{(\underline{t_0}+1)} = \varphi_{\underline{t_0}} = 1. \tag{3.9}$$

- This completes the description of the earnings process.

3.2 Discussion

- The process described by equations (3.2) to (3.8) is very flexible and nests the majority of specifications considered in the literature that feature heterogeneous returns to experience.

- A number of features are worth highlighting.

- First, the process is the sum of a permanent, a persistent, and a purely transitory component, a decomposition that has been suggested as early as Friedman's (1957) seminal study of individual consumption choices.

- A second important feature of the earnings process described above is the rich specification of age effects.

- It is the central result of this paper that a priori restrictions on age heteroscedasticity in the distribution of earnings shocks are a model misspecification that produces an upward bias in the estimate of profile heterogeneity.

- A flexible approach to modeling age heteroscedasticity is using polynomials, as in equations (3.6), (3.7), and (3.8).

- A third feature worth emphasizing is the presence of time effects in innovation variances.
- There is a large literature emphasizing the need to control flexibly for age and time effects when estimating empirical life-cycle models of conditional first moments of the earnings distribution, as reviewed above.
- The age and time structure of the model in (3.2) to (3.8) is an application of similar ideas to second moments of life-cycle earnings dynamics.
- Indeed, changes of innovation variances over the life cycle can be driven by either age or time effects.
- For consistent estimation of the former, it is thus crucial to control for the latter.
- As a consequence, the covariance structure needs to be disaggregated to the cohort level, which imposes large demands on the data.

- The model could be enriched further, for example, by adding an MA(q) component or allowing for ARCH or GARCH in the distribution of shocks.

- I do not consider the former for two major reasons.

- First, introducing a MA(q) component would break point identification without changing the main result of the paper that omission of age effects causes an omitted variable bias of slope heterogeneity.

- Second, in empirical implementations I have found the MA(q) component to be insignificant.

- I do not allow for ARCH or GARCH because it would carry the process out of the family of processes that can be estimated from autocovariance structures.

- More importantly, the type of heteroscedasticity specified in equations (3.6), (3.7), and (3.8) can generate complex variance dynamics themselves, and it is neither clear that adding ARCH or GARCH would improve model validity nor that its parameters would be point identified.

3.3 Estimation

- The model generates theoretical autocovariances

$$\text{cov}(\widehat{y}_{ibt}, \widehat{y}_{ib,t+k}) = p_t * p_{t+k} * \left\{ \begin{array}{c} [\widetilde{\sigma}_\alpha^2 + (2h_{bt} + k) * \sigma_{\alpha\beta} + h_{bt} * (h_{bt} + k) * \sigma_\beta^2] \\ + [\widetilde{\sigma}_{u_0}^2 + f^u(h_{bt}, \delta_0, \ldots, \delta_{J_v})] \end{array} \right\}$$

$$+ \rho^k * \text{Var}(z_{ibt}) + 1(k=0) * \varphi_t^2 * \left(\sum_{j=0}^{J_\varepsilon} h_{bt}^j * \phi_j \right). \qquad (3.10)$$

where k is the order of the lag, $f^u(h_{bt}, \delta_0, \dots, \delta_{J_\nu})$ is a polynomial of order $(J_\nu + 1)$ that is linear in the $\delta_j' s$, $1(k = 0)$ is an indicator function for the variance elements, and the term $\text{Var}(z_{ibt})$ follows the recursion

$$\text{Var}(z_{it_0(b)}) = (\lambda_{t_0(b)})^2 * \sigma_{\xi_0}^2, \qquad (3.11)$$

$$\text{Var}(z_{ibt}) = \rho^2 * \text{Var}(z_{ibt-1}) + \lambda_t^2 * \left(\sum_{j=0}^{J_\xi} h_{bt}^j * \gamma_j \right) \quad \text{for all } t > t_0(b). \qquad (3.12)$$

- In stationary models, equation (3.11) can be shown to have a closed-form solution that is highly nonlinear in model parameters.
- With factor loadings on the persistent shocks, the resulting process is nonstationary and does not have a closed-form solution.
- As a consequence, this expression has to be evaluated numerically.
- In principle, one can estimate the model by matching M appropriately chosen moments, where M is the number of parameters.
- This is the approach commonly used to prove identification theoretically.
- However, it is statistically inefficient and selects the “targets” fairly arbitrarily.
- Hence, I follow the majority of the literature and adopt a Minimum Distance Estimator (MD).

- Let $\hat{C}_b$ be the estimated covariance matrix for a cohort born in year b.
- A typical element $\hat{c}_{btk}$ is the cohort-specific covariance between residual earnings in period t with residual earnings k periods apart.
- Collecting nonredundant elements of $\hat{C}_b$ in a vector $\hat{C}_b^{\mathrm{vec}}$ and stacking them yields the vector of empirical moments to be matched, denoted $\hat{C}^{\mathrm{vec}}$.
- Each element $\hat{c}_{btk}$ in $\hat{C}^{\mathrm{vec}}$ has a theoretical counterpart described by (3.10).
- Denoting the parameter vector by θ and observables by Z, I write the stacked version of these *theoretical* autocovariance matrices as $G(\theta, Z)$.

- To be clear, Z is composed of observable objects entering equation (3.10), such as age, birth year, time, the lag, and various nonlinear functions of these variables.

- The (MD) estimator for θ solves

$$\widehat{\theta} = \min_{\widetilde{\theta}} [\widehat{C}^{\text{vec}} - G(\widetilde{\theta}, Z)]' W [\widehat{C}^{\text{vec}} - G(\widetilde{\theta}, Z)], \tag{3.13}$$

- where W is some positive definite weighting matrix.

