Supplement to "HIP, RIP, and the robustness of empirical earnings processes"

(*Quantitative Economics*, Vol. 10, No. 3, July 2019, 1279–1315)

Florian Hoffmann
Vancouver School of Economics, University of British Columbia

Appendix A: Sample construction

Constructing a quarterly panel of earnings from the IABS

The IABS reports average daily labor earnings for each employment spell of workers who are subject to compulsory social insurance contributions. According to the German Data and Transmission Act (DEÜV), employers must report at least once a year all labor earnings and some additional information such as education, training status, etc. for this group of employees. Reported earnings are gross earnings after the deduction of the employer's social security contributions. The German Employment Agency combines these data with its own information on unemployment benefits collected by individuals. Employment and unemployment spells are recorded with exact start and end dates. A spell ends for different reasons, usually due to a change in the wage paid by the firm or a change in the employment relationship. If no such change occurs, a firm has to report one spell per year. The reported average daily earnings for employment spells are total labor earnings for a spell divided by its duration in days.

To generate a panel data set that follows workers over the life cycle, one needs to choose the level of time aggregation. Theoretically, one can generate time series at the daily frequency, but given sample sizes and empirical frequencies of earnings changes, this is neither practical nor desirable. Instead, I study wage dynamics at the quarterly level. This involves aggregation of the data if a worker has more than one spell for some quarters, and disaggregation for spells that are longer than two quarters. More precisely, I keep spells that start and end in different quarters and compute the quarterly wage as the product of the reported daily earnings for this spell and the number of days of the quarter. As a consequence, spells that start and end in the same month are dropped, and spells that cross several quarters are artificially split into multiple spells, one for each quarter.[1] One rationale of choosing this approach rather than averaging all spells within a quarter is to avoid smoothing out productivity variation across jobs.[2] For the same reason, I also only keep the main job of a worker, defined by the highest paid job

Florian Hoffmann: florian.hoffmann@ubc.ca

[1]For example, a spell that takes 1 year, starting on January 1 and ending on December 31, is split into four spells, each with the same quarterly earnings.

[2]Given the lower job mobility rates in Germany compared to the US, the bias from time aggregation will be smaller than in quarterly US data.

Available at http://qeconomics.org. https://doi.org/10.3982/QE863

held during a quarter. I deflate earnings by the quarterly German CPI provided by the German Federal Statistics Office.

Censoring

Once the wage income of a worker exceeds the contribution assessment ceiling, it is replaced by the ceiling, thus introducing a censoring problem.[3] The fraction of censored observations varies strongly across education groups, providing a further motivation for estimating earnings processes for each group separately. The IABS provides an education variable with 6 categories, ranging from “no degree at all” to “university degree,” which I aggregate up to three categories, “high-school dropouts,” “secondary degree,” and “some post-secondary degree.” While I drop the last group from the analysis because of its high fraction of top-coded earnings, censoring still needs to be addressed in the other two education groups. The standard approach in studies using the PSID, such as Meghir and Pistaferri (2004) and Hryshko (2012), is to drop top-coded earnings records, introducing a sample selection problem that potentially leads to a bias in the empirical autocovariances that are matched by the model. In particular, with older workers being more likely to be at the top of the earnings distribution, dropping top-coded observations can lead to a downward bias in covariances between earnings early and late in the life cycle—the moments that provide important identification variation for the parameters. Furthermore, in contrast to missing observations, top-coded earnings records contain valid information, namely that the individual has a large positive earnings residual relative to the comparison group. For this reason, I adopt the imputation procedure in Dustmann, Ludsteck, and Schoenberg (2009), which is a static Tobit model that controls for observables with maximum flexibility and adds a random draw from some distribution.[4] While this procedure cannot determine which individuals with top-coded earnings should be allocated a particularly high residual, it captures the important fact that top-coded individuals have a larger residual component than their comparison group. The conclusions drawn in this paper are unaffected by following the literature and dropping top-coded observations altogether.

Structural break

Since 1984, it is mandatory for firms to also report one-time payments, potentially generating a discrete increase in measured earnings inequality. Steiner and Wagner (1998) showed that it is only earnings in the upper percentiles of the cross-sectional

[3]This ceiling is adjusted annually. In some cases, recorded earnings exceed the ceiling, most likely because of bonus payments and other one-time payments. In order to avoid my results to be driven by these outliers, I replace these records with the upper contribution limit.

[4]Dustmann, Ludsteck, and Schoenberg (2009) performed numerous specification checks and cross-validations with the major German survey panel data set, the SOEP, and conclude that this procedure works best among other imputation procedures. Card, Heining, and Kline (2013) adopted the same methodology to their data. Haider (2001), estimating earnings processes from the PSID, used a static imputation/interpolation procedure as well for a subset of censored observations.

distribution that are significantly affected by this change. Since I study life-cycle earnings dynamics for workers who are observed from the time of labor market entry on, those included in my sample in 1984 are relatively young, with the oldest individual being 29 years old in this year. Together with my focus on the lower educated, it is unlikely that my earnings data are significantly affected by the change in data collection.

I use several approaches to rigorously test for a structural break in the autocovariance structure. I first run a regression of the variance of residual log-income on a high-order polynomial in time and an indicator variable that is one for observations recorded past 1984, *using only those individuals who are present in the sample before 1984*.[5] For those with a secondary degree, the estimate for the dummy is 0.0013 with a standard deviation of 0.002. The R-squared is 0.86, suggesting that the regression specification approximates the evolution of the variances over time quite well. For those without a secondary degree, the corresponding estimate is −0.039 with a standard deviation of 0.016, implying that there is a significant discontinuous *decrease* in measured variances in years after the structural break. However, an R-squared of 0.47 indicates that the regression specification misses a considerable part of the evolution of variances over time. With estimates being negative, the result is more likely to be driven by experience effects. I thus reestimate the regressions for both samples, *but adding the cohorts entering the labor market after 1984*. This allows me to precisely estimate experience profiles in variances. The estimates for the break-dummy for the two samples are now 0.0018 with a standard deviation of 0.004 and 0.0002 with a standard deviation of 0.012, respectively. In both cases, the specification can explain over 80% of the variation in the data. Taken together, these results suggest that the auto-covariances matched in the estimation below are not affected by the structural break in 1984, and I thus include all cohorts I observe from the age of labor market entry.

Appendix B: Standard errors

In this section of the Online Appendix, I briefly address the issue of computing standard errors of the EWMD-estimator. In the following, assume that the model is well specified in the sense that $C^{\text{vec}} = G(\theta_0, Z)$, where θ_0 is the true parameter value, and let $J_{\theta_0}(Z) = \frac{\partial G(\theta_0, Z)}{\partial \theta_0'}$ be the Jacobian of $G(\widetilde{\theta}, Z)$ at $\widetilde{\theta} = \theta_0$. Then the asymptotic distribution of the EWMD estimator is $N(\theta_0, \frac{1}{\sqrt{N}} V_{\theta_0})$, with

$$V_{\theta_0} = \left(J_{\theta_0}' J_{\theta_0}\right)^{-1} * \left(J_{\theta_0}' * \Omega * J_{\theta_0}\right) * \left(J_{\theta_0}' J_{\theta_0}\right)^{-1}. \tag{1}$$

Here, Ω is the asymptotic covariance matrix of $\widehat{C}^{\text{vec}}$, which since $\widehat{C}^{\text{vec}}$ are autocovariances, can be interpreted as a matrix of fourth-order moments of residual log wages. This matrix has size $[\dim(\widehat{C}^{\text{vec}})]^2$ and can be estimated consistently from *individual-level* panel data on earnings. It is common to estimate Ω directly from the raw data and to plug it into the formula (1) to obtain an estimate of V_{θ_0}. Unfortunately, when relying

[5] I use a sixth-order polynomial as coefficients on higher-order terms are insignificant.

on a sample with many observations per individual and many cohorts, computing this matrix is infeasible because its size grows quadratically in the length of observed life cycles. This problem is exacerbated in my case because I use quarterly rather than annual earnings data and because the data are administrative in nature so that there are legal restrictions on the size of empirical objects that can be transferred outside of research data centers.

