HIP, RIP and the Robustness of Empirical Earnings Processes *

Florian Hoffmann
Vancouver School of Economics
University of British Columbia

Version: April 2013

Abstract

Parameter estimates from earnings processes are key inputs into life-cycle models with heterogeneous agents, but it remains an open question how to model earnings dynamics appropriately. In this study I interpret the wide range of estimates of key parameters in the literature obtained from the same data, such as the variance of individual heterogeneity or persistence of income shocks, as reflecting a fundamental misspecification problem in two commonly used families of models – HIP- and RIP-models. I show that to obtain credible estimates of these parameters it is crucial to control flexibly for age- and time effects in innovation variances, including a rich specification of initial conditions. Starting from a model that is well-specified and that nests HIP- and RIP-models, I investigate the robustness of key parameters across specifications. To isolate the model-specific identifying variation of a parameter, I compare across specifications the results from novel numerical comparative statics that perturb the parameter around its estimated value. Since identification of my preferred model requires covariance structures that are disaggregated to the cohort-level, I rely on administrative social-security data from Germany on quarterly earnings that follow workers from labor market entry until 27 years into their career. 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 (i) estimates of key parameters fluctuate widely across specifications, (ii) permanent and persistent shocks as well as intercept-heterogeneity are always significant while transitory shocks are not, (iii) a persistent initial condition matches the complex earnings dynamics early in the life-cycle, (iv) slope-heterogeneity is highly significant in a standard HIP-process but vanishes once one controls for age-effects appropriately and (v) slope-heterogeneity introduces a problem of "over-fitting". These results are unchanged when I allow slopes to vary over the life-cycle and when I estimate the model from an education group with a drastically different covariance structure.

*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 and continuous support throughout this project. I also thank Michael Baker, David Green, Gueorgui Kambourov, Thomas Lemieux, Jean-Marc Robin, Aloysius Siow 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. Financial support from Philip Oreopoulos, Shouyong Shi's Canada Research Chair Fund, the CLSRN fellowship, the University of Toronto Travel Grant and the IAB is gratefully acknowledged. All remaining errors are mine.

1. Introduction

As heterogeneous agents life-cycle models have become a standard empirical tool in a wide range of economic research such as labor economics, macroeconomics, economics of education and public finance, there has been a resurgance of interest in the estimation of earnings processes. Econometric models of earnings dynamics are used to quantify the role of skill heterogeneity relative to the role of exogenous shocks with different persistences in shaping individual life-cycle earnings trajectories. Parameter estimates serve as crucial inputs into structural life-cycle models of labor market dynamics and determine the extent to which earnings fluctuations map into wealth or consumption inequality and which types of insurances can arise endogenously in decentralized markets.[1] It remains an open question of how to model earnings processes appropriately.[2] Estimates of the same parameters, such as the variance of individual heterogeneity or persistence of income shocks, vary widely across studies that rely on the same data but postulate different econometric models. While it is commonly accepted that intercept heterogeneity and shocks with *some* persistence are important, there is no agreement about the quantitative importance of heterogeneity in the returns to labor market experience and the actual persistence of shocks.

Two prominent models are HIP- and RIP-processes, the former of which usually finds significant slope heterogeneity and moderate persistence of shocks while the latter allows for intercept heterogeneity only and commonly finds that earnings evolve according to a unit-roots process.[3] Recent research by Guvenen (2009) and Hryshko (2012) suggests that the source of disagreement in the literature is the difficulty to empirically distinguish between these hypotheses. In this paper I interpret these results as a reflection of a more fundamental problem, that is estimates of key parameters are not robust to model-misspecification, implying that minor modifications in the empirical processes may result in very different conclusions about the quantitative role of risk and heterogeneity. Omission of a relevant variance component may thus lead to severe biases in all model estimates. For example, the commonly observed convexity of life-cycle variance profiles can be interpreted as evidence in favor of slope-heterogeneity, but it can also be generated by age-dependence of risk. Hence, abstracting from one component may bias the estimates of the other, a type of omitted variable bias in non-linear models.

My study makes two main contributions to the vast literature on earnings dynamics. First, starting from an empirically well-specified model that nests RIP- and HIP-processes and subsequently estimating more restrictive models commonly used in the literature, I conduct the first systematic study of robustness of model parameters across specifications. Second, I explore the common underlying factors that are driving the sensitivity of parameter estimates. To this end I conduct numerical comparative statics exercises that perturb a set of parameters

[1]Recent examples for 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 (2007) and Low, Meghir and Pistaferri (2010) for consumption, Abbott, Gallipoli, Meghir and Violante (2013) for education, Erosa, Kambourov and Fuster (2011) for labor supply, Huggett and Kaplan (2012) for human capital, and Farhi and Werning (2012) 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.

[2]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]"HIP" stands for "Heterogeneous Income Profiles" and "RIP" stands for "Restricted Income Profiles". These labels were introduced by Guvenen (2007, 2009).

around their estimates while holding everything else constant. Carrying out these exercises for each specification helps isolate the *model-specific* identifying variation for a particular set of parameters. If this variation is not found to depend on the specification, parameter estimates should be expected to be robust. Given its prevalence in the literature, I restrict my investigation of robustness to the family of parametric earnings processes that can be point-identified from autocovariance matrices without any distributional assumptions.

Persistence of shocks is one of the central objects of interest in the study of life-cycle income dynamics as it determines the insurability of individual-level income fluctuations and thus the relationship between income-, life-cycle- and consumption-inequality. As reflected by the distinction between HIP- and RIP-processes, it has become customary to estimate persistence from an ARMA(1,q)-process and to substitute the AR(1)-component by a unit-roots process if one cannot reject non-stationarity. Recent research by Baker and Solon (2003) and Hryshko (2012) show this distinction to be arbitrary as one can identify both a unit-roots process and a process with moderate persistence simultaneously. I adopt this approach and consider models that decompose earnings into a component that reflects observed and unobserved heterogeneity and a component that reflects risk and allows for permanent, persistent and transitory shocks. Furthermore, following a large literature that emphasizes the need to control for age- and time-effects when studying the first moments of life-cycle income dynamics, I allow innovation variances of the model components to directly depend on labor market experience and calendar time.[4] The resulting model is rich enough to match all features of the autocovariance structure in my data up to sampling error, thus minimizing biases in parameter estimates potentially plaguing more restrictive specifications, while being parsimonious enough to be applicable to structural heterogeneous agent modeling.

In the presence of age- and time-effects in second moments, identification requires moments of life-cycle earnings dynamics that are disaggregated to the cohort-level. Publicly available survey panel data such as the PSID have relatively few observations per cohort-age cell, forcing the researcher to aggregate. I therefore rely on a sufficiently large administrative data set from Germany that follows individuals from time of labor market entry up until 27 years into their careers and that allows me to generate quarterly rather than annual panels. A further advantage of these data is that a worker's education is observed, in contrast to administrative data from the US or Canada. As a consequence, I can construct education-specific autocovariance matrices of earnings for which the life-cycle is defined by actual, rather than potential experience, thus avoiding a common incidental parameter problem that arises if individuals are not observed from labor market entry on or if one needs to aggregate over education groups.

I find that the largest education group in the German labor market displays an autocovariance structure of labor market earnings that shares the main qualitatively features of the North American counterpart, most importantly those that are commonly used to identify slope-heterogeneity and the persistence of shocks. To strenghten external validity of my results, I therefore focus my analysis on this education group. My preferred

[4]Examples for papers that dicuss the importance of age- and time-effects in studies of inequality are Storesletten, Telmer and Yaron (2004b), Heathcote, Storesletten and Violante (2005) and Lemieux (2006).

model matches only 65 parameters to a covariance structure with over 56,000 elements. Yet, the model fits all features of the covariance structure, such as the evolution of variances over the life-cycle and over time, almost perfectly. This implies that a model that allows for a flexible specification in observed heterogeneity, labor market experience and time, can generate rather complex patterns of earnings dynamics. The near-perfect match of my model together with the focus of this paper also explains why I do not conduct Monte-Carlo simulations; They would merely replicate the findings from the robustness exercises while requiring distributional assumptions.

I find that most components of my preferred model are important determinants of life-cycle earnings dynamics. Both, permanent and persistent shocks have a significant impact on the evolution of earnings over the life-cycle. However, there is no evidence for transitory shocks in earnings, consistent with Baker and Solon (2003) who estimate a similar model using Canadian administrative tax data. This may suggest that what is commonly interpreted as transitory shocks in survey panel data is in fact measurement error.[5] I also find that age- and time-effects are important for matching salient features of the cohort-specific autocovariance structures. Interestingly, my estimated time effects imply that the well-documented fanning out of the German earnings distribution is almost entirely driven by an increase in the variance of the persistent rather than the permanent component.[6] Furthermore, my numerical comparative statics exercises show that a particular type of age-effect that is commonly neglected in the literature, the initial condition of the persistent component, is particularly important for fitting earnings dynamics of workers who are at an early stage of their career. In contrast, I do not find robust evidence in favor of slope heterogeneity.

My investigation of robustness uncovers a number of striking regularities. First, estimates of key parameters that are of particular interest for structural life-cycle modeling vary widely across specifications. For example, persistence of the AR-component is low in some specifications, but statistically indistinguishable from one in standard RIP-specifications. Similarly, heterogeneity in returns to experience is significant in some specifications, but not in others. Most importantly, it is highly significant in a standard HIP-model but only marginally significant in the preferred specification. Second, when estimating various nested models, I find that the exclusion of the persistent initial condition has a particularly large effect on the estimated heterogeneity in slopes. If one omits this component, the observed earnings dynamics early in the life-cycle are matched by the HIP-component, a type of omitted variable bias with a quantitatively large impact. I support this conjecture using numerical comparative statics. Third, no matter the specification I estimate, I find the significance of slope heterogeneity to depend crucially on whether I allow it to be correlated with intercept heterogeneity or not. In particular, when re-estimating all model specifications, but with the restriction that intercepts and slopes are uncorrelated, estimates of the latter tend to zero, often hitting the non-negativity constraint. This suggests that inclusion of a HIP-component may generate "overfitting" as it does not match any economically meaningful features of the

[5] The variance of transitory shocks cannot be point-identified in the presence of measurement error. See e.g. Meghir and Pistaferri (2004). Since the data used herein and those used in Baker and Solon (2003) are administrative and are the basis for the calculation of social security contributions and taxes respectively, it is reasonable to assume that measurement error is negligible.

[6] For a detailed study of trends in earnings inequality in Germany, see e.g. Dustmann, Ludsteck and Schoenberg (2009).

covariance-structure. My conclusions remain unchanged when I estimate a specification that allows individual slopes to vary over the life-cycle.

Since one may be worried that my results are an artifact of the wage structure in my sample, I repeat my analysis using data from an education group with a different autocovariance structure. The observed changes of the wage structure for this group across cohorts are too complex to be matched by my model, thus generating the interesting situation in which the benchmark specification itself is misspecified. While estimates of most parameters are statistically different from those obtained from the main sample, the investigation of robustness uncovers the same regularities. In particular, estimates of key parameters are not robust to exclusion of most model components, with the initial condition of the persistent component playing the most important role, and slope heterogeneity is not a robust feature of the autocovariance matrix.

This paper contributes to a large and growing literature that uses panel data, often the PSID, to estimate life-cycle earnings processes from empirical covariance structures of earnings. Early studies, such as Lillard and Weiss (1979) and Hause (1980) assume that individuals are not only heterogeneous with respect to their average income as measured by an individual fixed effect, but also with respect to the slope of their earnings-experience profile. MaCurdy (1982) tests this hypothesis explicitly and rejects it. As a consequence, subsequent papers in the literature impose the assumption of no profile heterogeneity a priori.[7] Baker (1997) however 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.[8]

This literature does not analyse explicitely the identifying variation for the main parameters of interest in overidentified models. Guvenen (2009) is the first to fill this gap, showing that slope heterogeneity in a standard HIP-model is identified from both, the convexity of experience profiles and the behavior of lag-profiles of autocovariances. His estimates support Baker's (1997) findings of significant slope heterogeneity and low persistence of shocks and emphasizes the upward bias in the persistence parameter when slope heterogeneity is not properly controlled for. In contrast, Hryshko (2012) argues that estimates of slope heterogeneity are not robust to the inclusion of shocks with different levels of persistence. While I follow these two studies in focussing on isolating the data features that can and should be used for identifying certain parameters in earnings processes, I emphasize the need to flexibly control for time- and age effects and non-degenerate initial conditions in the persistent component, thus building on recent findings in more descriptive analyses, such as Heathcote et al. (2005) and Heisz et al. (2012). This allows me to address identification of key parameters across a wide range of model specifications and enables me to investigate the importance of specification error when relying on more restrictive specifications. The model estimated in this paper is similar to Baker and Solon (2003), who use administrative data from Canada. The focus of their study is to quantify the role of persistent and permanent shocks in driving recent trends in residual inequality and is therefore very different from mine.[9] Furthermore,

[7]Examples are Abowd and Card (1989) and Meghir and Pistaferri (2004).

[8]Haider (2001) finds similar results in a slightly more general model.

[9]Using earnings processes to study the sources of trends in earnings dynamics has been proposed for example by Gottschalk and

their data impose various important limitations that are absent from the German data, most importantly the significantly shorter time span and the lack of educational information, the latter of which rules out constructing individual careers from labor market entry on.[10]

Given a sufficient number of observations per individual and cohort, all parameters in standard earnings processes can be identified from earnings data alone. A small literature follows a different, and potentially more powerful, approach that exploits the joint dynamics of consumption and earnings to derive overidentifying restrictions.[11] A drawback of this approach is the lack of high-quality administrative panel data that simultaneously record earnings and consumption dynamics. Computational issues when estimating Dynamic Programming impose further limitations on the types of earnings processes that can be considered. 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.

My work is also related to a small literature that attempts to gain explicit economic interpretations for earnings shocks, relying on the idea that life-cycle variation in earnings are reflections of choices. Abowd and Card (1989) study the relationship between hours and earnings changes, while Hagedorn and Manovskii (2010) quantify the amount of earnings variation explained by inter-firm mobility. Similarly, Low, Meghir and Pistaferri (2010) and Altonji, Smith and Vidangos (2013) estimate reduced-form and semi-structural selection correction models to explore the relationship between earnings changes and mobility between employment states and firms. Hoffmann (2010) and Pavan (2011) introduce earnings processes into fully structural dynamic programming models, while Huggett, Ventura and Yaron (2011) derive overidentifying restrictions from a human capital model. Again, data quality and computational tractability limits the type of earnings processes that can be considered.

My study restricts its attention to the class of parametric earnings processes that are estimated from autocovariance structures alone. This covers the large majority of specifications used in the literature and provides estimates for the parameters that are relevant as inputs into economic life-cycle models of labor market outcomes. A number of recent studies either utilize more moments than autocovariances or consider non-parametric specifications. Meghir and Pistaferri (2004), while abstracting from slope heterogeneity, persistent shocks and time effects, allow for ARCH-effects in the transitory and permanent innovations. Browning, Ejrnaes and Alvarez (2010) extend this framework and estimate processes in which the majority of parameters are random variables themselves.[12] An advantage of my preferred specification is that it keeps the size of the state-space tractable and that it focusses estimation on parameters that are relevant as inputs into quantitative heterogeneous agent models, while approximating variance dynamics over the life-cycle using a flexible specification of age effects.

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.

[10]If earnings are not observed from the time of labor market entry on, the estimated initial condition is a combination of the true initial condition and the accumulated history of shocks for those who enter the labor market at a younger age. This is likely to lead to an upward bias in the estimated role of initial conditions.

[11]Well-cited examples are Hall and Mishkin (1982), Guvenen (2007), Blundell, Pistaferri and Preston (2009), Guvenen and Smith (2010), and Heathcote, Storesletten and Violante (2012).

[12]Horowitz and Markatou (1996), Hirano (2002) and Bonhomme and Robin (2009) consider semi-parametric earnings processes.

2. Data

I use the confidential version of the IABS, a 2%-extract from German administrative social security records for the years 1975 to 2004 that is collected by the German Federal Employment Agency. The IABS 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, it is followed for the rest of the sample period. A new random sample of labor market entrants is added each year.

For the purpose of this study, using these data instead of publicly available panel data has at least five advantages. 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. Second, given the large number of observations in the sample I can construct *cohort-specific* autocovariances, enabling me to estimate models of second moments of residual earnings that allow for both age- and time-effects. This contrasts sharply with studies relying on the PSID where sample size requires 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, with large sample sizes for each education group I can perform separate analyses for each education group. Fifth, earnings records are provided by firms under a thread of legal sanctions for misreporting and can be expected not to be plagued by measurement error, in contrast to commonly used survey panel data.