- As demonstrated by Altonji and Segal (1996), using W can introduce sizable small-sample biases, and it has become customary to use the identity matrix instead.

- In this case, $\hat{\theta}$ in (3.13) becomes the Equally Weighted Minimum Distance Estimator (EWMD).

- A seldomly used, though very useful result, is the equivalence between EWMD estimation and nonlinear least squares (NLS).

- I heavily rely on this equivalence in my discussion of identification because regression models have been studied extensively and are commonly viewed as transparent and intuitive.

- It also guides how to estimate standard errors when autocovariance structures are large.

- To see equivalence of (EWMD) and (NLS), define the regression error $\hat{\chi}_{btk}(\tilde{\theta}, Z_{btk}, \hat{c}_{btk}) = \hat{c}_{btk} - G(\tilde{\theta}, Z_{btk})$.

- Here, $\hat{c}_{btk}$ is an element in $\hat{C}^{\text{vec}}$ uniquely determined by cohort, year, and lag.

- Similarly, $G(\tilde{\theta}, Z_{btk})$ is the theoretical counterpart, the nonlinear function of parameters and observables given by equation (3.10).

- The level of observation is *cohort–year–lag*.

- By definition, $\hat{\theta}$ solves

$$\widehat{\theta} = \min_{\widetilde{\theta}} \sum_{btk} \widehat{\chi}^2_{btk}(\widetilde{\theta}, Z_{btk}, \widehat{c}_{btk}) \tag{3.14}$$

which is the (NLS)-estimation criterion, whereby one regresses autocovariances on the nonlinear parametric function $G(\theta, Z)$.

- A consistent estimator of $\sqrt{\text{var}(\hat{\theta})}$, the standard error of the $\hat{\theta}$, is readily available, but depends on the matrix of fourth-order moments of residual earnings.

- This matrix has size $\left[\dim\left(\hat{C}^{\text{vec}}\right)\right]^2$.

- Given the length of my data and its administrative nature, using a consistent estimator is infeasible.

- Instead, I use cluster-robust standard errors of the NLS-estimator in (3.14), where clusters are defined by birth cohort.

- Since this involves data that are aggregated to the cohort-year-lag level rather than individual-level earnings panel data, there is clearly an information loss, and consistent estimation of $\sqrt{\text{var}(\hat{\theta})}$ will require additional assumptions.

- In the appendix, I describe under which assumptions this approach delivers an asymptotically valid estimator of $\text{var}(\hat{\theta})$.

3.4 Identification

- Since NLS and EWMD estimation are identical, the estimator $\hat{\theta}$ solves the system of $\dim(\theta)$ first-order conditions

$$J_{\widehat{\theta}}(Z)' * \left[\widehat{C}^{\text{vec}} - G(\widehat{\theta}, Z)\right] = 0, \tag{3.15}$$

  where $J_\theta(Z) = \frac{\partial G(\theta,Z)}{\partial \theta'}$ is the Jacobian of $G(\tilde{\theta}, Z)$ at $\tilde{\theta} = \theta$, a matrix of size $\dim(Z) \times \dim(\theta)$.

- If the model structure is linear in parameters, that is, $G(\theta, Z) = Z' * \theta$, then the (NLS) estimator is equivalent to OLS: $\hat{\theta} = (Z' * Z)^{-1} * Z' * \hat{C}^{\text{vec}}$.

- Notice that the level of observation is an element in the covariance structure, *not* individual earnings.

- For general nonlinear models, there is no closed-form solution, but sufficient conditions for local point identification and consistency of the NLS-estimator $\hat{\theta}$ have been established and are as follows:

(i) $p\lim(\widehat{C}) = C.$

(ii) $C = G(\theta, Z).$

(iii) $\operatorname{rank}(J_\theta) = \dim(\theta).$

- As argued above, the EWMD estimator is the NLS estimator of the model

$$C^{\text{vec}} = G(\theta, Z) + \chi, \tag{3.16}$$

where χ is an i.i.d. error term.

- Now suppose that the model is well specified.
- As indicated by the notation above, it is assumed that the covariance structure is disaggregated to the cohort level.
- It is also assumed that recorded life cycles are sufficiently long for an order condition for identification to be satisfied.

- Then the conditions for parametric identification have the following key implications:

  (Implication 1) The parameters $\tilde{\sigma}_{\alpha}^{2}$ and $\tilde{\sigma}_{u_0}^{2}$ cannot be separately identified.

  (Implication 2) If $\rho < 1$ all other model parameters are locally point identified.

  (Implication 3) Age profiles of variances are uninformative about HIP.

  (Implication 4) Age profiles of high-order autocovariances are also uninformative about HIP.

$$\text{cov}(\widehat{y}_{ibt}, \widehat{y}_{ib,t+k}) \approx \sigma_{\alpha}^{2} + (2h_{bt} + k)\sigma_{\alpha\beta} + h_{bt} * (h_{bt} + k)\sigma_{\beta}^{2} + f^{u}(h_{bt}, \delta_0, \ldots, \delta_{K_v}). \quad (3.17)$$

  (Implication 5) A credible source of identification of HIP are the tails of lag profiles.

$$\text{cov}(\widehat{y}_{ibt}, \widehat{y}_{ib,t+k}) - \text{cov}(\widehat{y}_{ibt}, \widehat{y}_{ib,t+[k+n]}) \approx \left(n * \sigma_{\alpha\beta} + h_{bt} * n * \sigma_{\beta}^{2}\right). \quad (3.18)$$

Example 3.1

- Restricting the identifying variation for slope heterogeneity to the behavior of lag profiles at high orders can be achieved via controlling flexibly for age effects in innovation variances, as is the case for the earnings process (3.2)–(3.8).