Without a plug-in estimator of Ω from *individual-level* data, one thus needs to ask if it is possible to obtain an asymptotically valid estimator of V_{θ_0} from data on autocovariances that are reported on the *cohort-age-lag level*. To address this question, define the vectorized error term $\chi^{\text{vec}} = (\widehat{C}^{\text{vec}} - C^{\text{vec}})$, which has the same covariance matrix Ω as $\widehat{C}^{\text{vec}}$. Its sample analogue is the vector of residuals $\widehat{\chi}^{\text{vec}}$ with generic element $\widehat{\chi}_{btk}$. From the theory of nonlinear regression, it is well known that without further restrictions on the distribution of χ^{vec} it is *not* possible to estimate Ω consistently from observations on $\widehat{\chi}^{\text{vec}}$ as there are more elements in Ω than observations in the aggregated data. At the same time, it is also known that one can estimate $V(\widehat{\theta})$ consistently without direct estimation of Ω if additional restrictions on the distribution of χ^{vec} are satisfied. This is the case if one can recover the object $(J_\theta' * \Omega * J_\theta)$ consistently, which has lower dimension than Ω.

More specifically, if it is possible to divide the sample into clusters such that (a) the $\widehat{\chi}_{btk}$ are uncorrelated across clusters, (b) the distribution of $\widehat{\chi}_{btk}$ is independent of the clustering variable conditional on Z, and (c) the number of clusters grows with sample size, then cluster-robust standard errors provide a consistent estimator of $V(\widehat{\theta})$. The question of obtaining an asymptotically valid estimator of V_{θ_0} thus reduces to the question of whether the covariance structure can be partitioned into clusters satisfying requirements (a) to (c).

To answer this question, one should notice that χ^{vec} has the interpretation of a sampling error that is uncorrelated with the explanatory variables Z_{btk} as long as the model is well specified, that is, as long as $C^{\text{vec}} = G(\theta_0, Z)$. Using cohort as clustering variable then produces an error term that satisfies all three requirements for the following reasons. First, a wage observation never enters the computation of covariance structures for different cohorts, so that sampling error will not be correlated across clusters. Importantly, under the assumption that the model is correctly specified, any correlation between cohorts that is not sampling error is controlled for. As an example, any correlation in autocovariances between cohorts that arise because of time effects is controlled for by inclusion of the factor loadings $(p_t, \lambda_t, \varphi_t)_{t \geq t_0}$. It thus follows that (a) is satisfied. At the same time, since the same residual wages enter the computation of multiple elements in $\widehat{C}^{\text{vec}}$ for the same cohort, the $\widehat{\chi}_{btk}$ cannot be uncorrelated across observations *within the same cohort*. Clustering takes care of this correlation, as long as it remains stable over time, which is requirement (b). In addition to the assumption that the model is well specified, this requires that the distribution of sampling error in the data does not change across cohorts. This is arguably the strongest assumption and represents the cost of estimating V_{θ_0} from aggregated rather than individual-level data. Assumption (c) is satisfied in my context since the IABS is a representative sample of the population and is updated regularly. This ensures that the number of cohorts grows as sample size grows. This finishes the justification of estimating V_{θ_0} using cluster-robust standard errors for (NLS), where clustering takes place on the cohort level.

Appendix C: Formal discussion of implication 2

In this section of the Online Appendix, I discuss how the rank condition for local identification can be verified, thereby providing a more formal treatment of implication 2.

Some algebraic results

It is helpful to start with a number of algebraic results that are omitted from the main text. First, define the permanent component as

$$P_{ibt} = p_t * [\alpha_i + \beta_i * h_{bt} + u_{ibt}]. \tag{2}$$

Then, using the assumptions on each of the random variables in P_t, we get

$$\begin{aligned}\operatorname{cov}(P_{ibt}, P_{ibt+k}) &= p_t * p_{t+k} * \begin{bmatrix} \operatorname{cov}\big(\alpha_i + \beta_i * h_{bt}, \alpha_i + \beta_i * (h_{bt+k} + k)\big) \\ + \operatorname{cov}(u_{ibt}, u_{ibt+k}) \end{bmatrix} \\ &= p_t * p_{t+k} * \big[\widetilde{\sigma}_\alpha^2 + (2h_{bt} + k) * \sigma_{\alpha\beta} + h_{bt} * (h_{bt} + k) * \sigma_\beta^2\big] \\ &\quad + p_t * p_{t+k} * \operatorname{var}(u_{ibt}). \end{aligned} \tag{3}$$

Using a backward recursion on the unit roots component and exploiting the independence of its shocks across periods yields:

$$\begin{aligned}\operatorname{var}(u_{ibt}) &= \widetilde{\sigma}_{u_0}^2 + \sum_{\tau=0}^{t-1} \operatorname{var}(\nu_{ibt-\tau}) \\ &= \widetilde{\sigma}_{u_0}^2 + \sum_{\tau=0}^{t-1} \sum_{j=0}^{J_\nu} (h_{bt-\tau})^j * \delta_j \\ &= \widetilde{\sigma}_{u_0}^2 + \sum_{j=0}^{J_\nu} \delta_j * \sum_{\tau=0}^{t-1} (h_{bt-\tau})^j \\ &\equiv \widetilde{\sigma}_{u_0}^2 + f^u(h_{bt}, \delta_0, \dots, \delta_{J_\nu}). \end{aligned} \tag{4}$$

Here, potential labor market experience $h_{bt-\tau}$ takes integer values in $\{1, \dots, t - t_0(b)\}$. Hence, the term $\sum_{\tau=0}^{t-1} (h_{bt-\tau})^j$ is a sum of integers to the power of j and is, applying standard results on sums of powers of integers, a polynomial of degree $(j + 1)$ with zero intercept. It follows that $f^u(h_{bt}, \delta_0, \dots, \delta_{J_\nu})$ is a polynomial of degree $(J_\nu + 1)$, intercept excluded, and that it is linear in the parameters $(\delta_0, \dots, \delta_{J_\nu})$. Hence, δ_j is the coefficient on a term that varies on the order of $(h_{bt})^{j+1}$.

Similarly, the persistent component

$$z_{ibt} = \rho * z_{ib,t-1} + \lambda_t * \xi_{ibt} \tag{5}$$

solves recursively for

$$z_{ibt} = \rho^{t-t_0(b)} * z_{ib,t-t_0(b)} + \sum_{k=0}^{t-t_0(b)-1} \rho^k * \lambda_{t-k} * \xi_{ibt-k}. \tag{6}$$

Thus,

$$\begin{aligned}\mathrm{Var}(z_{ibt}) &= \rho^{2*(t-t_0(b))} * (\lambda_{t_0(b)})^2 * \sigma^2_{\xi_0} + \sum_{\tau=0}^{t-t_0(b)-1} \rho^{2*\tau} * (\lambda_{t-\tau})^2 * \mathrm{var}(\xi_{ibt-\tau}) \\ &= \rho^{2*h_{bt}} * (\lambda_{t_0(b)})^2 * \sigma^2_{\xi_0} + \sum_{\tau=0}^{h_{bt}-1} \rho^{2*\tau} * (\lambda_{t-\tau})^2 * \sum_{j=0}^{J_\xi} (h_{bt})^j * \gamma_j \\ &= \rho^{2*h_{bt}} * (\lambda_{t_0(b)})^2 * \sigma^2_{\xi_0} + \sum_{j=0}^{J_\xi} \gamma_j * \sum_{\tau=0}^{h_{bt}-1} \rho^{2*\tau} * (\lambda_{t-\tau})^2 * (h_{bt})^j. \end{aligned} \tag{7}$$

Identification

To derive identification results, I exploit the additive composition of the covariance structure into permanent, persistent and transitory components. First, I study the permanent component and show that, if viewed in isolation, its parameters are identified under the conditions stated in the main text. Next, I show that adding the transitory component does not cause a failure of identification as long as one additional normalization on factor loadings is imposed. Finally, I study under which conditions adding the AR(1) term does not generate a failure of the rank condition. In the following discussion, I abstract from the trivial case with $p_t = 0$ or $\lambda_t = 0$ for all t. It is also understood that a partial derivative with respect to one parameter depends directly on some of the other parameters. Any statement below about these derivatives hold for any value of the full parameter vector, unless noted otherwise.