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 to rule out earnings dynamics to be 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 at least once.[13] 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.[14] Since top-coded earnings observations contain valid information, namely that an individual has a large positive earnings residual relative to the comparison group, I follow Haider in

[13]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.

[14]"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.

using an imputation procedure rather than dropping these observations.[15][16] Third, I use a novel and important sample restriction that only keeps individual labor market careers observed from labor market entry on and therefore avoids an incidental parameters problem.[17] Since earnings histories are left-censored in 1975 I drop individuals who are observed in that year.[18] 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.[19] 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. A consequence of this restriction is that there are more than two cohorts observed for each experience level. Further details of sample construction are given in the appendix.

Sample Sizes Sample sizes for the two education groups and for each cohort are reported in panel A 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 youngest cohort in the dropout group is born in 1957 and enters the labor market in 1976. The youngest cohort in the other education group is born in 1955 and enters the labor market in 1978. In total there are 414,231 income observations for the first and 4,752,287 income observations for the second education group. Panel B reports sample sizes by experience in years instead. Approximately 35 thousand individuals with no degree are observed from their first year in the labor market on, compared with 323 thousand individuals for the other education group. Half of these entrants are still observed after 11 years for the first and 14 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 at all, 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.

[15] A comparison group is defined by age, year of birth, and education. The average earnings are below the contribution limit for any comparison group in the sample.

[16] 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, exactly those moments that provide important identification variation for the parameters. Most importantly, as it compresses the wage structure artificially, it is 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.

[17] It is this sample restriction that avoids earnings data to be affected by the structural break in 1984 as documented by, e.g. Steiner and Wagner (1998). See the appendix for a discussion.

[18] 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.

[19] These are 19 years for high school dropout and 23 years for those with a formal secondary degree.

3. Descriptive Analysis

In this section I provide a detailed graphical analysis of the autocovariance structure and its evolution across cohorts, a helpful first step to detect the main empirical regularities to be explained by the model. Corresponding empirical first moments of log-labor income by education and labor market experience are listed in appendix table 2. Similar to the US, log-earnings profiles are concave in labor market experience and strictly monotone in education. Growth rates in earnings differ significantly across the two education groups. High-school dropouts start their career with very low earnings which are more than doubled after 24 years of experience. Most of this growth takes place over the first 5 years. In contrast, earnings of those with a secondary degree grow gradually by approximately 55 percent over the first 24 year of labor market experience.[20]

In the analysis of the covariance structures of residual earnings, I adopt the approach in Baker and Solon (2003) and consider econometric models of demeaned log-earnings. I denote log-earnings in period t of individual i born in year b in education group e by y^e_{ibt} and assume that they follow the education-specific process

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

where μ^e_{bt} is a set of education specific cohort-time fixed effects, and $\widehat{y}^e_{ibt}$ is the residual whose property I will study below. The residuals are demeaned log-earnings observations, where averages are taken over cohorts, time, and education. This procedure adjusts for age and cohort effects in a more flexible way than the conventional approach that relies on regressions with cohort-specific age-polynomials. The cohort-specific autocovariance structures studied below are the sample analogues of $cov(\widehat{y}^e_{ibt}, \widehat{y}^e_{ib,t+k})$, where k is the order of the lag.

Figure 1 plots autocovariances at different lags against experience for the secondary degree group, where $experience^e_{bt} = t - b - \underline{t}^e_b$, with $\underline{t}^e_b$ denoting the labor market entry year of individuals with education e and born in year b. 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, 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

[20] Average earnings during apprenticeship training are significantly below those without a secondary degree. Including training spells for those with a secondary degree results in earning growth that is quite large early in the life-cycle. See for example Adda et al. (2011) for a detailed analysis.

present lag-profiles at different levels of experience for the secondary-degree group in appendix figure 1. Again, I split the full sample into four cohort groups. Auto-covariances are gradually and monotonically decreasing, eventually converging to some positive constant, and 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 auto-covariance structure over time and across cohorts, is also a well known feature of North-American data.[21] 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.[22]

In figure 2 and appendix figure 2 I repeat the exercise for the dropout group. Its auto-covariance structure differs significantly from the one discussed above. 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 autocovariance 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.

Recent research studying variance dynamics over the life-cycle emphasizes the need to flexibly control for time-effects. Since the convexity in variance profiles plays a central role in the identification of the HIP-component, it is useful to study if it is preserved once time-effects in second moments are removed. I therefore regress the variances on experience dummies and a set of time- or cohort-fixed-effects. Appendix figure 3 plots the estimates for the experience dummies for the two education groups, after controlling for either cohort or time effects. For comparison I also plot variance profiles that do not control for either of these effects and that are aggregated over all cohorts. They resemble the corresponding cohort specific profiles documented above. Controlling for cohort effects strengthens the convexities, possibly because earnings data for older workers are drawn over-proportionally from the sample of cohorts that have generally smaller autocovariances. Controlling for time effects has quite different implications. For the high education group, the increase of variances later in the life-cycle is not preserved, and the convexity is replaced by a profile that is declining initially and that remains flat thereafter. For the dropout sample the rate of decline also becomes larger, and variances eventually approach zero. This underlines the importance of modeling time-effects flexibly. Qualitatively, these results are remarkably similar to the findings from US-data documented in Heathcote et al. (2005) and Guvenen (2009).

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

4. Econometric Framework, Estimation and Identification

4.1. The Econometric Model

A general parametric earnings process that nests the majority of models considered in the literature is given by the additive decomposition of log-earnings y_{ibt},

$$y_{ibt} = f(X_{ibt}, \Pi_i) + P_{ibt} + Z_{ibt} + \Xi_{ibt}, \tag{4.1}$$

where the term $f(X_{ibt}, \Pi_i)$ specifies the relationship between observables X_{ibt} and log-earnings y_{ibt} and allows the vector of parameters Π_i to vary across individuals. Heterogeneity of intercepts or returns to experience are examples of cases in which the distribution of Π_i is non-degenerate. The three remaining terms naturally decompose the variation of residual log-earnings into stochastic processes of different persistences: P_{ibt} is a unit-roots component, Z_{ibt} is a stochastic process with moderate persistence to be estimated, and Ξ_{ibt} are purely transitory shocks. This equation nests HIP- and RIP-models, which are two families of earnings processes dominating the literature. "Restricted Income Profiles" (RIP) assume that individuals are heterogenous with respect to their log-earnings intercepts, but not their returns to experience. A common specification imposes the two restrictions $f(X_{ibt}, \Pi_i) = g(X_{ibt}, \theta) + \alpha_i$ and $Z_{ibt} = 0$, where $\Pi_i = (\theta, \alpha_i)$ and where α_i captures heterogeneity in intercepts across individuals, whereas all other parameters θ are assumed to be constant in the population. According to this view, any increases of residual inequality over the life-cycle are due to permanent unobserved shocks P_{ibt}. "Heterogeneous Income Profiles" (HIP) postulate instead that returns to experience vary across individuals, while shocks that accumulate over the life-cycle are of moderate persistence. With $h_{ibt} \subseteq X_{ibt}$ denoting labor market experience, a common assumption is $f(X_{ibt}, \Pi_i) = g(X_{ibt}, \theta) + \alpha_i + \beta_i * h_{ibt}$ and $P_{ibt} = 0$. In this specification, a systematic fanning out of the residual earnings distribution over the life-cycle is generated by slope heterogeneity.

It is clear from equation (4.1) that HIP- and RIP-specifications impose strong restrictions on the parametric earnings process. It therefore does not come as a surprise that estimates of the same parameters are often found to be sensitive across specifications. Examples are the studies by Baker (1997) and Guvenen (2009), who demonstrate estimates of the persistence of shocks to be biased upwards when not controlling for a HIP component, and Hryshko (2012), who shows estimates of slope heterogeneity to be biased upwards if one omits the unit-roots component. A major impediment to studying robustness of parameters of interest more systematically, possibly by starting from less restictive models, is the quality of commonly used and publicly available panel data sets such as the PSID. In this paper I adopt the strategy of Baker and Solon (2003) and collect high-quality administrative data with large sample sizes in the cross-section and the time-series and to consider processes that match all features of the empirical covariances structure well. In particular, I estimate the following model for the residuals $\widehat{y}^e_{ibt}$ computed according to (3.1), where I suppress the superscript for education for notational convenience:

$$\widehat{y}_{ibt} = p_t * [\alpha_i + \beta_i * h_{ibt} + u_{ibt}] + z_{ibt} + \varepsilon_{ibt} \tag{4.2}$$

with

$$\begin{aligned} u_{ibt} &= u_{ib,t-1} + \nu_{ibt} && (4.3) \\ z_{ibt} &= \rho * z_{ib,t-1} + \lambda_t \xi_{ibt}. && (4.4) \end{aligned}$$

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

$$\begin{aligned} var\,(\alpha_i) &= \widetilde{\sigma}^2_{\alpha}; \;\; var\,(\beta_i) = \sigma^2_{\beta}; \;\; cov(\alpha_i, \beta_i) = \sigma_{\alpha\beta} && (4.5) \\ var(\nu_{ibt}) &= \sum\nolimits_{j=0}^{J_\nu} h^j_{ibt} * \delta_j; \;\; var(u_{it_0(b)}) = \widetilde{\sigma}^2_{u_0} && (4.6) \\ var\,(\xi_{ibt}) &= \sum\nolimits_{j=0}^{J_\xi} h^j_{ibt} * \gamma_j; \;\; var(z_{it_0(b)}) = \lambda_{t_0(b)} * \sigma^2_{\xi_0} && (4.7) \\ var\,(\varepsilon_{ibt}) &= \sigma^2_{\varepsilon}. && (4.8) \end{aligned}$$

No further distributional assumptions are required. Identification requires normalization of factor loadings p_t and λ_t for some t. Given the limited number of cohorts that are present in the sample prior to 1980 I set $p_t = \lambda_t = 1$ for all $t < 1980$ to increase precision.[23]

Model (4.2) corresponds to the general process (4.1) with $f(X_{ibt}, \Pi_i) = \mu_{bt} + p_t * [\alpha_i + \beta_i * h_{ibt}]$, $P_{ibt} = p_t * u_{ibt}$, $Z_{ibt} = z_{ibt}$ and $\Xi_{ibt} = \varepsilon_{ibt}$, where the adjustment for education-specific cohort-time effects is performed in a first-stage fixed effects regression. This model nests RIP- and HIP-processes by allowing for heterogeneity in the returns to experience, a random walk process that updates heterogeneous intercepts α_i over the life-cycle, an AR(1)-process with persistence ρ, and a purely transitory shock that cannot be separated from measurement error, ε_{ibt}. In contrast to the majority of models estimated in the literature, each dynamic error component has non-degenerate initial conditions, given by $\widetilde{\sigma}^2_{u_0}$ and $\sigma^2_{\xi_0}$. An important feature of the model are the flexible specifications for age- and time-effects in innovation variances. Age-effects are introduced by way of polynomials of order J_ν for permanent shocks and J_ξ for persistent shocks and are a simple way to generate variance dynamics as emphasized by Meghir and Pistaferri (2004). The order of these polynomials need to be determined empirically. In contrast, time-effects enter through the set of factor loadings $\{p_t, \lambda_t\}^{2004}_{t=1980}$, where p_t multiplies the permanent component and can be interpreted as a skill price, while the factor loading λ_t enters the persistent component indirectly through its multiplication with the shock v_{ibt}. This distinction is crucial as it allows the impact of λ_t on earnings dynamics to fade gradually over time, as can be expected from business-cycle shocks or firm closures.

[23] This assumption is consistent with the lack of noticable trends in earnings inequality over the 1970's in Germany. See for example Dustmann et al. (2009).

Initial conditions in the persistent component $\left(\lambda_{t_0(b)} * \sigma^2_{\xi_0}\right)$ will vary across cohorts indexed by b because different cohorts enter the labor market in different years $t_0(b)$.

There are many extensions or alternative non-nested specifications one may consider instead. I have chosen the structure (4.2) to (4.8) for several reasons. First, it provides an intuitive decomposition of the earnings variation over the life-cycle into three components with different degrees of persistence: permanent, moderate, and purely transitory. All three components present labor market risks with different degrees of insurability and play a prominent role in heterogeneous agents models. Second, features of the model that are uncommon in the literature, such as age heteroscedasticity and cohort-specific persistent initial conditions, can be easily motivated economically. For example, search frictions can generate dispersion in initial earnings that are gradually eliminated through the process of job search, thus generating earnings dynamics that look like a process with a persistent initial condition in the reduced form.[24] Third, the model matches all dimensions of the autocovariance structure well while being parsimonious enough to be used in heterogenous agents modelling. For the secondary-degree group, the model can explain 98 percent of the total variation in 56 thousand autocovariance elements with only 65 parameters, and the remaining variation strongly resembles sampling error in the empirical second moments. Importantly, I do not consider an MA(q)-component simply because it is not significant given the model structure: An AR(1)-component with moderate persistence together with slope-heterogeneity provides a sufficiently good match to the slope of lag-profiles, even at very low orders. This is consistent with results in Baker and Solon (2003), who do not find evidence for transitory shocks in form of an MA(q)-component in Canadian tax data.[25] On the other hand, a flexible specification for time-effects are essential to generate the change of the autocovariance structure across cohorts observed in the data, and slope heterogeneity together with age dependence in the variances of the permanent and persistent component introduces enough heterogeneity across agents to fit the life-cycle profiles within cohorts. At the same time, although the number of parameters of the model is much higher than in standard formulations of HIP- and RIP-models, the number of state variables is not. Indeed, compared to Guvenen (2009) there is only one additional state variable - the state of the random walk. The model is therefore tractable enough to be used in the calibration of heterogenous agents models, in contrast to earnings processes with larger state-spaces, such as MA(q)-, ARCH- and GARCH-models.

4.2. Estimation

The model generates theoretical autocovariances

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

[24]Postel-Vinay and Turon (2010) study the autocovariance structure generated by a structural equilibrium search model.

[25]The lack of evidence for measurement error and transitory shocks has also motivated my choice of not allowing the variance of this component to depend on age or time. In contrast, Guvenen (2009) and Hryshko (2012) "load" the transitory component, but not the permament component with a time factor. When estimating their specifications below as a point of comparison I adapt their model assumptions.

$$+\rho^k * Var\left(z_{ibt}\right) + 1(k=0) * \sigma_\varepsilon^2 \tag{4.9}$$

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

$$\begin{aligned} Var(z_{it_0(b)}) &= \lambda_{t_0(b)} * \sigma_{\xi_0} \\ Var\left(z_{ibt}\right) &= \rho^2 * Var\left(z_{ibt-1}\right) + \lambda_t^2 * \left(\sum\nolimits_{j=0}^{J_\xi} h_{ibt}^j * \gamma_j\right) \quad \text{for all } t > t_0(b). \end{aligned} \tag{4.10}$$

In stationary models, equation (4.10) can be shown to have a closed form solution that is highly non-linear in model parameters. However, 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 chosen 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 (4.9). Denoting the parameter vector by θ and observables by Z, I write the stacked version of these *theoretical* autocovariance matrices as $G\left(\theta, Z\right)$. The (MD)-Estimator for Θ solves

$$\widehat{\theta} = \min_{\widetilde{\theta}} \left[\widehat{C}^{vec} - G\left(\widetilde{\theta}, Z\right)\right]' W \left[\widehat{C}^{vec} - G\left(\widetilde{\theta}, Z\right)\right] \tag{4.11}$$

where W is some positive definite weighting matrix.[26] 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 (4.11) becomes the Equally Weighted Minimum Distance Estimator (EWMD). While its asymptotic distribution is well known, the particular features of my model together with the large number of moments to be matched introduces two challenges that are non-standard.[27] First, to calculate standard errors for the parameter estimates, one needs to estimate the covariance matrix $\widehat{\Omega}$ of $\widehat{C}^{vec}$, a matrix of size $\left[\dim(\widehat{C}^{vec})\right]^2$. In contrast to studies that rely on annual data and that aggregate over cohorts, $\dim(\widehat{C}^{vec})$ is large in my estimation - over $56,000$ in the secondary-degree group and over $64,000$ in the dropout-group. As a consequence of the confidential nature of the data it is infeasible to compute $\widehat{\Omega}$ directly.[28] I solve this issue by using the fact that the EWMD-estimator is a non-linear least-squares estimator (NLS), where one regresses autocovariances on the non-linear parametric function $G\left(\theta, Z\right)$. It is therefore possible to compute standard errors of $\widehat{\theta}$ without computing $\widehat{\Omega}$ by using the

[26] Asymptotically, the optimal choice of W is the inverse of a matrix consistently estimating the covariance matrix of $\widehat{C}^{vec}$.