- Conversely, if one does not allow for age effects even though they are important, then assumption (ii) is violated and slope heterogeneity will also be identified from the shape of age profiles, as discussed in Guvenen (2009).

- In this case, empirical estimates of the HIP component confound slope heterogeneity with age effects in variances of various types of shocks.

- This can be framed in terms of a classical omitted variable bias.

- To illustrate this point, suppose that the true earnings process is a simple version of (3.2)–(3.8), described by

$$\begin{aligned}
\widehat{y}_{ibt} &= \alpha_i + \beta_i * h_{bt} + u_{ibt}, \\
u_{ibt} &= u_{ib,t-1} + \nu_{ibt}, \\
\mathrm{var}(\alpha_i) = \widetilde{\sigma}_\alpha^2; \qquad \mathrm{var}(\beta_i) &= \sigma_\beta^2; \qquad \mathrm{cov}(\alpha_i, \beta_i) = 0, \\
\mathrm{var}(\nu_{ibt}) = h_{bt} * \delta_1; \qquad \mathrm{var}(u_{it_0(b)}) &= 0.
\end{aligned} \tag{3.19}$$

- This combines a HIP model and a unit roots process with linear age effects in innovation variances. The autocovariance structure (3.10) reduces to

$$\mathrm{cov}(\widehat{y}_{ibt}, \widehat{y}_{ib,t+k}) = \widetilde{\sigma}_\alpha^2 + \sigma_\beta^2 * [h_{bt} * (h_{bt} + k)] + \delta_1 * \left[\frac{h_{bt} * (h_{bt} + 1)}{2}\right]. \tag{3.20}$$

- This model is linear in parameters so that the EWMD estimator is equivalent to OLS.

- Estimation is performed on aggregate covariance structures, and I therefore drop the index i on the right-hand side.

- The level of observation is the kth order autocovariance in year t for individuals of birth cohort b.

- Now suppose one erroneously neglects the age effect in innovation variances, corresponding to the a priori restriction $\delta_1 = 0$.

- Defining $z_{bt} = \frac{h_{bt}*(h_{bt}+1)}{2}$, $x_{btk} = h_{bt} * (h_{bt} + k)$, and $\hat{c}_{btk} = \widehat{\text{cov}}(\hat{y}_{ibt}, \hat{y}_{ib,t+k})$, the parameter estimate for σ_β^2 is given by $\hat{\sigma}_\beta^2 = \frac{\sum_{btk}(x_{btk}-\bar{x})*\hat{c}_{btk}}{\sum_{btk}(x_{btk}-\bar{x})^2}$ and the omitted-variable bias formula for OLS implies that asymptotically

$$\widehat{\sigma}_\beta^2 - \sigma_\beta^2 = \delta_1 * \frac{\text{cov}(x_{btk}, z_{bt})}{\text{var}(x_{btk})}. \tag{3.21}$$

- Since $\text{cov}(x_{btk}, z_{bt}) > 0$, the bias is positive if $\delta_1 > 0$: If variances increase over the life cycle quadratically due to an increase in the dispersion of permanent shocks, and if heteroscedasticity is not properly controlled for, then the EWMD estimator mistakenly assigns all of the convexity in the experience profile to the estimate of slope heterogeneity $\hat{\sigma}_\beta^2$.

Example 3.2

- It is helpful to demonstrate graphically the predictions of various model parts on the autocovariance structure.
- To this end, I compute theoretical experience profiles corresponding to various model components, using the parameter estimates from a similar model in Baker and Solon (2003).
- Results are shown in the six panels of the Online Appendix, Figure 1.

4. Data and Descriptive Analysis

4.1 Sample Construction

- How important quantitatively is the bias in estimates of slope heterogeneity when failing to properly control for age effects in innovation variances?
- This is an empirical question and requires data.
- The discussion of identification above suggests that two data features are crucial for addressing this question convincingly.
- First, one requires panel data with many earnings observations per worker.
- Second, the attrition rate from the sample needs to be small.
- Optimally, one would also like to have a sample with an externally valid covariance structure.
- A data set that satisfies all of these requirements is the confidential version of the IABS, a 2% extract from German administrative social security records for the years 1975 to 2004.

4.2 Sample sizes

- After imposing all sample restrictions, the oldest cohort in the secondary degree group, which is the education group I will focus on for reasons explained below, is born in 1955 and enters the labor market in 1978.

- The oldest cohort in the other education group is born in 1957 and enters the labor market in 1976.

- In total, there are 4,752,287 income observations for the first and 414,231 income observations for the second education group.

4.3 Descriptive analysis

- Figure 1 plots autocovariances at different lags against potential experience h for the secondary degree group.

- Separate figures are provided for four different cohort groups, all of which display similar qualitative patterns in their covariance structures.