To start, it is useful to notice that $\mathrm{cov}(P_{ibt}, P_{ibt+k})$ as written in (3) is a linear regression model. The term $(p_t * p_{t+k})$ is an interaction of a set of time fixed effects, measured at t and $(t+k)$. It is interacted with an intercept, the variables $(2h_{ibt}+k)$ and $(h_{bt} * (h_{bt}+k))$ and a polynomial of degree $(J_\nu + 1)$ in h_{bt} that has intercept $\widetilde{\sigma}^2_{u_0}$. The coefficients on these variables are, in order, $\widetilde{\sigma}^2_\alpha$, $\sigma_{\alpha\beta}$, σ^2_β, the set of parameters $(\delta_0, \ldots, \delta_{J_\nu})$, and $\widetilde{\sigma}^2_{u_0}$. The parameters $\widetilde{\sigma}^2_\alpha$ and $\widetilde{\sigma}^2_{u_0}$ enter this expression additively and are thus not separately identified. This results is stated in implication 1. Furthermore, without normalization even their sum and the term $(p_t * p_{t+k})$ are not identified because they enter (3) as $(p_t * p_{t+k}) * \sigma^2_\alpha$, where $\sigma^2_\alpha = \widetilde{\sigma}^2_\alpha + \widetilde{\sigma}^2_{u_0}$. A natural normalization is treating skills in period $\underline{t_0}$ as Numeraire, so that $p_{\underline{t_0}} = 1$. With variation in k, none of the other parameters multiply variables that are collinear, so that the Jacobian of $\mathrm{cov}(P_{ibt}, P_{ibt+k})$ with respect to its parameters has full rank. As a consequence, a failure of the rank condition, if any, must come from the other variance components.

Next, write the transitory component as

$$\Xi_{ibt} = \varphi_t * \varepsilon_{ibt}. \tag{8}$$

Its contribution to the covariance structure is

$$\operatorname{cov}(\Xi_{ibt}, \Xi_{ibt+k}) = 1(k=0) * \varphi_t^2 * \left(\sum_{j=0}^{J_\varepsilon} h_{bt}^j * \phi_j \right). \tag{9}$$

This too is a linear regression model. It is a triple interaction of a dummy variable that is equal to one for variances and zero for covariances, a set of time fixed effects with coefficients φ_t^2, and a polynomial of degree J_ε in h that is linear in the parameters $(\phi_0, \dots, \phi_{J_\varepsilon})$. Because of the presence of a constant term ϕ_0 that is interacted with time fixed effects in both, (9) and (3), additional normalizations have to be imposed. Again, one choice is initializing time effects by setting $\varphi_{\underline{t_0}}^2 = 1$. This however is not sufficient because in the presence of age heteroscedasticity, age profiles of variances for different cohorts do not provide the variation to separate φ_t^2 from p_t^2. Informally, one may think of moments with $k = 0$ as "taken up" by the transitory component. Identification of the dynamic processes must thus come from covariance terms. Absent the AR(1) component, the covariance structure for $k \geq 1$ is determined by $\operatorname{cov}(P_{bt}, P_{bt+k})$, which is a linear regression model that involves interactions of time fixed effects. Every such covariance term involves products of factor loadings $p_t * p_{t+k}$ with $t > \underline{t_0}$. To set the scale of these products, one more factor loading on the permanent component needs to be normalized, and I set $p_{\underline{t_0}+1} = 1$. Now, since $1(k = 0)$ is not in the span of any of the variables in (3), the interaction terms in (9) are neither. Hence, the rank condition applied to the sum of (3) and (9) is satisfied. It must thus be true that any failure of the rank condition, if any, comes from the persistent component (7). This component is given by

$$\begin{aligned} \operatorname{cov}(z_{ibt}, z_{ibt+k}) &= \rho^k * \operatorname{Var}(z_{ibt}) \\ &= \rho^{(2*h_{bt}+k)} * (\lambda_{t_0(b)})^2 * \sigma_{\xi_0}^2 + \sum_{j=0}^{J_\xi} \gamma_j * S_j(h_{bt}, k, t), \end{aligned} \tag{10}$$

where

$$S_j(h_{bt}, k, t) = \sum_{\tau=0}^{h_{bt}-1} \rho^{(2*\tau+k)} * (\lambda_{t-\tau})^2 * (h_{bt})^j. \tag{11}$$

To understand the behavior of this term it is informative to start with the time-stationary case where $\lambda_t^2 = 1$ for all t. In this case, it is convenient to view (7) as a difference equation and write the solution, which exists as long as $\rho \in (0, 1)$, as $V_Z(h, k)$. It is possible to derive this solution analytically, but for the study of identification this does not yield any new insights. The major result is that $V_Z(h, k)$ cannot be a polynomial. Indeed, any such solution must satisfy equation (7), which cannot be written in polynomial form as long as $\rho < 1$. This remains true if $\sigma_{\xi_0}^2 = 0$, implying that the function $S_j(h, k)$ is not spanned by the space of polynomials in (h, k) of order $J = \max\{J_\varepsilon, J_\nu\}$ for any finite J. Identification follows almost immediately. First, the γ_j are coefficients on functions of (h, k) that are not in the span of polynomials of finite order, and thus not collinear with any terms in $\operatorname{cov}(P_{ibt}, P_{ibt+k})$ and $\operatorname{cov}(\Xi_{ibt}, \Xi_{ibt+k})$. Second, derivatives with respect to ρ involve the logarithmic function, which is not in the span of any of these terms either. The same is true for the derivative with respect to $\sigma^2_{\xi_0}$, which is simply $\rho^{(2*h_{bt}+k)}$. The rank condition is thus satisfied. If $\rho = 1$, instead then z_{ibt} is a unit roots process that is indistinguishable from u_{ibt} and identification fails.

Checking the rank condition for the nonstationary case follows the same line of arguments. Start with the case with $\rho < 1$. For the same reason as for the unit roots process, one needs to impose more than one normalization on factor loadings. The necessity of the restriction $\lambda_{\underline{t_0}} = 1$ is obvious, but the restriction $\lambda_{(\underline{t_0}+1)} = 1$ is not. The first two factor loadings are identified from the two oldest cohorts only, and so are the initial conditions $\sigma^2_{\xi_0}$ and σ^2_α. At the same time, because $\text{cov}(z_{ibt}, z_{ibt+k}) = \rho^k * \text{Var}(z_{ibt})$ the lag profiles for these cohorts impose restrictions on ρ, but not on any parameters in $\text{Var}(z_{ibt})$. As a consequence, without the restriction $\lambda_{(\underline{t_0}+1)} = 1$ the parameters of the AR(1)-process are not identified. Conversely, the autocovariances of orders $k = 1, 2$ for the two oldest cohorts, evaluated at $\underline{t_0}$ and $(\underline{t_0} + 1)$, pin down the initial conditions once this restriction is imposed. The remaining arguments for establishing identification are then a straightforward extension of those used above.

First, the model remains linear in the $\gamma_j' s$, but they are now coefficients on a function that varies nonpolynomially in three instead of two variables, namely (h, k, t). The model is also linear in the $(\lambda_{t-\tau})^2$ since the derivative of $\text{cov}(z_{ibt}, z_{ibt+k})$ with respect to $(\lambda_{t-\tau})^2$ does not depend on $(\lambda_{t-\tau})^2$ itself. Importantly, this derivative is a function that varies non-polynomially in (h, t). This is because the same factor loading enters at different positions of the summation in $S_j(h, k, t)$, depending on the age and the calendar year the covariance is calculated for. This also implies directly that, as long as $\rho < 1$, the partial derivatives with respect to the $\gamma_j' s$, which are the $S_j(h, k, t)$, and the factor loadings are not perfectly collinear. That is, the $S_j(h, k, t)$ are a combination of the $(\lambda_{t-\tau})^2$, but not a linear one since the coefficients are nonlinear functions. The derivative with respect to ρ involves logarithms that are not in the span of any of the polynomial terms of the other variance components.

If instead $\rho = 1$, identification fails without additional restrictions on parameters. To see this, notice that in this case $\text{cov}(z_{ibt}, z_{ibt+k}) = \text{Var}(z_{ibt})$ for all $k \geq 0$. Hence, each additional year adds one restriction on the autocovariance structure per cohort, but also one additional factor loading. As a consequence, it is generally impossible to separate age from time effects in the AR(1) process. A sufficient condition for identification is $\lambda_t = 1$ for all $t \in \{\underline{t_0}, \underline{t_0} + J_\xi + 2\}$. In this case, there are sufficiently many periods where variation in $\text{Var}(z_{ibt})$ is entirely due to age effects, and this is sufficient to pin down γ_j. This is the case even in the presence of the unit roots process. Indeed, $\text{cov}(u_{ibt}, u_{ibt+k}) = p_t * p_{t+k} * \text{var}(u_{ibt})$ so that the multiplicative nature of the factor loadings imposes more than just one restriction on autocovariances per cohort and per additional year. It introduces variation in the lag that is sufficient to recover the parameters of $\text{var}(u_{ibt})$.

