HIP, RIP and the Robustness of Empirical Earnings Processes *

Florian Hoffmann
Vancouver School of Economics
University of British Columbia

Version: August 2018

Abstract

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.

*This study uses the weakly anonymous IAB Employment Sample. Data access was provided via on-site use at the Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB) and remote data access. I thank Philip Oreopoulos, Shouyong Shi and Victor Aguirregabiria for their help. I also thank Michael Baker, Paul Beaudry, David Green, Gueorgui Kambourov, Thomas Lemieux, Jean-Marc Robin, Aloysius Siow, Kjetil Storesletten and seminar participants at the CEPA seminar at the University of Toronto, the UBC empirical micro lunch seminar and the 2012 conference on the use of administrative data at the IAB for helpful suggestions, and Benedikt Hartmann, Daniela Hochfellner, Peter Jacobebbinghaus and other staff members at the FDZ-IAB for their hospitality. Daniel Shack has provided excellent research assistance. The usual disclaimer applies.

1. Introduction

How much do returns to labor market experience differ between individuals? An answer to this question is not only important for evaluating models of human capital accumulation as an empirical tool for studying life-cycle labor market dynamics, but also for quantifying the importance of individual heterogeneity for earnings and consumption inequality. Indeed, heterogeneity in earnings growth rates, often referred to as profile- or slope heterogeneity, can generate sizeable and permanent increases of earnings inequality over the life-cycle.[1] It also has important qualitative and quantitative predictions for individual-level consumption behavior and thereby on the welfare effects of income inequality.[2] For these reasons, quantifying profile heterogeneity convincingly and transparently is of central interest to a wide range of economic research.[3]

The dominating methodological framework for estimating heterogeneous returns to experience is a Mincer (1974) earnings equation with random coefficients and a dynamic error structure, often referred to as a HIP-process.[4] While the modeling details differ substantially across studies, two broad identification results for this class of models are well established. First, parameter identification requires panel data; and second, under the assumption that the model is well-specified, profile heterogeneity and most other structural parameters can be point identified and estimated from the covariance structure of earnings. What remains unclear however is which particular features of covariance structures can and should be used for identification of slope heterogeneity and other model parameters and how sensitive parameter estimates are to model misspecification. It is thus popular to view the procedure of matching earnings processes to covariance structures as a "black box".[5] Whether existing estimation approaches deliver credible estimates of heterogeneity in the returns to experience is therefore unknown. Only recently has a small literature developed that attempts to resolve this issue. For example, Guvenen (2009) shows that profile heterogeneity generates a non-linear relationship between labor market experience and residual income inequality that fits empirical age-profiles of residual variances well. It has therefore become common to use this feature of the data as a calibration target in work that structurally estimates human capital models.[6]

In this paper I make two distinct contributions to the literature on quantifying profile heterogeneity and mod-

[1]See for example Haider (2001) and Haider and Solon (2006). I use "heterogeneous returns to labor market experience", "profile heterogeneity", "slope heterogeneity", and "heterogeneous growth rates" interchangeably.

[2]See for example Guvenen (2007), Primiceri and van Rens (2009) and Guvenen and Smith (2014). Summary papers of the heterogeneous-agents literature by Guvenen (2011) and of the consumption literature by Meghir and Pistaferri (2011) highlight the importance of earnings processes in structural modeling of life-cycle choices.

[3]Recent examples of quantitative life-cycle models with heterogeneous agents in which estimates from earnings processes are key inputs are Storesletten, Telmer and Yaron (2004a), Heathcote, J., K. Storesletten and G. Violante (2014) and Low, Meghir and Pistaferri (2010) for consumption, Abbott, Gallipoli, Meghir and Violante (2018) for education, Erosa, Kambourov and Fuster (2016) for labor supply, Huggett and Kaplan (2016) for human capital, and Farhi and Werning (2013) and Fukushima (2010) for public finance. Alvarez and Jermann (2000) and Krueger and Perri (2005) study the types of insurance mechanisms that are supported in decentralized markets depending on the persistence of exogenous shocks.

[4]"HIP" stands for "Heterogeneous Income Profiles". Models without random coefficients are called "RIP", which stands for "Restricted Income Profiles". These labels were introduced by Guvenen (2007, 2009).

[5]For example, a recent study of earnings dynamics in administrative data from the US by Guvenen, Karahan, Ozkan and Song (2015) criticizes the approach of matching covariance structures as "too opaque and a bit mysterious".

[6]Examples are Guvenen, Kuruscu and Ozkan (2014) and Huggett, Ventura and Yaron (2011).

eling earnings dynamics. First, for a large class of earnings processes I show that age profiles of variances and higher-order auto-covariances of residual earnings do not contain valid information for identifying heterogeneous returns to human capital accumulation. The reason is that *any* shape of variance profiles can be rationalized by age effects in the variances of transitory shocks. Similarly, a combination of age-heteroscedastic persistent and permanent shocks can match a wide range of age-profiles in higher-order auto-covariances. As a consequence, models that do not feature age-dependent heteroscedasticity tend to deliver biased estimates of slope heterogeneity. This is not merely a statistical issue since age effects in second moments are generated by various economic theories. Examples are search models, where reallocation of workers to better firm-matches via search generates a decline in residual variances over the life-cycle, or models featuring career progression, where promotions and demotions become more likely in the middle of a life-cycle. In fact, even the most basic human-capital accumulation process with less than full depreciation of the human capital stock can generate age-heteroscedasticity in residual earnings. Without theoretical guidance regarding the functional form of age effects, it is best to leave it unrestricted and to model it flexibly.

On the other hand, the shape of lag-profiles at high orders provides a credible source of identification for slope heterogeneity, even in the presence of age effects. It is here where HIP imposes strong and unique predictions on the covariance structure. Intuitively, slope heterogeneity does not generate significant earnings differences among inexperienced workers, but its effect becomes increasingly strong as individuals accumulate labor market experience. At the same time, earnings of the young already partially reflect differences in earnings growth and are thus predictive of earnings differences many years later. In combination, this implies that the HIP-component imposes strong and testable restrictions on high-order autocovariances. It is hard to think of any other mechanism that can generate this pattern, which renders the shape of lag-profiles at their tails a clean source of identifying variation for profile heterogeneity. Conversely, in the absence of rich dynamics in these tails it is unlikely that slope heterogeneity is important.

Remarkably, these results can be derived simply by checking a collinearity-condition. This remains true even in the most general specification considered in this paper, which introduces time effects in addition to age-effects. More specifically, since I consider earnings processes that can be estimated by matching covariance structures, identification can be explored by exploiting the well-established but rarely used equivalence of equally weighted minimum distance estimation – the dominating methodological approach in the literature – with the non-linear least squares estimator. Viewing estimation and parametric identification from this perspective has the advantage that one can rely on the well-understood econometric theory of parametric regression analysis. Concepts such as omitted variable bias and multicollinearity carry over directly, rendering identification transparent and intuitive. This goes a long way in opening the "black box" of estimating earnings processes.

As a second contribution I explore whether the omitted variable bias in estimates of profile heterogeneity from exluding age effects in innovation variances is likely to be important in practice. Since credible identification of slope heterogeneity needs to come from the shape of the right tail of lag profiles this requires data with long worker-level time series, large sample sizes and low attrition rates. Open-source panel data sets, such as the Panel Study of Income Dynamics (PSID) or the National Longitudinal Survey of Youth (NLSY), do not satisfy these criteria and are thus not well suited for estimating HIP earnings processes that flexibly control for age effects. Instead, I rely on a large administrative panel data set from German social security records. Individuals in these data are followed from time of labor market entry up until 27 years into their careers, and the spell-based recording enables me to generate samples on the quarterly rather than the annual frequency.[7] A further advantage of the data is that it provides information on educational attainment, in contrast to administrative data from North-America. Several central findings from my empirical analysis rely crucially on this information, for two major reasons. On the one hand, the largest education group in the German labor market displays an auto-covariance structure of labor market earnings that shares the main qualitative features with the North American counterpart. On the other hand, the covariance structures are quite different across education groups, which permits carrying out a thorough robustness exercise.

The central empirical result coming out of this exercise is that omitting age effects in innovation variances can lead to a severe upward bias in estimates of profile heterogeneity. Estimating the standard HIP-model without age effects as it is commonly specified in the literature on the sample of the largest education group delivers a variance of earnings growth rates that is economically and statistically highly significant. In sharp contrast, when I estimate my preferred specification, which features age effects and fits the data exceptionally well with relatively few parameters, the estimated slope heterogeneity decreases by a factor of thirteen and becomes statistically indistinguishable from zero.[8] Since one may be worried that my results are an artifact of the earnings structure in my sample, I repeat my analysis using data from an education group with a very different covariance structure. Again I find that a standard HIP specification yields significant slope heterogeneity while the inclusion of age effects drives estimates to zero.

I complement my empirical analysis with a Monte-Carlo analysis that replicates, in the simulated data, the cohort- and age-structure of the German data. The two central questions I address with this analysis are (i) whether a HIP-component, if present in the true data generating process, can be recovered precisely from a finitely sized sample if one models age effects in innovation variances flexibly and (ii) whether estimates of the HIP-component are systematically biased upward if such age effects are omitted. Both questions are answered assertively in the positive. Hence, samples of similar sizes like the IABS data are sufficient for implementation of the approach to identification of HIP suggested in this paper.

[7] Administrative data on earnings are increasingly used in economics. The high frequency at which earnings are recorded in the German data is one of its distinct features. Studies on worker mobility across firms, occupations and employment states have highlighted the problems associated with time aggregation for quite some time, and it has become standard in this literature to use data on the monthly frequency. In contrast, I am not aware of any study in the earnings literature that uses data on a higher than the annual frequency. This is an important omission because just as time aggregation will yield downward biased estimates of worker mobility rates, it will also miss possibly economically significant fluctuations in individual earnings.

[8] My benchmark specification also includes time effects. In total it has 62 parameters that are matched to cohort-specific covariance structures with over 56,000 elements. The model fits all of its features, such as the evolution of variances over the life-cycle and over time, almost perfectly.

2. Relation to Literature

The question of how dispersed individual earnings growth rates are has been explored at least since Mincer's (1958) work on human capital investments. Ben-Porath (1967) formulates more explicitly a model of human capital investments in which differences in the accumulation rate between individuals can generate heterogeneous slopes of experience-earnings profiles. Seminal studies by Lillard and Weiss (1979) and Hause (1980) were among the first to quantify this type of heterogeneity using panel data on labor income. The econometric models formulated in these studies have the interpretation of a Mincer (1974) earnings regression with random slopes and an added dynamic structure for the error term. MaCurdy (1982) carries out model specification tests by estimating various specifications for the dynamics of the error term. He concludes that slope heterogeneity is not an important component of life-cycle earnings dynamics. Subsequent papers in the literature, such as Abowd and Card (1989) and Meghir and Pistaferri (2004), adopt this view. However, Baker (1997) shows that MaCurdy's test for slope heterogeneity has low power in small samples and documents evidence for slope heterogeneity and modest persistence of shocks, a result that has been corroborated by Haider (2001) and Guvenen (2009).

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. Some progress towards understanding the sources of identifying variation for the main parameters of interests in earnings processes has been made recently, however. Guvenen (2009) shows that slope heterogeneity in a standard HIP-model without age effects is identified from both the convexity of age profiles and the behavior of lag-profiles of earnings covariances. In line with his identification result, he also establishes that HIP models can replicate the age-profile of residual earnings inequality, which he documents to be convex, better than RIP models. However, as Hryshko (2012) shows, these results are not robust to a slight modification in the error process, hinting at the sensitivity of key parameters to model misspecification.[9] Compared to these works 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 non-linear 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, i.e. 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 non-parametric

[9] This is directly implied by my identification results since a unit roots process generates linear age effects in age-profiles of covariances.

time-effects.

There is a large literature that emphasizes the importance of controlling flexibly for age and time-effects when studying the sources of individual earnings variation over time and the life-cycle, such as Storesletten, Telmer and Yaron (2004b), Heathcote, Storesletten and Violante (2005) and Lemieux (2006). Recent analyses of the covariance structure of earnings have also incorporated age effects in various components of the earnings process and find that they are important. Examples are Karahan and Ozkan (2013) and Lopez-Daneri (2016), who establish the presence of age effects in a RIP-process estimated from the PSID, and Sanchez and Wellschmied (2017), who find substantial asymmetries and age effects in the distribution of earnings shocks in German administrative data. Baker and Solon (2003) and Blundell, Graeber and Mogstad (2015) estimate, using administrative data from Canada and Norway respectively, rich HIP-processes that incorporate flexible specifications of age- and time effects.[10] The empirical model I estimate on the German administrative data is similar to theirs. The prime difference between these studies and mine is the focus. While they use 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.[11] The first exploits the joint dynamics of consumption and earnings for parameter identification. In most applications, such as in Hall and Mishkin (1982), Blundell, Pistaferri and Preston (2009) or Heathcote, Storesletten and Violante (2012), these dynamics merely provide overidentifying restrictions on the parameters of the earnings process. Indeed, Arellano, Blundell and Bonhomme (2017) consider a model of consumption in which the transition density of the earnings process is non-parametrically identified from earnings data alone. They also propose an attractive two-step estimator in which this density is recovered in a first step. In some cases however, consumption data may be necessary to achieve identification. For example, Guvenen (2007) and Guvenen and Smith (2014) consider models in which individuals learn about their abilities so that individual ability differences are not fully reflected in earnings, at least at an early stage of the life-cycle. Browning and Ejrnaes (2014) and Hryshko (2014) specify earnings processes in which different variance components are correlated. Alan, Browning and Ejrnaes (2018) estimate a model in which heterogeneity in preference parameters and the parameters describing the earnings process are co-dependent. In these cases, identification relies explicitly on the comovement between consumption and earnings. At the same time, there are significant drawbacks from relying on consumption data. Examples are the lack of high-quality administrative panel data that simultaneously record earnings and

[10]These studies build on an earlier literature that use earnings processes to quantify the sources of trends in earnings dynamics, innovated by Gottschalk and Moffitt (1994). Similar exercises, though with more restrictive models and less rich data, have been carried out for various countries, e.g. by Moffitt and Gottschalk (2002) for the US, Biewen (2005) for Germany, and Dickens (2002) for the UK.

[11]A recent study by Daly, Hryshko and Manovskii (2017) that does not fall within these two areas finds that estimates of earnings processes are substantially affected by non-random missings in both, public-use panel data and high-quality administrative data. Non-random missing data can explain why estimates from earnings measured either in levels or in growth often differ substantially.

consumption dynamics, the potential computational burden of estimating structural decision-theoretic models of consumption in the presence of an earnings process with many state variables, and the need to rely on strong parameteric assumptions. As highlighted by Meghir and Pistaferri (2011), relying on large administrative data sets to estimate flexible earnings processes, the approach followed in this paper, should be seen as complementary. Furthermore, 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. For example, it is possible that the estimate of profile heterogeneity that best fits observed consumption behavior generates earnings dynamics that are at odds with the tail behavior of lag-profiles of earnings covariances. This provides a powerful overidentifying restriction.

The second area of active research departs from the conventional approach to earnings dynamics by going beyond autocovariance structures for estimation. For example, Meghir and Pistaferri (2004) allow for ARCH-effects in the transitory and permanent innovations, and Browning, Ejrnaes and Alvarez (2010) extend this framework to a processes in which the majority of parameters are random variables. In both cases, identification needs to rely on more information than second moments. There has also been some progress on non-parametric identification of earnings transition densities, such as Horowitz and Markatou (1996), Hirano (2002), Bonhomme and Robin (2009), Lochner and Shin (2014), Arellano, Blundell and Bonhomme (2017), and Hu, Moffitt and Sasaki (2018). De Nardi, Fella and Pardo (2018) explore the implications of modelling earnings dynamics non-parametrically for consumption behavior. Guvenen, Karahan, Ozkan and Song (2015) highlight the need to move beyond second moments in a detailed study of earnings growth in US administrative data. 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 in which 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.[12]

[12]It is important to note that the model formulated below does not impose any restrictions on higher-order moments. The theoretical and empirical results established in my study are thus independent from the behaviour of higher-order moments, such as skewness or excess kurtosis. If one is willing to impose additional distributional assumptions, the approach followed here could thus be combined with a second stage that identifies parameters governing higher order moments. Suppose for example that one assumes that excess kurtosis in the distribution of earnings changes, as documented in Guvenen et al (2015), is driven by the distribution of the returns to experience, β_i. If one postulates that β_i has a population distribution with, say, three parameters, written $f_\beta\left(\phi_1, \phi_2, \phi_3\right)$, then one requires at least three moment restrictions. Let $E_\beta^k\left(\phi_1, \phi_2, \phi_3\right)$ be the $k-th$ central moment of f_β. One restriction is the normalization, also used in my estimation, that $E_\beta^1\left(\phi_1, \phi_2, \phi_3\right)=0$. Since the framework proposed below recovers the variance of β, written σ_β^2, without explicit distributional assumptions, it can be used as a first-stage estimator for constructing a second moment condition $E_\beta^2\left(\phi_1, \phi_2, \phi_3\right)=\sigma_\beta^2$. A third restriction must come from higher-order moments. My study explores identification of σ_β^2 without any explicit distributional assumptions about $f_\beta\left(\phi_1, \phi_2, \phi_3\right)$. It is in this sense that the procedure of matching covariance structures is a semi-parametric estimator.