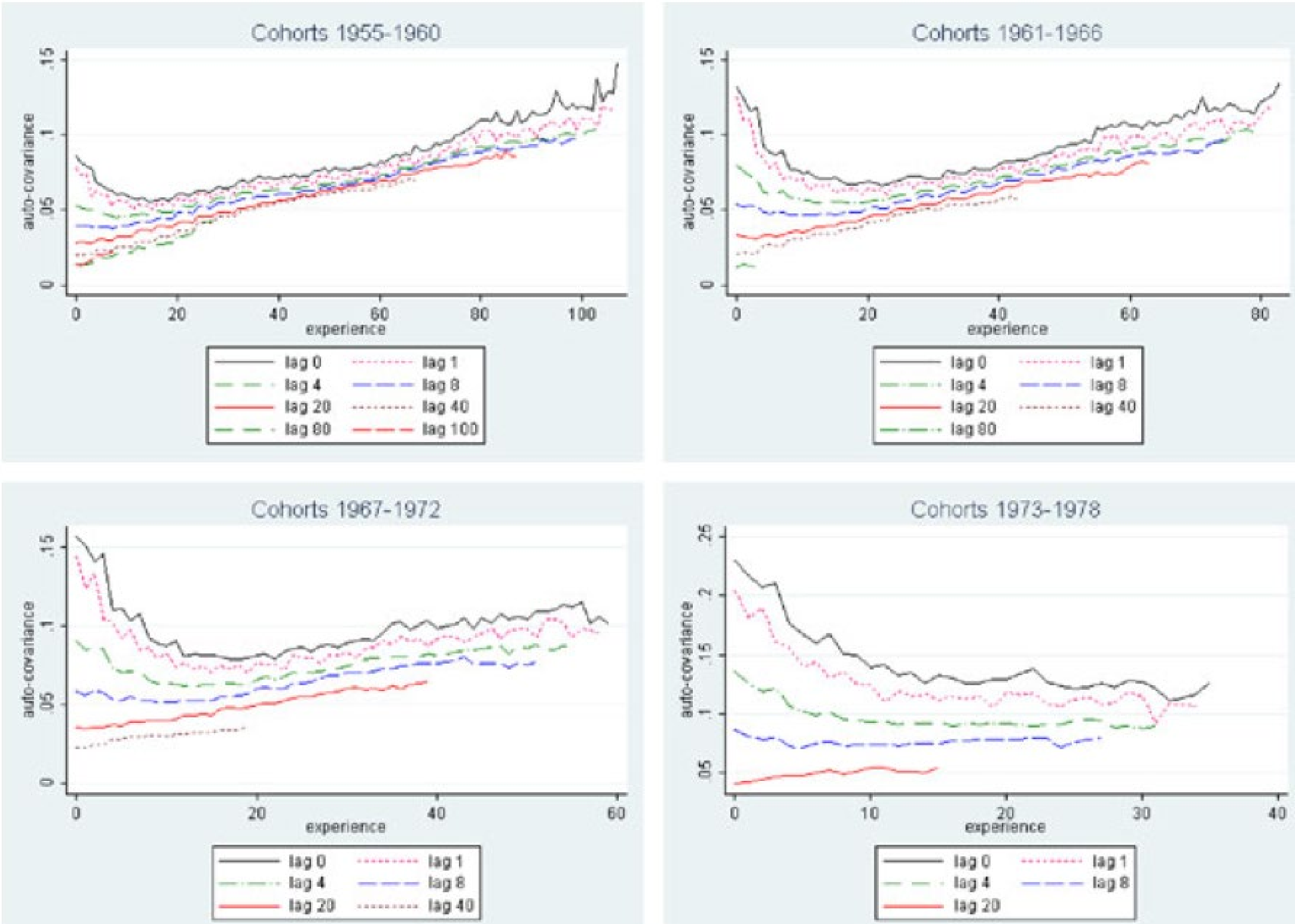
- First, autocovariances are converging gradually towards a positive constant as the lag increases, consistent with a random effects model that incorporates an AR process.

- Second, variance and autocovariance profiles at low lags decline over the first 20 to 30 quarters and increase slowly and steadily afterwards.

- Third, starting at a lag of approximately 20 quarters, the profiles become linear and strictly increasing, a possible evidence for the presence of a random walk component in earnings innovations.

- Fourth, earnings inequality as measured by the variance of log-earnings residuals is significantly larger for younger cohorts, and the same is true for higher-order covariances.

Figure 1. Life cycle profiles of autocovariances at different lags, by cohorts. Sample: Secondary Degree Group.