Appendix D: Constrained optimization and computational issues

The MD-estimator does not impose any nonnegativity constraints on the estimates of variance parameters such as σ^2_β. If the model is misspecified, or if a variance parameter is zero while the match can be improved by choosing a negative value, these constraints may be violated. As long as a variance is summarized by a single parameter, one can easily avoid this problem by iterating over standard errors instead, or by using some positive transformations of the underlying parameters. However, variances of permanent and persistent shocks are polynomials in age, and parameters $\{\delta_j\}_{j=0}^{J_\nu}$ and $\{\gamma_j\}_{j=0}^{J_\xi}$ need to be allowed to be negative as long as $\text{var}(\nu_{ibt})$ and $\text{var}(\xi_{ibt})$ evaluated at any age are restricted to be nonnegative. The MD estimator therefore becomes the solution of a constrained minimization problem for which the constraints are linear in parameters. With an objective function that is continuously differentiable and with linear constraints, there are a number of numerical algorithms that work well in theory. After experimenting extensively with different algorithms, I have found that a SQP algorithm works best in the sense that it is least sensitive to initial values, and converges quite quickly to a solution.[6] If a variance parameter hits the constrained, calculation of standard errors becomes problematic. In this case, I restrict the parameter to zero and reestimate the model.

Appendix E: How robust are the conclusions? Results from the dropout sample

In this Online Appendix, I document and discuss results from estimating the model of earnings dynamics on the sample of high school dropouts. This exercise is interesting for two main reasons. First, the covariance structure of earnings for this group displays different features than the corresponding structure for the secondary degree sample or for the US labor market, thereby enabling me to explore the robustness of my results. Second, while the preferred model matches well the covariance structure of the more educated, it is clearly misspecified for the high school dropouts, as shown in the Appendix, Figure 5.[7] Instead of modifying the model to improve its fit—a promising approach would be to allow all parameters to vary freely across cohort groups—I investigate whether the conclusions drawn from the main sample hold when one starts from a misspecified model.

Parameter estimates for various specifications are shown in the Appendix, Table 4. This table has the same structure as Table 2. Results are thus directly comparable. Including the factor loadings, there are 62 parameters in the benchmark specification that

[6]To evaluate if a numerical solution is a candidate for a global minimizer, I use several approaches. First, since there are fast and robust numerical algorithms for unconstrained least-squares estimation, I start with solving this problem. Only if some of the constraints are violated do I reestimate the parameters. If the minimized value of the estimation criterion from the constrained routine is significantly larger than the one from the unconstrained routine, I interpret it as a sign that a global constrained minimum has not been found, and I start with a different initialization and/or a different solver.

[7]Inspection of this figure shows that the model's problems to fit the data is primarily driven by a significant change in the covariance structure for recent cohorts. Most importantly, cohorts born after 1967 experience an increase in low-order autocovariance early in the life cycle that peaks at a value higher than any covariances of older cohorts. At the same time, covariance structures late in the life cycle or at large lags appear to remain fairly stable across cohorts. This suggests that intercohort changes can only be explained by an increase in the variance of the persistent or transitory component. The model is not rich enough to account for these rather complex changes.

are estimated on a sample of 64,278 moments. Estimates are shown in column 1 of the table and, in the case of the factor loadings, in the lower panel of the Appendix, Figure 4. There are two major differences in parameter estimates of the benchmark specification compared to results from the secondary degree group. First, a Wald-test for the joint significance of $(\sigma_\beta^2, \sigma_{\alpha\beta})$ cannot reject the null hypothesis of no heterogeneity in earnings growth rates in either specification. Second, the persistent component as captured by the heteroscedastic AR(1) process plays a significantly larger role. The estimated initial condition of the AR(1) process is much larger than in the secondary degree sample. In the Appendix, Figures 2 and 3, which correspond to the experience and lag-profiles plotted in Figures 1 and 2, show that the large role of a persistent initial condition is driven by the high intercepts of lag profiles. Given the steep initial decline of the lag profiles, one may be surprised by the insignificance of $\sigma_{\alpha\beta}$. However, this decline is rather rapid and ends in a constant lag-profile later in the life cycle, which is consistent with a large persistent initial condition of the earnings process coupled with imperfect persistence. Other parameters such as the estimated variance of the intercept σ_α^2 and the persistence of the AR(1)-process ρ are quite similar to those from the secondary degree sample.

The robustness exercises documented in the rest of the table reveal patterns that are remarkably consistent with those found in the main sample. In particular, a standard HIP process yields highly significant estimates of slope heterogeneity. At the same time, the large inequality at the beginning of the life cycle is now primarily matched by intercept heterogeneity, with an estimate of σ_α^2 that is five times as large as the corresponding estimate from the full model. Furthermore, results in columns 4 to 6 imply that omission of any of the components in the benchmark specification has substantial effects on the estimates of slope heterogeneity $(\sigma_\beta^2, \sigma_{\alpha\beta})$. Again, exclusion of the persistent initial condition produces the most dramatic omitted variable bias.

Taken together, these conclusions are consistent with those found from the secondary degree sample. As the covariance structures for these two samples are quite different, the results documented in this paper are unlikely to be an artifact of one particular data set.

One interesting conclusion from this Appendix section is that controlling for age effects is important even if the model is misspecified. This hints at the difference between fitting selected moments well and estimating parameters consistently. Consistent estimation is generally possible even in misspecified models—if the identifying variation is chosen appropriately. Indeed, it is common to have a low R^2 in microeconometric studies that rely on experimental data for consistent parameter estimation. In the context of this paper, estimates of profile heterogeneity will be biased upwards if they are estimated from life-cycle variance profiles, even if the model matches these profiles perfectly. Conversely, the estimates are likely to be consistent if they are identified from the tails of lag profiles, no matter how poor the model fit is.

Appendix F: Finite sample performance: A Monte Carlo simulation

In this Appendix section, I investigate using Monte Carlo simulation whether the key parameters of my model of earnings dynamics can be recovered without any systematic biases from simulated samples of sizes similar to the IABS data. Related to this, I explore whether controlling for age effects in innovation variances, and in particular allowing for an initial condition in age profiles of second moments, tends to produce estimates of HIP that are biased toward zero. This may be an issue because it is hard to distinguish empirically between age heteroscedasticity and profile heterogeneity, as shown in the theoretical section. More specifically, once one models age heteroscedasticity flexibly, HIP is identified from the tail of lag profiles, which is a second-order feature of the data.

Simulation protocol

Every Monte Carlo exercise carried out in the following simulates 1000 data on individual-level life-cycle earnings dynamics. Sample sizes and attrition rates in each of these simulated data sets are the same as in the actual IABS data. To focus on the joint estimation of age effects in innovation variances and HIP, I abstract from time effects and simulate time-stationary earnings processes. As a consequence, I compute from each panel data on earnings one aggregate covariance structure rather than covariance structures that are disaggregated to the cohort level. This works against estimating the parameters of interest precisely since the number of observations grow faster than the number of parameters as one disaggregates to the cohort level. The parameters of the earnings process are then estimated 1000 times, once on the covariance structures computed from each of the simulated individual level panel data on earnings dynamics.

Parameters describing the Monte Carlo simulation are displayed in the upper panel of the Appendix, Table 5. Unless noted otherwise, I simulate a stationary version of the benchmark earnings process, which features HIP, an AR(1) process with age-varying innovation variances, a homoscedastic unit roots process, and a transitory component. With the exception of the HIP component, I use the estimates from the time-stationary specification in column 5 of Table 2 as parameter values. To have substantial HIP, I replace the estimates of permanent heterogeneity and the persistence of the AR(1) process by the estimates of the Hryshko-specification in column 6 of the same table. The table also displays the values of each parameters used as initial conditions in the nonlinear numerical estimation routine. These are relatively far away from their true values, but chosen on “intuitive” criteria. For example, as an initial guess for σ_α^2 I choose an approximate long-run average of autocovariances, and for the initial condition of the AR(1) process, I use the approximate difference between the intercept and the long-run value of lag profiles of labor market entrants. Age effects and HIP are initialized at zero.