[27] For a characterization of the asymptotic distribution, see e.g. Abowd and Card (1989).

[28] All statistics that are based on individual-level data need to be computed at the Research Data Centers (RDC) of the IAB. Aggregated statistics are allowed to be used off-site as long as the number of observations used to compute them pass a certain threshold, and as long as the administrative burden from checking this requirement is not too large. Both criteria are met by the auto-covariance structures matched in my estimation procedure. Given the large size of $\widehat{\Omega}$ I am neither allowed to use it off-site, nor can it be computed on-site given the computational resources.

appropriate formulae from NLS-estimation. To account for sampling error that is correlated arbitrarily across observations because different autocovariances for a cohort rely on the same data I use cluster-robust standard errors, where clustering takes place on the cohort-level.[29] Second, variances of permanent and persistent shocks as specified in equations (4.6) and (4.7) have to satisfy a non-negativity constraint, while parameters $\{\delta_j\}_{j=0}^{J_\nu}$ and $\{\gamma_j\}_{j=0}^{J_\xi}$ need to be allowed to be negative.[30] I therefore estimate the parameters by using a constrained optimization routine. A further discussion of both these issues is provided in the appendix.

4.3. Identification

The EWMD-estimator of the model (4.2) - (4.8) is equivalent to NLS-estimation with empirical covariances as dependent variables and $G(\theta, Z)$ as the non-linear regression model.[31] The estimator $\widehat{\theta}$ solves the system of $\dim(\theta)$ first-order conditions

$$J_{\widehat{\theta}}(Z)' * \left[\widehat{C}^{vec} - G\left(\widehat{\theta}, Z\right)\right] = 0, \tag{4.12}$$

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 $\dim(Z) \times \dim(\theta)$. This system generally does not have a closed-form solution, but applying standard results for NLS-estimation implies the following assumptions to be sufficient for local identification and consistency of $\widehat{\theta}$: (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 the model $G(\theta, Z)$ to be correctly specified. The last assumption requires the Jacobian to have full rank at θ, thereby guaranteeing local point-identification. While these assumptions are rather abstract, they have a number of immediate implications. First, since $\widetilde{\sigma}_\alpha^2$ and $\widetilde{\sigma}_{u_0}^2$ enter the equation (4.9) additively, one cannot identify these two parameters separately and assumption (iii) is violated. 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. I therefore estimate a "combined initial condition" for the permanent component $\sigma_\alpha^2 = \widetilde{\sigma}_\alpha^2 + \widetilde{\sigma}_{u_0}^2$. Second, $\widehat{\theta}$ will depend on the model specification $G(\theta, Z)$. Generally, one should view parameter estimates as "credible" only if the model passes a mis-specification test. A priori it is not clear how large the biases and inconsistencies in the estimates will be if the model is misspecified. It is a primary objective of this paper to explore this issue quantitatively. Third, the Jacobian J_θ evaluated at the estimates $\widehat{\theta}$ can be used to analyse the unique data features matched by a parameter. My numerical comparative statics exercise will heavily rely on this fact.

[29] Heteroscedasticity robust standard errors and clustered standard errors also involve outer products of the vectors of sampling errors, which have the same dimension as the variance-covariance matrix of $\widehat{C}^{vec}$. As is well known, the "sandwich estimators" reduce the dimensionality of this problem.

[30] When computing theoretical moments from the model I restrict these variances to be non-negative up to age 63, the mandatory retirement age in Germany, even though these age groups are never observed in the data.

[31] For a discussion of EWMD-estimation in terms of NLS-estimation, see Cameron and Trivedi (2005). To make the dependence on observables explicit, it is convenient to define two sets of fixed effects: Let I_{τ_1} be an indicator variable equal to one if $t = \tau_1$, and I_{τ_2} be an indicator variable equal to one if $t + k = \tau_2$. Then we can write $p_t * p_{t+k} = \left(\prod_\Upsilon (p_{\tau_1})^{I_{\tau_1}}\right) * \left(\prod_\Upsilon (p_{\tau_2})^{I_{\tau_2}}\right)$ and $\lambda_t^2 = \prod_\Upsilon \left(\lambda_{\tau_1}^2\right)^{I_{\tau_1}}$, where Υ is the set of all time periods observed in the data. Hence, the matrix of observables is given by $(1, \{I_{\tau_1}, I_{\tau_2}\}, h, b, k, 1_{k=0})$ together with higher order monomials of several of these variables.

Unfortunately, apart from the simplest earnings processes it is not possible to derive closed-form expressions for $\widehat{\theta}$ and it becomes difficult to determine the model-specific data variation identifying a particular parameter of interest. One approach, followed e.g. by Guvenen (2009) and Hryshko (2012), is to derive closed-form expressions in the exactly identified case where one matches as many selected autocovariances as there are parameters. Results from this approach may not carry over to the fully efficient MD-estimator. A more general and novel approach is to rely on equations (4.12) and to investigate specification error numerically. To describe this approach, let $J^l_{\widehat{\theta}}(Z)$ be the $l-th$ element of the Jacobian $J_{\widehat{\theta}}(Z)$ conditional on Z. The $l-th$ element of a first-order Taylor approximation of $\left[\widehat{C}^{vec} - G\left(\widehat{\theta}, Z\right)\right]$ around the estimate $\widehat{\theta}$ yields

$$\begin{aligned} \widehat{C}^{vec} - G\left(\widetilde{\theta}, Z\right) &\approx \widehat{C}^{vec} - G\left(\widehat{\theta}, Z\right) - J^l_{\widehat{\theta}}(Z) * \left(\widetilde{\theta} - \widehat{\theta}\right) \\ &\Rightarrow \left[\widehat{C}^{vec} - G\left(\widetilde{\theta}, Z\right)\right] - \left[\widehat{C}^{vec} - G\left(\widehat{\theta}, Z\right)\right] \approx -J^l_{\widehat{\theta}}(Z) * \left(\widetilde{\theta}^l - \widehat{\theta}^l\right). \qquad (4.13) \end{aligned}$$

This equation measures the approximate change in the unexplained variation of $\widehat{C}^{vec}$ as one moves $\widehat{\theta}^l$ to some counterfactual parameter value $\widetilde{\theta}^l$, holding everything else constant. Since consistency of the parameter estimate together with the identification assumption (iii) implies that $J^l_{\widehat{\theta}}(Z)$ has full rank, this type of counterfactual experiment is feasible. With everything else held constant, this exercise extracts the data feature that is matched by the parameter θ^l *alone* and therefore isolates its identifying variation.[32]

Equation (4.13) clarifies two issues that play an important role in the numerical exploration of identification. First, the effect of moving θ^l away from its estimated value depends on the observables Z that enter the model. I will therefore plot the deterioration of the model match in the counterfactual exercise for entire experience profiles of autocovariances at different lags.[33] Second, the counterfactual depends on the model specification $G(\theta, Z)$. Hence, comparing the effect of changing $\widehat{\theta}^l$ to some value $\widetilde{\theta}^l$ across specifications will clarify if a particular parameter is identified off similar variation in different models. The results in Guvenen (2009) and Hryshko (2012) suggest that it may not. The following example demonstrates this approach for a case that admits an analytical expression for the EWMD-estimator:

Example 4.1. *Consider a much simplified version of the earnings process (4.2) - (4.8) that allows the variance of the permanent shock to vary linearly in age:*

$$\begin{aligned} \widehat{y}_{it} &= \alpha_i + \beta_i * h_{it} + u_{it} \qquad (4.14) \\ u_{it} &= u_{i,t-1} + \nu_{it} \\ var(\alpha_i) &= \widetilde{\sigma}^2_\alpha; \;\; var(\beta_i) = \sigma^2_\beta; \;\; cov(\alpha_i, \beta_i) = 0 \\ var(\nu_{it}) &= h_{it} * \delta_1; \;\; var(u_{i0}) = 0. \end{aligned}$$

The autocovariance structure is given by

$$c_{it} = \sigma^2_\alpha + h_{it} * (h_{it} + k)\,\sigma^2_\beta + \delta_1 * \frac{h_{it} * (h_{it} + 1)}{2} \qquad (4.15)$$

[32] In OLS, the Frisch-Waugh theorem is based on a similar thought experiment and delivers a closed-form solution due to the linearity of the underlying model.

[33] To keep the number of figures managable, I will average these effects over cohorts.

and the EWMD-estimator reduces to OLS. Now suppose one erroneously neglects heteroscedasticity which corresponds to the a-priori restriction $\delta_1 = 0$. Defining $z_{it} = \frac{h_{it}*(h_{it}+1)}{2}$ and $x_{it} = h_{it} * (h_{it} + k)$, the parameter estimate for σ_β^2 is given by $\widehat{\sigma}_\beta^2 = \frac{\sum_{ibt}(x_{ibt}-\overline{x})*\widehat{c}_{ibt}}{\sum_{ibt}(x_{ibt}-\overline{x})^2}$ and the omitted-variable bias formula for OLS implies that asymptotically

$$\widehat{\sigma}_\beta^2 - \sigma_\beta^2 = \delta_1 * \frac{cov(x_{ibt}, z_{ibt})}{var(x_{ibt})}. \tag{4.16}$$

Since $cov(x_{ibt}, z_{ibt}) > 0$ 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 OLS-estimator mistakenly assigns all of the convexity in the experience profile to the estimate of slope heterogeneity $\widehat{\sigma}_\beta^2$. This source of a bias will also be uncovered by the numerical comparative statics exercise described above: If one estimates model (4.15) with the a-priori restriction $\delta_1 = 0$ even so in reality $\delta_1 > 0$, yielding parameter estimates $\left(\widehat{\sigma}_\alpha^2, \widehat{\sigma}_\beta^2\right)$, and then conducts the counterfactual experiment in which one sets $\sigma_\beta^2 = 0$, the resulting autocovariances will all have value $\widehat{\sigma}_\alpha^2$. In contrast, if one conducts the same experiment from estimates of the unrestricted model $\left(\widehat{\sigma}_\alpha^2, \widehat{\sigma}_\beta^2, \widehat{\delta}_1\right)$, some of the convexity of the autocovariance profiles will be preserved.■

Even though the exploration of misspecification error needs to be addressed numerically, analyzing equation (4.9) helps uncover data features that are likely to be matched by a particular set of parameters in the full model. To facilitate the discussion I assume that $p_t = \lambda_t = 1$ for all t. If $\rho = 1$ the model fails to be identified. If $\rho < 1$ one can use the fact that $Var(z_{ibt})$ is bounded above by $\max_{b,t}\{var(\widehat{y}_{ibt})\}$ to derive the following approximation for large k,[34]

$$cov(\widehat{y}_{ibt}, \widehat{y}_{ib,t+k}) \approx \sigma_\alpha^2 + (2h_{ibt} + k)\,\sigma_{\alpha\beta} + h_{ibt} * (h_{ibt} + k)\,\sigma_\beta^2 + f^u(h_{ibt}, \delta_0, ..., \delta_{K_\nu}). \tag{4.17}$$

In the following I refer to this term as the "permanent component". Since $f^u(h_{ibt}, \delta_0, ..., \delta_{K_\nu})$ is a polynomial of degree $(K_\nu + 1)$ in h_{ibt}, this expression is linear in parameters and can be estimated by OLS. Several results follow immediately. First, the permanent component is the only model feature that can match high-order autocovariances. Second, since none of the observables are multicollinear, all parameters entering this linear regression equation are *globally* point-identified. Third, age-effects in the variance of the innovations to the permanent component can be separated from slope heterogeneity because the former generate a relationship between experience and higher-order covariances that is stable in the order of the lag, while the latter generates a direct relationship between the lag and the autocovariances. In other words, as long as one chooses sufficiently high K_ν, heteroscedastic permanent shocks can approximate any continuous age-profiles in autocovariances at large k, corresponding to the lower envelope of the profiles plotted in figures 1 and 2, but the slope of lag-profiles

[34] 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 $\max_{b,t}\{var(\widehat{y}_{ibt})\} < 1$. As can be seen from figures 1 to 4, in my sample $\max_{b,t}\{var(\widehat{y}_{ibt})\} < 0.5$. Thus, $\rho^k * Var(z_{ibt})$ will vanish quickly as k increases.

is always zero. In contrast, if $\sigma_{\alpha\beta} < 0$ and $\sigma^2_\beta >> 0$ slope heterogeneity produces convex experience- and lag-profiles. In fact, with flexible age-effects in the variances of permanent shocks, the slope of lag-profiles is the only variation in the data that helps identifying $\left(\sigma_{\alpha\beta}, \sigma^2_\beta\right)$. This contrasts sharply with a simple HIP-process for which these parameters will also be identified off the convexity in experience profiles at any lag.

Since slope heterogeneity imposes strong restrictions on the slope of lag-profiles at large k, it is possible to develop an "eye-ball" test for its relevance. Fixing h 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

$$cov(\widehat{y}_{ibt}, \widehat{y}_{ib,t+k}) - cov(\widehat{y}_{ibt}, \widehat{y}_{ib,t+[k+n]}) \approx \left(n * \sigma_{\alpha\beta} + h * n * \sigma^2_\beta\right). \tag{4.18}$$

It follows directly that (i) negatively sloped lag-profiles at large k can only be explained by $\sigma_{\alpha\beta} < 0$ - even in the generalized HIP-model (4.2) - (4.8) - and that (ii) lag-profiles that converge to a constant are only consistent with $\sigma^2_\beta \leq \frac{|\sigma_{\alpha\beta}|}{\max\{h_{ibt}\}}$. 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.

Given that the permanent component is the only part of the model that has implications for the behavior of autocovariances at high orders, the persistent component will match the remaining unexplained variation at low orders. It can only be expected to be important if the covariance structure at low lags differs significantly from the one at high lags. In a specification without age and time effects in the variances, the AR(1)-component in (4.9) is given by

$$\rho^k * Var\left(z_{ibt}\right) = \rho^k * \left[\rho^{2h_{ibt}} * \sigma^2_{\xi 0} + \left(1 - \rho^{2h_{ibt}}\right) * \frac{\gamma_0}{1-\rho^2}\right]. \tag{4.19}$$

At low lags, this expression imposes strong restrictions on the shape of experience profiles. If the initial condition is assumed to be zero ($\sigma_{\xi_0} = 0$) as is commonly the case in the literature, the implied experience profile at k is concave and approches the constant $\frac{\delta_0}{1-\rho^2}$ from below, which is at odds with the covariance structure in the IABS and the PSID. However, if $\frac{\delta_0}{1-\rho^2} > \sigma_{\xi_0} > 0$, then the experience profile is convex and approaches the constant $\frac{\delta_0}{1-\rho^2}$ from above. Hence, an AR(1)-process with a positive initial condition can in principle explain the decline in variances early in the life-cycle observed in many frequently used panel data sets.

To demonstrate graphically some of these issues, I plot the autocovariance structure generated by different variance components in appendix figure 4, using parameters from Baker and Solon (2003).[35] Each line in a panel of the figure represents the experience-profiles of $k-th$ 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 the observed convexity. The interaction between the lag and experience identifying σ^2_β is reflected in high-order autocovariances increasing

[35]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): $\sigma^2_\alpha = 0.134$, $\sigma^2_\beta = 0.00009$; $\sigma_{\alpha\beta} = -0.0031$; $\delta_0 = 0.007$; $\sigma^2_{\xi 0} = 0.167$; $\rho = 0.54$; $\gamma_0 = 0.09$; $\gamma_1 = -0.005$; $\gamma_2 = 0.0001$; $\gamma_3 = 2.21 * e(-6)$; $\gamma_4 = 2.1 * e(-9)$. I set all factor loadings equal to one.

relative to the 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 experience profiles of autocovariances generated by a heteroscedastic AR(1)-process with an initial condition. Given 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 of the secondary-degree sample in the IABS as documented in figure 1.