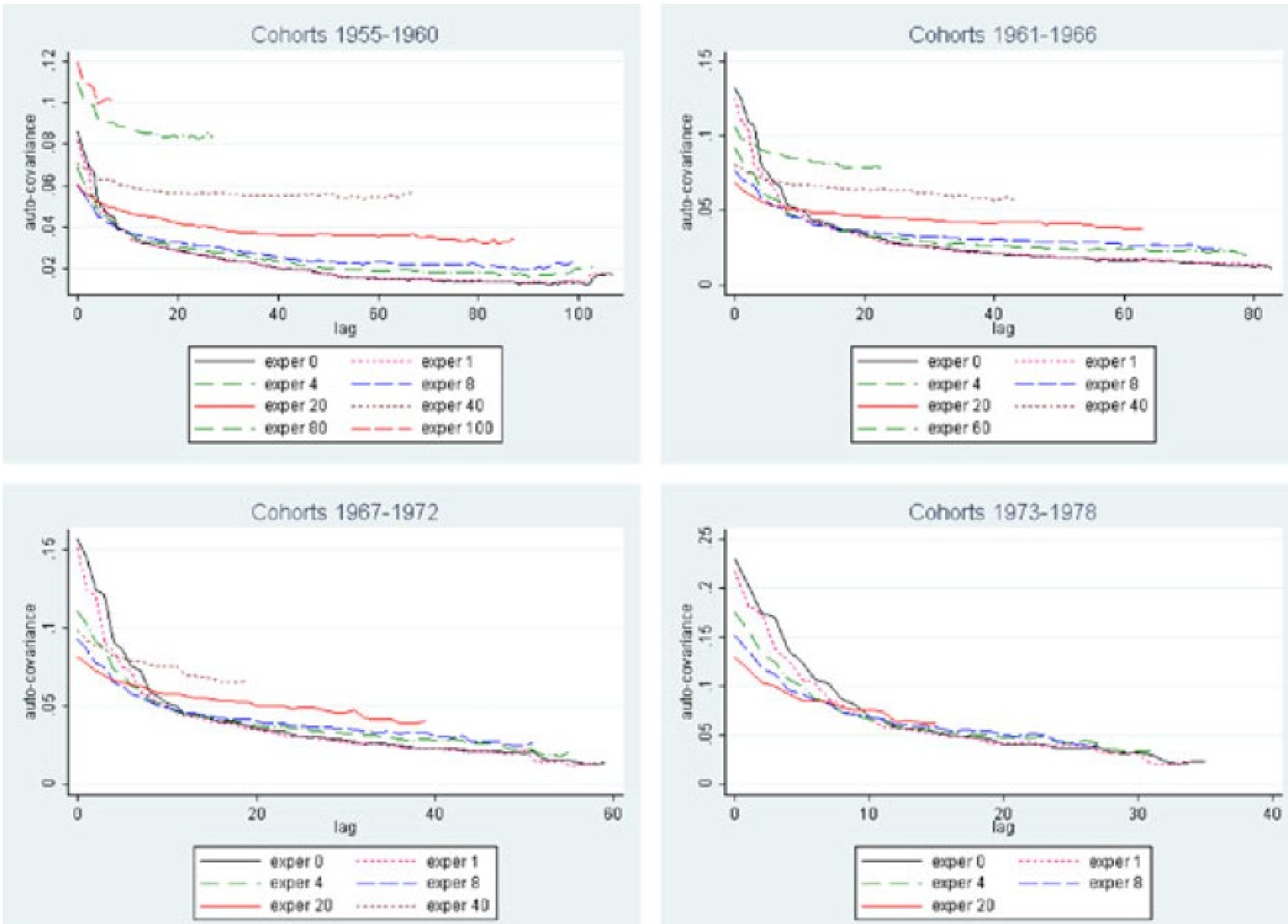
- Earnings processes do not only have implications for the shape of life-cycle profiles of autocovariances, but also for the relationship between autocovariances and the lag, holding constant labor market experience.

- I present lag profiles at different levels of experience for the secondary-degree group in Figure 2.

- Again, I split the full sample into four cohort groups. Autocovariances are gradually and monotonically decreasing, eventually converging to some positive constant.

- Other than for small lags, the profiles for older workers within cohort lie significantly above those for younger workers.

Figure 2. Lag profiles of autocovariances for different experience groups, by cohorts. Sample: Secondary Degree Group.



5. Empirical Results

- In this section, I explore quantitatively how omission of age effects in innovation variances can affect estimates of profile heterogeneity.
- I start with showing that a slightly more restrictive model than (3.2) to (3.8) can be viewed as well specified in the sense that it fits the main empirical features of the covariance structure exceptionally well.
- This benchmark specification delivers estimates of slope heterogeneity that are not significantly different from zero.
- Afterwards, I demonstrate that imposing restrictions on this benchmark specification that are common in the literature dramatically alters this conclusion.
- I use Monte Carlo analysis to demonstrate that (i) the model parameters can be estimated precisely from data of the same size and structure as the IABS even if age heteroscedasticity is modeled flexibly and that (ii) the central result of the paper that failing to control for such age effects produces substantial biases in estimates of HIP can be replicated in simulated data.

- A pretesting stage is required to determine the order of the age polynomials that govern the life-cycle variance dynamics of the process.
- This stage yields insignificant age effects for the unit roots process and the transitory component of the earnings process.
- This result can be anticipated from inspecting Figures 1 and 2.
- For the lower envelope of empirical age profiles is close to linear, consistent with a homoscedastic unit-roots process, and lag profiles are smooth around a lag of zero, suggesting that a transitory component is unlikely to be important.

- Given these results, I treat a specification that restricts $\delta_j = \varphi_j = 0$ for all $j > 0$ in equations (3.6) and (3.8) as my benchmark.

- The parameters δ_0 and φ_0 can then be interpreted, respectively, as the variance of permanent and transitory shocks for any age group.

- In contrast, I find robust and significant age effects in the persistent component, and I use a polynomial of order 4, corresponding to $J_\xi = 4$ in equation (3.7).