3. Econometric Framework, Estimation and Identification

3.1. The Econometric Model

Let y^e_{ibt} 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

$$y^e_{ibt} = \mu^e_{bt} + \widehat{y}^e_{ibt}, \tag{3.1}$$

where μ^e_{bt} represents a set of education specific cohort-time fixed effects and $\widehat{y}^e_{ibt}$ is the error term. The focus of this study will be on the life-cycle dynamics of $\widehat{y}^e_{ibt}$. 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. For example, if age effects in conditional first moments of log-earnings are highly non-linear, then improperly controlling for them will spuriously generate age effects in conditional second moments of the residual.[13] The required flexibility is achieved by using the non-parametric 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, $\widehat{y}^e_{ibt}$.

To describe the dynamics of $\widehat{y}^e_{ibt}$ 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 $\widehat{y}_{ibt}$ 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 $(p_t, \lambda_t, \varphi_t)_{t \geq \underline{t_0}}$ are factor loadings, one for each component. They allow the process of $\widehat{y}_{ibt}$ to change over time, so that different cohorts are subject to different life-cycle earnings dynamics.

[13] From now on I will use "age" and "potential labor market experience" interchangeably since they correlate perfectly once one imposes a normalization on the age of labor market entry, as I do in the estimation.

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:

$$\begin{aligned} var\left(\alpha_i\right) &= \widetilde{\sigma}_{\alpha}^2; \;\; var\left(\beta_i\right) = \sigma_{\beta}^2; \;\; cov(\alpha_i, \beta_i) = \sigma_{\alpha\beta} && (3.5) \\ var(\nu_{ibt}) &= \sum\nolimits_{j=0}^{J_\nu} \left(h_{bt}\right)^j * \delta_j; \;\; var(u_{it_0(b)}) = \widetilde{\sigma}_{u_0}^2 && (3.6) \\ var\left(\xi_{ibt}\right) &= \sum\nolimits_{j=0}^{J_\xi} \left(h_{bt}\right)^j * \gamma_j; \;\; var(z_{it_0(b)}) = \left(\lambda_{t_0(b)}\right)^2 * \sigma_{\xi_0}^2 && (3.7) \\ var\left(\varepsilon_{ibt}\right) &= \sum\nolimits_{j=0}^{J_\varepsilon} \left(h_{bt}\right)^j * \phi_j. && (3.8) \end{aligned}$$

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 $(p_t, \lambda_t, \varphi_t)_{t \geq \underline{t_0}}$ need to be normalized for some periods. The following restrictions are sufficient for identification:[14]

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

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. All three components present labor market risks with different degrees of insurability and play a prominent role in structural models of consumption- and savings decisions and in heterogeneous agents models. A precise interpretation of these shocks is difficult because they are modeled as unobserved components and because there is limited evidence on how they map into measureable characteristics and events. A number of recent studies have made significant progress on this issue, however. For example, Polachek, Das and Thamma-Apiroam (2015) show that permanent individual differences in earnings growth relate to differences in cognitive ability, personality traits, and family background. Various structural studies of life-cycle earnings- and mobility dynamics, such as Low, Meghir and Pistaferri (2010), Hoffmann (2010) and Pavan (2011), show that transitions across firms and between occupations generate substantial and persistent changes in residual earnings. Postel-Vinay and Turon (2010) document that a canonical job search model with job-to-job transitions can produce an earnings process with a persistence that is consistent with the data. Altonji, Smith and Vidangos (2013) establish a similar result and propose health shocks

[14]See section 3.4 and appendix C.

as another source of persistent earnings changes. Guiso, Pistaferri and Schivardi (2005) and Lamadon (2016) find in matched employer-employee data that a sizeable part of persistent or permanent firm-level productivity shocks are passed on to workers, while transitory shocks are not. There is also growing evidence, summarized in a recent paper by Davis and von Wachter (2012), that job displacement is a source of highly persistent earnings loss. On the other hand, one-time bonus payments and a temporary absence from work are often cited as an example of transitory shocks, though there is less tangible evidence on this hypothesis. Another interpretation of purely transitory earnings variation is measurement error, consistent with the small estimates of its variance found in administrative data, as in Baker and Solon (2003) and as reported for the German data below. One may therefore conjecture that any economically meaningful shocks have at least some persistence.

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).[15] What is perhaps surprising at first is that point identification, holding fixed the degree of the polynomial, can be achieved even though multiple error components are allowed to be heteroscedastic in age. This result relies crucially on exploiting information in the entire covariance structure, not only its age-variance profiles.

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.[16] The specification chosen here allows for maximum flexibility. Each of the three variance components have their own factor loading. Since there are no distributional- or functional form restrictions on these loadings, the specification of time-effects is essentially non-parametric. Also notice that the factor loading λ_t enters the persistent component indirectly through its multiplication with the shock ξ_{ibt}, so that its impact on earnings dynamics fades gradually over time. A pattern of this form can be expected from the effects of business-cycle shocks or firm closures on earnings.[17] Furthermore, initial conditions in the persistent component $\left(\lambda_{t_0(b)} * \sigma^2_{\xi_0}\right)$ vary across cohorts indexed by b because different cohorts enter the labor market in different years.

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 an

[15]This may be viewed as a non-parametric series method to approximating the age-structure of autocovariances. Notice however that it holds the degree of the polynomials fixed.

[16]Alternatively, one can disaggregate the data to the age-time level, as in Abowd and Card (1989) and Blundell, Graeber and Mogstad (2015).

[17]This specification is adopted from Baker and Solon (2003).

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.[18] 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

$$\begin{aligned} cov(\widehat{y}_{ibt}, \widehat{y}_{ib,t+k}) &= p_t * p_{t+k} * \left\{ \begin{array}{c} \left[\widetilde{\sigma}_{\alpha}^2 + (2h_{bt}+k) * \sigma_{\alpha\beta} + h_{bt} * (h_{bt}+k) * \sigma_{\beta}^2\right] \\ + \left[\widetilde{\sigma}_{u_0}^2 + f^u(h_{bt}, \delta_0, ..., \delta_{J_\nu})\right] \end{array} \right\} \\ & \quad + \rho^k * Var\left(z_{ibt}\right) + 1(k=0) * \varphi_t^2 * \left(\sum\nolimits_{j=0}^{J_\varepsilon} h_{bt}^j * \phi_j\right). \end{aligned} \tag{3.10}$$

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

$$\begin{aligned} Var(z_{it_0(b)}) &= \left(\lambda_{t_0(b)}\right)^2 * \sigma_{\xi_0}^2 \\ Var\left(z_{ibt}\right) &= \rho^2 * Var\left(z_{ibt-1}\right) + \lambda_t^2 * \left(\sum\nolimits_{j=0}^{J_\xi} h_{bt}^j * \gamma_j\right) \quad \text{for all } t > t_0(b). \end{aligned} \tag{3.11}$$

In stationary models, equation (3.11) can be shown to have a closed form solution that is highly non-linear in model parameters. With factor loadings on the persistent shocks, the resulting process is non-stationary 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 $\widehat{C}_b$ be the estimated covariance matrix for a cohort born in year b. A typical element $\widehat{c}_{btk}$ is the cohort-specific covariance between residual earnings in period t with residual earnings k periods apart. Collecting non-redundant elements of $\widehat{C}_b$ in a vector $\widehat{C}_b^{vec}$ and stacking them yields the vector of empirical moments to be matched, denoted $\widehat{C}^{vec}$. Each element $\widehat{c}_{btk}$ in $\widehat{C}^{vec}$ 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 $G\left(\theta, Z\right)$. To be clear, Z is composed of observable objects

[18] The variance of transitory shocks is not point identified in the presence of measurement error. This result generalizes to MA(q)-processes for arbitrary q (Meghir and Pistaferri, 2004). However, in administrative social security data it is plausible to assume that measurement error is sufficiently small to equate transitory movements in earnings with true worker-level fluctuations in productivity. It is for this reason that the lack of evidence for purely transitory movement in earnings in both the IAB-data and the Baker-Solon data, briefly mentioned above, can be interpreted as evidence that any earnings shocks have some persistence.

entering equation (3.10), such as age, birth year, time, the lag, and various non-linear functions of these variables. The (MD)-Estimator for θ solves

$$\widehat{\theta} = \min_{\widetilde{\theta}} \left[\widehat{C}^{vec} - G\left(\widetilde{\theta}, Z\right)\right]' W \left[\widehat{C}^{vec} - G\left(\widetilde{\theta}, Z\right)\right] \tag{3.12}$$

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, $\widehat{\theta}$ in (3.12) becomes the Equally Weighted Minimum Distance Estimator (EWMD).

A seldomly used, though very useful result, is the equivalence between EWMD-estimation and non-linear 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 $\widehat{\chi}_{btk}\left(\widetilde{\theta}, Z_{btk}, \widehat{c}_{btk}\right) = \widehat{c}_{btk} - G\left(\widetilde{\theta}, Z_{btk}\right)$. Here, $\widehat{c}_{btk}$ is an element in $\widehat{C}^{vec}$ uniquely determined by cohort, year, and lag. Similarly, $G\left(\widetilde{\theta}, Z_{btk}\right)$ is the theoretical counterpart, the non-linear function of parameters and observables given by equation (3.10). The level of observation is $cohort - year - lag$. By definition, $\widehat{\theta}$ solves

$$\widehat{\theta} = \min_{\widetilde{\theta}} \sum_{btk} \widehat{\chi}^2_{btk}\left(\widetilde{\theta}, Z_{btk}, \widehat{c}_{btk}\right) \tag{3.13}$$

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

A consistent estimator of $\sqrt{var\left(\widehat{\theta}\right)}$, the standard error of the $\widehat{\theta}$, is readily available, but depends on the matrix of fourth-order moments of residual earnings.[19] This matrix has size $\left[\dim(\widehat{C}^{vec})\right]^2$. 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 (3.13), 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 $\sqrt{var\left(\widehat{\theta}\right)}$ will require additional assumptions. In the appendix I describe under which assumptions this approach delivers an asymptotically valid estimator of $var\left(\widehat{\theta}\right)$.

3.4. Identification

Viewing estimation of earnings processes via matching covariance structures through the lens of non-linear least squares has one central advantage: Identification can be discussed in terms of concepts that are familiar from parametric regression models. Concepts such as omitted variable bias, control variables, or multicollinearity can be applied directly, and sufficient conditions for local point identification are readily available. The question of how to credibly identify slope heterogeneity can be answered by exploring if it produces any *unique* prediction on the data, i.e. a prediction that is hard to generate by any other plausible mechanism. This is the central point I address in this section.

[19] For an in-depth treatment of MD-estimation, see chapter 6.7 in Cameron and Trivedi (2005).

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

$$J_{\widehat{\theta}}(Z)' * \left[\widehat{C}^{vec} - G\left(\widehat{\theta}, Z\right)\right] = 0, \tag{3.14}$$

where $J_{\theta}(Z) = \frac{\partial G(\theta,Z)}{\partial \theta'}$ is the Jacobian of $G\left(\widetilde{\theta}, Z\right)$ at $\widetilde{\theta} = \theta$, a matrix of size $\dim(Z) \times \dim(\theta)$. If the model structure is linear in parameters, i.e. $G(\theta, Z) = Z' * \theta$, then the (NLS)-estimator is equivalent to OLS: $\widehat{\theta} = (Z' * Z)^{-1} * Z' * \widehat{C}^{vec}$. Notice that the level of observation is an element in the covariance structure, *not* individual earnings.

For general non-linear models there is no closed-form solution, but sufficient conditions for local point identification and consistency of the NLS-estimator $\widehat{\theta}$ have been established and are as follows:[20]

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

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

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

Assumption (i) requires consistent estimation of the autocovariance structure, while assumption (ii) postulates that the model $G(\theta, Z)$ is correctly specified. The last assumption requires the Jacobian to have full rank at θ. Given that a consistent non-parametric estimator for the covariance structure is readily available, the first assumption is satisfied. One should thus view the second assumption as critical. For if the model $G(\theta, Z)$ is ill specified, the estimator $\widehat{\theta}$ is inconsistent even if the rank condition (iii) is satisfied. Since the explanatory variables entering $G(\theta, Z)$ are usually limited to age, time, the order of the lag, and possibly education, model specification manifests itself in functional form restrictions on how these observables enter the model prediction. For example, a standard HIP model satisfies the rank condition, but — as shown below — it is ill-specified because it inherently confounds estimates of slope heterogeneity with age effects in innovation variances. As a consequence, if one does not introduce age effects, condition (ii) is violated. The model (3.2) to (3.8) is particularly attractive from this point of view because it does not impose any arbitrary functional form restriction on the relationship between age or time and autocovariances.

To gain some intuition for the identification assumptions it is helpful to notice that they have direct analogues in the theory of linear regression. Specifically, assumption (i) corresponds to a random-sampling assumption since this guarantees that the covariance structure C can be estimated consistently. Assumption (ii) corresponds to the linearity-in-parameters assumption combined with conditional independence of the error term. Indeed, as argued above, the EWMD-estimator is the NLS estimator of the model

$$C^{vec} = G(\theta, Z) + \chi \tag{3.15}$$

[20] See for example Cameron and Trivedi (2005).

where χ is an iid error term. Finally, assumption (iii) implies that the explanatory variables cannot be perfectly collinear.

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.[21] Then the conditions for parametric identification have the following key implications:

(Implication 1) The parameters $\widetilde{\sigma}^2_\alpha$ and $\widetilde{\sigma}^2_{u_0}$ cannot be separately identified. Inspection of equation (3.10) shows that $\widetilde{\sigma}^2_\alpha$ and $\widetilde{\sigma}^2_{u_0}$ enter the model additively, thus violating assumption (iii). This is a problem of multicollinearity. Intuitively, a random walk process changes individuals' intercepts permanently. If such a shock occurs immediately before labor market entry it cannot be distinguished from pre-labor market skills that are captured by α_i. In the following I estimate a "combined initial condition" for the permanent component $\sigma^2_\alpha = \widetilde{\sigma}^2_\alpha + \widetilde{\sigma}^2_{u_0}$.

(Implication 2) If $\rho < 1$ all other model parameters are locally point identified. With a consistent estimator of C readily available, and under the assumption that the model is well specified, establishing identification requires checking the rank condition (iii). For general non-linear models this is difficult, especially if numerical computation or even simulation of $G(\theta, Z)$ is involved. However, for the class of earnings processes considered here it turns out to be quite straightfoward because the model is "close to" linear in parameters. More specifically, apart from the AR(1)-term, the theoretical covariance structure involves time fixed effects, polynomials in h_{bt} and k, and their interactions. The parameters of the model are coefficients on these terms. If $\rho < 1$ the AR(1)-part of the model introduces some non-linearity, which turns out to be crucial for identification as it guarantees that there is no perfect collinearity with the other terms. If $\rho = 1$ the AR(1)-process generates a collinearity problem, and identification fails unless one imposes additional restrictions on the factor loadings. Details are given in appendix C.

Three points regarding this identification result are worth highlighting. First, it is common to prove identification of earnings processes by deriving a set of $\dim(\theta)$ equations involving the parameters and population moments that solve uniquely for θ. For the rich specification considered here, this is tedious. In contrast to this approach, I rely directly on the rank-condition for the EWMD-problem, where all, rather than $\dim(\theta)$, restrictions imposed by the model on the covariance structure are exploited. This does not only deliver tractability but also yields new insights as it clarifies which sources of variation identify which parameters. Second, it seems surprising that with just a few normalizations on factor loadings one can identify three sets of factor loadings and three sets of polynomials in age, something that is impossible

[21] The standard order condition for NLS estimation is satisfied as long as $\max\{J_\varepsilon, J_\nu, J_\xi\} < \max\{h_{bt}\}$. Recorded life-cycles must be longer than the longest polynomial in age.

when estimating models of conditional first moments of earnings. The fundamental reason is that a life-cycle of T periods provides only T *first* moments, but $\frac{1}{2} * T * (T+1)$ autocovariance elements. As shown above and in appendix C the behavior of off-diagonal elements as a function of the lag is a crucial source of identification. Third, inter-cohort variation, that is, variation in autocovariances conditional on calendar time and lag, is fundamental for establishing identification of both time- and age-effects at that level of generality. In practice, this requires covariance structures that are disaggregated to the cohort-level, thereby imposing large demands on the data.