5. Results

In this section I present the main results of the paper. As documented above, the autocovariance structure of the secondary-degree group shares many of the qualitative features of the North American counterpart, particularly those that have been highlighted as evidence in favor of slope heterogeneity. In contrast, the autocovariance structure for the dropout-group has very different characteristics. I therefore focus the discussion on the secondary-degree group and view the results for the high-school dropouts, presented in a separate section, as a robustness check for my findings.[36]

5.1. Main Results

Parameter estimates of the model described by equations (4.2) to (4.8) for the secondary-degree group are presented in the first column of table 1 and, for the two sets of factor loadings, in appendix figure 5. The model fit is shown in figure 3. All parameters but those on the higher-order terms in the variance of the permanent shock component are statistically significant at least on the 5%-significance level and are precisely estimated.[37] There is substantial heterogeneity in the intercept and the initial condition of the persistent component, with estimated variances of $\widehat{\sigma}^2_\alpha = 0.03$ and $\widehat{\sigma}^2_{\xi 0} = 0.057$ respectively. The persistence of shocks to the AR(1)-process on the quarterly level is estimated at $\widehat{\rho} = 0.863$, a fairly low value. While the estimated variance of the returns

[36] A remaining issue is the the degree of the polynomials in age that describe the variances of the permanent and the persistent shocks. After experimenting with a wide range of degrees, I choose $J_\nu = J_\xi = 4$, for two main reasons. First, adding additional degrees to the permanent component does not significantly improve the model fit. Second, higher orders of the polynomial in the persistent component introduces numerical inaccuracies since $Var(z_{ibt})$ is badly scaled. Exponentials of the experience-variables grow quickly and eventually take on very large values, while the parameters are matched to autocovariances that are often close to zero. The unit-roots process with polynomial age-heteroscedasticity is badly scaled as well, but can be rescaled accordingly given its linearity-in-parameters.

[37] This also applies to all factor loadings. To avoid clutter in figure 7 I do not plot the confidence intervals.

to experience $\widehat{\sigma}_\beta^2$ is significant on the 5%-level, it is smaller than conventional estimates from the PSID.[38] 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 significant on at least the 5% level. In contrast, there is no evidence for age effects in innovations to the permanent component.

The evolution of the two sets of factor loadings plotted in appendix figure 5 helps identify whether the trends 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. While this result may be surprising at first sight, it is mirrored in the lag-profiles of appendix figure 1: An increase in the permanent component will shift these profiles upward *at any lag*, while an increase in the persistent component will mainly act on the autocovariances at low orders. However, as shown in the figure, lag-profiles converge to similar values for all cohorts, implying that the variance of the permanent component has not increased substantially.

Since the estimation is equivalent to non-linear least squares regression, the R^2 is a valid measure of the goodness of fit. With a value of .98 the model fits the data almost perfectly. Given that I am matching 56,072 autocovariances with only 65 parameters, the majority of which are factor loadings, this is quite remarkable. A graphical illustration of the match, shown in figure 3, provides additional information. Each of the panels plot theoretical against empirical autocovariances for four cohort groups, keeping constant the order of the lag.[39] 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.

The lack of evidence for age effects in the variances of the permanent component motivates estimation of the model with homoscedastic permanent shocks, corresponding to the restriction $\delta_j = 0$ for $j > 0$. Results are shown in column 2 of the same table. The majority of estimates change substantially, in most cases moving outside of the 99%-confidence interval of the estimates in column 1. Several results are worth highlighting: First, $\widehat{\sigma}_\beta^2$ decreases by 75 percent and becomes insignificant on any conventional level, so that heterogeneity in slopes, the HIP component, is not found to be important anymore. Second, $\widehat{\sigma}_{\alpha\beta}$, while still being highly significant, decreases by $\frac{2}{3}$ in absolute value. Interestingly, this finding clarifies that heterogeneity in slopes can be insignficant even though its covariance with the intercept is significant. Indeed, there is no intrinsic restriction by the model that rules out this possibility. It is therefore important to document a test statistic for the *joint* significance of the two parameters, which is provided at the bottom of the table. Given the precision of $\widehat{\sigma}_{\alpha\beta}$ and the sample size, it is not surprising that one rejects the null-hypothesis $\left(\sigma_{\alpha\beta}, \sigma_\beta^2\right) = (0,0)$. Third, the persistence of the

[38] Guvenen (2009) shows that even a low level of slope-heterogeneity can translate into significant life-cycle inequality.

[39] An alternative would be to clean the autocovariances from cohort effects much like in figure 5, but this would mask the ability of the model to fit inter-cohort changes.

AR(1)-process increases to a value of $\widehat{\rho} = 0.88$. Fourth, dispersion in the intercept decreases, while the variance of the initial condition of the persistent component increases substantially.

These result already hint at the sensitivity of key parameters to minor changes in the empirical specification of the earnings process. While one would expect that restricting the specification of the unit-roots process assigns more weight to both intercept and slope heterogeneity the opposite is the case. Both, $\widehat{\sigma}^2_\beta$ and $\widehat{\sigma}^2_\alpha$ are significantly smaller than in the unrestricted model. In contrast, model estimates assign relatively more weight to the persistent initial condition in generating inequality at labor market entry. The source of the sensitivity of estimates will be investigated systematically below. A useful first step is to analyze the model fit. The R^2 of the restricted model is .964 and hence only slightly lower than for the full model. However, a Wald-test for joint significance of the $\delta_j = 0$ for $j > 0$ rejects the Null-hypothesis of homoscedasticity, implying that controlling for age-effects in the variance of permanent innovations is important. A graphical illustration of the match is shown in appendix figure 6, following the structure of figure 3. Comparing figures 3 and appendix figure 6 uncovers that most of the deterioration of the model fit takes place because it misses the concavity of the life-cycle profiles at large lags. This result turns out to be important for an understanding of the sensitivity of $\widehat{\sigma}^2_\beta$ across specifications. Since age-effects in the *persistent* component are controlled for, the main source of identification of $\left(\sigma_{\alpha\beta}, \sigma^2_\beta\right)$ is the shape of lag-profiles, which is decreasing and convex. Equation (4.18) implies that $\sigma_{\alpha\beta} < 0$ and $\sigma^2_\beta > 0$ in this case. However, because of equation (4.17) this also generates the overidentifying restriction that experience profiles at large lags are convex, in sharp contradiction with the observed concavity. The tension between convexities in lag-profiles and concavities in high-order experience profiles can be relaxed if one allows for flexible age-effects in permanent innovations. Without age-effects, the unit roots process generates auto-covariances that are linear in experience. As a consequence, the concavities in the auto-covariance profiles cannot be generated by the model, and a positive σ^2_β would further deteriorate the fit. Hence, $\widehat{\sigma}^2_\beta$ must be close to zero, $\widehat{\sigma}_{\alpha\beta}$ must decrease in absolute value, and large intercepts of lag-profiles can only be explained by shifting weight from the permanent to the persistent initial condition.

Given the insignificance of $\widehat{\sigma}^2_\beta$ in column (2), I also estimate the full model with the a-priori restriction $\left(\sigma_{\alpha\beta}, \sigma^2_\beta\right) = (0, 0)$. Parameter estimates are listed in column (3) of table 1, and the fit of the model is presented in appendix figure 7. Compared to column (2), the variance of intercept heterogeneity $\widehat{\sigma}^2_\alpha$ decreases further to a value of .013, which is just over a third of the corresponding estimate in column 1. The persistence ρ is estimated at .906, compared to the value of .863 in the unrestricted model. Again, all higher-order terms for the variance of the permanent component are insignificant, but an F-test rejects the hypothesis $\delta_j = 0$ for $j > 0$ at all conventional significance levels. One wonders immediately if the low estimate of $\widehat{\sigma}^2_\alpha$ is a coincidence. Comparing figures 3 and appendix figure 7 suggests it is not. The most noticeable difference is the inability of the restricted model to fit the intercepts of the various life-cycle autocovariance-profiles. In a model with $\sigma_{\alpha\beta} = 0$, a higher $\widehat{\sigma}^2_\alpha$ will unambigously increase the intercepts at any lag. Since the intercepts are significantly smaller at high lags, this may improve the fit for small k at the expense of deteriorating the fit at large k. Eliminating the restriction

$\sigma_{\alpha\beta} = 0$ relaxes this tension as now a negative $\widehat{\sigma}_{\alpha\beta}$ can generate low intercepts at high k even if $\widehat{\sigma}^2_\alpha$ is large. In other words, $\sigma_{\alpha\beta}$ helps to "free up" the parameter σ^2_α so that intercept heterogeneity can be estimated to be large even if autocovariances converge to a very small value at high orders. Given the generally negative slope of lag-profiles it is not surprising that (a) $\widehat{\sigma}_{\alpha\beta}$ is quite robust across specifications and precisely estimated and that (b) $\widehat{\sigma}^2_\alpha$ is larger in specifications where $\sigma_{\alpha\beta}$ is not restricted to be zero.

5.2. Robustness

Next I explore the sensitivity of key parameter estimates across various specifications, all of which are nested within my preferred model (4.2) to (4.8). If estimates of the same parameters differ significantly across specifications, it becomes likely that results from more restrictive specifications commonly estimated in the empirical literature are plagued by sizable biases. Results for the secondary-degree group are shown in the top panel of table 2. For comparison I reproduce estimates from the full model as discussed above in the first column. The next two columns show results for two models that are particularly popular in the literature: a HIP-model as considered by Guvenen (2009) and a RIP-model that only features intercept heterogeneity, a homoscedastic AR(1)-process, and transitory shocks, the latter of which cannot be separated from measurement error. Together, I view these three specifications as benchmarks. The last four columns of the table display results from specifications that exclude one family of parameters from the full model: heteroscedasticity in column 4, time effects in column 5, the initial condition of the persistent component in column 6, and a combination of these restrictions in column 7.[40] Comparing the resulting estimates to those from the full model will clarify which model features are likely to be particularly relevant for obtaining "credible" estimates of key parameters.

Perusing the table, two results are immediately noticeable. First, none of the parameters are robust across specifications. For example, the estimated heterogeneity in slopes $\widehat{\sigma}^2_\beta$ varies between zero for the homoscedastic version of the full model and a highly significant .006 for Guvenen's HIP-model while the estimated persistence of the AR(1)-process $\widehat{\rho}$ varies between .76 for Hryshko's augmented HIP-process and a value not significantly different from 1. All other model parameters vary widely across specifications as well. However, intercept heterogeneity σ^2_α, the initial condition of the AR(1)-process $\sigma^2_{\xi 0}$ and the age effects of persistent innovations $(\gamma_j)^4_{j=1}$ are highly significant in all specifications. A homoscedastic unit roots process, represented by the parameter δ_0, is found to be significant in most specifications as well. Second, the model fit deteriorates substantially as one moves towards more restrictive specifications. While the full model matches the data almost perfectly with an R^2 of .98, a simple AR(1)-process explains less than 60 percent of the variation in the empirical covariance structure. Among the components of the full model, exclusion of time effects has the largest effect on the model match, with the R^2 dropping to a value of .84. Exclusion of the initial condition of the AR(1)-process or the age

[40]The latter is a stationary version of Hryshko's (2012) model that allows for slope-heterogeneity, an AR(1)-process and a unit-roots process, thus merging HIP and RIP. C

effects in the persistent and permanent component also have large effects. Moving to the HIP-models decreases the fit even further to about three quarters of the data variation.[41]

A striking result is the discrepancy in results between the full model and a standard HIP-model, as shown in the first two columns. Compared to the full model, the HIP-process restricts age-effects in innovation variances, time-effects in the permanent component, and the persistent initial condition to be zero. These restrictions lead to a sizeable upward bias in the estimated heterogeneity in intercepts and slopes. Most importantly, $\widehat{\sigma}^2_\beta$ is three times as large as in the full model and significant on the 1%- rather than the 5%-level. Somewhat surprisingly, the estimated persistence of the AR(1)-shocks is much larger in the HIP-model than in the full model and it is in fact not significantly different from one. Thus, allowing for heterogeneity in slopes alone does not rule out perfect persistence in shocks. 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 $\widehat{\rho}$.

Moving from the full model to a HIP-process means removing several model components at once. To isolate the model feature that has a particularly large impact on the estimates of the key parameters, I show estimates from models that remove only one of the components - age-effects, time-effects, and initial conditions for the persistence component - in columns 4 to 6. Each of the models yield estimates of ρ and σ^2_β whose confidence intervals do not include the corresponding estimates from the HIP-model. Furthermore, while exclusion of age-effects has the strongest effect on the estimated heterogeneity in slopes and intercepts, removal of the persistent initial condition has the strongest effect on the estimated persistence.

5.3. Exploring Identification - Numerically

A natural next step is to investigate whether the documented sensitivity of key parameter estimates reflects some systematic specification error common to popular, but overly restrictive models. Optimally one would like to derive closed-form expressions for the model-dependent MD-estimators that can be compared across specifications. However, for all but the most restrictive specification such solutions do not exist. To understand the sensitivity of the results presented in tables 1 and 2, I rely instead on a numerical analysis of identification that implements *for each specification* the counterfactual exercise that is implicit in equation (4.13). This exercise consists of comparing the model-generated covariance structure at the estimates $\widehat{\theta}$ with the the covariance structure generated by some perturbation $\widetilde{\theta}$ that keeps all but a subset of parameters at their estimated values.

The results for the full model are depicted in figures 4 and 5. The first of these figures plots predicted experience-profiles of predicted auto-covariances when evaluated at the estimates. Each of the four panels of the figure corresponds to a different order of the lag. To minimize the number of figures, I average the auto-covariance elements $\widehat{c}_{btk}$ over cohorts.[42] In addition to the model prediction I also show in the same figure the experience-

[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] The conclusions are the same when performing the numerical comparative statics for each cohort separately. Alternatively one could "clean" the series from cohort effects by estimating and removing cohort fixed effects. Given the non-linearity of the model,

profiles from the following counterfactual exercises: (1) no heterogeneity in slopes: $\sigma_\beta^2 = 0$; (2) no correlation between intercept and slope heterogeneity: $\sigma_{\alpha\beta} = 0$; (3) no AR(1)-component, but a persistent initial condition: $\gamma_j = 0, j = \{0, 1, ..., 4\}$; (4) no initial condition for the AR(1)-process: $\sigma_{\xi 0}^2 = 0$. I add the results from a fifth counterfactual - the removal of age-effects in innovation variances, associated with $\gamma_j = \delta_j = 0, j = \{1, ..., 4\}$ - to the third and fourth panel of the figure.[43] Finally, since equation (4.9) implies that slope heterogeneity is identified in the full model only from the shape of lag-profiles, I plot estimated and counterfactual averaged *lag*-profiles in figure 5 for four different experience groups. The only parameter other than σ_β^2 and $\sigma_{\alpha\beta}$ that has a direct effect on the relationship between theoretical auto-covariances and the lag is the persistence of the AR(1)-process ρ. I therefore only investigate the numerical effects of perturbing $\left(\sigma_{\alpha\beta}, \sigma_\beta^2\right)$ around their estimates, holding everything else constant.

The first two panels of figure 4 uncover a dominant role of the AR(1)-process in matching low-order autocovariances: Keeping all other parameters at their estimated values, the removal of the persistent initial condition $\sigma_{\xi 0}^2$ eliminates virtually the entire decline of these moments early in the life-cycle. On the other hand, the predicted decline is too large if one removes the AR(1)-process instead, as shown by the third counterfactual. The last two panels of the figure show that the effect of the persistent component vanishes fairly quickly as removing it has a negligible effect on predicted autocovariance after a lag of about 20 quarters. Taken together these results suggest that the persistent component is strongly identified from the life-cycle profiles of autocovariances at low orders. Most importantly, convexities and initial declines of variance profiles can be explained by a rich specification of the persistent component.

With age-effects in the variance of the permanent component, slope heterogeneity can only be identified from the shape of lag-profiles. Equation (4.18) implies that at large lags, the approximate rate of decline is given by $\left(\sigma_{\alpha\beta} + h * \sigma_\beta^2\right)$, where $h = 0$ for labor market entrants. Figure 5 shows that the predicted behavior of lag-profiles at labor market entry for large enough orders of the autocovariance are entirely driven by $\sigma_{\alpha\beta}$ - removing it while keeping everything else constant leaves a flat profile. In contrast, the initial decline of these profiles is not affected by slope heterogeneity at all, suggesting that it is matched by the persistence of the initial condition of the AR(1)-process, ρ. This clarifies that $\sigma_{\alpha\beta}$ is identied from a robust data feature - the decrease of lag-profiles at moderate levels of the lag, especially for labor market entrants. However, in the data the lag profiles approach a constant, which is consistent with a negative $\sigma_{\alpha\beta}$ only if $\sigma_\beta^2 > 0$. Not surprisingly, setting $\sigma_\beta^2 = 0$ for older workers generates autocovariance profiles that are too steep and too small. While slope heterogeneity is solely identified from lag-profiles, they also impose strong restrictions on the shape of experience profiles. This is documented in figure 4. Most importantly, in the absence of age-effects in innovation variances, experience profiles would be strongly convex, as shown by the red line in the two bottom panels of the figure. However, in the data these

this would also remove some of the time and age effects that are matched by the estimation procedure, thus rendering interpretation of the associated numerical comparative statics difficult.