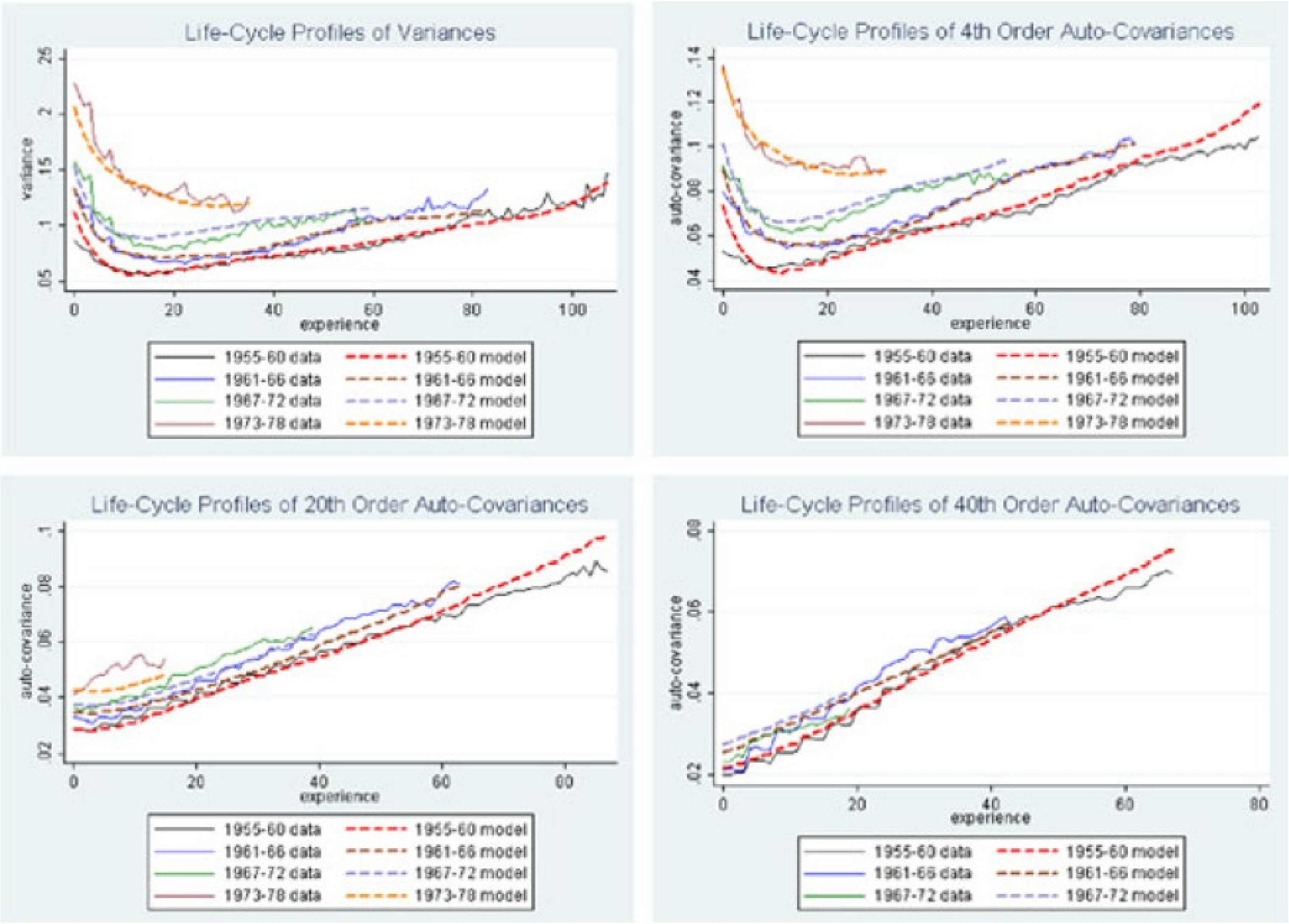
5.1 Estimates from the benchmark specification

- Parameter estimates for the benchmark specification are shown in the first column of Table 1.
- The model fit is shown in Figure 3.
- Each of the panels plot theoretical against empirical autocovariances for four cohort groups, keeping constant the lag order.
- The exercise is carried out for life-cycle profiles of autocovariances at a lag of 0, 4, 20, and 40 quarters.

Table 1. Parameter estimates for baseline specifications: Secondary degree group.

		(1) Benchmark specification	(2) No slope heterogeneity
Intercept heterogeneity	σ_α^2	0.023	0.012
		(0.004)	(0.001)
Slope heterogeneity	$\sigma_\beta^2 * 10^3$	0.0005	–
		(0.0014)	
Cov (intercept; slope)	$\sigma_{\alpha\beta}^2 * 10$	−0.001	–
		(0.0003)	
Persistence of AR(1)	ρ	0.880	0.906
		(0.006)	(0.005)
AR(1) error structure			
Initial condition	$\sigma_{\xi 0}^2$	0.092	0.080
		(0.014)	(0.005)
Intercept	γ_0	0.007	0.004
		(0.002)	(0.001)
experience	γ_1	$-3.16 * e(-4)$	$-1.49 * e(-4)$
		$(1.17 * e(-4))$	$(3.59 * e(-5))$
experience²	γ_2	$7.68 * e(-6)$	$3.18 * e(-6)$
		$(3.42 * e(-6))$	$(1.21 * e(-6))$
experience³	γ_3	$-9.16 * e(-8)$	$-3.73 * e(-8)$
		$(4.63 * e(-8))$	$(1.65 * e(-8))$
experience⁴	γ_4	$3.95 * e(-10)$	$1.61 * e(-10)$
		$(2.17 * e(-10))$	$(7.53 * e(-11))$
Variance of permanent shocks	$\delta_0 * 10$	0.007	0.004
		(0.001)	(0.001)
Variance of measurement error	ϕ_0	0.004	0.006
		(0.001)	(0.001)
Number of moments		**56,072**	
R^2		0.964	0.959
Wald test for slope heterogeneity (P-value)		0.000	–

Figure 3. Fit of benchmark model: secondary degree group.



- As can be seen from the figures, the model can generate qualitatively and quantitatively all the features of the autocovariance structure highlighted above, most importantly its evolution over the life cycle and over time.

- With EWMD estimation being equivalent to NLS, the R^2 is an informative summary measure of the goodness-of-fit.

- As can be expected from the graphical illustration, this value is very high: Over 96% of the variation in autocovariances can be explained by the model.

- This is quite remarkable given that I am matching 56,072 autocovariances with only 62 parameters.

- All parameter estimates but the variance of slopes, $\hat{\sigma}_{\beta}^{2}$, are significant on the 10%, and with few exceptions, on the 1% level.

- There is substantial heterogeneity in the intercept and the initial condition of the persistent component, with estimated variances of $\hat{\sigma}_{\alpha}^{2} = 0.023$ and $\hat{\sigma}_{\xi 0}^{2} = 0.092$, respectively.

- The estimated persistence of shocks to the AR(1) process on the quarterly level is $\hat{\rho} = 0.88$, a fairly low value.

- Age effects in the variance of the persistent component as captured by the polynomial specification is estimated to be important, with all four coefficients on the monomials in experience being statistically significant.