(Implication 3) Age profiles of variances are uninformative about HIP. Variances correspond to elements with $k = 0$. Equation (3.10) implies that their age-profiles can be matched perfectly by allowing the polynomial $\sum_{j=0}^{J_\varepsilon} h_{ibt}^j * \phi_j$, which is the transitory component, to have sufficiently high order. This means that any type of non-linearity in age-profiles of variances can be explained by age-effects in the variance of transitory shocks. Since there is no theory stating that dispersion of transitory shocks is constant over the life-cycle, any a-priori functional form restrictions on this component are arbitrary.

(Implication 4) Age profiles of high-order autocovariances are also uninformative about HIP. To see this implication it is convenient to assume that $p_t = \lambda_t = 1$ for all t.[22] If $\rho < 1$ one can use the fact that $Var\left(z_{ibt}\right)$ is bounded above by $\max_{b,t}\left\{var(\widehat{y}_{ibt})\right\}$ to derive the following approximation for large k,[23]

$$cov(\widehat{y}_{ibt}, \widehat{y}_{ib,t+k}) \approx \sigma_\alpha^2 + (2h_{bt} + k)\,\sigma_{\alpha\beta} + h_{bt} * (h_{bt} + k)\,\sigma_\beta^2 + f^u(h_{bt}, \delta_0, ..., \delta_{K_\nu}). \tag{3.16}$$

Since $f^u(h_{bt}, \delta_0, ..., \delta_{K_\nu})$ is a polynomial of degree $(K_\nu + 1)$ in h_{bt} that is linear in the $\delta_j' s$, this expression is linear in parameters and can be estimated by OLS. None of the observables are multicollinear if h_{bt} and k vary across observations. All parameters entering this linear regression equation are therefore *globally* point-identified. However, holding k constant — corresponding to age-profiles of high-order autocovariances — does generate multicollinearity between the variables multiplying the parameters of the HIP-component and the parameters of the heteroscedastic unit-roots process. This clarifies the special role of the lag profiles for identification.

(Implication 5) A credible source of identification of HIP are the tails of lag-profiles. Slope heterogeneity imposes strong restrictions on the slope of lag-profiles at large k, which can be used to develop an "eye-ball" test for its relevance. Fixing h_{bt} at some arbitrary value and k at a large value, the difference of autocovariances between two lag values k and $k+n$ is given by

$$cov(\widehat{y}_{ibt}, \widehat{y}_{ib,t+k}) - cov(\widehat{y}_{ibt}, \widehat{y}_{ib,t+[k+n]}) \approx \left(n * \sigma_{\alpha\beta} + h_{bt} * n * \sigma_\beta^2\right). \tag{3.17}$$

[22] Relaxing this assumption does not have an effect on this implication.

[23] Since $\widehat{y}_{ibt}^e = y_{ibt}^e - \mu_{bt}^e$, where μ_{bt}^e is the average log-wage of cohort b with education e in period t, and since y_{ibt}^e is in logs, $\left|\widehat{y}_{ibt}^e\right|$ is rarely observed to be above 1 in *any* data set that is commonly used for the estimation of earnings processes. Hence, it is reasonable to assume that $\max_{b,t}\left\{var(\widehat{y}_{ibt})\right\} < 1$. In the sample used for empirical implementation of the model I find that $\max_{b,t}\left\{var(\widehat{y}_{ibt})\right\} < 0.5$. Thus, $\rho^k * Var\left(z_{ibt}\right)$ will vanish quickly as k increases.

The only parameters entering this expression are those for the HIP-component, and they multiply variables that are not perfectly collinear. It is in this sense that HIP generates unique predictions on the tails of lag-profiles. The restriction on the tails is important because the persistent AR(1) component can explain negatively sloped lag profiles at low orders. In contrast, if lag-profiles at large k are downward sloping it must be the case that $\sigma_{\alpha\beta} < 0$. Furthermore, convexity can only be explained by $\sigma_{\beta}^2 > 0$. Conversely, if lag-profiles converge to a constant, then $\sigma_{\beta}^2 \leq \frac{|\sigma_{\alpha\beta}|}{\max\{h_{bt}\}}$, which is likely to be very small. Combined, these results suggest that as long as empirical *lag-profiles* do not display noticeable and robust convexities, slope heterogeneity is unlikely to be important even if *experience-profiles* are convex.■

In combination these results show that the only credible source of identification of slope-heterogeneity is the behavior of lag profiles at their tails. This is discouraging for two reasons. First, slope heterogeneity has the unique prediction that these tails are convex, which is a second-order feature of the empirical moments. Using consumption data in addition to earnings data will not overcome this issue because this prediction is generated by any heterogeneous agents model with a HIP earnings process. Second, the tails of lag-profiles are constructed from earnings data for the same individual at two different points in time that are far apart. They are thus most likely affected by endogenous attrition. It is for this reason that I rely on the administrative IABS data in the empirical implementation since they have large sample size, partially addressing the first issue, and since they follow individuals for long periods of time because of administrative reasons, partially addressing the second issue.

The remainder of this section discusses some further issues via two examples:

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

$$\begin{aligned} \widehat{y}_{ibt} &= \alpha_i + \beta_i * h_{bt} + u_{ibt} \\ u_{ibt} &= u_{ib,t-1} + \nu_{ibt} \\ var\left(\alpha_i\right) &= \widetilde{\sigma}_{\alpha}^2; \quad var\left(\beta_i\right) = \sigma_{\beta}^2; \quad cov(\alpha_i, \beta_i) = 0 \\ var(\nu_{ibt}) &= h_{bt} * \delta_1; \quad var(u_{it_0(b)}) = 0. \end{aligned} \tag{3.18}$$

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

$$cov(\widehat{y}_{ibt}, \widehat{y}_{ib,t+k}) = \widetilde{\sigma}^2_\alpha + \sigma^2_\beta * [h_{bt} * (h_{bt} + k)] + \delta_1 * \left[\frac{h_{bt} * (h_{bt} + 1)}{2}\right]. \tag{3.19}$$

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 $\widehat{c}_{btk} = \widehat{cov}(\widehat{y}_{ibt}, \widehat{y}_{ib,t+k})$, the parameter estimate for σ^2_β is given by $\widehat{\sigma}^2_\beta = \frac{\sum_{btk}(x_{btk}-\overline{x})*\widehat{c}_{btk}}{\sum_{btk}(x_{btk}-\overline{x})^2}$ and the omitted-variable bias formula for OLS implies that asymptotically

$$\widehat{\sigma}^2_\beta - \sigma^2_\beta = \delta_1 * \frac{cov(x_{btk}, z_{bt})}{var(x_{btk})}. \tag{3.20}$$

Since $cov(x_{btk}, z_{bt}) > 0$ 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 $\widehat{\sigma}^2_\beta$.■

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).*[24] *Results are shown in the six panels of appendix figure 1. Each line in a panel represents the experience-profiles of kth order autocovariances. The first panel plots the covariance structure implied by a random walk with a random effect. This is a line with intercept $\sigma^2_\alpha = 0.134$ and slope $\delta_0 = 0.\ 007$. In the second panel, I replace the random walk component by slope heterogeneity. With a relatively large estimate for $|\sigma_{\alpha\beta}|$, the experience profiles have negative slopes, while σ^2_β introduces some convexity. The interaction between the lag and experience identifying σ^2_β is reflected in high-order autocovariances having larger slopes in absolute value than low-order autocovariances. The third panel of the figure displays the covariance structure when one combines the first two panels. Given the large estimate for δ_0, experience profiles are strictly increasing, and slope heterogeneity generates the convexity of these profiles and introduces a non-trivial relationship between autocovariances and the order of the lag. Next I plot a homoscedastic AR(1)-process with a non-zero initial condition. The long-run value of its variance is given by $\frac{\gamma_0}{1-\rho^2} = \frac{0.09}{1-0.54^2} = 0.127$. Given that the initial condition $\sigma^2_{\xi 0} = 0.167$ is larger than this value, convergence to the long-run value is from above, and the experience profile is convex. The next panel demonstrates*

[24] I compute experience profiles up to the largest potential experience level observed in their data, which is 33. The parameters values are taken from table 4 in Baker and Solon (2003): $\sigma^2_\alpha = 0.134$, $\sigma^2_\beta = 0.00009$; $\sigma_{\alpha\beta} = -0.0031$; $\delta_0 = 0.007$; $\sigma^2_{\xi 0} = 0.167$; $\rho = 0.54$; $\gamma_0 = 0.09$; $\gamma_1 = -0.005$; $\gamma_2 = 0.0001$; $\gamma_3 = 2.21 * \exp(-6)$; $\gamma_4 = 2.1 * \exp(-9)$. I set all factor loadings equal to one.

experience profiles of autocovariances generated by a heteroscedastic AR(1)-process with an initial condition. With the parameter values used, these profiles are convex and U-shaped. The final panel combines all five panels and demonstrates very clearly the points discussed above: The profiles are dominated by the properties of the AR(1)-process at low lags, while they quickly converge to a lower envelope that is entirely dominated by the permanent component of the process. The final graph is remarkably similar to the empirical covariance structure used in the main part of my empirical analysis.■

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. The IABS is collected by the German Federal Employment Agency and is representative of the population of workers who are subject to compulsory social insurance contributions or who collect unemployment benefits. This amounts to approximately 80% of the German workforce, excluding self-employed and civil servants. Once an individual is drawn, he is followed for the rest of the sample period.

A number of advantages of using these data instead of publicly available panel data or administrative panel data from other countries are worth discussing in some more detail. First, I can generate unusually long worker-specific earnings histories; I observe up to 120 earning records on the *quarterly* level for the same worker. Indeed, the spell-based sampling design of the IABS that allows construction of quarterly rather than annual panel data is one of its distinct features. Second, given the large number of observations in the sample I can construct *cohort-specific* covariance structures, enabling me to estimate models of second moments of residual earnings that allow for both age and time-effects. This contrasts with studies relying on the PSID where small sample sizes require aggregation of autocovariances over cohorts. Third, since employees are observed from the time of labor market entry, I can flexibly model initial conditions of wage processes. Fourth, in contrast to North American administrative data, the IABS provides a well-defined education variable. Consequently, I can perform separate analyses for each education group because of the large sample sizes. Since covariance structures are quite different between these groups, I exploit this data feature to test the external validity and generalizability of my results. Importantly, the covariance structure of the largest education group strongly resembles covariance structures documented in various papers for the United States, Canada and Great Britain, as discussed above. Fifth, earnings records are provided by firms under a threat of legal sanctions for misreporting and are unlikely to be plagued by measurement error.

There are also a number of drawbacks of the data, most importantly the top coding of earnings at the social insurance contribution limit, a structural break in the earnings records in 1984, and the lack of a variable that records the hours worked. Most of these issues can be addressed directly by applying sample restrictions that are common in the literature. First, I only keep full-time work spells for the main job held during a quarter to rule out that earnings dynamics are driven by hours changes along the intensive margin, and I drop individuals with unstable employment histories, defined as those who are absent from the data for at least 3 consecutive years.[25] Second, to minimize the fraction of top-coded earnings I drop highly educated workers, defined as those with a technical college or university degree. This leaves two large education groups, subsequently referred to as "high-school dropout" and "high-school degree" samples, with fractions of top-coded earnings observations that are low and similar to the ones in commonly used survey data.[26] Since top-coded earnings observations contain valid information, namely that an individual has a large positive earnings residual relative to some comparison group, I follow Haider (2001) and Card, Kline and Heining (2013) in using an imputation procedure rather than dropping these observations.[27] Third, I use a novel and important sample restriction that only keeps individual labor market careers observed from labor market entry and therefore avoids an incidental parameters problem. Since earnings histories are left-censored in 1975 I drop individuals who are observed in that year.[28] Some employees entering the labor market after 1975 do so at a fairly high age for possibly endogenous reasons. Hence, I only keep a sample of workers who start their career at education-specific mass points of age-at-labor-market-entry. These are 19 years for high school dropout and 23 years for those with a formal secondary degree. Finally, I restrict the sample to male workers whose entire career is recorded in Western Germany. Due to fairly small sample sizes at the highest experience levels, I also drop observations for which experience exceeds 108 quarters in the secondary-degree sample and 100 quarters in the dropout-sample. Further details of sample construction

[25] The first restriction is similar to the hours restrictions used by most of the studies that rely on the PSID. See for example Haider (2001), Guvenen (2009) and Hryshko (2012). The IABS contains a variable recording whether the job is full- or part-time. In my sample of male employed workers who are observed on their main job held during a quarter, only approx. 2.5% percent of all spells are part-time. Here, the main job is defined as the job that generates the highest earnings during a quarter. The share of part-time workers in the raw data, that is before any sample restrictions are imposed, is 7.5% among male workers and 35% among female workers.

[26] "High-school dropouts" are individuals who do not obtain a formal secondary degree. "High-school graduates" are defined as those who hold on to a formal secondary degree, including those with an apprenticeship degree. Because of the importance of the apprenticeship system in the German labor market, this group covers over 70% of the employed. The fraction of censored observations is 0.5% in the high-school dropout sample and 4.7% in the high school degree sample. In comparison, it is 55.2% in the education group that is dropped from the sample.

[27] A more common approach is to drop top-coded earnings records. This introduces a sample selection problem that potentially leads to a bias in the empirical auto-covariances. 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. It is therefore likely to lead to a downward bias in parameters that generate a fanning out of the wage distribution over the life-cycle: Permanent shocks and slope heterogeneity. I have reestimated all specifications in this paper using this approach instead. The conclusions remain unaltered.

[28] Labor market entry is defined as the period a worker has completed his highest degree and is recorded to have positive earnings. This drops apprenticeship spells from the data.

are given in appendix A.

4.2. Sample Sizes

Sample sizes for the two education groups and for each cohort are reported in the left panel of appendix table 1. These are sums over both, individuals and time. As younger cohorts have shorter time series by construction of the sample, their sample sizes are significantly smaller than those for older cohorts. 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. The right panel of the table reports sample sizes by experience in years instead. Approximately 323,000 individuals with a secondary degree are observed from their first year in the labor market, compared with 35,000 individuals for the other education group. Half of these entrants are still observed after 14 years for the first and 11 years for the second education group. In all cases, far more than 10% of the initial sample are still present after 20 years. Sample sizes decrease quickly as we approach the highest observed experience levels because less and less cohorts contribute to these observations. For example, only 3 cohorts reach an experience level of 24 years in the group with a secondary educational degree. In total, there are 824,962 earnings observations for these 3 groups. If there was no attrition, these groups should contribute 824,962/(24+1) = 32,998 observations to each experience group. Given that over 26,000 observations are left after 24 years, the attrition rate is quite low.

4.3. Descriptive Analysis

Estimation of Mincer earnings regressions with random coefficients and a dynamic error structure can be cast in terms of non-linear least squares estimation on covariance structures, as argued above. Each parameter will thus be identified from some particular statistical variation in the empirical covariance structure. A detailed graphical analysis, carried out in this subsection, will give a first impression of the types of variation that are featured by the covariance structures.[29]

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 twenty to thirty quarters and increase slowly and steadily afterwards. As highlighted by Guvenen (2009) this convexity is consistent with heterogeneous returns to experience, i.e. the "HIP-component", but it can potentially be generated by other mechanisms as well,

[29] Corresponding empirical first moments of log-labor income are listed in appendix table 2.

such as age-dependence in the innovation variances. 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.

Earnings processes do not only have implications for the shape of life-cycle profiles of auto-covariances, but also for the relationship between auto-covariances 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. Auto-covariances 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.