[43] I add the results from this counterfactual only to the third and fourth panel of the figure to keep the other two panels transparent. Also, this counterfactual is particularly important for the behavior of the model at higher values of the lag.

profiles are slightly concave at large k, a feature that can be explained by the model only if permanent shocks are heteroscedastic. Hence, age-effects in these innovations counterbalance the convexities generated by σ_β^2. This immediately explains why $\widehat{\sigma}_\beta^2$ is zero in homoscedastic versions of the model as shown in column (2) of table 1 or column (4) of table 2.

Next I implement similar numerical comparative statics exercises for the rudimentary HIP-model, with results shown in appendix figures 8 and 9. As this model does not control for age-effects in innovation variances or for persistent initial conditions, both the shape of experience- and lag-profiles help identify $\left(\sigma_{\alpha\beta}, \sigma_\beta^2\right)$. Somewhat surprisingly, when removing the AR(1)-component one obtains experience profiles that are slightly convex but almost flat. Since intercept- and slope-heterogeneity are the only remaining components, this finding implies that the latter is still mostly identified from the lag-profiles. Indeed, as one sets $\sigma_{\alpha\beta} = 0$ the lag-profiles become almost flat at any level of experience, as shown in appendix figure 9. Crucial to this result is the omission of a persistent initial condition, a consequence of which is that the persistence ρ has no predictive power for the autocovariances of labor market entrants. Hence, the negative slope of the lag-profile for this group will be matched entirely by $\sigma_{\alpha\beta}$, yielding an estimate that lies outside of the confidence interval of the corresponding estimate from the full model. At the same time, this also allows the high intercepts of experience- and lag-profiles to be matched by a large σ_α^2 without deteriorating the model fit at large lags. As clarified by the counterfactual exercises with respect to slope heterogeneity documented in the two figures, σ_β^2 mainly acts to neutralize the effect of a large $\sigma_{\alpha\beta}$ on autocovariances for older individuals.

Together, these results suggest that controlling for an initial persistent component is key for obtaining credible estimates of persistence and heterogeneity in intercepts and slopes. Comparing the results in columns (4) to (6) of Table 2 with those for the unrestricted specification in column (1) clarifies that it is the a-priori restriction of setting this component to zero which moves key parameter estimates closer to those obtained from a standard HIP-model. Thus, the numerical exploration of identification is particularly interesting for this specification, and its results are displayed in appendix figures 10 and 11. Since this specification controls for age effects in innovations to the permanent component, the parameters $\left(\sigma_{\alpha\beta}, \sigma_\beta^2\right)$ are only identified from the slope of lag-profiles. The only difference to the full model is that the persistence ρ does not have any bite on the lag-profiles when the AR(1)-process is assumed to be degenerate at labor market entry. As a consequence, $\sigma_{\alpha\beta}$ is the only parameter that can match the decline of autocovariance in its order for the group of labor market entrants. This is clearly reflected in strikingly different counterfactual effects of setting $\sigma_{\alpha\beta}$ to zero in the full model and the model without a persistent initial condition, as can be seen when comparing the first panels of figure 5 and appendix figure 11. In particular, while most of the initial decline of lag-profiles for this group is explained by the persistence parameter ρ in the full model, the entire decline is matched by $\sigma_{\alpha\beta}$ in the more restrictive model. Yet, for older cohorts the counterfactual effects are quite similar, which can only be explained by $\widehat{\sigma}_\beta^2$ adjusting to counterbalance the strong effect of the large $|\widehat{\sigma}_{\alpha\beta}|$. The numerical counterfactuals for experience-profiles documented in appendix figure 10 show that restricting $\sigma_{\xi 0}^2 = 0$ a-priori has also major effects on the variation identifying various other parameters. Most importantly, while the model generates highly non-linear experience effects, the removal of the AR(1)-process altogether while holding everything else constant generates experience profiles that are nearly linear. This is not true in the full model where the initial decline of these profiles is almost entirely matched by the persistent initial condition, as documented in figure 4. Hence, even slight modifications of the model specification can signficantly alter any of the parameter estimates.

5.4. The Effect of imposing $cov(\alpha_i, \beta_i) = 0$

Inspection of the theoretical auto-covariance structure in (4.9) together with the results from the numerical comparative statics exercises suggest that $\sigma_{\alpha\beta}$ is firmly identified from the behavior of lag-profiles at higher orders. In contrast, the parameter capturing heterogeneity in slopes σ^2_β seems to play the role of freeing up $\sigma_{\alpha\beta}$. One may interpret this result as "overfitting" - a situation where a parameter is identified from purely statistical artifacts that have no intrinsic economic meaning. This can also explain why $\sigma_{\alpha\beta}$ is always found to be highly significant and precisely estimated, even in cases where a specification yields no significant estimates of slope heterogeneity. An implication of this hypothesis is that σ^2_β should be insignificant once one imposes the restriction $\sigma_{\alpha\beta} = 0$ a priori. To test this conjecture, I reestimate all specifications shown in table 2, but imposing the restriction $\sigma_{\alpha\beta} = 0$. Results are displayed in table 3. The results are striking - with the exception of the rudimentary HIP-process, the estimates $\widehat{\sigma}^2_\beta$ are statistically indistinguishable from zero in all specifications. In four out of six cases, the estimate hits the boundary of zero, powerfully demonstrating that the parameter does not help improve the model fit in any dimension.[44] The relative decline in the model fit from imposing $\sigma_{\alpha\beta} = 0$ relative to the unrestricted case is particulary large in model specifications where $\widehat{\sigma}^2_\beta$ was estimated to be large - the HIP-model and the full model without a persistent initial condition.

5.5. How Robust are the Conclusions: Results from the Low Education Group

In this section I replicate the empirical analysis using 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, thus enabling me to explore the robustness of my results. Second, while the preferred model matches well the autocovariance structure of the more educated, it is clearly ill-specified for the high-school dropouts, as shown in appendix figure 12.[45] Instead of modifying the model to improve its fit - a promising approach would be to allow all parameters

[44]To compute the standard errors for the other parameters I re-estimate the model in this case imposing $\sigma^2_\beta = 0$.

[45]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.

to vary freely across cohort groups - I investigate whether the conclusions drawn from the main sample hold when one starts from a misspecified model. Parameter estimates for various specifications are shown in columns 4 to 6 of table 1 and the lower panels of tables 2 and 3. Including the factor loadings, there are 66 parameters that are estimated on a sample of 64,278 moments. There are two major differences in parameter estimates of the full model compared to results from the main sample. 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, a result that is robust to the exclusion of age-effects in the variance of permanent innovations. Second, the persistent component as captured by the heteroscedastic AR(1)-process plays a significantly larger role. Most importantly, the estimated initial condition of the AR(1)-process is much larger than in the secondary-degree sample, consistent with the higher intercepts of lag profiles in the dropout-sample as documented in appendix figures 1 and 2. Given the steep initial decline of 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, consistent with a large persistent initial condition of the earnings process. Other parameters such as the estimated variance of the intercept σ_α^2 and the persistence of the AR(1)-process ρ are surprisingly similar to those from the secondary-degree sample. Estimated factor loadings for the full model are shown in appendix figure 5. As in the main sample, the increase in within-group income inequality is almost entirely driven by an increase in the variance of the persistent rather than the permanent component.

The robustness exercises documented in the lower panel of table 2 reveal patterns that are remarkably similar to those found in the main sample. Most importantly, 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. In fact, any of the specifications shown in the table that yield significant estimates for slope heterogeneity are associated with large intercept heterogeneity as well, thus reproducing the strong correlation between $\widehat{\sigma}_{\alpha\beta}$ and $\widehat{\sigma}_\alpha^2$ *across* specifications uncovered in the main sample. Furthermore, comparing results in columns 4 to 6 clarifies that the sensitivity of estimates of slope heterogeneity $\left(\sigma_\beta^2, \sigma_{\alpha\beta}\right)$ is almost entirely driven by the exclusion of the persistent initial condition.

To further explore the sources of identifying variation for a subset of parameters, I again rely on comparative statics exercises that generate counterfactual covariance structures from various scenarios. Results are displayed in appendix figures 13 to 18. Like in the secondary-degree sample, most of the earnings dynamics early in the life-cycle are explained by the persistent initial condition together with the heteroscedastic AR(1)-process. As shown in appendix figure 13, setting the parameter estimates of these components to zero would miss the entire decline of the experience profiles for young workers. In the full model in which the HIP-component is prefectly multi-collinear with the heteroscedastic random walk, it is the slope of lag-profiles at high orders that identifies $\left(\widehat{\sigma}_{\alpha\beta}, \widehat{\sigma}_\beta^2\right)$. Given these profiles converge to a constant, it is not surprising that slope-heterogeneity is found to be unimportant, which is mirrored in the negligible counterfactual effect on the model-generated lag-profiles displayed in appendix figure 14. This counterfactual effect is strikingly different when using the estimates from a simple HIP-model, as shown in the next two appendix figures. The high auto-covariances early in the life-cycle are now matched by both, a high dispersion in intercepts across individuals and the variance in the AR(1)-shock that has an initial condition of zero. The subsequent decline observed in the data is partially generated by a negative estimate of $\sigma_{\alpha\beta}$. At the same time, since a negative $\sigma_{\alpha\beta}$ is essentially the only parameter that can match the strength of the decline of lag-profiles it is not surprising that removing it generates counterfactual profiles that are flat early in the life-cycle. Again, slope heterogeneity σ^2_β seems to free up the parameter $\sigma_{\alpha\beta}$ by counterbalancing its effect on lag-profiles later in the life-cycle. As can be seen from the last two appendix figures, similar conclusions are reached when replicating these counterfactuals, but starting from a model that differs from the benchmark specification only by excluding the persistent initial condition. Similar to the findings from the secondary-degree sample, the counterfactual exercises indicate that $\widehat{\sigma}^2_\beta$ merely allows the parameter $\sigma_{\alpha\beta}$ to improve the fit of the model early in the life-cycle without deteriorating it for older individuals. To explore this hypothesis further, I also estimate various models that impose a-priori the restriction $\sigma_{\alpha\beta} = 0$ for the dropout-sample. Results are shown in the lower panel of Table 3. Again it is found that $\widehat{\sigma}^2_\beta$ is equal to zero in any specification once one does not allow for a correlation of intercept- and slope-heterogeneity.

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

6. Slope Heterogeneity over the Life-Cycle

A potential concern with the lack of robust evidence for slope heterogeneity is that the specification of this component may be overly restrictive. A reasonable hypothesis is that heterogeneity in slopes is particularly important early in the life-cycle when average growth rates in earnings are the largest. I test this hypothesis by introducing a spline in potential labor market experience, thus allowing for a different distribution of returns to experience at different parts of the life-cycle. Defining H as a cutoff point that separates the life-cycle into an early and a late stage with potentially different amounts of heterogeneity in growth rates, I consider the following specification for β_i that explicitely depends on experience h_{ibt} :

$$\beta_i = \begin{cases} \beta_{i,1} \text{ if } & h_{ibt} \leq H \\ \beta_{i,2} \text{ if } & h_{ibt} > H \end{cases} \tag{6.1}$$

with first and second moments of their joint distribution given by

$$E(\beta_{i,1}) = E(\beta_{i,2}) = 0; var(\beta_{i,1}) = \sigma^2_{\beta,1}; var(\beta_{i,2}) = \sigma^2_{\beta,2}; cov(\beta_{i,1}, \beta_{i,2}) = \sigma_{1,2}. \tag{6.2}$$

This model nests the specification considered above by allowing $\sigma_{1,2} \neq 1$. To rule out estimates of these parameters to be driven by the overfitting documented above, I impose the restriction $cov(\alpha_i, \beta_{i,1}) = cov(\alpha_i, \beta_{i,2}) = 0$. This specification allows individual-specific slopes to be correlated over the life-cycle while restricting intercepts and slopes to be uncorrelated.[46]

Results for both samples and for different cutoff levels H are shown in appendix table 3. Comparing the results with those documented in column 1 of Table 3 clarifies that estimates of intercept heterogeneity $\widehat{\sigma}^2_\alpha$ and persistence of the AR(1)-process $\widehat{\rho}$ are very similar and do not depend on the choice of H. Furthermore, the goodness of fit increases marginally, if at all. Interestingly, in none of the samples is there evidence for a significant positive correlation between slopes early and late in the life-cycle. In fact, for the secondary degree sample the opposite is true as the results point towards a strong negative correlation.[47] The variance of slopes in the population is found to be insignificant in the second stage of the life-cycle no matter the sample or the choice of H, often hitting the non-negativity constraint. The same is true for $\widehat{\sigma}^2_{\beta,1}$, with the exception of the secondary-degree sample when $H = 40$. Taken together, these results suggest that there is only weak evidence for slope heterogeneity early in the life-cycle and no evidence for slope heterogeneity late in the life-cycle.

7. Conclusion

There is wide disagreement about the sources of life-cycle earnings dynamics and the quantitative importance of risk and worker heterogeneity for earnings inequality. In this study I argue this disagreement is driven by specification error. Starting from a parametric process that flexibly models the evolution of the autocovariance structure of earnings over the life-cycle and over time, and relying on a numerical exploration of model-specific identifying variation, I conduct the first systematic study of robustness of parameter estimates across specifications. My preferred model fits an empirical autocovariance matrix computed from German administrative data, containing over 56,000 elements, almost perfectly with just 65 parameters.

I find that estimates of key parameters, such as the persistence of exogenous shocks or the heterogeneity in returns to experience, vary widely across specifications. At the same time, a number of findings are qualitatively robust. First, heterogeneity in average residual earnings is always highly significant. Second, both permanent and persistent shocks are important features of earnings dynamics, implying that allowing for only one type of non-transitory shock, a common restriction in HIP- and RIP specifications, is arbitrary and likely to yield biased estimates of *all* model parameters. On the other hand, there is no evidence for transitory shocks. Third, the rich dynamics of earnings early in the life-cycle, similar to those documented for the PSID, can almost entirely be

[46] The terms involving the σ^2_β in (4.9) are now given by

$$h_{ibt} * (h_{ibt} + k) * \left[\begin{array}{c} \sigma^2_{\beta,1} * 1(h_{ibt} \leq H \cup (h_{ibt} + k) \leq H) \\ +\sigma^2_{\beta,2} * 1(h_{ibt} > H \cup (h_{ibt} + k) > H) \\ +\sigma_{1,2} * \begin{pmatrix} 1 - 1(h_{ibt} \leq H \cup (h_{ibt} + k) \leq H) \\ -1(h_{ibt} > H \cup (h_{ibt} + k) > H) \end{pmatrix} \end{array} \right] \tag{6.3}$$

instead of $h_{ibt} * (h_{ibt} + k)\, \sigma^2_\beta$.

[47] Equation (6.3) in the previous footnote suggests this result to be driven by the fact that $\sigma_{1,2}$ enters covariance terms only that are sufficiently far apart, thus partially matching the slope of lag-profiles at high orders.

explained by a persistent rather than a permanent initial condition. This is in sharp contrast with recent research on HIP-models that interprets the documented initial decline of autocovariances as evidence in favor of slope heterogeneity. In fact, as shown in this paper, once age-effects in innovation variances are properly controlled for, slope heterogeneity cannot be identified from experience-profiles and one needs to rely on the slope of *lag-profiles* at high orders. This underlines the importance of estimating earnings processes from administrative data that follow individuals for a long time since profile heterogeneity needs to be identified from data features that are strongly affected by sample attrition in survey data.

Regarding the question of which type of earnings process, RIP or HIP, is favored by the data I find that estimates of profile heterogeneity are extremely sensitive to minor modifications in the model specifications. While a standard HIP-specification produces strong evidence in favor of slope heterogeneity, its significance vanishes once I move towards my preferred model. I identify the omission of a particular age effect in innovation variances, the initial condition of the AR(1)-process, to be the main driving force of this sensitivity. I also present results that suggest estimates of slope heterogeneity to be suffering from "overfitting" as it merely helps to free up other parameters. This result remains valid when considering a specification that allows slopes to vary over the life-cycle and when reestimating all models using data for an education group with a very different covariance structure.[48]

One may be concerned that my results apply only to the German labor market. However, 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. Furthermore, I reach similar conclusion if I rely on an empirical autocovariance structure with very different properties, suggesting that my results apply more generally. A second concern may be that my preferred specification is too restrictive. For example, while my model is unusually rich in its formulation of age- and time-effects and various initial conditions, it does not allow for the type of heterogeneity considered in Browning et al (2010). However, an extensive exploration of the model fit shows that it matches all features of cohort-specific autocovariance structures apart from sampling error. Therefore, allowing variance components to depend on the observables "calendar time" and "labor market experience" can go a long way in controlling for heterogeneity in earnings dynamics and in generating rich variance dynamics. This can be seen as good news for heterogeneous agent modeling as my model has a fairly small number of state variables.