- The variance of the transitory component, while statistically significant, is very small, with a value of $\varphi_0 = 0.004$.

- Since both, the variance of earnings intercepts, $\hat{\sigma}_\alpha^2$, and of the transitory component, φ_0, translate one-to-one into log-earnings inequality, their magnitude can be directly related to overall log-earnings inequality.

- With a sample average of 0.094 for the 1488 variance elements in the autocovariance structure, the permanent component can explain approximately one quarter (0.023/0.094) of the total variation in log earnings in the group of the high school educated.

- Controlling for age, permanent inequality has remained almost unchanged, while persistent inequality has nearly quadrupled.

- Turning to the HIP component, the estimated heterogeneity in slopes $\hat{\sigma}_{\beta}^{2}$ is insignificant on any conventional level and very small in magnitude, but its covariance with intercept heterogeneity $\hat{\sigma}_{\alpha\beta}$ is highly significant.

- At first sight, this finding is counterintuitive, but inspection of equation (3.10) clarifies that there is no intrinsic restriction by the model that forces $\hat{\sigma}_{\alpha\beta}$ to be insignificant whenever $\hat{\sigma}_{\beta}^{2}$ is.

- It is therefore important to document a test statistic for the joint significance of the two parameters.

- The null hypothesis $(\sigma_{\alpha\beta}, \sigma_{\beta}^{2}) = (0,0)$ is rejected on the 1% level.

- The HIP hypothesis is about the heterogeneity of returns to human capital accumulation, σ_{β}^{2}, and not about its covariance with the intercept term.

- Since the results in column (1) of the table do not provide any evidence in favor of this hypothesis, I also estimate the benchmark specification with the a-priori restriction $(\sigma_{\alpha\beta}, \sigma_{\beta}^{2}) = (0,0)$.

- The estimates are listed in column (2) of the same table.

- The R^2 decreases by only 0.005, indicating that omission of the HIP component has no noticeable effect on the model fit.

- However, a number of estimates change substantially; most of all the variance of intercept heterogeneity σ2α, which decreases by a half to a value of 0.012.

- Estimates of earnings processes commonly rely on annual, rather than quarterly data.
- I therefore compute the map from my parameter estimates to their annual counterparts, which does not have a closed form.
- To this end, I simulate quarterly worker-level panel data of log-earnings in a first step, using the model, its parameter estimates, and a data structure that is identical to the one in my sample.
- In a second step, I translate these data into earnings levels, aggregate them to the annual level, transfer them back into log earnings and estimate the model on the resulting covariance structure of annual log earnings.
- The most interesting estimate coming out of this exercise is the persistence of the AR(1) process.
- On the quarterly level, this number has been estimated to be 0.88. This translates into a persistence of 0.632 on the annual frequency.

5.2 HIP and age effects: Results from misspecified models

- The discussion of identification above, in particular implications (4) and (5), predict that omission of age effects will yield inconsistent estimates of profile heterogeneity.

- This is a standard omitted variable bias because data variation that is consistent with various channels, such as age-dependent risk, contributes to identification of HIP.

- I now investigate the quantitative importance of this bias.

- Parameter estimates for a standard HIP model as favored in the heterogeneous agent literature are shown in column (2) of Table 2.

- This is a model with intercept and slope heterogeneity, an $AR(1)$-process without an initial condition, a purely transitory component, and factor loadings on both persistent and transitory shocks.

- There are no factor loadings for the HIP component.

- For a direct comparison with the benchmark specification, I reproduce its estimates in column (1) of the table.

- I also show results for a simple RIP model that does not have any time or age effects in column (3) of the table.

- The next four columns of the table explore which components of the benchmark model have a particularly large effect on the estimates of the HIP component.

Table 2. Robustness of parameter estimates: Secondary degree group.

		Full model, HIP, and RIP			Restrictions on benchmark spefication			
		(1)	(2)	(3)	(4)	(5)	(6)	(7)
		Benchmark Specification	AR(1)—HIP (Guvenen)	Simple AR(1)	Homoscedastic	Stationary	Zero initial condition for AR(1)	Models (4)–(6) combined (Hryshko, stationary)
Intercept heterogeneity	σ^2_α	0.023	0.053	0.027	0.015	0.028	0.034	0.051
		(0.004)	(0.004)	(0.002)	(0.001)	(0.003)	(0.004)	(0.004)
Slope heterogeneity	$\sigma^2_\beta * 10^3$	0.0005	0.006	–	0.000	0.002	0.002	0.005
		(0.0014)	(0.003)		(0.001)	(0.002)	(0.001)	(0.002)
Cov (intercept; slope)	$\sigma_{\alpha\beta} * 10$	−0.001	−0.005	–	−0.0006	−0.002	−0.003	−0.005
		(0.0003)	(0.0008)		(0.0001)	(0.001)	(0.001)	(0.0007)
Persistence of AR(1)	ρ	0.880	0.996	0.982	0.905	0.857	0.755	0.757
		(0.006)	(0.004)	(0.003)	(0.004)	(0.006)	(0.01)	(0.017)
AR(1) error structure								
Initial condition	$\sigma^2_{\xi 0}$	0.092	–	–	0.081	0.136	–	–
		(0.014)			(0.007)	(0.01)		
Intercept	γ_0	0.007	0.001	0.002	0.002	0.026	0.0212	0.011
		(0.002)	(0.0006)	(0.0002)	(4.31 * e(−4))	(0.007)	(0.005)	(0.002)
experience	γ_1	−3.16 * e(−4)	–	–	–	−0.001	−0.002	–
		(1.17 * e(−4))				(2.75 * e(−4))	(3.27 * e(−4))	
experience²	γ_2	7.68 * e(−6)	–	–	–	2.98 * e(−5)	4.45 * e(−5)	–
		(3.42 * e(−6))				(1.02 * e(−5))	(8.86 * e(−6))	
experience³	γ_3	−9.16 * e(−8)	–	–	–	−3.67 * e(−7)	−5.42 * e(−7)	–
		(4.63 * e(−8))				(1.48 * e(−7))	(9.93 * e(−8))	
experience⁴	γ_4	3.95 * e(−10)	–	–	–	1.64 * e(−9)	2.29 * e(−9)	–
		(2.17 * e(−10))				(7.34 * e(−10))	(3.92 * e(−10))	

(*Continues*)

Table 2. Robustness of parameter estimates: Secondary degree group, Cont'd.