A number of these empirical facts are consistent with the North-American evidence. Guvenen (2009) documents a decrease of the variances over the first five years of a life-cycle and an increase afterwards. Non-stationarity of the earnings structure, with a significant increase in the covariance structure over time and across cohorts, is also a well known feature of North-American data.[30] Negatively sloped lag-profiles at low lags have been found in US earnings data as well, but there is some evidence that they are not monotonically declining for highly educated older workers.[31]

The qualitative similarity of the covariance structure of German male workers with a secondary degree to the covariance structure reported in US data is the main reason for my empirical focus on this group. All empirical results for the dropout group are documented in the appendix. As shown in appendix figures 2 and 3 the covariance structure for those without a secondary degree differs substantially from the covariance structure for the secondary-degree group. Most importantly, there is little evidence for convexities in the experience profiles, and convergence of experience and lag-profiles takes place over the first five years of a career. High-order autocovariances are very close to zero and remain so for the entire life-cycle. However, similar to the secondary-degree group, high-school dropouts have experienced a significant fanning out of the wage structure as reflected in the increase of covariance profiles, but only early in the life-cycle and at small lags. Hence, in contrast to the higher educated workers, there is a significant compression of the wage distribution over the life-cycle for all cohorts.

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.

[30] See e.g. Gottschalk and Moffitt (1994), Haider (2001), Baker and Solon (2003), and Blundell, Pistaferri and Preston (2009).
[31] See e.g. Guvenen (2009)

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 modelled 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.

In the following I focus my discussion on the results for the secondary-degree group since its covariance structure of earnings is qualitatively similar to the North American counterpart. I view the results for the high-school dropouts, presented in the appendix, as an extensive robustness check. 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.[32] Given these results I treat a specification that restricts $\delta_j = \phi_j = 0$ 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).[33]

5.1. Estimates from the Benchmark Specification

Parameter estimates for the benchmark specification are shown in the first column of table 1 and, for the two sets of factor loadings, in the top panel of appendix figure 4.[34] 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.[35] The exercise is carried out for life-cycle profiles of autocovariances at a lag of 0, 4, 20 and 40 quarters. As can be seen from the figures, the model can generate qualitatively and quantitatively all the features of the auto-covariance 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 percent of the variation

[32] As discussed in the modeling section, this is also the reason why I am not including an MA(q)-component or factor loadings on the transitory component.

[33] Increasing J_ξ does not improve the model fit significantly. One concern is that the Wald-tests carried out in the pretesting stage, with the result that age effects in the variances of permanent and transitory shocks are jointly insignificant, have low power. I have investigated this issue using Monte-Carlo analysis and found that this is not a problem. Importantly, the simulation results indicate that the estimator detects age heteroscedasticity in the distribution of permanent shocks if it is present in the data generating process. All results from the pretesting stage are available upon request.

[34] The absolute value of $\widehat{corr(\alpha_i, \beta_i)}$ is contained in $[-1, 1]$ for all my estimates of $\left(\widehat{\sigma}^2_\alpha, \widehat{\sigma}^2_\beta, \widehat{\sigma}_{\alpha\beta}\right)$. However, using the estimates displayed in the tables to calculate the correlation coefficient yields $\widehat{corr(\alpha_i, \beta_i)} < -1$ in some cases. This is due to rounding error.

[35] An alternative would be to clean the autocovariances from cohort effects much like in appendix figure 1, but this would mask the ability of the model to fit inter-cohort changes.

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, $\widehat{\sigma}_{\beta}^{2}$, are significant on the 10%-, and with few exceptions, on the 1%−level.[36] There is substantial heterogeneity in the intercept and the initial condition of the persistent component, with estimated variances of $\widehat{\sigma}_{\alpha}^{2} = 0.023$ and $\widehat{\sigma}_{\xi 0}^{2} = 0.092$ respectively. The estimated persistence of shocks to the AR(1)-process on the *quarterly* level is $\widehat{\rho} = 0.88$, 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 $\phi_0 = .004$. Since both, the variance of earnings intercepts, $\widehat{\sigma}_{\alpha}^{2}$, 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 .094 for the $1,488$ variance elements in the autocovariance structure, the permanent component can explain approximately one quarter (0.023/.094) of the total variation in log earnings in the group of the high school educated.[37]

The evolution of the two sets of factor loadings plotted in appendix figure 4 helps identify whether the trend in the wage structure towards a higher level of income inequality is driven by an increase in the dispersion of the permanent or the persistent component. The empirical results are quite striking. Controlling for age, permanent inequality has remained almost unchanged, while persistent inequality has nearly quadrupled. As highlighted by Haider and Solon (2006), this implies that life-cycle inequality has grown much less than cross-sectional inequality.

Turning to the HIP component, the estimated heterogeneity in slopes $\widehat{\sigma}_{\beta}^{2}$ is insignificant on any conventional level and very small in magnitude, but its covariance with intercept heterogeneity $\widehat{\sigma}_{\alpha\beta}$ 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 $\widehat{\sigma}_{\alpha\beta}$ to be insignificant whenever $\widehat{\sigma}_{\beta}^{2}$ is. It is therefore important to document a test statistic for the *joint* significance of the two parameters, provided at the bottom of the table. The null-hypothesis $\left(\sigma_{\alpha\beta}, \sigma_{\beta}^{2}\right) = (0,0)$ 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 $\left(\sigma_{\alpha\beta}, \sigma_{\beta}^{2}\right) = (0,0)$. The estimates are listed in column (2) of the same table. The R^2 decreases by only .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 $\widehat{\sigma}_{\alpha}^{2}$, which decreases by a half to a value of .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

[36] This also applies to all factor loadings. To avoid clutter in appendix figure 3 I do not plot the confidence intervals.
[37] The variance elements correspond to elements with a lag of zero ($k = 0$). Their range is [.047, .266].

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. Results are shown in appendix table 3 for the main parameters of interest and for various specifications.[38] Estimates corresponding to column 1 of table 1 are listed in column 1 of the appendix table. The time-aggregated transitory component now has variance of zero since it is assumed to be *iid* across individuals, age and time. The estimated intercept heterogeneity in the annual data is almost identical to its quarterly counterpart. In fact, if earnings were constant within a year, then the two estimates should be identical. 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 .88. As shown in the table, this translates into a persistence of .632 on the annual frequency.

5.2. HIP and Age Effects: Results from Mis-specified 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 *wihout* an initial condition, a purely transitory component, and factor loadings on both persistent and transitory shocks.[39] 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. The discrepancy between the estimates of the two specifications is quite striking. Compared to the full model, profile heterogeneity as captured by $\widehat{\sigma}^2_\beta$ is thirteen times as large and becomes highly statistically significant. The R^2 decreases dramatically to a value of .764, indicating that the model is severely misspecified. This is particularly remarkable given that the model still has 56 parameters, compared with 62 parameters in the benchmark specification. Another interesting result is that the estimated persistence of the AR(1)-shocks is much larger in the HIP-model than in the full model. In fact, it is not significantly different from one. This can be explained by the near-linearity of experience profiles at high orders, which can be generated by a simple unit roots process. It follows that controlling for slope heterogeneity does not necessarily imply that shocks will be estimated to be less than persistent.[40]

[38] I do not show standard errors in this table since the data are generated exactly once under the assumption that the model is correctly specified and that sampling error is absent.

[39] The process is identical to the one estimated in Guvenen (2009). Importantly, this allows for factor loadings on the transitory component as well, which are found to be insignificant in my benchmark specification, but not in the more restrictive specification shown here.

[40] Comparison with estimates in column 7 of the table clarifies that it is the introduction of a unit-roots component into the HIP-process that has a major effect on $\widehat{\rho}$.

I also show results for a simple RIP-model that does not have any time- or age effects in column (3) of the table. Not surprisingly, this model delivers an estimated persistence that is much higher than in the benchmark model as well. The amount of variation in earnings associated with persistent and transitory shocks is much larger than in Guvenen's HIP-model, with $\widehat{\gamma}_0$ and $\widehat{\phi}_0$ being substantially larger in in column (3) (RIP) than in column (2) (HIP) of the table. At the same time, intercept heterogeneity is much smaller in the RIP model. This demonstrates very clearly that the a-priori choice of an earnings process has first-order effects on the quantitative importance for earnings inequality one assigns to risk and to heterogeneity.

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. I consider four nested versions of my benchmark specification: A model with homoscedastic shocks in column (4), a time-stationary model in column (5), a model without an initial condition for the AR(1)-process in column (6), and a model that combines all of these restrictions in column (7). The last specification is equivalent to Hryshko's (2012) combined HIP-RIP process.[41] The conclusion one can draw from this analysis is quite clear. The HIP component is found to be significant only in specifications that do not feature an initial condition for the AR(1)-process. This is a particular type of age effect, where one allows the variance of the persistent component to differ between labor market entrants and the rest of the employees. Its estimated importance for earnings dynamics is consistent with a large literature that documents a persistent impact of initial job placement on career advancement.[42] At the same time, the result that its omission causes a — potentially large — upward bias in the estimated profile heterogeneity is new. It can be understood in terms of the discussion of identification in the previous section. Both model components can generate age effects in autocovariance profiles, in this particular case a decline at the beginning of the life-cycle. Once one does not control for the initial condition, all of this decline will be associated with slope heterogeneity, thereby biasing the estimate. In the full model, the two channels can be separated because the effect of the persistent initial condition eventually vanishes as experience and the lag increase, while the effect of slope heterogeneity becomes stronger.

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 how well my estimation performs in finitely-sized samples. A detailed discussion of both exercises are included in appendix E and F. Here I briefly summarize the main findings.

[41] The fit of Hryshko's (2012) model is lower than the fit of Guvenen's (2009) model because I consider a version of the former that excludes time effects.

[42] See e.g. von Wachter and Bender (2006), Kahn (2010), Oreopoulos, von Wachter and Heisz (2012), and Altonji, Kahn and Speer (2014).

How Robust are the Conclusions? Results from the High School Dropout Sample The results from estimating my benchmark specification on the sample of high school dropouts are documented in appendix table 4, which has the same structure as table 2 for the main sample. Generally the results 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, as discussed above. 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 simulation protocol is described in appendix F, and results are shown in appendix table 5. 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 1,000 Monte Carlo repetition is small relative to the magnitude of the true parameter values.

6. Concluding Discussion

The dispersion of individual returns to experience is an important parameter in life-cycle models of career- or consumption choices. It is common to estimate it by matching a HIP-process to the empirical covariance structure of residuals from a Mincer earnings regression with random coefficients. In this study I argue that such an approach to identification and estimation tends to produce an upward bias in profile heterogeneity if age effects in innovation variances are not controlled for. This is because the age structure of covariances is one source of identifying variation for slope heterogeneity in the absence of age heteroscedasticity, while various economic models suggest that the latter is an important source of life-cycle earnings variation. Once one models age effects semiparametrically, the only remaining identifying source for slope heterogeneity is the shape of the tails of lag profiles. It is here where profile heterogeneity makes particularly strong and unique empirical predictions.

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.

While I have established theoretically that estimation of a HIP-process without age effects in innovation variances will inevitably yield estimates of slope heterogeneity that are biased upwards, it is not clear a priori whether the bias is quantitatively important. To investigate this issue I rely on German administrative data that follow workers for a long time and that record their earnings on the quarterly frequency, thus satisfying the large demands on the data. As my results show, the bias from omitting age effects can be substantial. In both samples I am using, slope heterogeneity is found to be significant if I estimate a standard HIP model but turn insignificant and very small in magnitude when controlling for age effects. This is not due to larger standard errors in my benchmark specification, but because of a decrease in estimates, in my main sample by a factor of more than ten. Whether the bias is particularly large in the German data can only be answered with additional evidence from other countries. However, a number of findings that I document suggest that external validity is strong. First, qualitatively the autocovariance structure in the main sample shares many of the features of its North American counterpart. This is reflected in estimates of standard RIP- and HIP-processes that are qualitatively similar to those obtained from US-data. Second, as I have shown I reach the same conclusions for two very different autocovariance structures.

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.

References

[1] Abbott, B., G. Gallipoli, C. Meghir and G. Violante (2018): "Education Policies and Intergenerational Transfers in Equilibrium." *Journal of Political Economy*, forthcoming.

[2] Abowd, J. M. and D. Card (1989): "On the Covariance Structure of Earnings and Hours Changes." *Econometrica*, vol. 57(2), pp. 411-445.

[3] Adda, J., C. Dustmann, C. Meghir and J.-M. Robin (2011): "Career Progression and Formal versus On-the-Job Training." mimeo, University College London.

[4] Alan, S., M. Browning and M. Ejrnaes (2018): "Income and Consumption: A Micro Semi-Structural Analysis with Pervasive Heterogeneity." *Journal of Political Economy*, forthcoming.

[5] Altonji, J. and L. Segal (1996): "Small Sample Bias in GMM-estimation of Covariance Structures" *Journal of Business and Economic Statistics*, vol. 14(3), pp. 353-66.

[6] Altonji, J., A. Smith Jr. and I. Vidangos (2013): "Modeling Earnings Dynamics", *Econometrica*, vol. 81(4), pp. 1395-1454.

[7] Alvarez, F. and U. Jermann (2000): "Efficiency, Equilibrium, and Asset Prices with Risk of Default." *Econometrica*, vol. 68(4), pp. 775-798.

[8] Arellano, M., R. Blundell and S. Bonhomme (2017): "Earnings and Consumption Dynamics: A Nonlinear Panel Data Framework." *Econometrica*, vol. 85(3), pp. 693-734.

[9] Baker, M. (1997): "Growth Rate Heterogeneity and the Covariance Structure of Life Cycle Earnings", *Journal of Labor Economics*, 15, pp. 338-375.

[10] Baker, M., and G. Solon (2003): "Earnings Dynamics and Inequality Among Canadian Men, 1976-1992: Evidence from Longitudinal Income Tax Records", *Journal of Labor Economics*, vol. 6(4), pp. 445-69.

[11] Ben-Porath, Y. (1967): "The Production of Human Capital and the Life Cycle of Earnings." *Journal of Political Economy*, vol. 75(4), pp. 352–65.

[12] Biewen, M. (2005): "The Covariance Structure of East and West German Incomes and its Implications for the Persistence of Poverty and Inequality." *German Economic Review*, 15, pp. 338-375.

[13] Blundell, R., L. Pistaferri and I. Preston (2009): "Consumption Inequality and Partial Insurance", *American Economic Review*, 98(5), pp. 1887-1921.

[14] Blundell, R., M. Graeber and M. Mogstad (2015): "Labor Income Dynamics and the Insurance from Taxes, Transfers, and the Family", *Journal of Public Economics*, 127, pp. 58-73.

[15] Bonhomme, S. and J.-M. Robin (2009): “Assessing the Equalizing Force of Mobility using Short Panels: France, 1990-2000”, *Review of Economic Studies*, 76(1), pp. 63-92.

[16] Browning, M., M. Ejrnaes and J. Alvarez (2010): ”Modelling Income Processes with Lots of Heterogeneity”, *Review of Economic Studies*, 77(4), pp. 1353-81.

[17] Browning, M. and M. Ejrnaes (2014): ”The Persistent-Transitory Representation for Earnings Processes”, *Quantitative Economics*, 5(3), pp. 555-581.

[18] Cameron, A. and P. Trivedi (2005): “Microeconometrics.” Cambridge University Press.

[19] Card, D., J. Heining and P. Kline (2013): “Workplace Heterogeneity and the Rise of West German Wage Inequality,“ *Quarterly Journal of Economics*, 128, pp. 967-1015.

[20] Cunha, F., & Heckman, J. (2007). ”The technology of skill formation” NBER Working Paper w12840).

[21] Currie, J., & Almond, D. (2011). ”Human capital development before age five.” Handbook of labor economics, vol. 4, pp. 1315-1486.

[22] Daly, M., D. Hryshko and I. Manovskii (2017): “Improving the Measurement of Earnings Dynamics.” mimeo, University of Alberta.

[23] Davis, S. and T. von Wachter (2012): “Recessions and the Costs of Job Loss.” *Brookings Papers on Economic Activity*.

[24] De Nardi, M., G. Fella and G. P. Pardo (2018): “Nonlinear Household Earnings Dynamics, Self-Insurance, and Welfare,“ NBER working paper 24326.

[25] Dickens, R. (2000): ”The Evolution of Individual Male Earnings in Great Britain: 1975-95”, *The Economic Journal*, vol. 110, pp. 27-49.

[26] Dustmann, C., J. Ludsteck and U. Schoenberg (2009): “Revisiting the German Wage Structure.” *Quarterly Journal of Economics*, vol. 124, pp. 843-881.

[27] Erosa, A., G. Kambourov and L. Fuster (2016): ”Towards a Micro-Founded Theory of Aggregate Labor Supply.” *Review of Economic Studies*, 83(3), pp. 1001-1039.

[28] Farhi, E. and I. Werning (2013): ”Insurance and Taxation over the Life-Cycle.” *Review of Economic Studies*, 80(2), pp. 596-635.

[29] Friedman, M. (1957): ”A Theory of the Consumption Function.” Princeton University Press, Princeton, NJ.