While earnings processes are purely statistical, their empirical estimates point towards economic models that are potentially most successful in explaining observed life-cycle labor market dynamics. As documented in this study, a persistent initial condition is crucial for explaining the earnings dynamics of those who are at the beginning of their career. As shown in Adda et al (2011) and Hoffmann (2010) this is the period of the life-cycle

[48] A remaining question is why Baker and Solon (2003), whose model is very similar to the one estimated here, find significant heterogeneity in slopes. One possible answer is that they do not explore its sensitivity when restricting the covariance between intercept- and slope-heterogeneity to be zero. I show this restriction to be crucial for understanding the identification of profile heterogeneity.

in which workers are particularly mobile across firms and occupations. Research by von Wachter and Bender (2006) and Oreopoulos et al. (2012) document evidence that labor market entrants are particularly vulnerable to aggregate economic changes. Card et al (2013) find that a significant part of the recent increase in German income inequality can be explained by a firm-specific component and a change in the matching process between firms and workers. As initial placement can be interpreted as a persistent initial condition and firm mobility as persistent shocks, these findings suggest that a focus on human capital models may be too narrow and that structural work that attempts to unify human capital theory and search theory is particularly promising for enriching our understanding of labor market dynamics.

References

[1] Abbott, B., G. Gallipoli, C. Meghir and G. Violante (2013): "Education Policies and Intergenerational Transfers in Equilibrium." mimeo, University of British Columbia.

[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] 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.

[5] Altonji, J., A. Smith Jr. and I. Vidangos (2013): "Modeling Earnings Dynamics", *Econometrica*, forthcoming.

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

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

[8] 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.

[9] 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.

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

[11] 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.

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

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

[14] Card, D., J. Heining and P. Kline (2013): “Workplace Heterogeneity and the Rise of West German Wage Inequality,“ *Quarterly Journal of Economics*, forthcoming.

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

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

[17] Erosa, A., G. Kambourov and L. Fuster (2012): ”Towards a Micro-Founded Theory of Aggregate Labor Supply.” mimeo, University of Toronto.

[18] Farhi, E. and I. Werning (2012): ”Insurance and Taxation over the Life-Cycle.” *Review of Economic Studies*, forthcoming.

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

[20] 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.

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

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

[23] Guvenen, F. and A. Smith (2010): ”Inferring Labor Income Risk from Economic Choices: An Indirect Inference Approach.” mimeo, University of Minnesota.

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

[25] Hagedorn, M. and I. Manovskii (2010): ”Search Frictions and Wage Dispersion.” mimeo, University of Minnesota.

[26] 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.

[27] 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.

[28] 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.

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

[30] 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.

[31] 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.

[32] Heathcote, J., K. Storesletten, and G. L. Violante (2012): "Consumption and Labor Supply with Partial Insurance: An Analytical Framework", mimeo, New York University.

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

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

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

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

[37] Huggett, M and G. Kaplan (2012): "The Money Value of a Man." mimeo, Georgetown University.

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

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

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

[41] 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.

[42] 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.

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

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

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

[46] 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.

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

[48] 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.

[49] 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.

[50] 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.

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

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

[53] 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			Dropout Group		
		(1.)	(2.)	(3.)	(4.)	(5.)	(6.)
		Full Model	**Full Model**, *homosc. unit roots*	**Full Model**, *no Slope Het.*	**Full Model**	**Full Model**, *homosc. unit roots*	**Full Model**, *no Slope Het.*
Intercept Heterogeneity	σ_{α}^{2}	0.031 (0.002)***	0.023 (0.004)***	0.013 (0.0005)***	0.024 (0.008)***	0.027 (0.007)***	0.022 (0.006)***
Slope Heterogeneity	$\sigma_{\beta}^{2} * 10^{3}$	0.002 (0.001)**	0.0005 (0.0014)	-	0.001 (0.004)	0.000 -	-
Cov (Intercept; Slope)	$\sigma_{\alpha\beta}^{2} * 10$	-0.003 (0.0003)***	-0.001 (0.0003)***	-	-0.001 (0.001)	-0.0004 (0.0001)	-
Persistence of AR(1)	ρ	0.863 (0.006)***	0.880 (0.006)***	0.906 (0.005)***	0.884 (0.008)***	0.883 (0.009)***	0.886 (0.006)***
AR(1) error structure							
Initial Condition	$\sigma_{\xi 0}^{2}$	0.057 (0.005)***	0.092 (0.014)***	0.080 (0.005)***	0.292 (0.025)***	0.283 (0.032)***	0.293 (0.028)***
Intercept	γ_0	0.003 (5.22*e(-4))***	0.007 (0.002)***	0.004 (5.01*e(-4))***	0.044 (0.004)***	0.044 (0.005)***	0.044 (0.004)***
experience	γ_1	-1.5*e(-4) (3.7*e(-5))***	-3.16*e(-4) (1.17*e(-4))***	-1.63*e(-4) (3.91*e(-5))***	-0.003 (2.51*e(-4))***	-0.003 (4*e(-4))***	-0.003 (2.63*e(-4))***
experience^2	γ_2	3.56*e(-6) (1.15*e(-6))***	7.68*e(-6) (3.42*e(-6))**	3.56*e(-6) (1.40*e(-6))**	8.07*e(-5) (6.62*e(-6))***	7.78*e(-5) (9.27*e(-6))***	8.01*e(-5) (7.31*e(-6))***
experience^3	γ_3	-3.63*e(-8) (1.48*e(-8))**	-9.16*e(-8) (4.63*e(-8))**	-3.52*e(-8) (1.97*e(-8))*	-9.06*e(-7) (7.62*e(-8))***	-8.65*e(-7) (1.05*e(-7))***	-8.97*e(-7) (8.8*e(-8))***
experience^4	γ_4	1.33*e(-10) (6.63*e(-11))**	3.95*e(-10) (2.17*e(-10))*	1.29*e(-10) (9.46*e(-11))	3.66*e(-9) (3.23*e(-10))***	3.48*e(-9) (4.29*e(-10))***	3.62*e(-9) (3.84*e(-10))***
Random Walk error structure							
Intercept	$\delta_0 * 10$	0.013 (0.002)***	0.007 (0.001)***	0.010 (0.001)***	0.022 (0.01)**	0.003 (0.001)**	0.020 (0.012)*
experience	$\delta_1 * 10^{3}$	-0.002 (0.021)	-	-0.016 (0.016)	-0.133 (0.119)	-	-0.119 (0.129)
experience^2	$\delta_2 * 10^{5}$	-0.038 (0.095)	-	0.003 (0.076)	0.318 (0.475)	-	0.281 (0.517)
experience^3	$\delta_3 * 10^{7}$	0.030 (0.163)	-	0.012 (0.132)	-0.356 (0.766)	-	-0.309 (0.831)
experience^4	$\delta_4 * 10^{9}$	0.002 (0.092)	-	-0.006 (0.075)	0.161 (0.421)	-	0.137 (0.456)
Measurement Error	σ_{ε}^{2}	0.000 -	0.004 (0.001)***	0.000 -	0.001 (0.001)	0.000 -	0.001 (0.001)
Number of Moments			**56,072**			**64,278**	
R^2		0.977	0.964	0.967	0.862	0.859	0.861
Wald Test for Slope Heterogeneity (P-Value)		0.000	0.000	-	0.657	0.659	-
Wald Test for Heteroscedastic Unit Roots (P-Value)		0.000	-	0.000	0.367	-	0.3

NOTES: This table shows parameter estimates for the benchmark specification as described in equations (4.2) to (4.8) of the paper, together with two more restrictive, but nested, models. The "secondary degree group" includes individuals who have a high-school and a vocational degree, while the "dropout group" includes individuals without a formal degree. All specifications allow for factor loadings on the permanent and persistent component, all of which are found to be signficant on the 1%-level. Estimated factor loadings for the full model are displayed in figure 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.

TABLE 2 - ROBUSTNESS OF PARAMETER ESTIMATES: FULL MODEL

PANEL A: SECONDARY DEGREE GROUP

		BENCHMARK SPECIFICATIONS			RESTRICTIONS ON FULL MODEL			
		(1.)	(2.)	(3.)	(4.)	(5.)	(6.)	(7.)
		Full Model	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}	0.031 (0.002)***	0.053 (0.004)***	0.027 (0.002)***	0.015 (0.001)***	0.030 (0.002)***	0.038 (0.003)***	0.051 (0.004)***
Slope Heterogeneity	$\sigma_{\beta}^{2}*10^{3}$	0.002 (0.001)**	0.006 (0.003)***	-	0.000 (0.001)	0.002 (0.001)*	0.004 (0.001)***	0.005 (0.002)***
Cov (Intercept; Slope)	$\sigma_{\alpha\beta}^{2}*10$	-0.003 (0.0003)***	-0.005 (0.0008)***	-	-0.0006 (0.0001)***	-0.002 (0.0003)***	-0.004 (0.001)***	-0.005 (0.0007)***
Persistence of AR(1)	ρ	0.863 (0.006)***	0.996 (0.004)***	0.982 (0.003)***	0.905 (0.004)***	0.868 (0.005)***	0.768 (0.001)***	0.757 (0.017)***
AR(1) error structure								
Initial Condition	$\sigma_{\xi 0}^{2}$	0.057 (0.005)***	-	-	0.081 (0.007)***	0.119 (0.011)***	-	-
Intercept	γ_0	0.003 (5.22*e(-4))***	0.001 (0.0006)*	0.002 (0.0002)***	0.002 (4.31*e(-4))***	0.016 (0.003)***	0.0199 (0.002)***	0.011 (0.002)***
experience	γ_1	-1.5*e(-4) (3.7*e(-5))***	-	-	-	-7.16*e(-4) (2.67*e(-4))***	-0.002 (1.64*e(-4))***	-
experience^2	γ_2	3.56*e(-6) (1.15*e(-6))***	-	-	-	1.53*e(-5) (1.05*e(-5))	4.06*e(-5) (5.36*e(-6))***	-
experience^3	γ_3	-3.63*e(-8) (1.48*e(-8))**	-	-	-	-1.32*e(-7) (1.62*e(-7))	-4.56*e(-7) (7.27*e(-8))***	-
experience^4	γ_4	1.33*e(-10) (6.63*e(-11))**	-	-	-	3.86*e(-10) (8.41*e(-10))	1.77*e(-9) (3.47*e(-10))***	-
Random Walk error structure								
Intercept	δ_0*10	0.013 (0.002)***	-	-	0.003 (0.001)***	0.011 (0.002)***	0.001 (0.002)	0.011 (0.0009)***
experience	δ_1*10^{3}	-0.002 (0.021)	-	-	-	0.026 (0.036)	0.090 (0.002)***	-
experience^2	δ_2*10^{5}	-0.038 (0.095)	-	-	-	-0.132 (0.169)	-0.313 (0.119)***	-
experience^3	δ_3*10^{7}	0.030 (0.163)	-	-	-	0.164 (0.297)	0.345 (0.217)	-
experience^4	δ_4*10^{9}	0.002 (0.092)	-	-	-	-0.059 (0.171)	-0.115 (0.129)	-
Measurement Error	σ_{ε}^{2}	0.000 -	0.018 (0.008)**	0.026 (0.003)***	0.000 -	0.005 (0.0003)***	0.003 (0.0004)***	0.005 (0.0006)***
Number of Moments					**56,072**			
R^2		0.977	0.764	0.592	0.919	0.840	0.910	0.738
Wald Test for Slope Heterogeneity (P-Value)		0.000	0.000	-	0.000	0.000	0.000	0.000

PANEL B: DROPOUT GROUP

		BENCHMARK SPECIFICATIONS			RESTRICTIONS ON FULL MODEL			
		(1.)	(2.)	(3.)	(4.)	(5.)	(6.)	(7.)
		Full Model	**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}	0.024 (0.008)***	0.134 (0.011)***	0.041 (0.003)***	0.039 (0.013)***	0.025 (0.005)***	0.118 (0.024)***	0.134 (0.01)***
Slope Heterogeneity	$\sigma_{\beta}^{2} * 10^{3}$	0.001 (0.004)	0.026 (0.003)***	-	0.004 (0.002)**	0.001 (0.004)	0.021 (0.008)***	0.027 (0.003)***
Cov (Intercept; Slope)	$\sigma_{\alpha\beta}^{2} * 10$	-0.001 (0.001)	-0.016 (0.002)***	-	-0.003 (0.0015)*	-0.002 (0.001)**	-0.013 (0.003)***	-0.016 (0.002)***
Persistence of AR(1)	ρ	0.884 (0.008)***	0.784 (0.027)***	0.808 (0.013)***	0.921 (0.006)***	0.871 (0.009)***	0.832 (0.015)***	0.756 (0.027)***
AR(1) error structure								
Initial Condition	$\sigma_{\xi 0}^{2}$	0.292 (0.025)***	-	-	0.273 (0.035)***	0.403 (0.03)***	-	-
Intercept	γ_0	0.044 (0.004)***	0.032 (0.017)*	0.039 (0.005)***	0.003 (0.001)***	0.111 (0.009)***	0.077 (0.015)***	0.038 (0.003)***
experience	γ_1	-0.003 (2.51*e(-4))***	-	-	-	-0.008 (0.001)***	-0.006 (0.001)***	-
experience^2	γ_2	8.07*e(-5) (6.62*e(-6))***	-	-	-	2.1*e(-4) (5.38*e(-5))***	1.54*e(-4) (3.22*e(-5))***	-
experience^3	γ_3	-9.06*e(-7) (7.62*e(-8))***	-	-	-	-2.48*e(-6) (8.1*e(-7))***	-1.75*e(-6) (3.83*e(-7))***	-
experience^4	γ_4	3.66*e(-9) (3.23*e(-10))***	-	-	-	1.05*e(-8) (4.08*e(-9))**	7.15*e(-9) (1.63*e(-9))***	-
Random Walk error structure								
Intercept	$\delta_0 * 10$	0.022 (0.01)**	-	-	0.000 -	0.017 (0.006)***	0.017 (0.023)	0.000 -
experience	$\delta_1 * 10^{3}$	-0.133 (0.119)	-	-	-	-0.093 (0.092)	-0.110 (0.296)	-
experience^2	$\delta_2 * 10^{5}$	0.318 (0.475)	-	-	-	0.203 (0.437)	0.276 (1.223)	-
experience^3	$\delta_3 * 10^{7}$	-0.356 (0.766)	-	-	-	-0.162 (0.776)	-0.315 (1.931)	-
experience^4	$\delta_4 * 10^{9}$	0.161 (0.421)	-	-	-	0.042 (0.455)	0.135 (1.035)	-
Measurement Error	σ_{ε}^{2}	0.001 (0.001)	0.004 (0.003)	0.000 (0.003)	0.023 (0.004)***	0.001 (0.001)	0.000 -	0.000 -
Number of Moments					**64,278**			
R^2		0.862	0.433	0.212	0.689	0.735	0.750	0.429
Wald Test for Slope Heterogeneity (P-Value)		0.657	0.000	-	0.002	0.044	0.000	0.000

NOTES: This table explores the robustness of parameter estimates. Results for the benchmark specification as described in equations (4.2) to (4.8) are shown in column 1. 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). The "secondary degree group" includes individuals who have a high-school and a vocational degree, while the "dropout group" includes individuals without a formal degree. *** 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 3 - ESTIMATES OF KEY PARAMETERS IN MODELS WITH $\sigma_{\alpha\beta} = 0$

PANEL A: SECONDARY DEGREE GROUP

		BENCHMARK SPECIFICATIONS		RESTRICTIONS ON FULL MODEL			
		(1.)	(2.)	(3.)	(4.)	(5.)	(6.)
		Full Model	**AR(1) - HIP *(Guvenen)***	**Homoscedastic**	**Stationary**	**Zero initial condition for AR(1)**	**Models (4)-(6)** *combined* ***(Hryshko, stationary)***
Intercept Heterogeneity	σ_α^2	0.017 (0.003)***	0.032 (0.002)***	0.010 (0.0006)***	0.016 (0.0006)***	0.014 (0.0006)***	0.030 (0.002)***
Slope Heterogeneity	$\sigma_\beta^2 * 10^3$	0.000 -	0.004 (0.0003)***	0.000 -	0.000 -	0.000 -	0.001 (0.001)
Persistence of AR(1)	ρ	0.906 (0.005)***	0.949 (0.011)***	0.920 (0.003)***	0.909 (0.005)***	0.865 (0.010)***	0.895 (0.009)***
Number of Moments				**56,072**			
R^2		0.962	0.696	0.914	0.834	0.854	0.651