Results

The results from the Monte Carlo analysis are shown in the bottom panel of the Appendix, Table 5. I simulate three different models, shown in three different sub-panels. Each subpanel in turn lists the results from estimating three different models. That is, holding fixed the model being simulated, I estimate three different models on each of the 1000 simulated covariance structures. The first model does not impose any restrictions on the parameters in the estimation. In this case, parameter estimates should not be systematically biased, but they may be estimated with less precision if the true underlying model is more restrictive than the estimated model. The second model imposes $(\sigma^2_\beta, \sigma_{\alpha\beta}) = (0, 0)$, that is, the absence of HIP, in the estimation, and the third model sets the initial condition of the AR(1)-process to zero. To avoid clutter in the table, I only show results for the key parameters, namely those describing individual heterogeneity $(\sigma^2_\alpha, \sigma^2_\beta, \sigma_{\alpha\beta})$ and the persistence and the initial condition of the AR(1) process $(\rho, \sigma^2_{\xi_0})$. The sampling distribution of their 1000 parameters estimates are summarized in two statistics, the average bias and the standard deviation.

The three subpanels differ by which of the assumptions used in the estimation are actually imposed in the simulation. More specifically, the first simulated model does not impose any restrictions on the parameter values. This simulation has two principle goals. First, it explores whether parameters of an unrestricted process can be recovered precisely and without bias from the simulated data. Indeed, as shown in column 1, there is no significant bias in any of the parameters. This is reassuring and suggests that given the sample sizes a process with HIP and a rich structure for age-dependent heteroscedasticity in earnings dynamics can be estimated precisely and without bias. The second goal is to investigate whether imposing erroneous assumptions introduces substantial biases in the estimates. To this end, the next two columns impose the restrictions described above, both of which are wrong. Not surprisingly, this leads to substantial biases in most parameter estimates. Perhaps most importantly, if the initial condition of the AR(1) process is omitted, a specification whose estimates are shown in column 3, the parameters of HIP are severely biased upwards in absolute value, and these biases are highly significant. Hence, even in the presence of HIP, omission of age effects introduces substantial biases. This reaffirms the central result of this paper.

The second simulation sets the parameters of the HIP component to zero: $(\sigma^2_\beta, \sigma_{\alpha\beta}) = (0, 0)$. As shown in columns 4 and 5, there are no significant biases no matter if this restriction is imposed in the estimation (column 5) or not (column 4). In contrast, once the initial condition of the AR(1) process is erroneously set to zero, the estimate of $\sigma_{\alpha\beta}$ tends to be biased away from zero, and this bias is significant. The omitted variable bias in σ^2_β has a similar magnitude as in simulation 1 when the persistent initial condition is erroneously set to zero in the estimation. However, the sampling distribution is now too dispersed for this bias to be significant. Two points need to be kept in mind when interpreting this result however. First, a Wald test of the null hypothesis that $(\sigma^2_\beta, \sigma_{\alpha\beta})$ are jointly zero given their sampling distribution would be rejected. Second, in this particular exercise, age effects in the variances of the AR(1) process other than the initial condition are still allowed for in the estimation. Omitting age heteroscedasticity altogether leads to larger biases in estimates of HIP, though this is not shown in the table.

Finally, the third simulation explores whether allowing for an initial condition of the AR(1) process in the estimation may lead to overfitting in the sense that estimates of HIP are systematically biased toward zero. More specifically, it answers the question of whether allowing for a persistent initial condition in the estimation when none exists leads to biased estimates of HIP. Results in columns 7 to 9 indicate that this is not the case. Significant biases arise only in the case of erroneously omitting HIP in the estimation, as shown in column 8. In this case, the average estimate of $\sigma^2_{\xi_0}$ is 0.065, compared to a true value of zero.

To summarize, the Monte Carlo analysis establishes three results. First, as long as one does not impose a wrong restriction on the parameters of the earnings process, all parameters can be recovered precisely and without bias using equally weighted minimum distance estimation, at least given the sample sizes and the length of the panels in the IABS. Second, erroneously omitting the persistent initial condition, which is a particular age effect in the innovation variances of the AR(1) process, leads to substantial upward biases in the estimates of profile heterogeneity. Third, controlling for a persistent initial condition if none exists does not introduce any biases in estimates of HIP.

References

Card, D., J. Heining, and P. Kline (2013), "Workplace heterogeneity and the rise of West German wage inequality." *Quarterly Journal of Economics*, 128, 967–1015. [2]

Dustmann, C., J. Ludsteck, and U. Schoenberg (2009), "Revisiting the German wage structure." *Quarterly Journal of Economics*, 124, 843–881. [2]

Guvenen, F. (2009), "An empirical investigation of labor income processes." *Review of Economic Dynamics*, 12 (1), 58–79. [18]

Haider, S. (2001), "Earnings instability and earnings inequality of males in the United States: 1967–1991." *Journal of Labor Economics*, 19 (4), 799–836. [2]

Hryshko, D. (2012), "Labor income profiles are not heterogeneous: Evidence from income growth rates." *Quantitative Economics*, 3, 177–209. [2, 18]

Meghir, C. and L. Pistaferri (2004), "Income variance dynamics and heterogeneity." *Econometrica*, 72 (1), 1–32. [2]

Steiner, V. and K. Wagner (1998), "Has earnings inequality in Germany changed in the 1980's?" *Zeitschrift fuer Wirtschafts- und Sozialwissenschaften*, 118, 29–59. [2]

Co-editor Peter Arcidiacono handled this manuscript.

Manuscript received 2 April, 2017; final version accepted 21 November, 2018; available online 9 January, 2019.

Appendix Table 1. Sample sizes by education group, cohort, and experience.

Panel A: Sample sizes by education group and cohort			Panel B: Sample sizes by education group and experience		
	Education group			Education group	
Cohort	*Main Sample*: Secondary Degree Group	*Robustness Sample*: High School Dropouts	Experience (in years)	*Main Sample*: Secondary Degree Group	*Robustness Sample*: High School Dropouts
1955	250,387	–	0	322,907	34,966
1956	258,708	–	1	317,001	29,831
1957	315,867	31,810	2	312,600	25,600
1958	306,775	29,055	3	311,899	23,824
1959	315,022	31,763	4	302,277	23,107
1960	303,563	27,712	5	290,854	22,560
1961	297,691	25,912	6	279,578	22,014
1962	286,116	27,374	7	267,364	21,312
1963	289,492	26,231	8	255,348	20,023
1964	278,860	27,494	9	242,766	18,918
1965	263,462	23,218	10	228,805	17,816
1966	243,002	20,521	11	214,144	16,877
1967	226,214	18,152	12	199,207	15,938
1968	207,863	18,576	13	183,078	15,171
1969	179,188	13,766	14	166,489	14,243
1970	150,000	13,882	15	150,119	13,224
1971	128,949	12,101	16	134,430	12,309
1972	105,640	11,328	17	118,782	11,422
1973	79,753	9582	18	103,754	10,345
1974	72,457	9464	19	89,320	9472
1975	63,300	8260	20	75,806	8482
1976	53,368	9045	21	62,466	7425
1977	43,998	9540	22	50,103	6318
1978	32,612	9445	23	37,565	5318
			24	26,427	4328
			25	15,797	–
			26	7485	–
Total	4,752,287	414,231		4,752,287	414,231

Appendix Table 2. Average labor income by education group and experience (in years).

Experience (in years)	Education group	
	Main Sample: Secondary Degree Group	*Robustness Sample*: High School Dropouts
0	8.556	7.807
1	8.631	8.013
2	8.686	8.263
3	8.731	8.450
4	8.768	8.550
5	8.802	8.596
6	8.836	8.637
7	8.865	8.670
8	8.891	8.701
9	8.916	8.728
10	8.937	8.755
11	8.957	8.773
12	8.973	8.791
13	8.988	8.805
14	9.000	8.821
15	9.012	8.835
16	9.022	8.839
17	9.034	8.840
18	9.044	8.848
19	9.051	8.859
20	9.061	8.862
21	9.072	8.872
22	9.081	8.870
23	9.087	8.871
24	9.098	8.884
25	9.111	–
26	9.111	–

Appendix Table 3. Parameter estimates for baseline specifications, annual data: secondary degree group.

		(1) Benchmark specification	(2) No slope heterogeneity
Intercept heterogeneity	σ^2_α	0.021	0.012
Slope heterogeneity	$\sigma^2_\beta * 10^3$	0.004	–
Cov (intercept; slope)	$\sigma_{\alpha\beta} * 10$	−0.003	–
Persistence of AR(1)	ρ	0.632	0.688
Initial condition of AR(1)	$\sigma^2_{\xi_0}$	0.078	0.072
Permanent shocks	$\delta_0 * 10$	0.024	0.015
Number of moments		**3644**	

Note: This table shows parameter estimates for the specifications in Table 1, but estimated from simulated annual data. It provides the numerical mapping from parameters of the benchmark earnings processes on the quarterly level to the corresponding parameters on the annual level. I first simulate individual-level panel data on the quarterly level, using the parameters in Table 1 together with the same data structure as the original IABS-data. I then aggregate the worker-level data to the annual level, compute the covariance matrices and estimate the earnings processes. The estimates for the key parameters are listed in this table. The variance of transitory shocks is zero because aggregation averages over four random draws on the quarterly level. I thus do not show it in the table.