[30] Fukushima, K. (2010): ”Quantifying the Welfare Gain from Flexible Dynamic Income Tax Systems.” mimeo, University of Wisconsin at Madison.

[31] Gottschalk, P., and R. A. Moffitt (1994): "The Growth of Earnings Instability in the U.S. Labor Market", *Brookings Paper on Economic Activity*, 25(2), pp. 217-272.

[32] Guiso, L., L. Pistaferri and F. Schivardi (2005): "Insurance within the Firm." *Journal of Political Economy*, vol. 113(5), pp. 1054-1087.

[33] Guvenen, F. (2007): "Learning Your Earning: Are Labor Income Shocks really very Persistent?", *American Economic Review*, 97(3), pp. 687-712.

[34] Guvenen, F. (2009): "An Empirical Investigation of Labor Income Processes", *Review of Economic Dynamics*, 12(1), pp. 58-79.

[35] Guvenen, F. (2011): "Macroeconomics with Heterogeneity: A Practical Guide.", *Federal Reserve Bank of Richmond Quarterly.*

[36] Guvenen, F. and A. Smith (2014): "Inferring Labor Income Risk and Partial Insurance from Economic Choices." *Econometrica*, vol. 82(6), pp. 2085-2129.

[37] Guvenen, F., Kuruscu, B., & Ozkan, S. (2014). "Taxation of Human Capital and Wage Inequality: A Cross-Country Analysis." *Review of Economic Studies*, 81(2), pp. 818-850.

[38] Guvenen, F., Karahan, F., Ozkan, S., & Song, J. (2015). "What do data on millions of US workers reveal about life-cycle earnings risk?" NBER Working Paper w20913.

[39] Haider, S. (2001): "Earnings Instability and Earnings Inequality of Males in the United States: 1967-1991", *Journal of Labor Economics*, 19(4), pp. 799-836.

[40] Haider, S. and G. Solon (2006): "Life-Cycle Variation in the Association between Current and Lifetime Earnings", *American Economic Review*, 96(4), pp. 1308-20.

[41] Hall, R. E., and F. S. Mishkin (1982): "The Sensitivity of Consumption to Transitory Income: Estimation from Panel Data on Households", *Econometrica*, 50(2), pp. 461-481.

[42] Hause, J. (1980): "The Fine Structure of Earnings and the On-the-Job Training Hypothesis." *Econometrica*, vol. 48(4), pp. $1013 - 1029$.

[43] Heathcote, J., K. Storesletten, and G. L. Violante (2005): "TwoViews of Inequality over the Life-Cycle", *Journal of the European Economic Association*, 3, pp. 765-775.

[44] Heathcote, J., K. Storesletten, and G. L. Violante (2008): "Insurance and Opportunities: A Welfare Analysis of Labor Market Risk", *Journal of Monetary Economics*, 55(3), pp. 501-525.

[45] Heathcote, J., K. Storesletten, and G. L. Violante (2014): ”Consumption and Labor Supply with Partial Insurance: An Analytical Framework”, *American Economic Review*, 104(7), pp. 2075-2126.

[46] Hirano, K. (2002): “Semiparametric Bayesian Inference in Autoregressive Panel Data Models”, *Econometrica*, 70(2), pp. 781-799.

[47] Hoffmann, F. (2010): ”An Empirical Model of Life-Cycle Earnings and Mobility Dynamics”, mimeo, University of British Columbia.

[48] Horowitz, J., and M. Markatou (1996): “Semiparametric Estimation of Regression Models for Panel Data”, *Review of Economic Studies*, 63, pp. 145-168.

[49] Hryshko, D. (2012): ”Labor Income Profiles are not Heterogeneous: Evidence from Income Growth Rates.” *Quantitative Economics*, 3, pp. 177-209.

[50] Hryshko, D. (2014): ”Correlated Income Shocks and Excess Smoothness of Consumption.” *Journal of Economic Dynamics and Control*, 48, pp. 41-62.

[51] Hu, Y, R. Moffitt and Y Sasaki (2018): ”Semiparametric Estimation of the Canonical Permanent-Transitory Model of Earnings Dynamics.” mimeo, Johns Hopkins University.

[52] Huggett, M., G. Ventura, and A. Yaron (2011): ”Sources of Lifetime Inequality.” *American Economic Review*, 101(7), pp. 2923-54.

[53] Huggett, M and G. Kaplan (2016): ”How Large is the Stock Component of Human Capital?” *Review of Economic Dynamics*, vol. 22, pp. 21-51.

[54] Kahn, L. (2010): ”The Long-Term Labor Market Consequences of Graduating from College in a Bad Economy.” *Labour Economics*, 17(2), pp. 303-316.

[55] Karahan, F. and S. Ozkan (2013): ”On the Persistence of Income Shocks over the Life Cycle: Evidence, Theory, and Implications.” *Review of Economic Dynamics*, 16(3), pp. 452–476.

[56] Lamadon, T. (2016): ”Productivity Shocks, Long-Term Contracts and Earnings Dynamics.” mimeo, University of Chicago.

[57] Lemieux, T. (2006): ”Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill?.” *American Economic Review*, 96(3), pp. 461-98.

[58] Lillard, L. and Y. Weiss (1979): “Components of Variation in Panel Earnings Data: American Scientists, 1960-70.” *Econometrica*, vol. 47(2), pp. 437 – 54.

[59] Lochner, L. and Y. Shin (2014): "Understanding Earnings Dynamics: Identifying and Estimating the Changing Roles of Unobserved Ability, Permanent and Transitory Shocks." NBER working paper 20068.

[60] Lopez-Daneri, Martin (2016): "Life-Cycle Patterns of Earnings Shocks," mimeo, Temple University.

[61] Low, H., C. Meghir and L. Pistaferri (2010): "Wage Risk and Employment Risk over the Life-Cycle." *American Economic Review*, 100(4), pp. 1432-67.

[62] MaCurdy, (1982): "The Use of Time Series Processes to Model the Error Structure of Earnings in a Longitudinal Data Analysis,." *Journal of Econometrics*, vol. 18, pp. 82-114.

[63] Meghir, C. and L. Pistaferri (2004): "Income Variance Dynamics and Heterogeneity." *Econometrica*, vol. 72(1), pp. 1-32.

[64] Meghir, C. and L. Pistaferri (2011): "Earnings, Consumption and Life-Cycle Choices." *Handbook of Labor Economics,* vol. 4(5).

[65] Mincer, J. (1958): "Investment in Human Capital and Personal Income Distribution." *Journal of Political Economy*, vol. 66(4), pp. 281-302.

[66] Mincer, J. (1974): "Schooling, Experience and Earnings." New York: National Bureau of Economic Research.

[67] Moffitt, R. A., and P. Gottschalk (2002): "Trends in the Transitory Variance of Earnings in the United States", *Economic Journal*, 112, pp. C68-C73.

[68] Oreopoulos, P., T. von Wachter and A. Heisz (2012): "The Short- and Long-Term Career Effects of Graduating in a Recession", *AEJ-Applied Economics*, vol. 4(1), pp. 1-29.

[69] Pavan, R. (2011): "Career Choice and Wage Growth", *Journal of Labor Economics*, vol. 29(3), pp. 549 - 587.

[70] Perri, F. and D. Krueger (2006): "Does Income Inequality lead to Consumption Inequality? Evidence and Theory." *Review of Economic Studies*, vol. 73(1), pages 163-193.

[71] Polachek, S. W., T. Das and R Thamma-Apiroam (2015): "Micro- and Macroeconomic Implications of Heterogeneity in the Production of Human Capital." *Journal of Political Economy*, vol. 123(6), pp. 1410-1455.

[72] Postel-Vinay, F. and H. Turon (2010): "On-The-Job Search, Productivity Shocks, and the Individual Earnings Process," *International Economic Review*, vol. 51(3), pages 599-629.

[73] Primiceri, G. and T. van Rens (2009): "Heterogeneous Life-Cycle Profiles, Income Risk and Consumption Inequality," *Journal of Monetary Economics*, vol. 56(1), pages 20-39.

[74] Sanchez, M. and F. Wellschmied (2017): "Modeling Life-Cycle Earnings Risk with Positive and Negative Shocks," mimeo, University of Bristol.

[75] Steiner, V. and K. Wagner (1998): "Has Earnings Inequality in Germany Changed in the 1980's?", *Zeitschrift fuer Wirtschafts- und Sozialwissenschaften,* 118, pp. 29-59.

[76] Storesletten, K, C. Telmer, and A. Yaron (2004a): "Consumption and Risk-Sharing over the Life-Cycle", *Journal of Monetary Economics*, 51, pp. 609-633.

[77] Storesletten, K, C. Telmer, and A. Yaron (2004b): "Cyclical Dynamics in Ideosyncratic Labor Market Risk", *Journal of Political Economy*, 112, pp. 695-717.

[78] von Wachter, T. and S. Bender (2006): "In the Right Place at the Wrong Time: The Role of Firms and Luck in Young Workers' Careers", *American Economic Review*, 96(5), pp. 1679-1705.

TABLE 1 - PARAMETER ESTIMATES FOR BASELINE SPECIFICATIONS: SECONDARY DEGREE GROUP

		(1.)	(2.)
		Benchmark Specification	**no Slope Heterogeneity**
Intercept Heterogeneity	σ_{α}^{2}	0.023 (0.004)***	0.012 (0.001)***
Slope Heterogeneity	$\sigma_{\beta}^{2} * 10^{3}$	0.0005 (0.0014)	-
Cov (Intercept; Slope)	$\sigma_{\alpha\beta} * 10$	-0.001 (0.0003)***	-
Persistence of AR(1)	ρ	0.880 (0.006)***	0.906 (0.005)***
AR(1) error structure			
Initial Condition	$\sigma_{\xi 0}^{2}$	0.092 (0.014)***	0.080 (0.005)***
Intercept	γ_0	0.007 (0.002)***	0.004 (0.001)***
experience	γ_1	-3.16*e(-4) (1.17*e(-4))***	-1.49*e(-4) (3.59*e(-5))***
experience^2	γ_2	7.68*e(-6) (3.42*e(-6))**	3.18*e(-6) (1.21*e(-6))***
experience^3	γ_3	-9.16*e(-8) (4.63*e(-8))**	-3.73*e(-8) (1.65*e(-8))***
experience^4	γ_4	3.95*e(-10) (2.17*e(-10))*	1.61*e(-10) (7.53*e(-11))***
Variance of Permanent Shocks	$\delta_0 * 10$	0.007 (0.001)***	0.004 (0.001)***
Variance of Measurement Error	ϕ_0	0.004 (0.001)***	0.006 (0.001)***
Number of Moments		**56,072**	
R^2		0.964	0.959
Wald Test for Slope Heterogeneity (P-Value)		0.000	-

NOTES: This table shows parameter estimates for the benchmark specification as described in equations (3.2) to (3.8) of the paper, together with a nested specification that sets slope heterogeneity to zero. 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. Both specifications allow for factor loadings on the permanent and persistent component, all of which are signficant on the 1%-level. Estimated factor loadings for the full model are displayed in the first panel of appendix figure 5. *** Significant on 1%-level; ** Significant on 5%-level; * Significant on 10%-level. Standard errors are clustered by cohort to account for arbitrary correlation of sampling error within cohort-groups.

TABLE 2 - ROBUSTNESS OF PARAMETER ESTIMATES: SECONDARY DEGREE GROUP

		FULL MODEL, HIP, AND RIP			RESTRICTIONS ON BENCHMARK SPECIFICATION			
		(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_α^2	0.023 (0.004)***	0.053 (0.004)***	0.027 (0.002)***	0.015 (0.001)***	0.028 (0.003)***	0.034 (0.004)***	0.051 (0.004)***
Slope Heterogeneity	$\sigma_\beta^2 * 10^3$	0.0005 (0.0014)	0.006 (0.003)***	-	0.000 (0.001)	0.002 (0.002)	0.002 (0.001)*	0.005 (0.002)***
Cov (Intercept; Slope)	$\sigma_{\alpha\beta} * 10$	-0.001 (0.0003)***	-0.005 (0.0008)***	-	-0.0006 (0.0001)***	-0.002 (0.001)***	-0.003 (0.001)***	-0.005 (0.0007)***
Persistence of AR(1)	ρ	0.880 (0.006)***	0.996 (0.004)***	0.982 (0.003)***	0.905 (0.004)***	0.857 (0.006)***	0.755 (0.01)***	0.757 (0.017)***
AR(1) error structure								
Initial Condition	$\sigma_{\xi 0}^2$	0.092 (0.014)***	-	-	0.081 (0.007)***	0.136 (0.01)***	-	-
Intercept	γ_0	0.007 (0.002)***	0.001 (0.0006)*	0.002 (0.0002)***	0.002 (4.31*e(-4))***	0.026 (0.007)***	0.0212 (0.005)***	0.011 (0.002)***
experience	γ_1	-3.16*e(-4) (1.17*e(-4))***	-	-	-	-0.001 (2.75*e(-4))***	-0.002 (3.27*e(-4))***	-
experience^2	γ_2	7.68*e(-6) (3.42*e(-6))**	-	-	-	2.98*e(-5) (1.02*e(-5))***	4.45*e(-5) (8.86*e(-6))***	-
experience^3	γ_3	-9.16*e(-8) (4.63*e(-8))**	-	-	-	-3.67*e(-7) (1.48*e(-7))***	-5.42*e(-7) (9.93*e(-8))***	-
experience^4	γ_4	3.95*e(-10) (2.17*e(-10))*	-	-	-	1.64*e(-9) (7.34*e(-10))**	2.29*e(-9) (3.92*e(-10))***	-
Variance of Permanent Shocks	$\delta_0 * 10$	0.007 (0.001)***	-	-	0.003 (0.001)***	0.010 (0.001)***	0.007 (0.001)***	0.011 (0.0009)***
Variance of Measurement Error	ϕ_0	0.004 (0.001)***	0.018 (0.008)**	0.026 (0.003)***	0.000 -	0.003 (0.001)***	0.001 (0.001)	0.005 (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

NOTES: 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). *** Significant on 1%-level; ** Significant on 5%-level; * Significant on 10%-level. Standard errors are clustered by cohort to account for arbitrary correlation of sampling error within cohort-groups.