PANEL B: DROPOUT GROUP

		BENCHMARK SPECIFICATIONS		RESTRICTIONS ON FULL MODEL			
		(1.)	(2.)	(3.)	(4.)	(5.)	(6.)
		Full Model	**AR(1) - HIP *(Guvenen)***	**Homoscedastic**	**Stationary**	**Zero initial condition for AR(1)**	**Models (4)-(6)** *combined* ***(Hryshko, stationary)***
Intercept Heterogeneity	σ_α^2	0.019 (0.004)***	0.041 (0.003)***	0.012 (0.002)***	0.016 (0.002)***	0.039 (0.007)***	0.041 (0.003)***
Slope Heterogeneity	$\sigma_\beta^2 * 10^3$	0.0002 (0.004)	0.000 -	0.000 -	0.000 -	0.000 -	0.000 -
Persistence of AR(1)	ρ	0.886 (0.007)***	0.828 (0.015)***	0.937 (0.006)***	0.877 (0.006)***	0.854 (0.011)***	0.808 (0.011)***
Number of Moments				**64,278**			
R^2		0.862	0.207	0.657	0.735	0.725	0.212

NOTES: This table shows results for all specifactions considered in table 4, but with the covariance of intercept- and slope-heterogeneity restricted to zero. Since the RIP-process does not feature slope heterogeneity by definition, it is not considered in this exercise. *** 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

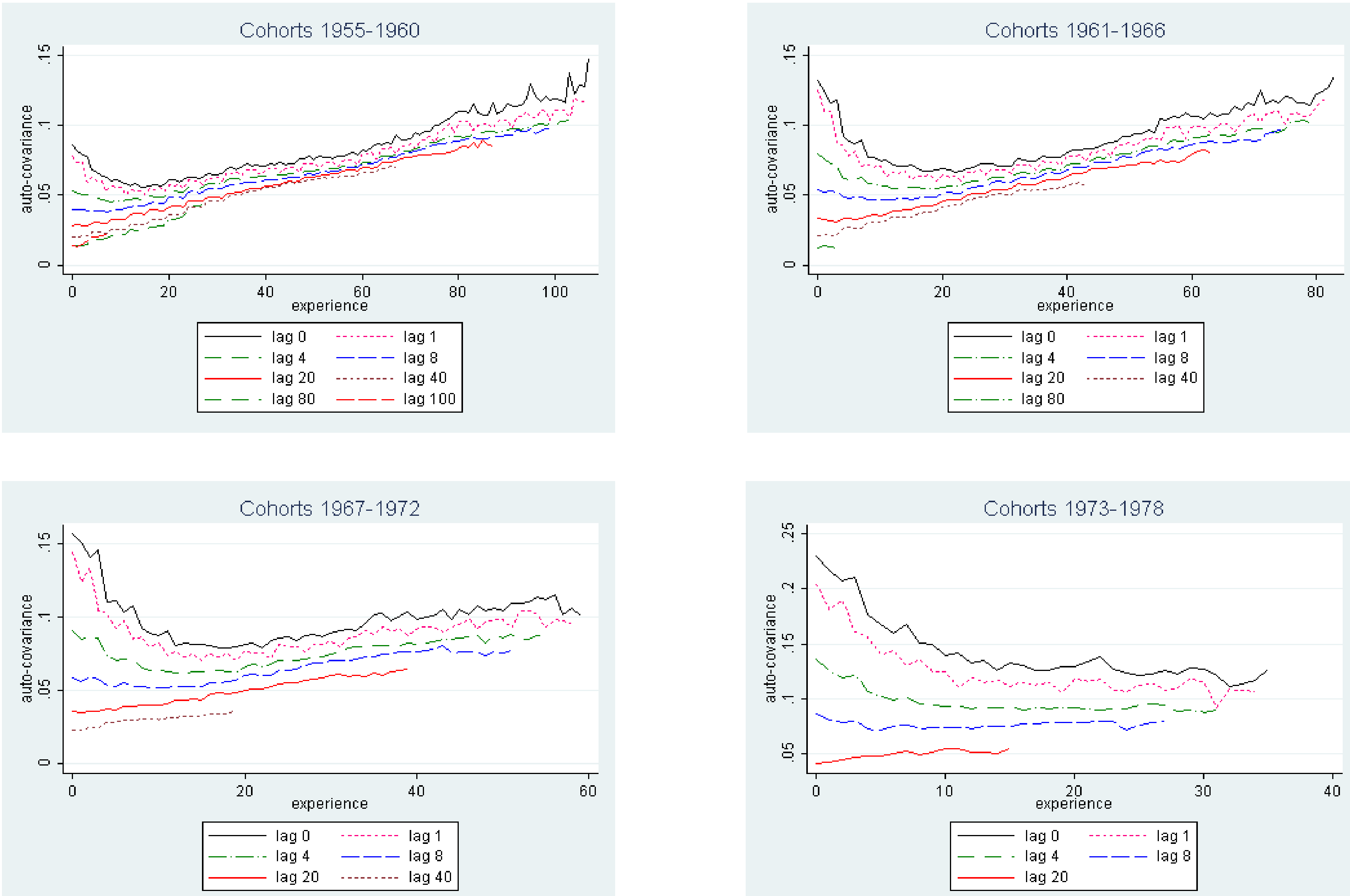


FIGURE 2 - LIFE-CYCLE PROFILES OF AUTO-COVARIANCES AT DIFFERENT LAGS, BY COHORTS

Sample: Dropout Group

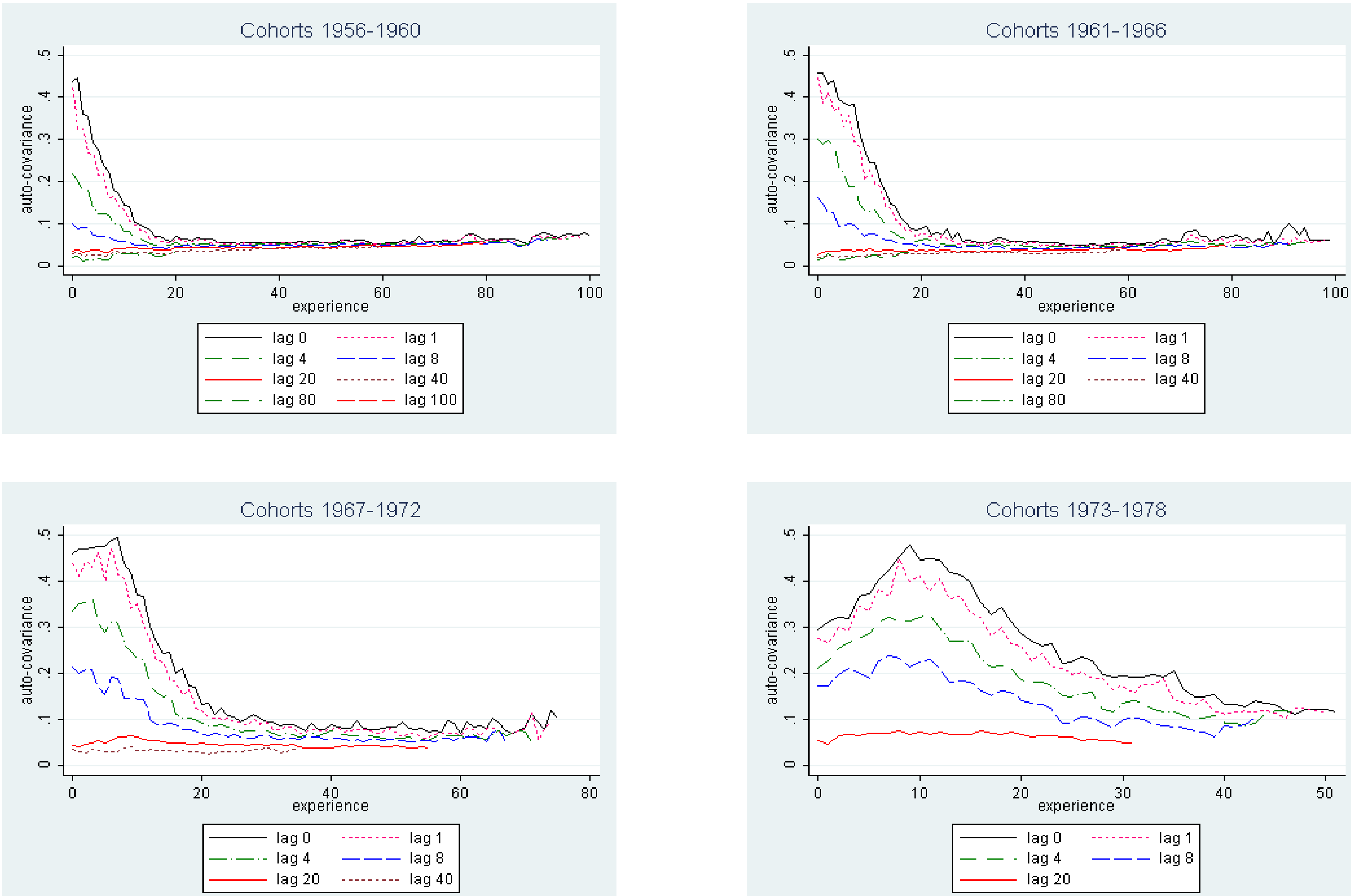


FIGURE 3 - FIT OF FULL MODEL: SECONDARY DEGREE GROUP

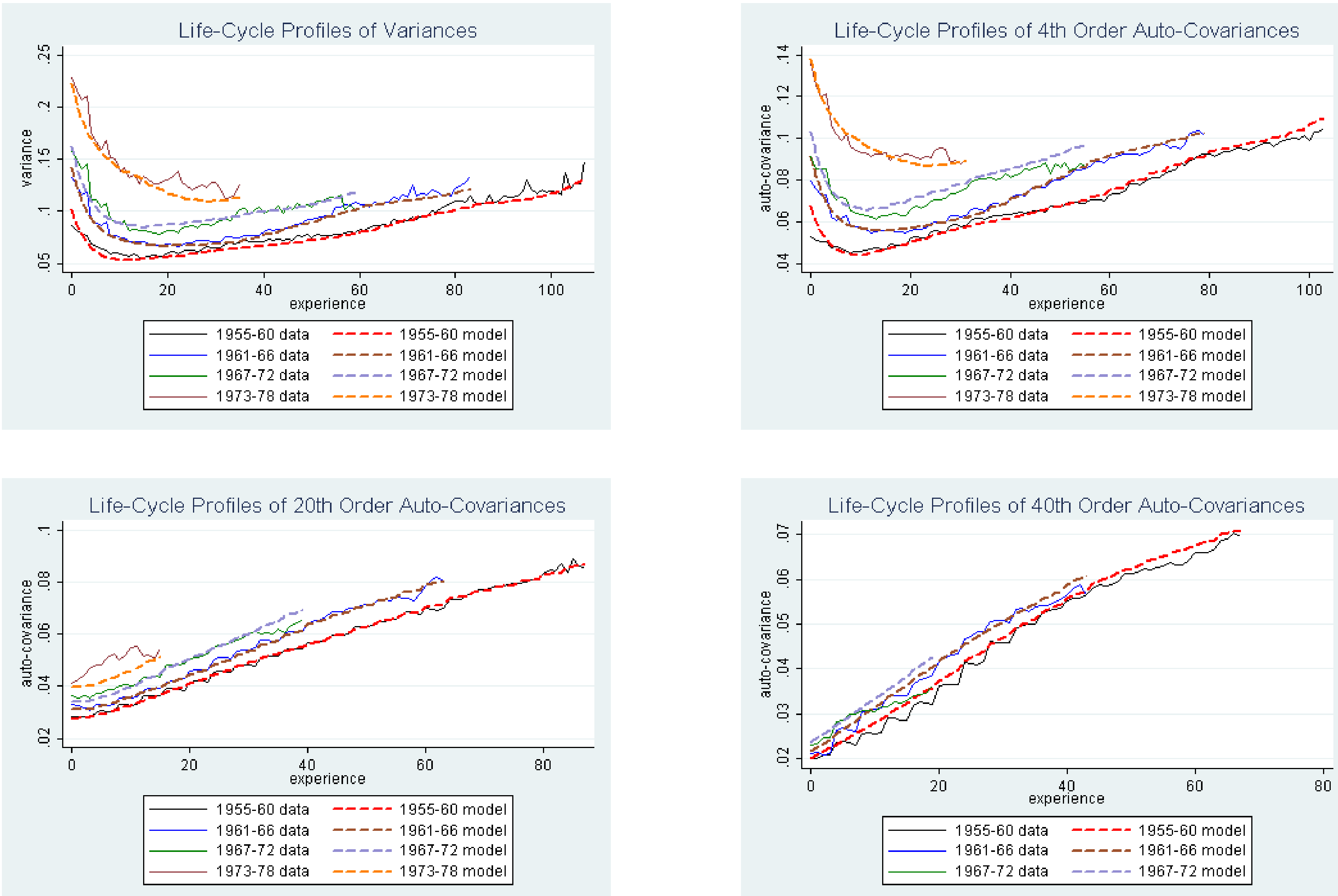
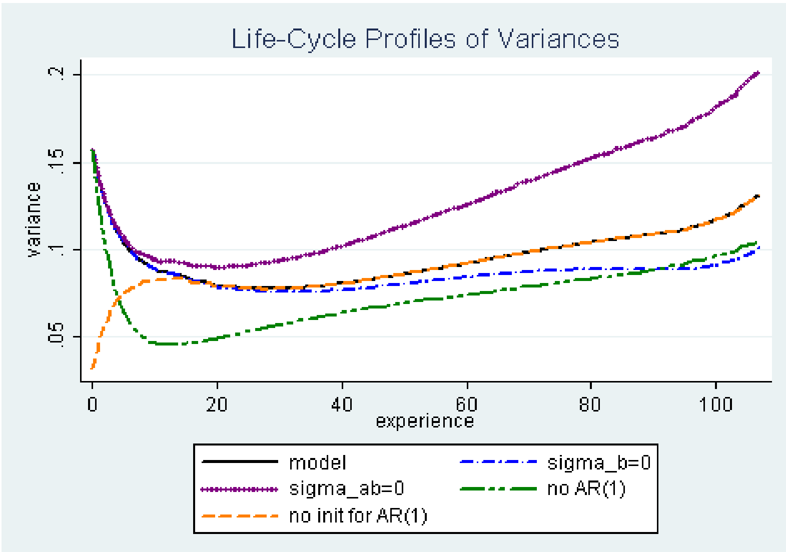
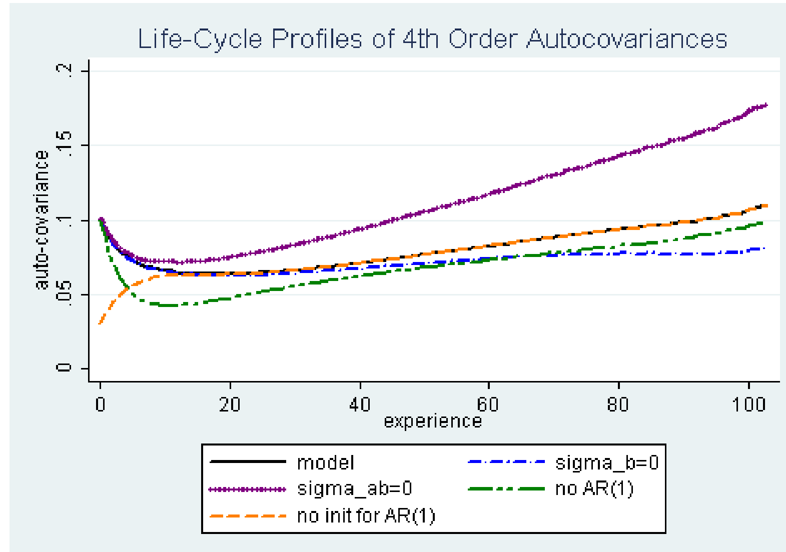
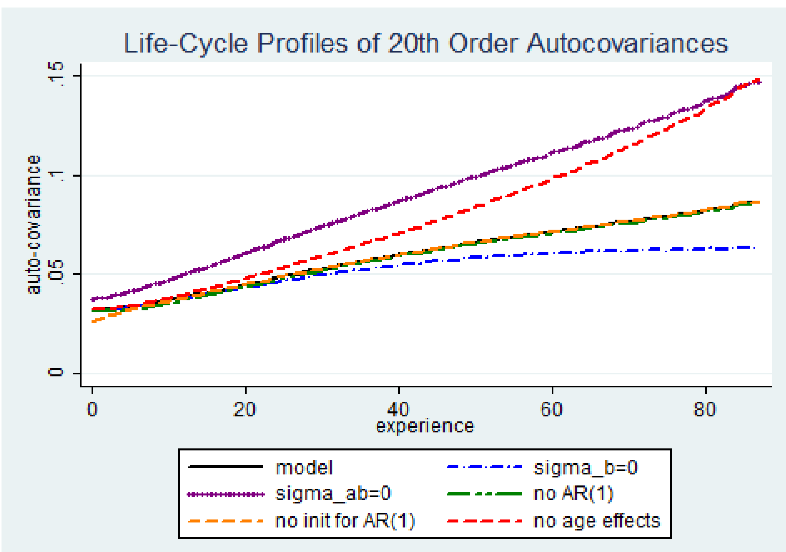