Appendix Table 4. Parameter estimates and robustness: high school dropout sample.

		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	0.027	0.134	0.041	0.039	0.068	0.111	0.134
		(0.007)	(0.011)	(0.003)	(0.013)	(0.009)	(0.019)	(0.01)
Slope heterogeneity	$\sigma^2_\beta * 10^3$	0.000	0.026	–	0.004	0.008	0.018	0.027
		–	(0.003)		(0.002)	(0.001)	(0.005)	(0.003)
Cov (intercept; slope)	$\sigma^2_{\alpha\beta} * 10$	−0.0004	−0.016	–	−0.003	−0.005	−0.011	−0.016
		(0.0001)	(0.002)		(0.0015)	(0.003)	(0.002)	(0.002)
Persistence of AR(1)	ρ	0.883	0.784	0.808	0.921	0.837	0.828	0.756
		(0.009)	(0.027)	(0.013)	(0.006)	(0.012)	(0.015)	(0.027)
AR(1) error structure								
Initial condition	$\sigma^2_{\xi 0}$	0.283	–	–	0.273	0.544	–	–
		(0.032)			(0.035)	(0.056)		
Intercept	γ_0	0.044	0.032	0.039	0.003	0.103	0.071	0.038
		(0.005)	(0.017)	(0.005)	(0.001)	(0.008)	(0.014)	(0.003)
experience	γ_1	−0.003	–	–	–	−0.007	−0.005	–
		$(4 * e(-4))$				(0.001)	(0.001)	
experience^2	γ_2	$7.78 * e(-5)$	–	–	–	$1.67 * e(-4)$	$1.43 * e(-4)$	–
		$(9.27 * e(-6))$				$(5.47 * e(-5))$	$(3.03 * e(-5))$	
experience^3	γ_3	$-8.65 * e(-7)$	–	–	–	$-1.93 * e(-6)$	$-1.64 * e(-6)$	–
		$(1.05 * e(-7))$				$(8.01 * e(-7))$	$(3.58 * e(-7))$	
experience^4	γ_4	$3.48 * e(-9)$	–	–	–	$8.22 * e(-9)$	$6.73 * e(-9)$	–
		$(4.29 * e(-10))$				$(3.86 * e(-9))$	$(1.5 * e(-9))$	

(*Continues*)

Appendix Table 4. *Continued.*

		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	$\delta_0 * 10$	0.003	–	–	0.000	0.002	0.001	0.000
		(0.001)			–	(0.002)	(0.002)	–
Variance of measurement error	ϕ_0	0.000	0.004	0.000	0.023	0.000	0.000	0.000
		–	(0.003)	(0.003)	(0.004)	(0.001)	–	–
Number of moments		**64,278**						
R^2		0.859	0.433	0.212	0.689	0.690	0.755	0.429
Wald test for slope heterogeneity (P-value)		0.659	0.000	–	0.002	0.000	0.000	0.000

Note: This table shows parameter estimates corresponding to those in Table 1 for the robustness sample, comprised of individuals without a formal educational degree. Results for the benchmark specification as described in equations (3.2) to (3.8) are shown in column 1. Estimated factor loadings, all of which are significant on the 1% level, are displayed in the second panel of Appendix Figure 5. 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.

APPENDIX TABLE 5. Results from Monte Carlo simulation.

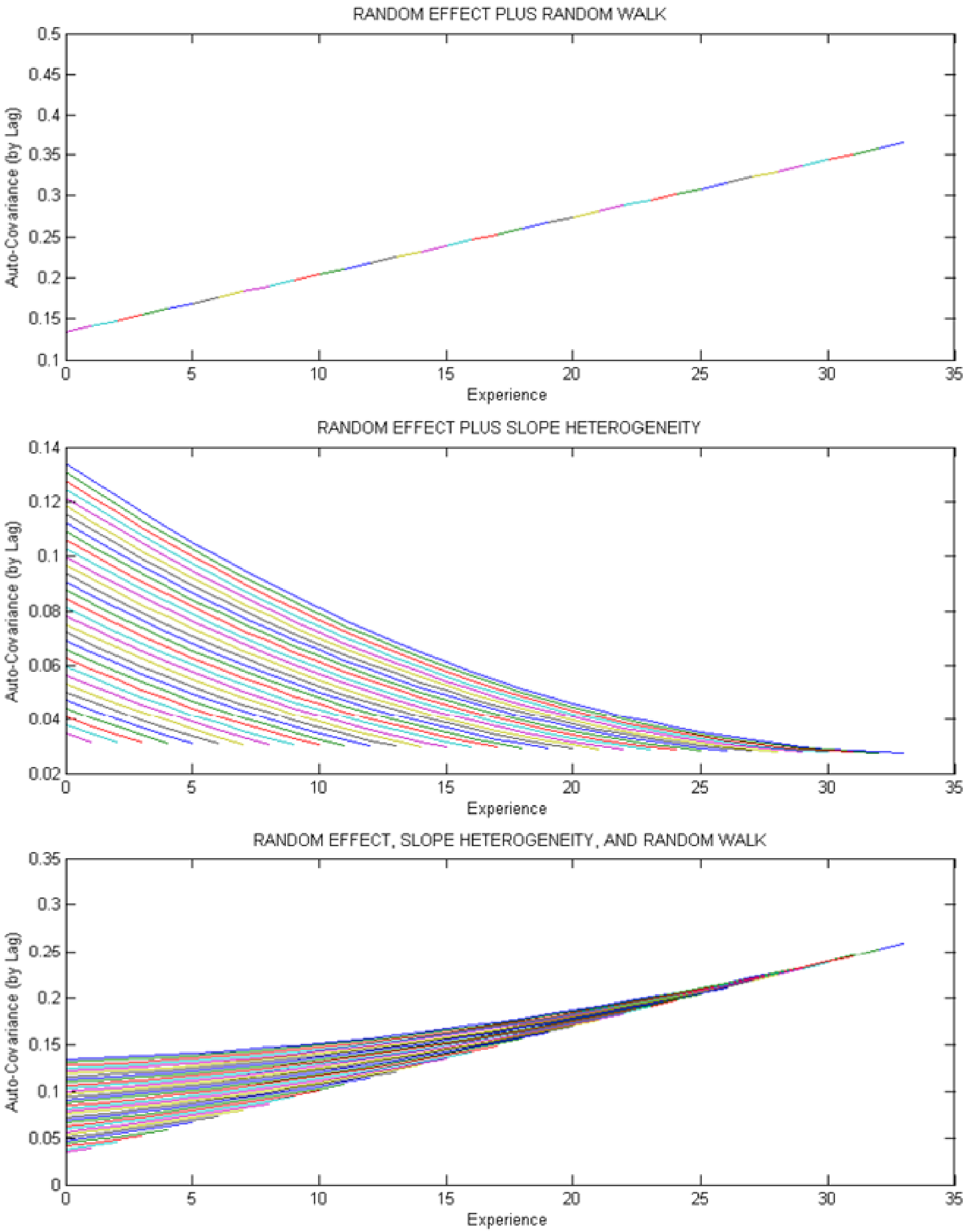
Panel A: Objects Held Constant Across Simulations

	Heterogeneity			AR(1) **Process**								
Parameter: description	Intercept	Slope	Covariance	Persistence	Initial Condition	4th-Order Polynomial in Experience					Unit roots	Measurement error
Parameter: notation	σ^2_α	σ^2_β	$\sigma_{\alpha\beta}$	ρ	$\sigma^2_{\xi_0}$	γ_0	γ_1	γ_2	γ_3	γ_4	δ_0	ϕ_0
True value	0.053	$5.00*e(-6)$	$-5.00*e(-4)$	0.757	0.136	0.026	−0.001	$2.98*e(-5)$	$-3.67*e(-7)$	$1.64*e(-9)$	0.001	0.003
Initial value in numerical optimization	0.03	0	0	0.95	0.05	0.05	0	0	0	0	0.001	0
Monte Carlo repetitions	**1000**											