		Full model, HIP, and RIP			Restrictions on benchmark spefication			
		(1)	(2)	(3)	(4)	(5)	(6)	(7)
		Benchmark Specification	AR(1)—HIP (Guvenen)	Simple AR(1)	Homoscedastic	Stationary	Zero initial condition for AR(1)	Models (4)–(6) combined (Hryshko, stationary)
Variance of permanent shocks	$\delta_0 * 10$	0.007	–	–	0.003	0.010	0.007	0.011
		(0.001)			(0.001)	(0.001)	(0.001)	(0.0009)
Variance of measurement error	ϕ_0	0.004	0.018	0.026	0.000	0.003	0.001	0.005
		(0.001)	(0.008)	(0.003)	–	(0.001)	(0.001)	(0.0006)
Number of moments					**56,072**			
R^2		0.964	0.764	0.592	0.919	0.803	0.889	0.738
Wald test for slope heterogeneity (P-value)		0.000	0.000	–	0.000	0.000	0.000	0.000

Note: This table explores the robustness of parameter estimates. Results for the benchmark specification as described in equations (3.2) to (3.8) are shown in column 1. Extensive pretesting indicated that age effects in the variances of transitory and permanent shocks are jointly insignificant. The benchmark specification thus allows for age effects in the variance of the persistent shocks only. Two specifications popular in the literature—a standard HIP-process as estimated in Guvenen (2009) and a simple RIP-process—are considered in the next two columns. The HIP-process allows for factor loadings on the permanent and the transitory (rather than the persistent) component. The four last columns explore the source of the sensitivity of parameter estimates by excluding various components from the full model: heteroscedasticity in column (4), factor loadings in column (5), an initial condition for the AR(1) process in column (6), and a combination of all these restrictions as considered in Hryshko (2012) in column (7). Standard errors are clustered by cohort to account for arbitrary correlation of sampling error within cohort-groups.

5.3 Further analysis: Robustness and a Monte Carlo analysis

- To explore further the interaction between controlling for age effects in innovation variances flexibly and the identification of HIP, I conduct two additional exercises.

- The first replicates the empirical analysis using a different sample, namely the workers in the IABS data who have no formal educational degree.

- The second investigates using Monte Carlo simulation on how well my estimation performs in finitely-sized samples.

How robust are the conclusions? Results from the high school dropout sample.

- Generally, the results from estimating my benchmark specification on the sample of high school dropouts are remarkably consistent with those found from the secondary-degree sample.

- In fact, they are even more extreme.

- The estimation of the benchmark specification delivers an estimate of zero for σ_β^2, while it is highly significant when estimating the more restrictive HIP specification.

- These results are interesting because the covariance structures for the two samples are quite different.

- Hence, the quantitative results documented in this paper are unlikely to be an artifact of one particular data set, and thus should have external validity.

A Monte Carlo analysis

- One concern with my quantitative results is that the EWMD estimator may be poorly behaved in samples of finite size, especially if one models age-heteroscedasticity flexibly.
- In particular, empirically it may be hard to distinguish between HIP and age-heteroscedasticity since identification of the former relies on the tail behavior of lag profiles, which is a second-order feature of the data.
- I address this concern with a Monte Carlo analysis.
- The main conclusion from this exercise is that a data set of the size of the IABS is sufficient to precisely estimate all model parameters.
- Most importantly, I do not find any systematic biases in the estimates of HIP and the parameters describing age heteroscedasticity, and the sampling variance across 1000 Monte Carlo repetition is small relative to the magnitude of the true parameter values.

6. Concluding Discussion

- The finding that heterogeneity in the returns to human capital accumulation needs to be identified from the joint distribution of earnings that are received many years apart may be discouraging, for two main reasons.
- On the one hand, lag profiles are most likely affected by endogenous sample attrition.
- On the other hand, patterns in the tails of lag profiles are second-order features of the data so that large sample sizes will be needed for precise parameter estimation.

- This however is not a methodological problem of matching autocovariances via EWMD or relying on earnings data only.

- Rather, it is a manifestation of the fact that it is difficult to statistically distinguish slope heterogeneity from other elements of earnings processes, such as heteroscedastic persistent shocks or a unit roots component.

- In practice, this means that data requirements for estimation of earnings processes are large, highlighting the importance of administrative data for future research.

- It is important to notice that my results do not imply that slope heterogeneity is unimportant generally.
- There may be samples and groups of workers for which the autocovariance structure of earnings is consistent with substantial profile heterogeneity.
- Instead, my results state that the HIP component will be estimated with an upward bias if age heteroscedasticity is not properly controlled for.
- This result carries over directly to any structural heterogeneous agents model in which individuals make choices about consumption or job search, as long as the underlying earnings process contains a HIP component.
- Hence, the behavior of the right tail of lag profiles of earnings covariances provide variation for a simple and powerful overidentifying test for HIP.