FIGURE 1 - LIFE-CYCLE PROFILES OF AUTO-COVARIANCES AT DIFFERENT LAGS, BY COHORTS

Sample: Secondary Degree Group

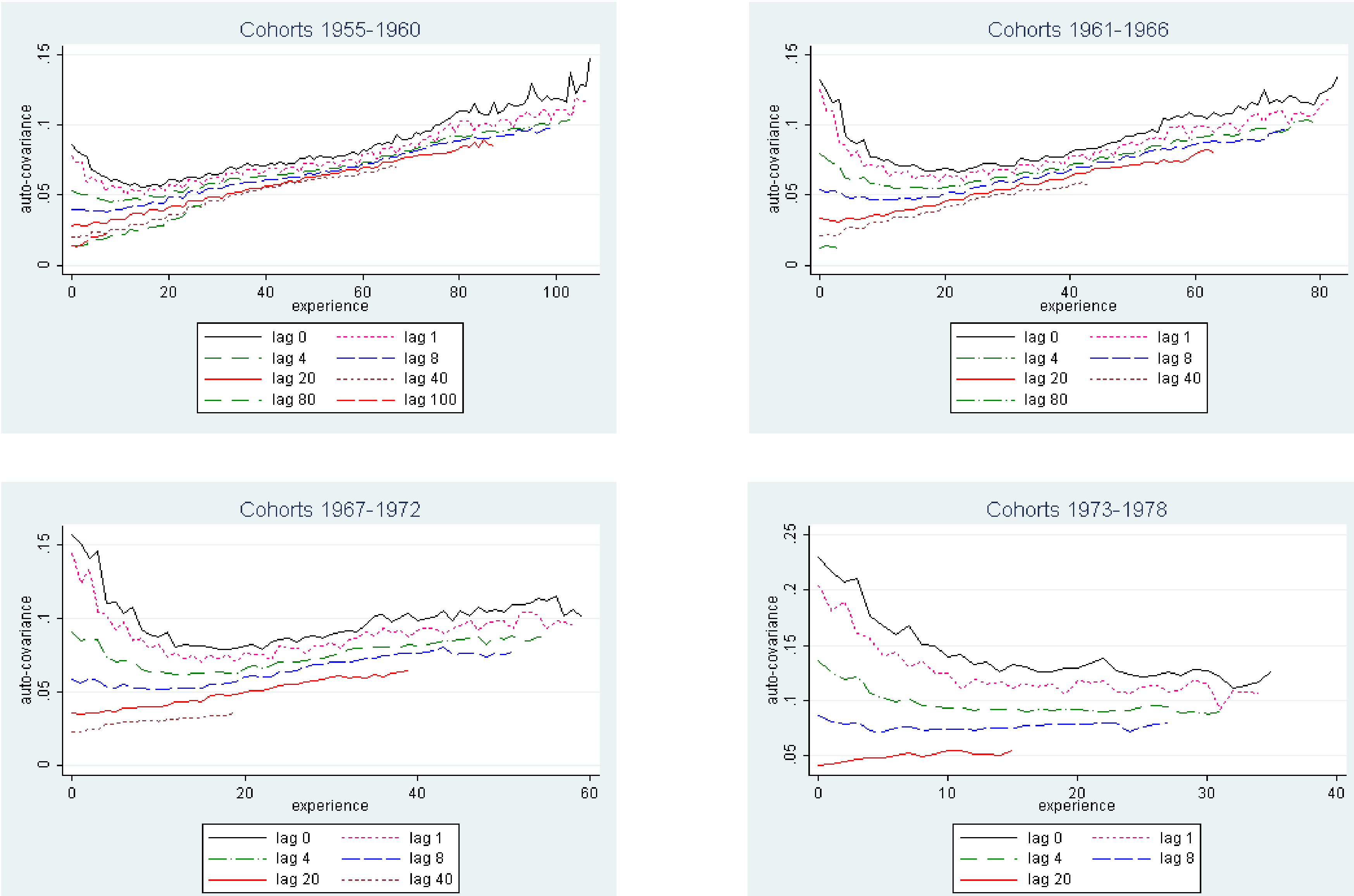


FIGURE 2 - LAG-PROFILES OF AUTO-COVARIANCES FOR DIFFERENT EXPERIENCE GROUPS, BY COHORTS

Sample: Secondary Degree Group

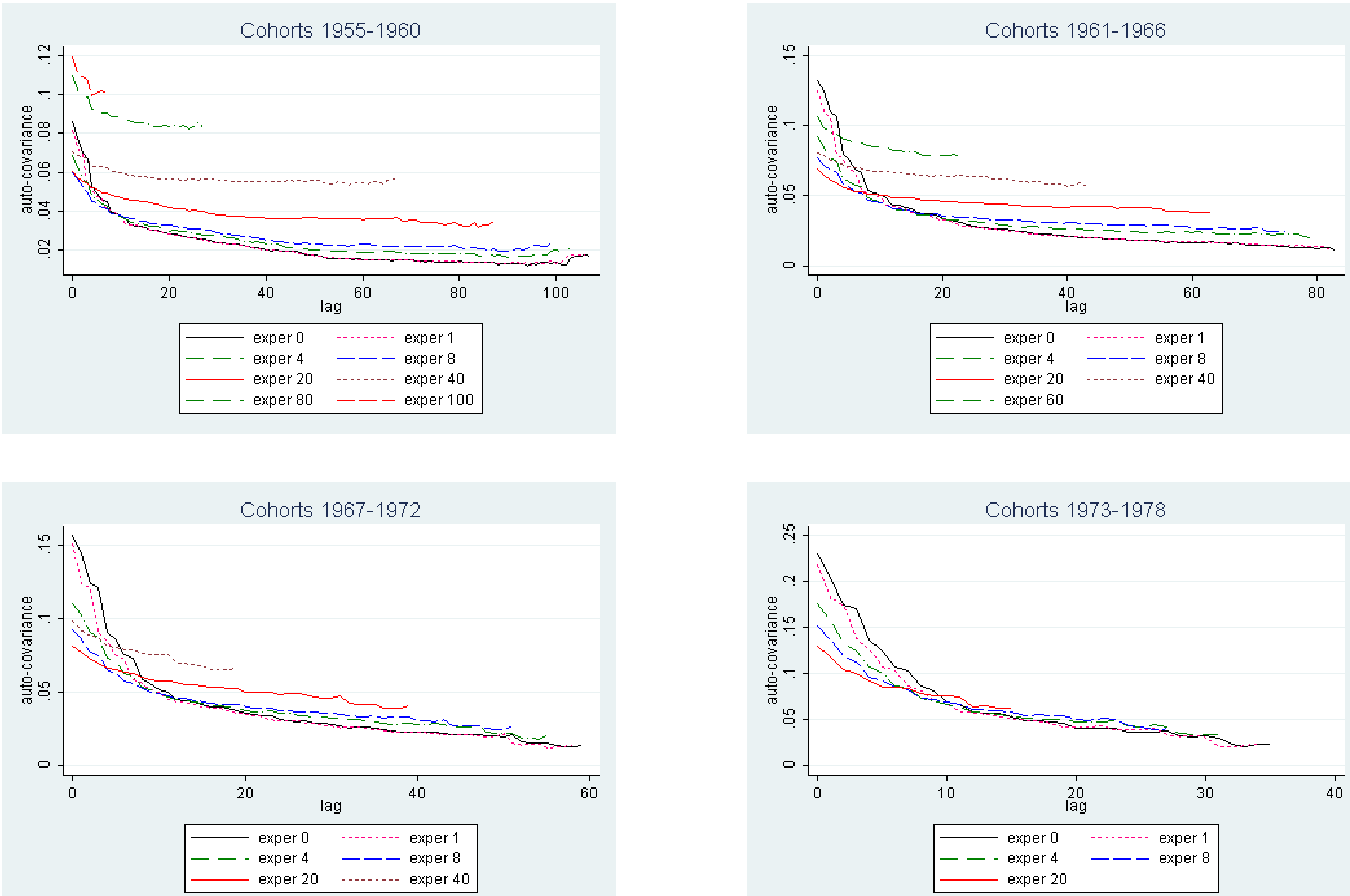
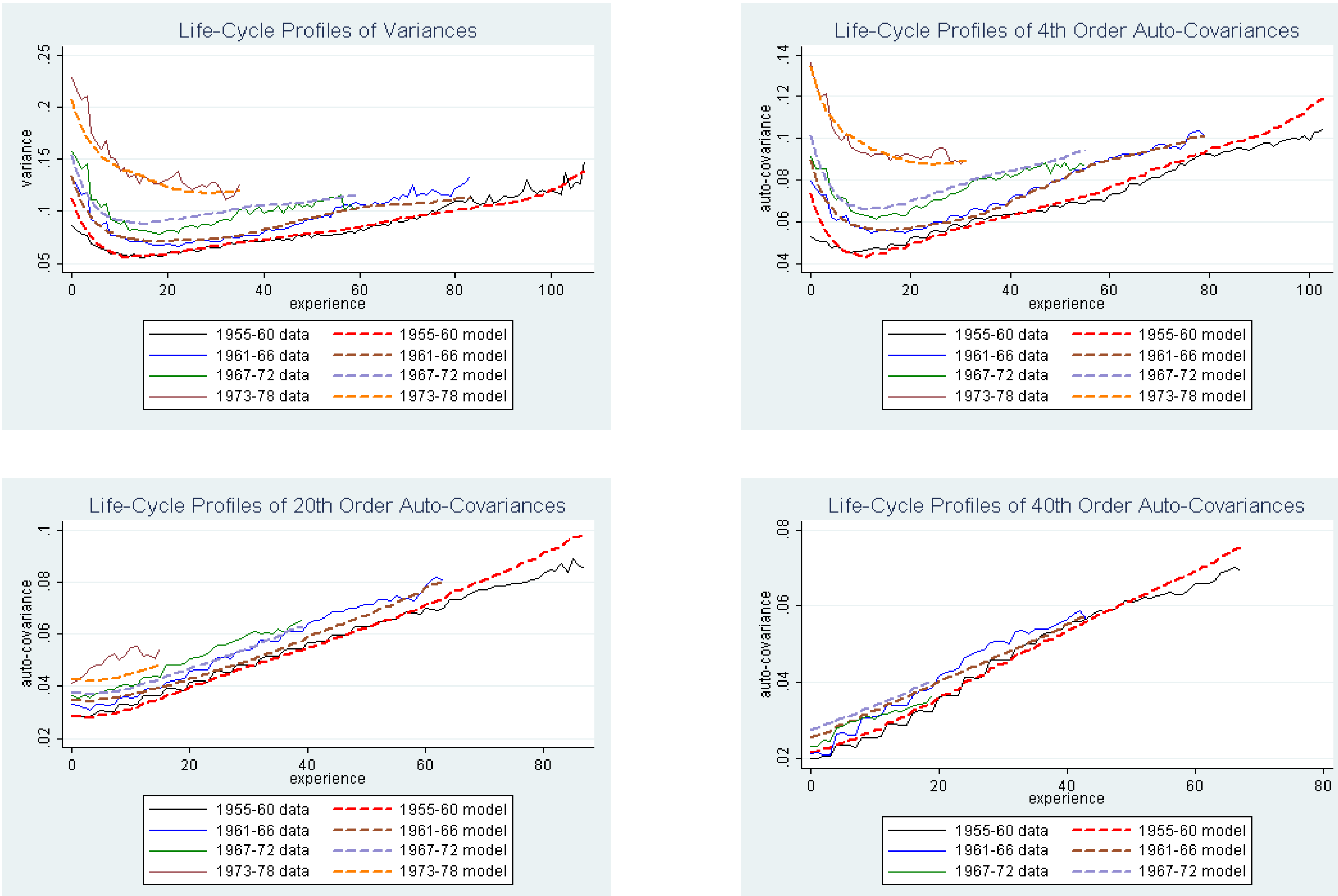


FIGURE 3 - FIT OF BENCHMARK MODEL: SECONDARY DEGREE GROUP



ONLINE 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, tranining status etc. for this group of employees. Reported earnings are gross earnings after the deducation 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 reporterd 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 disaggretation 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 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

[1]For example, a spell that takes one year, starting on January 1st and ending on December 31st, 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.

[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.

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 auto-covariances 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 et al. (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) show that it is only earnings in the upper percentiles of the cross-sectional 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 .0013 with a standard deviation of .002. The R-squared is .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 -.039 with a standard deviation of .016, implying that there is a significant discontinous *decrease* in measured variances in years after the structural break. However, an R-squared of .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 .0018 with a standard deviation of .004 and .0002 with a standard deviation of .012, respectively. In both cases, the specification can explain over 80 percent 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.

[4] Dustmann et al. (2009) perform 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) adopt the same methodology to their data. Haider (2001), estimating earnings processes from the PSID, uses a static imputation/interpolation procedure as well for a subset of censored observations.

[5] I use a 6th-order polynomial as coefficients on higher-order terms are insignificant.

B. Standard Errors

In this section of the 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 $C^{vec} = G(\theta_0, Z)$, 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 $G\left(\widetilde{\theta}, Z\right)$ at $\widetilde{\theta} = \theta_0$. Then the asymptotic distribution of the EWMD-estimator is $N\left(\theta_0, \frac{1}{\sqrt{N}} V_{\theta_0}\right)$, with

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

Here, Ω is the asymptotic covariance matrix of $\widehat{C}^{vec}$, which, since $\widehat{C}^{vec}$ are autocovariances, can be interpreted as a matrix of forth-order moments of residual log-wages. This matrix has size $\left[\dim(\widehat{C}^{vec})\right]^2$ 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 (B.1) to obtain an estimate of V_{θ_0}. Unfortunately, when relying 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 $\chi^{vec} = \left(\widehat{C}^{vec} - C^{vec}\right)$, which has the same covariance matrix Ω as $\widehat{C}^{vec}$. Its sample analogue is the vector of residuals $\widehat{\chi}^{vec}$ with generic element $\widehat{\chi}_{btk}$. From the theory of non-linear 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 $\widehat{\chi}^{vec}$ 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 $V\left(\widehat{\theta}\right)$ 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 $V\left(\widehat{\theta}\right)$. 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 $C^{vec} = G(\theta_0, Z)$. 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 $\widehat{C}^{vec}$ 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.

C. Formal Discussion of Implication 2

In this section of the 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

$$P_{ibt} = p_t * [\alpha_i + \beta_i * h_{bt} + u_{ibt}]\,. \tag{C.1}$$

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

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

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

$$\begin{aligned} var\,(u_{ibt}) &= \widetilde{\sigma}_{u_0}^2 + \sum\nolimits_{\tau=0}^{t-1} var\,(\nu_{ibt-\tau}) \\ &= \widetilde{\sigma}_{u_0}^2 + \sum\nolimits_{\tau=0}^{t-1} \sum\nolimits_{j=0}^{J_\nu} (h_{bt-\tau})^j * \delta_j \\ &= \widetilde{\sigma}_{u_0}^2 + \sum\nolimits_{j=0}^{J_\nu} \delta_j * \sum\nolimits_{\tau=0}^{t-1} (h_{bt-\tau})^j \\ &\equiv \widetilde{\sigma}_{u_0}^2 + f^u\,(h_{bt}, \delta_0, ..., \delta_{J_\nu})\,. \end{aligned} \tag{C.3}$$

Here, potential labor market experience $h_{bt-\tau}$ takes integer values in $\{1, ..., t - t_0(b)\}$. Hence, the term $\sum\nolimits_{\tau=0}^{t-1} (h_{bt-\tau})^j$ 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 $f^u\,(h_{bt}, \delta_0, ..., \delta_{J_\nu})$ is a polynomial of degree $(J_\nu + 1)$, intercept excluded, and that it is linear in the parameters $(\delta_0, ..., \delta_{J_\nu})$. Hence, δ_j is the coefficient on a term that varies on the order of $(h_{bt})^{j+1}$.

Similarly, the persistent component

$$z_{ibt} = \rho * z_{ib,t-1} + \lambda_t * \xi_{ibt} \tag{C.4}$$

solves recursively for

$$z_{ibt} = \rho^{t-t_0(b)} * z_{ib,t-t_0(b)} + \sum\nolimits_{k=0}^{t-t_0(b)-1} \rho^k * \lambda_{t-k} * \xi_{ibt-k}. \tag{C.5}$$

Thus,

$$\begin{aligned} Var\left(z_{ibt}\right) &= \rho^{2*(t-t_0(b))} * \left(\lambda_{t_0(b)}\right)^2 * \sigma_{\xi_0}^2 + \sum\nolimits_{\tau=0}^{t-t_0(b)-1} \rho^{2*\tau} * \left(\lambda_{t-\tau}\right)^2 * var\left(\xi_{ibt-\tau}\right) \\ &= \rho^{2*h_{bt}} * \left(\lambda_{t_0(b)}\right)^2 * \sigma_{\xi_0}^2 + \sum\nolimits_{\tau=0}^{h_{bt}-1} \rho^{2*\tau} * \left(\lambda_{t-\tau}\right)^2 * \sum\nolimits_{j=0}^{J_\xi} \left(h_{bt}\right)^j * \gamma_j \\ &= \rho^{2*h_{bt}} * \left(\lambda_{t_0(b)}\right)^2 * \sigma_{\xi_0}^2 + \sum\nolimits_{j=0}^{J_\xi} \gamma_j * \sum\nolimits_{\tau=0}^{h_{bt}-1} \rho^{2*\tau} * \left(\lambda_{t-\tau}\right)^2 * \left(h_{bt}\right)^j. \end{aligned} \tag{C.6}$$

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 $cov\left(P_{ibt}, P_{ibt+k}\right)$ as written in (C.2) 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 $\widetilde{\sigma}_{u_0}^2$. The coefficients on these variables are, in order, $\widetilde{\sigma}_\alpha^2$, $\sigma_{\alpha\beta}$, σ_β^2, the set of parameters $(\delta_0, ..., \delta_{J_\nu})$, and $\widetilde{\sigma}_{u_0}^2$. The parameters $\widetilde{\sigma}_\alpha^2$ and $\widetilde{\sigma}_{u_0}^2$ 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 (C.2) as $(p_t * p_{t+k}) * \sigma_\alpha^2$, where $\sigma_\alpha^2 = \widetilde{\sigma}_\alpha^2 + \widetilde{\sigma}_{u_0}^2$. 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 $cov\left(P_{ibt}, P_{ibt+k}\right)$ 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

$$\Xi_{ibt} = \varphi_t * \varepsilon_{ibt}. \tag{C.7}$$

Its contribution to the covariance structure is

$$cov\left(\Xi_{ibt}, \Xi_{ibt+k}\right) = 1(k=0) * \varphi_t^2 * \left(\sum\nolimits_{j=0}^{J_\varepsilon} h_{bt}^j * \phi_j\right). \tag{C.8}$$