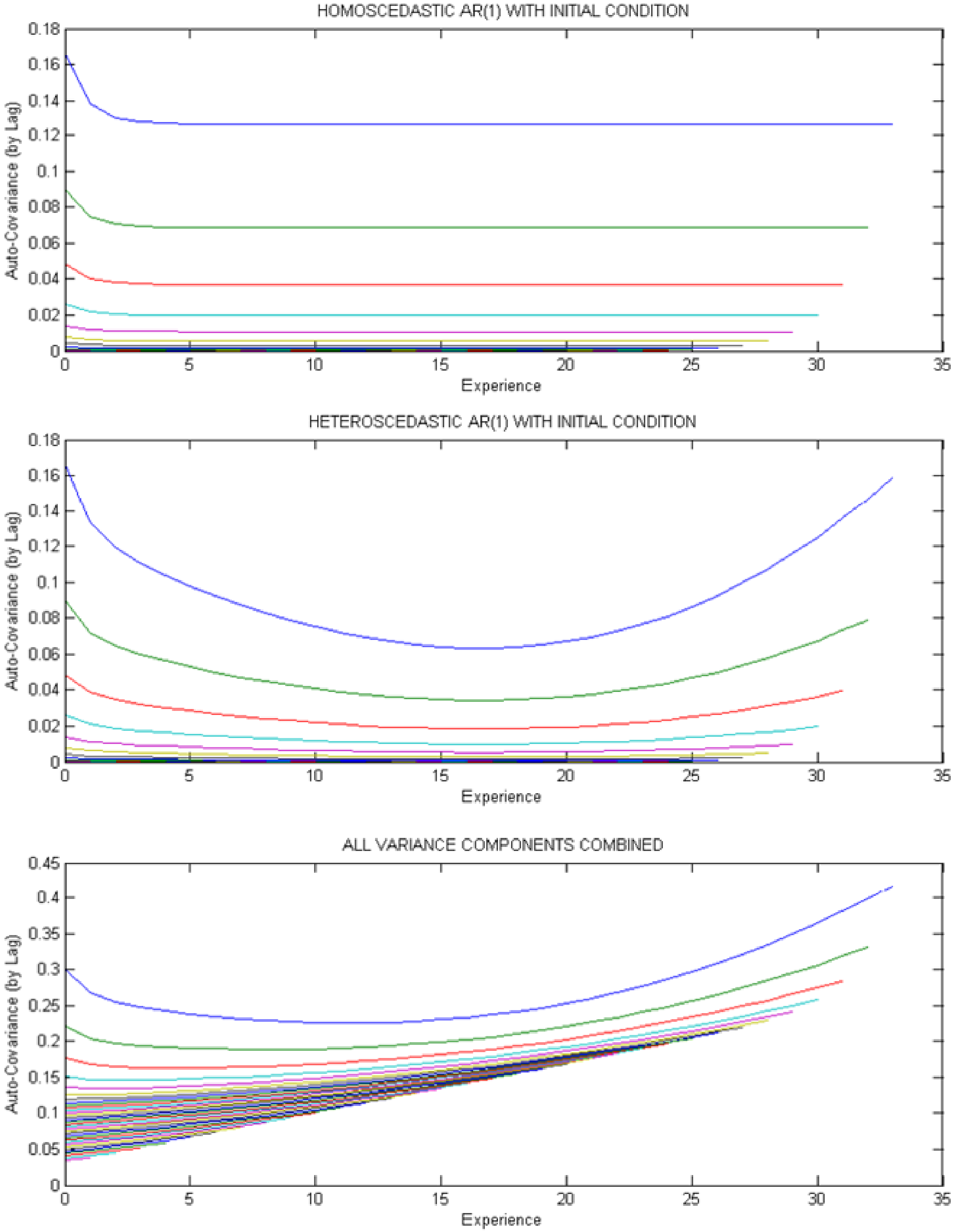
Panel B: Simulation Results

			Simulation 1: No restrictions on parameters			Simulation 2: $\sigma^2_\beta = \sigma_{\alpha\beta} = 0$			Simulation 3: $\sigma^2_{\xi_0} = 0$		
			(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
			Estimation: parameter restrictions			*Estimation: parameter restrictions*			*Estimation: parameter restrictions*		
			None	$\sigma^2_\beta = \sigma_{\alpha\beta} = 0$	$\sigma^2_{\xi_0} = 0$	None	$\sigma^2_\beta = \sigma_{\alpha\beta} = 0$	$\sigma^2_{\xi_0} = 0$	None	$\sigma^2_\beta = \sigma_{\alpha\beta} = 0$	$\sigma^2_{\xi_0} = 0$
Parameter estimates	σ^2_α	Bias	$-4.13*e(-5)$	−0.046	0.004	$-3.93*e(-5)$	$-3.09*e(-5)$	0.004	$-3.18*e(-5)$	−0.051	$-2.92*e(-5)$
		Std.	0.001	0.002	0.001	0.001	0.001	0.001	0.001	0.003	0.001
	σ^2_β	Bias	$2.10*e(-8)$	$-5.00*e(-6)$	$1.01*e(-6)$	$2.98*e(-8)$	0	$1.03*e(-6)$	$2.78*e(-8)$	$-5.00*e(-6)$	$2.84*e(-8)$
		Std.	$8.62*e(-7)$	0	$8.55*e(-7)$	$1.24*e(-6)$	0	$1.24*e(-6)$	$8.67*e(-7)$	0	$8.61*e(-7)$
	$\sigma_{\alpha\beta}$	Bias	$4.99*e(-7)$	$5.00*e(-4)$	$-4.54*e(-5)$	$3.88*e(-7)$	0	$-4.55*e(-5)$	$4.35*e(-7)$	$5.00*e(-4)$	$3.99*e(-7)$
		Std.	$1.83*e(-5)$	0	$1.7*e(-5)$	$2.01*e(-5)$	0	$1.86*e(-5)$	$1.73*e(-5)$	0	$1.69*e(-5)$
	ρ	Bias	0.001	0.21	−0.02	0.001	0.001	−0.021	0.001	0.22	0.001
		Std.	0.006	0.003	0.007	0.008	0.011	0.01	0.008	0.003	0.008
	$\sigma^2_{\xi_0}$	Bias	$-2.12*e(-4)$	−0.05	−0.136	$-4.26*e(-4)$	$-5.00*e(-4)$	−0.136	$7.95*e(-5)$	0.065	0
		Std.	0.002	0.001	0	0.003	0.003	0	0.002	0.003	0

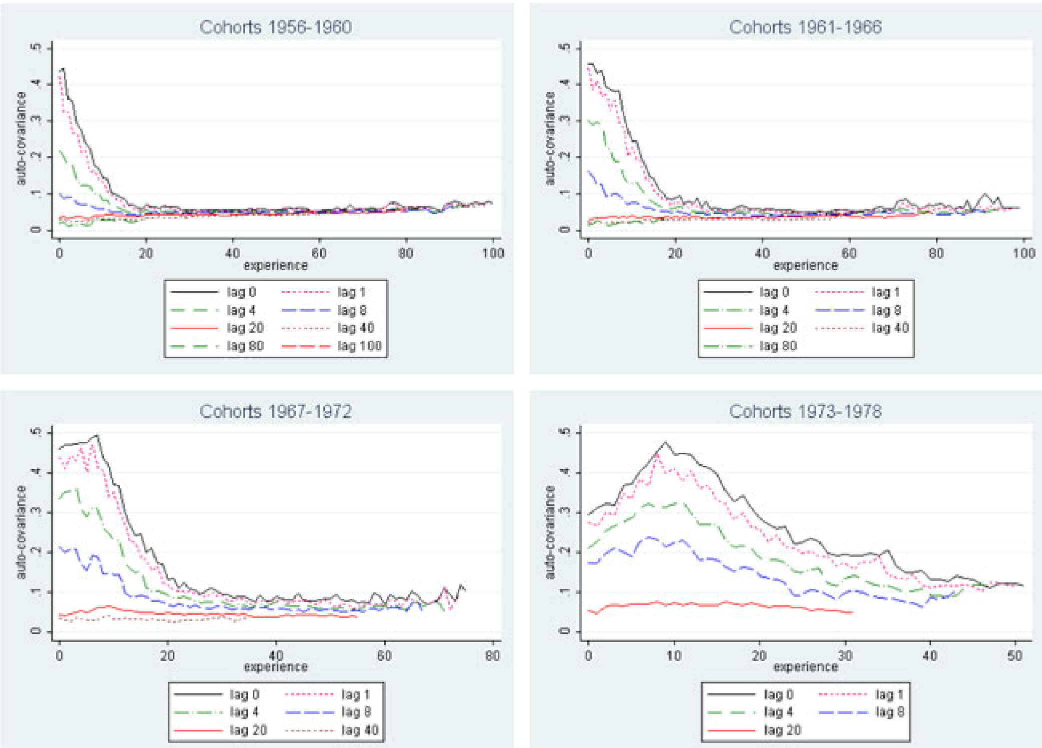
Note: This table shows results from Monte Carlo simulations of the benchmark model without time effects. Panel A lists the parameter values, initial values in the optimization routine used for estimation, and the number of Monte Carlo repetitions held constant across simulations, unless noted otherwise. Panel B shows results from 3 simulation exercises. In simulation 1, I do not impose any restrictions on the parameters. Simulation 2 sets the HIP component to zero, and simulation 3 eliminates the initial condition of the AR(1) process from the model. Each simulation proceeds as follows. I simulate 1000 data sets on the individual experience level, thereby replicating the data structure of the IABS. For each of these data, I compute the covariance structure and estimate 3 models on it. The first model, shown in columns (1), (4), and (7) imposes no parameter restrictions in the estimation. The second model, shown in columns (2), (5), and (8), does not allow for a HIP component, and the third model, shown in columns (3), (6), and (9), does not allow for an initial condition for the AR(1) process. Each column reports the bias and the standard deviation for the 1000 estimates of 5 key parameters.