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 $(\phi_0, ..., \phi_{J_\varepsilon})$. Because of the presence of a constant term ϕ_0 that is interacted with time fixed effects in both, (C.8) and (C.2), 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 $cov\left(P_{bt}, P_{bt+k}\right)$, 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 (C.2), the interaction terms in (C.8) are neither. Hence, the rank condition applied to the sum of (C.2) and (C.8) is satisfied. It must thus be true that any failure of the rank condition, if any, comes from the persistent component (C.6). This component is given by

$$\begin{aligned} cov\left(z_{ibt}, z_{ibt+k}\right) &= \rho^k * Var\left(z_{ibt}\right) \\ &= \rho^{(2*h_{bt}+k)} * \left(\lambda_{t_0(b)}\right)^2 * \sigma^2_{\xi_0} + \sum\nolimits_{j=0}^{J_\xi} \gamma_j * S_j\left(h_{bt}, k, t\right), \end{aligned} \tag{C.9}$$

where

$$S_j\left(h_{bt}, k, t\right) = \sum\nolimits_{\tau=0}^{h_{bt}-1} \rho^{(2*\tau+k)} * \left(\lambda_{t-\tau}\right)^2 * \left(h_{bt}\right)^j. \tag{C.10}$$

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 (C.6) as a difference equation and write the solution, which exists as long as $\rho \in (0, 1)$, as $V_Z\left(h, k\right)$. 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 $V_Z\left(h, k\right)$ cannot be a polynomial. Indeed, any such solution must satisfy equation (C.6), which cannot be written in polynomial form as long as $\rho < 1$. This remains true if $\sigma^2_{\xi_0} = 0$, implying that the function $S_j\left(h, k\right)$ is not spanned by the space of polynomials in (h, k) of order $J = \max\left\{J_\varepsilon, J_\nu\right\}$ 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 $cov\left(P_{ibt}, P_{ibt+k}\right)$ and $cov\left(\Xi_{ibt}, \Xi_{ibt+k}\right)$. 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 non-stationary 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 $\lambda_{\left(\underline{t_0}+1\right)} = 1$ 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 $cov\left(z_{ibt}, z_{ibt+k}\right) = \rho^k * Var\left(z_{ibt}\right)$ the lag profiles for these cohorts impose restrictions on ρ, but not on any parameters in $Var\left(z_{ibt}\right)$. As a consequence, without the restriction $\lambda_{\left(\underline{t_0}+1\right)} = 1$ 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 $\left(\underline{t_0} + 1\right)$, 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 non-polynomially 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 $cov\,(z_{ibt}, z_{ibt+k})$ 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 non-linear 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 $cov\,(z_{ibt}, z_{ibt+k}) = Var\,(z_{ibt})$ 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 $t \in \left\{\underline{t_0}, \underline{t_0} + J_\xi + 2\right\}$. In this case, there are sufficiently many periods where variation in $Var\,(z_{ibt})$ 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, $cov\,(u_{ibt}, u_{ibt+k}) = p_t {*} p_{t+k} {*} var\,(u_{ibt})$ 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 $var\,(u_{ibt})$.

D. Constrained Optimization and Computational Issues

The MD-estimator does not impose any non-negativity constraints on the estimates of variance parameters such as σ_β^2. 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 $var(\nu_{ibt})$ and $var\,(\xi_{ibt})$ evaluated at any age are restricted to be non-negative. The MD-estimator therefore becomes the solution of a constrained minimization problem for which the contraints 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 re-estimate the model.

[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.

E. How Robust are the Conclusions? Results from the Dropout Sample

In this 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 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 mis-specified model.

Parameter estimates for various specifications are shown in 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 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 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 $\left(\sigma_\beta^2, \sigma_{\alpha\beta}\right)$ 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. 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 $\left(\sigma_\beta^2, \sigma_{\alpha\beta}\right)$. 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

[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 inter-cohort 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.

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.

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 towards 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 $1,000$ 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 $1,000$ 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 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 non-linear 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 appendix table 5. I simulate three different models, shown in three different sub-panels. Each sub-panel 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 $1,000$ 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 $\left(\sigma_\beta^2, \sigma_{\alpha\beta}\right) = (0,0)$, 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 $\left(\sigma_\alpha^2, \sigma_\beta^2, \sigma_{\alpha\beta}\right)$ and the persistence and the initial condition of the AR(1)-process $\left(\rho, \sigma_{\xi_0}^2\right)$. The sampling distribution of their $1,000$ parameters estimates are summarized in two statistics, the average bias and the standard deviation.

The three sub-panels 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: $\left(\sigma_\beta^2, \sigma_{\alpha\beta}\right) = (0,0)$. As shown in columns 4 and 5, there are no significant biases no matter if this restriction is imposed in the estimation (col. 5) or not (col. 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_β^2 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 $\left(\sigma_\beta^2, \sigma_{\alpha\beta}\right)$ 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 alltogether 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 towards 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 .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.

APPENDIX TABLE 1 - SAMPLE SIZES BY EDUCATION GROUP, COHORT AND EXPERIENCE

PANEL A: SAMPLE SIZES BY EDUCATION GROUP AND COHORT

COHORT	EDUCATION GROUP — ***Main Sample:*** Secondary Degree Group	***Robustness Sample:*** High School Dropouts
1955	250,387	-
1956	258,708	-
1957	315,867	31,810
1958	306,775	29,055
1959	315,022	31,763
1960	303,563	27,712
1961	297,691	25,912
1962	286,116	27,374
1963	289,492	26,231
1964	278,860	27,494
1965	263,462	23,218
1966	243,002	20,521
1967	226,214	18,152
1968	207,863	18,576
1969	179,188	13,766
1970	150,000	13,882
1971	128,949	12,101
1972	105,640	11,328
1973	79,753	9,582
1974	72,457	9,464
1975	63,300	8,260
1976	53,368	9,045
1977	43,998	9,540
1978	32,612	9,445
TOTAL	4,752,287	414,231

PANEL B: SAMPLE SIZES BY EDUCATION GROUP AND EXPERIENCE

EXPERIENCE (IN YEARS)	EDUCATION GROUP — ***Main Sample:*** Secondary Degree Group	***Robustness Sample:*** High School Dropouts
0	322,907	34,966
1	317,001	29,831
2	312,600	25,600
3	311,899	23,824
4	302,277	23,107
5	290,854	22,560
6	279,578	22,014
7	267,364	21,312
8	255,348	20,023
9	242,766	18,918
10	228,805	17,816
11	214,144	16,877
12	199,207	15,938
13	183,078	15,171
14	166,489	14,243
15	150,119	13,224
16	134,430	12,309
17	118,782	11,422
18	103,754	10,345
19	89,320	9,472
20	75,806	8,482
21	62,466	7,425
22	50,103	6,318
23	37,565	5,318
24	26,427	4,328
25	15,797	-
26	7,485	-
	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_{α}^{2}	0.021	0.012
Slope Heterogeneity	$\sigma_{\beta}^{2} * 10^{3}$	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)	$\sigma_{\xi 0}^{2}$	0.078	0.072
Permanent Shocks	$\delta_{0} * 10$	0.024	0.015
Number of Moments		**3,644**	

NOTES: 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 SPECIFICATION			
		(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_α^2	0.027 (0.007)***	0.134 (0.011)***	0.041 (0.003)***	0.039 (0.013)***	0.068 (0.009)***	0.111 (0.019)***	0.134 (0.01)***
Slope Heterogeneity	$\sigma_\beta^2 * 10^3$	0.000 -	0.026 (0.003)***	-	0.004 (0.002)**	0.008 (0.001)***	0.018 (0.005)***	0.027 (0.003)***
Cov (Intercept; Slope)	$\sigma_{\alpha\beta}^2 * 10$	-0.0004 (0.0001)	-0.016 (0.002)***	-	-0.003 (0.0015)*	-0.005 (0.003)***	-0.011 (0.002)***	-0.016 (0.002)***
Persistence of AR(1)	ρ	0.883 (0.009)***	0.784 (0.027)***	0.808 (0.013)***	0.921 (0.006)***	0.837 (0.012)***	0.828 (0.015)***	0.756 (0.027)***
AR(1) error structure								
Initial Condition	$\sigma_{\xi 0}^2$	0.283 (0.032)***	-	-	0.273 (0.035)***	0.544 (0.056)***	-	-
Intercept	γ_0	0.044 (0.005)***	0.032 (0.017)*	0.039 (0.005)***	0.003 (0.001)***	0.103 (0.008)***	0.071 (0.014)***	0.038 (0.003)***
experience	γ_1	-0.003 (4*e(-4))***	-	-	-	-0.007 (0.001)***	-0.005 (0.001)***	-
experience^2	γ_2	7.78*e(-5) (9.27*e(-6))***	-	-	-	1.67*e(-4) (5.47*e(-5))***	1.43*e(-4) (3.03*e(-5))***	-
experience^3	γ_3	-8.65*e(-7) (1.05*e(-7))***	-	-	-	-1.93*e(-6) (8.01*e(-7))***	-1.64*e(-6) (3.58*e(-7))***	-
experience^4	γ_4	3.48*e(-9) (4.29*e(-10))***	-	-	-	8.22*e(-9) (3.86*e(-9))**	6.73*e(-9) (1.5*e(-9))***	-
Variance of Permanent Shocks	$\delta_0 * 10$	0.003 (0.001)**	-	-	0.000 -	0.002 (0.002)	0.001 (0.002)	0.000 -
Variance of Measurement Error	ϕ_0	0.000 -	0.004 (0.003)	0.000 (0.003)	0.023 (0.004)***	0.000 (0.001)	0.000 -	0.000 -
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

NOTES: 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). *** Significant on 1%-level; ** Significant on 5%-level; * Significant on 10%-level. 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

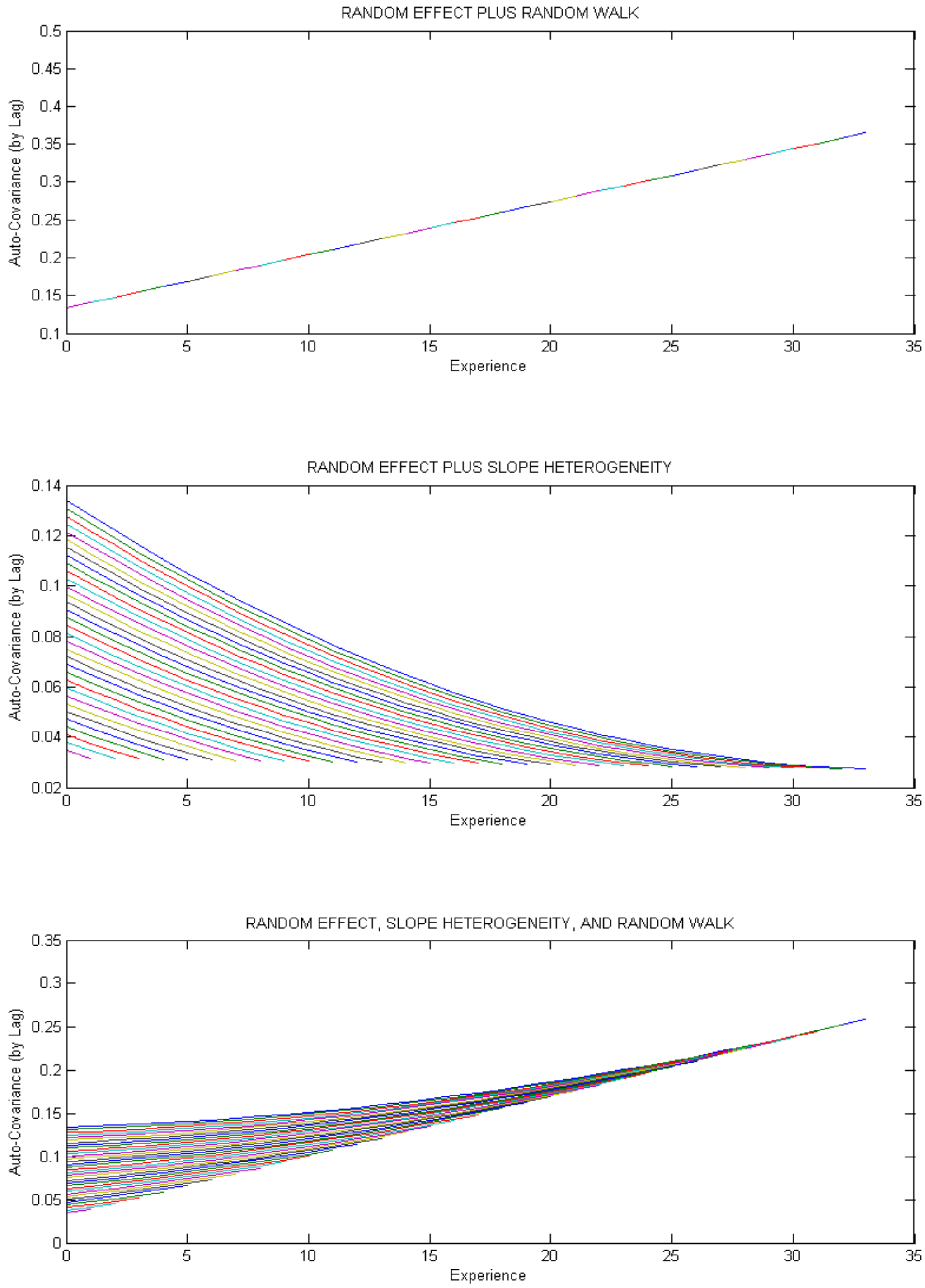
	Heterogeneity			AR(1)-Process							Unit Roots	Measurement Error
Parameter: Description	Intercept	Slope	Covariance	Persistence	Initial Condition	4th-Order Polynomial in Experience						
Parameter: Notation	σ_{α}^{2}	σ_{β}^{2}	$\sigma_{\alpha\beta}$	ρ	$\sigma_{\xi 0}^{2}$	γ_0	γ_1	γ_2	γ_3	γ_4	δ_0	ϕ_0
True 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**						

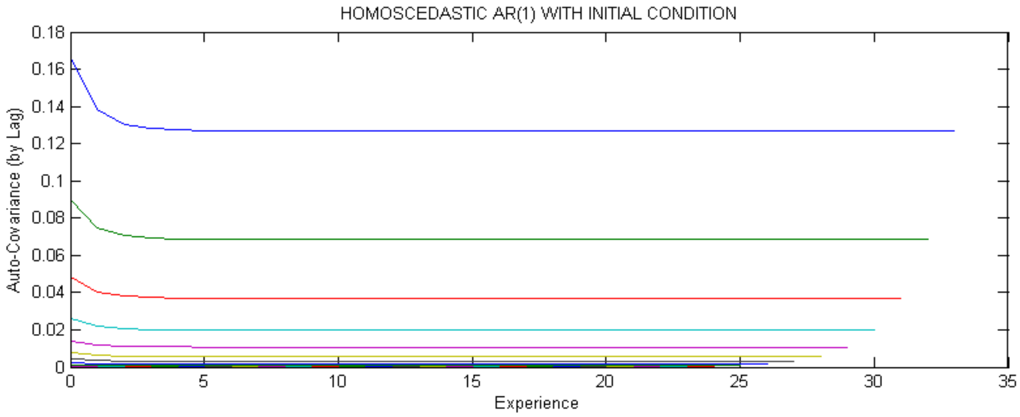
PANEL B: SIMULATION RESULTS

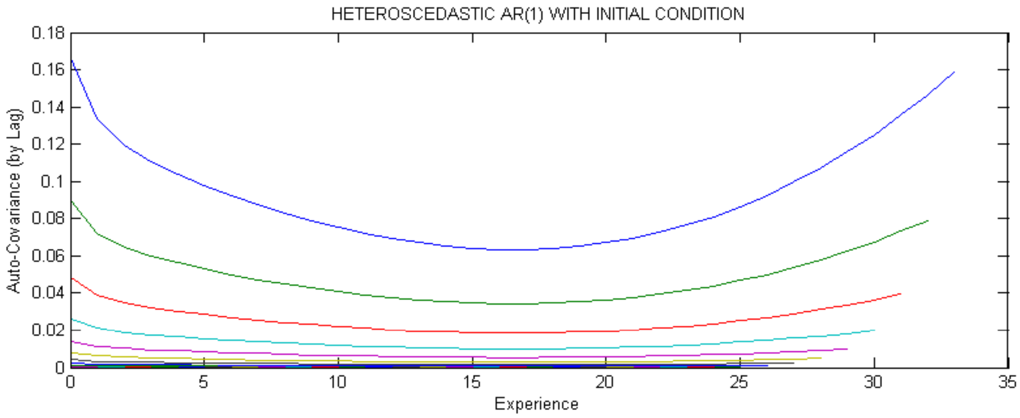
			Simulation 1: No Restrictions on Parameters			Simulation 2: $\sigma_{\beta}^{2}=\sigma_{\alpha\beta}=0$			Simulation 3: $\sigma_{\xi 0}^{2}=0$		
			(1.)	(2.)	(3.)	(4.)	(5.)	(6.)	(7.)	(8.)	(9.)
			Estimation: Parameter Restrictions			*Estimation: Parameter Restrictions*			*Estimation: Parameter Restrictions*		
			None	$\sigma_{\beta}^{2}=\sigma_{\alpha\beta}=0$	$\sigma_{\xi 0}^{2}=0$	**None**	$\sigma_{\beta}^{2}=\sigma_{\alpha\beta}=0$	$\sigma_{\xi 0}^{2}=0$	**None**	$\sigma_{\beta}^{2}=\sigma_{\alpha\beta}=0$	$\sigma_{\xi 0}^{2}=0$
Parameter Estimates	σ_{α}^{2}	**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_{β}^{2}	**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
	$\sigma_{\xi 0}^{2}$	**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