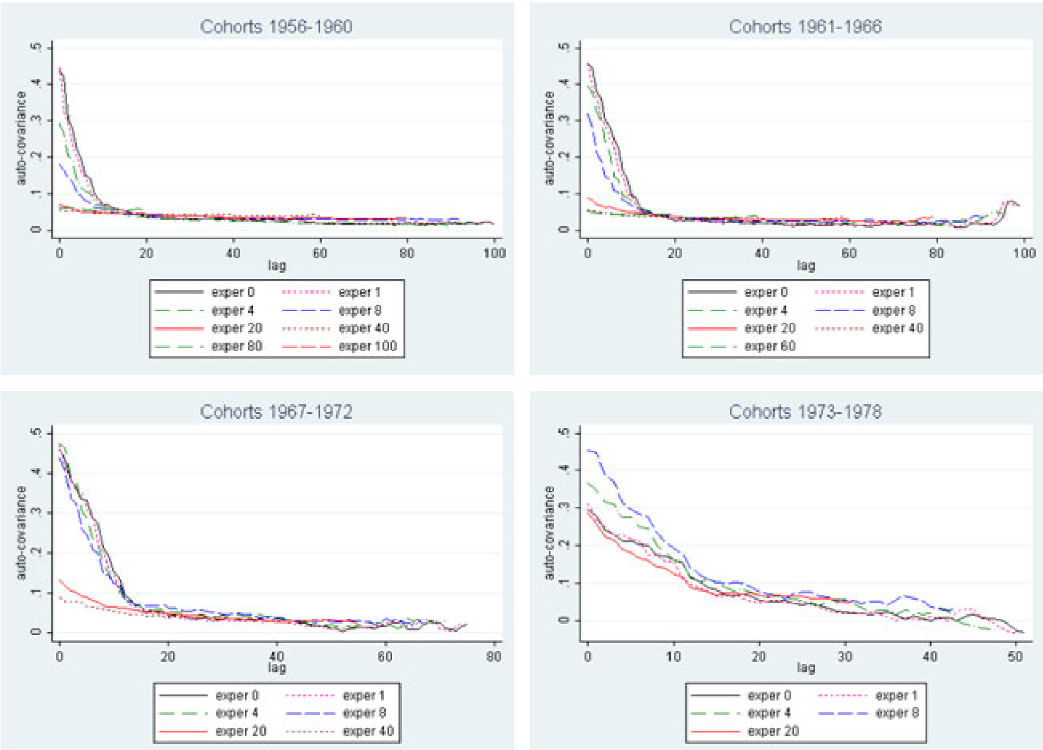
Appendix Figure 1. Variance components with Baker–Solon estimates, stationary part.



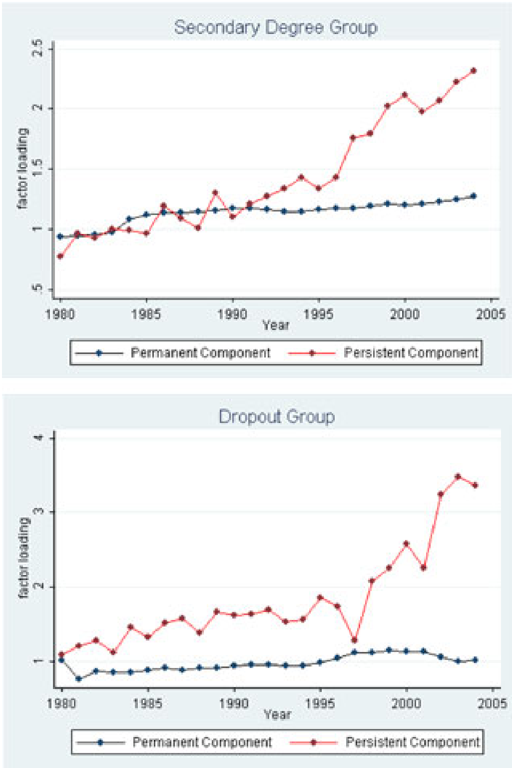
Appendix Figure 1. *Continued.*



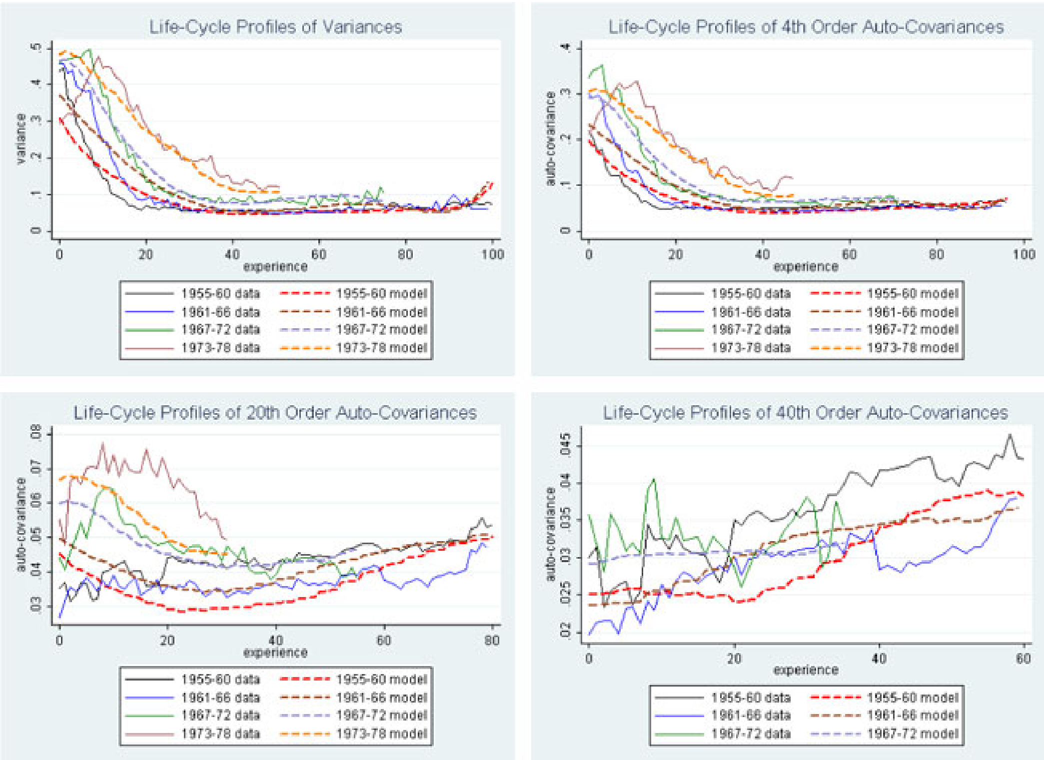
Appendix Figure 2. Life-cycle profiles of autocovariances at different lags, by cohorts. Sample: Dropout group.



Appendix Figure 3. Lag profiles of autocovariances for different experiencegroups, by cohorts. Sample: Dropout group.



Appendix Figure 4. Estimated factor loadings for the full model.



Appendix Figure 5. Fit of benchmark model: dropout group.