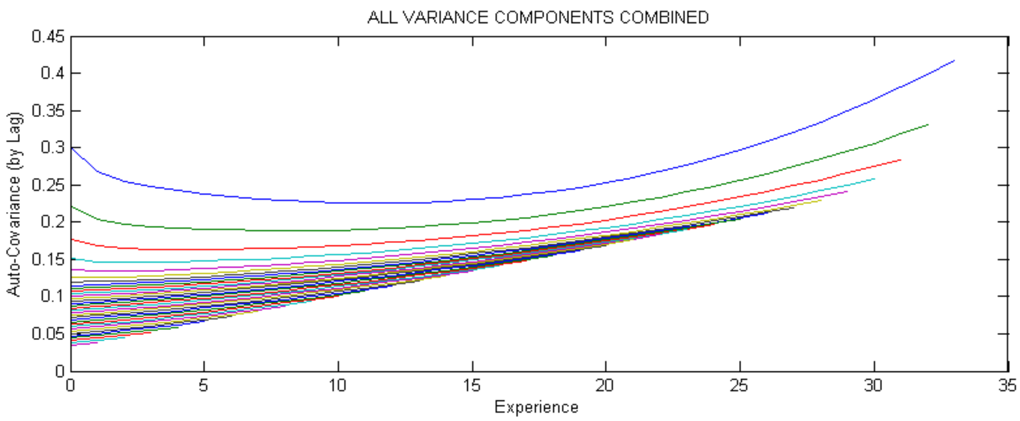
NOTES: This table shows results from Monte Carlo Simulations of the benchmark model without time effects. Panel A lists the parameter values, initial values in the optimization routine used for estimation, and the number of Monte-Carlo repetitions held constant across simulations, unless noted otherwise. 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 1,000 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



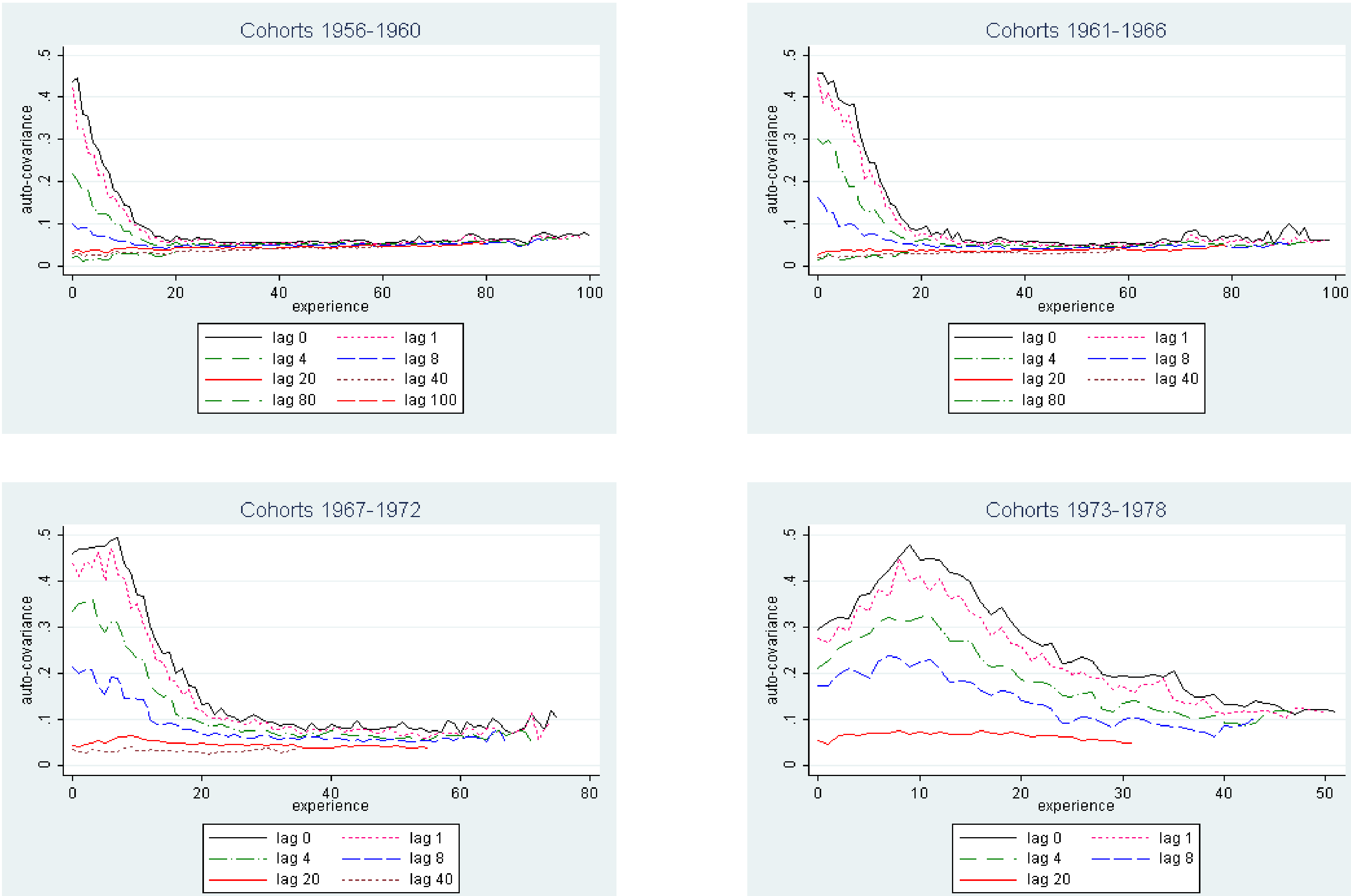

HOMOSCEDASTIC AR(1) WITH INITIAL CONDITION
Auto-Covariance (by Lag)
Experience
0
0.02
0.04
0.06
0.08
0.1
0.12
0.14
0.16
0.18
0
5
10
15
20
25
30
35


HETEROSCEDASTIC AR(1) WITH INITIAL CONDITION
Auto-Covariance (by Lag)
Experience
0
0.02
0.04
0.06
0.08
0.1
0.12
0.14
0.16
0.18
0
5
10
15
20
25
30
35


ALL VARIANCE COMPONENTS COMBINED
Auto-Covariance (by Lag)
Experience
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0
5
10
15
20
25
30
35

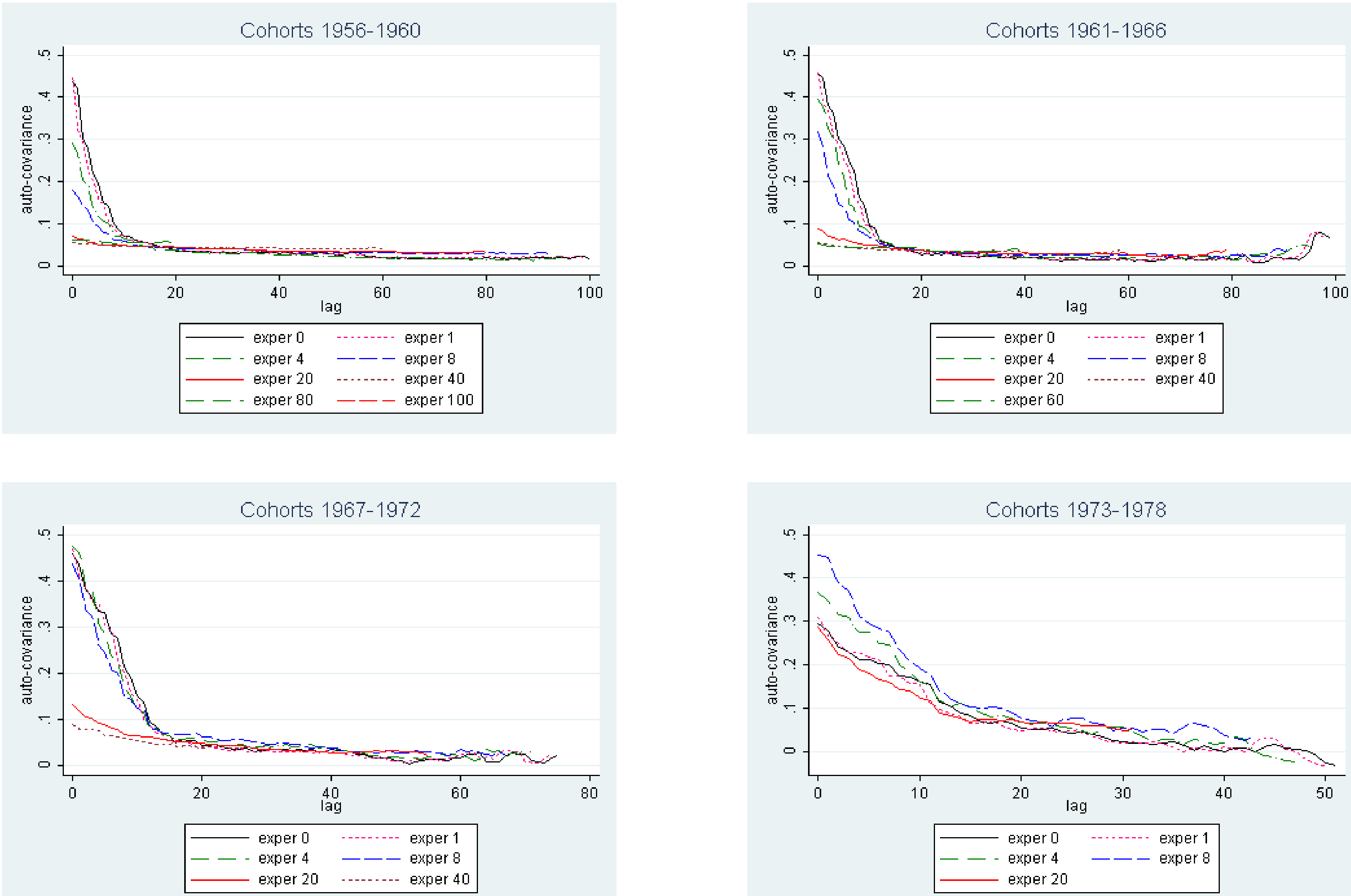
APPENDIX FIGURE 2 - LIFE-CYCLE PROFILES OF AUTO-COVARIANCES AT DIFFERENT LAGS, BY COHORTS

Sample: Dropout Group

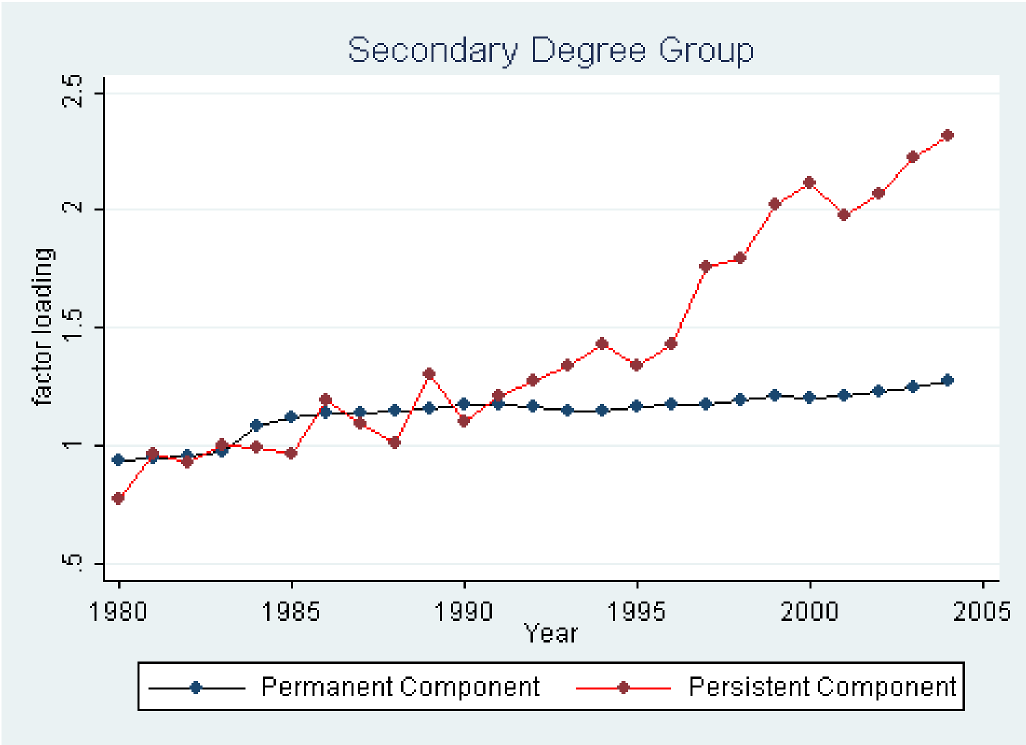


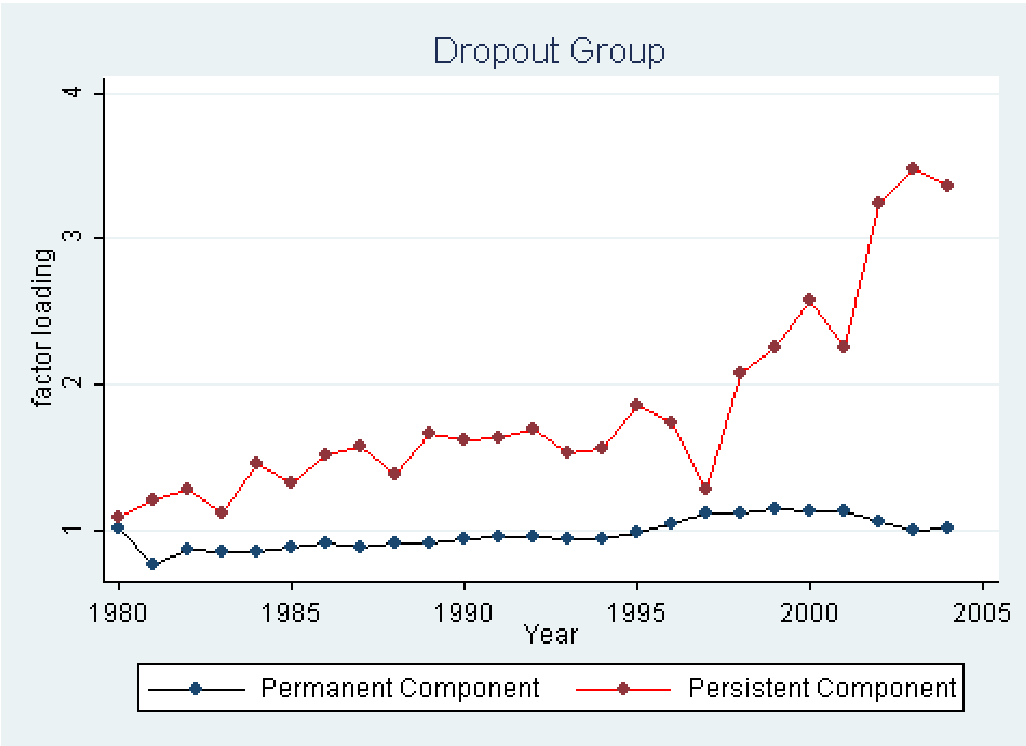
APPENDIX FIGURE 3 - LAG-PROFILES OF AUTO-COVARIANCES FOR DIFFERENT EXPERIENCE GROUPS, BY COHORTS

Sample: Dropout Group



APPENDIX FIGURE 4 - ESTIMATED FACTOR LOADINGS FOR THE FULL MODEL





APPENDIX FIGURE 5 - FIT OF BENCHMARK MODEL: DROPOUT GROUP