FIGURE 4 - COUNTERFACTUAL EXPERIENCE-PROFILES IN FULL MODEL: SECONDARY DEGREE GROUP







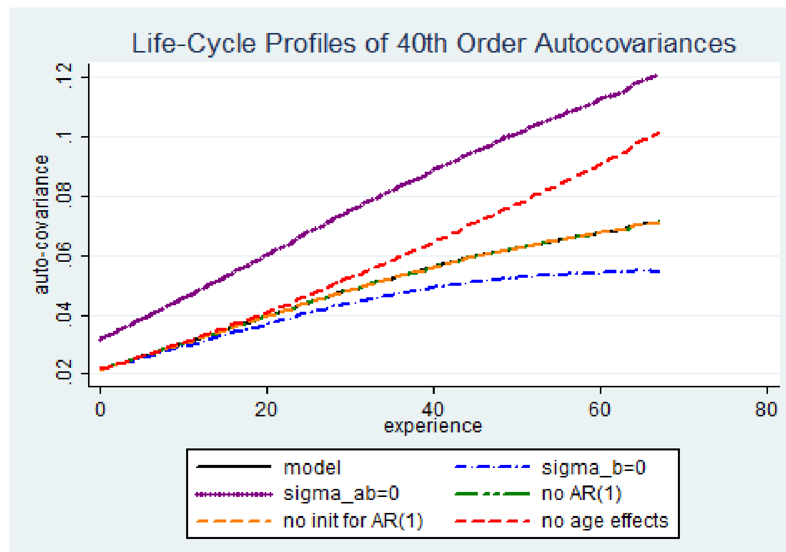
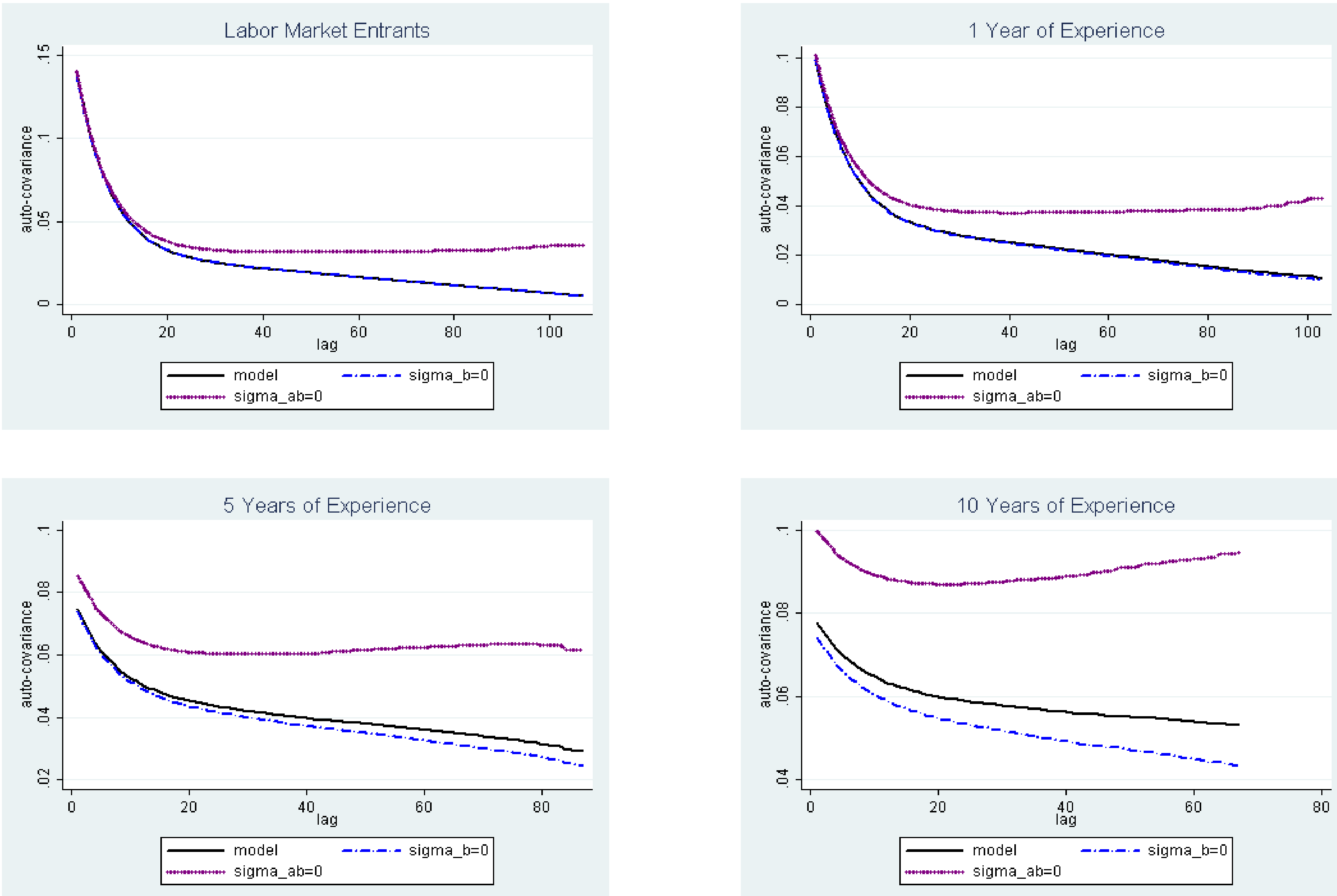


FIGURE 5 - COUNTERFACTUAL LAG-PROFILES IN FULL MODEL: SECONDARY DEGREE GROUP



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 to choose this approach rather than averaging all spells within a quarter is to avoid smoothing out productivity variation across jobs. 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. 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.[2] The fraction of censored observations varies strongly across education groups, providing a further motivation to estimate earnings processes for each group separately. The IABS provides an education variable with 6 categories, ranging from "no degree at all" to "university degree", which I aggregate up to three categories, "High-School Dropouts", "Secondary Degree" and "Some Post-Secondary Degree". While I drop the last group from the analysis because of its high fraction of top-coded earnings, censoring still needs to be addressed in the other two education groups. The standard approach in studies using the PSID, such as Meghir and Pistaferri (2004) and Hryshko (2012), is to drop top-coded earnings records, introducing a

[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 daily wage.

[2] 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.

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, those 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.[3] 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 alltogether.

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*.[4] 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 reesimate 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 on.

[3]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. Haider (2001), estimating earnings processes from the PSID, uses a static imputation/interpolation procedure as well for a subset of censored observations.

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

B. Standard Errors

If the MD-estimator $\widehat{\theta}$ is consistent, its asymptotic distribution is given by $N\left(\theta, \frac{1}{\sqrt{N}}V_\theta\right)$, where $N(.,.)$ is a joint Normal distribution and $\frac{1}{\sqrt{N}}V_\theta$ is the asymptotic covariance matrix of $\widehat{\theta}$. For EWMD, one can show that

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

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 $\dim(Z) \times \dim(\theta)$ - and Ω is the asymptotic covariance matrix of $\widehat{C}^{vec}$. To obtain standard errors for the estimates, one needs to estimate Ω consistently by computing $\widehat{\Omega} = \widehat{var}(\widehat{C}^{vec})$. Since $\widehat{C}^{vec}$ are autocovariances, $\widehat{\Omega}$ is the matrix of forth-order moments of residual log-wages. This matrix has size $\left[\dim(\widehat{C}^{vec})\right]^2$. Given the large number of elements in $\widehat{C}^{vec}$,over $56,000$ in the secondary-degree group and over $64,000$ in the dropout-group, and given the administrative nature of the data, computation of $\widehat{\Omega}$ is infeasible. I solve this problem by using the fact that the EWMD-estimator is a non-linear least-squares estimator (NLS), where one regresses autocovariances on the non-linear parametric function $G(\theta, Z)$. In analogy to heteroscedasticity-robust Huber-White standard errors or cluster-robust standard errors where one does not need to estimate the covariance matrix of the regression error to obtain standard errors of regression estimates, one can compute standard errors of $\widehat{\theta}$ without computing $\widehat{\Omega}$.

To see this, define the regression error $\widehat{\chi}_{btk} = \widehat{c}_{btk} - G\left(\widetilde{\theta}, Z_{btk}\right)$, where $\widehat{c}_{btk}$ is an element in $\widehat{C}^{vec}$ uniquely identified by cohort, year, and lag, and where $G\left(\widetilde{\theta}, Z_{btk}\right)$ is the theoretical counterpart. By definition, $\widehat{\theta}$ minimizes $\sum_{btk} \widehat{\chi}^2_{btk}$ and thus solves a standard (NLS)-estimation criterion. Since the independent variable $\widehat{c}_{btk}$ is a computed statistic, and since identification requires that $c_{btk} = G(\theta, Z_{btk})$, one should interpret χ_{btk} as a sampling error. By construction of the autocovariances, χ_{btk} cannot be independent across observations *within the same cohort* since the same residual wages enter the computation of multiple such moments to be matched. However, a wage observation never enters the computation of autocovariance structures for different cohorts. Hence, $\widehat{\chi}_{btk}$ is independent across cohorts, and one can obtain $\widehat{var}(\widehat{\theta})$ by using cluster-robust standard errors for (NLS), where clustering takes place on the cohort-level. This allows sampling error to be freely correlated within cohorts.[5]

[5] Heteroscedasticity robust standard errors and clustered standard errors also involve outer products of the vectors of sampling errors, which have the same dimension as the variance-covariance matrix of $\widehat{C}^{vec}$. However, as is well known, the "sandwich estimators" reduce the dimensionality of this problem.

C. 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 $var\left(\xi_{ibt}\right)$ 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.

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

PANEL A: SAMPLE SIZES BY EDUCATION GROUP AND COHORT

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

PANEL B: SAMPLE SIZES BY EDUCATION GROUP AND EXPERIENCE (IN YEARS)

EXPERIENCE (IN YEARS)	EDUCATION GROUP	
	High-School Dropouts	Secondary Degree
0	34,966	322,907
1	29,831	317,001
2	25,600	312,600
3	23,824	311,899
4	23,107	302,277
5	22,560	290,854
6	22,014	279,578
7	21,312	267,364
8	20,023	255,348
9	18,918	242,766
10	17,816	228,805
11	16,877	214,144
12	15,938	199,207
13	15,171	183,078
14	14,243	166,489
15	13,224	150,119
16	12,309	134,430
17	11,422	118,782
18	10,345	103,754
19	9,472	89,320
20	8,482	75,806
21	7,425	62,466
22	6,318	50,103
23	5,318	37,565
24	4,328	26,427
25	-	15,797
26	-	7,485
	414,231	4,752,287

APPENDIX TABLE 2 - AVERAGE LABOR INCOME BY EDUCATION GROUP AND EXPERIENCE (IN YEARS)

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

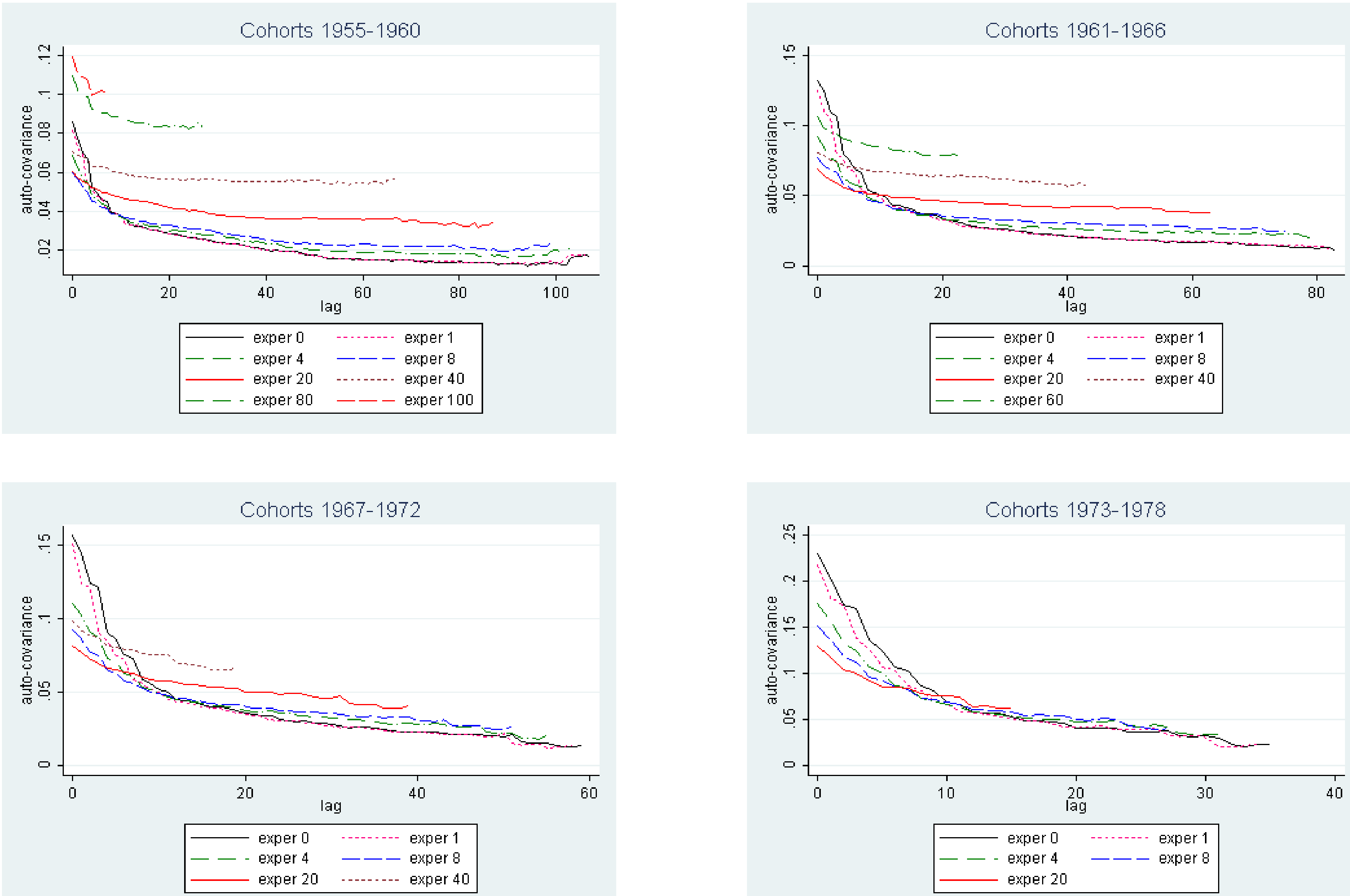
APPENDIX TABLE 3 - SLOPE HETEROGENEITY OVER THE LIFE-CYCLE

		Secondary Degree Group			Dropout Group		
		(1.)	(2.)	(3.)	(4.)	(5.)	(6.)
		H = 20 (quarters)	H = 40 (quarters)	H = 60 (quarters)	H = 20 (quarters)	H = 40 (quarters)	H = 60 (quarters)
Intercept Heterogeneity	σ_{α}^{2}	0.013 (0.001)***	0.013 (0.001)***	0.014 (0.001)***	0.020 (0.005)***	0.019 (0.004)***	0.019 (0.004)***
Slope Heterogeneity: Experience < H (in quarters)	$\sigma_{\beta,1}^{2}$	0.000 -	0.002 (0.0004)***	0.002 (0.0002)	0.001 (0.005)	0.001 (0.005)	0.003 (0.005)
Slope Heterogeneity: Experience >= H (in quarters)	$\sigma_{\beta,2}^{2}$	0.000 -	0.000 -	0.000 -	0.000 -	0.0003 (0.004)	0.0004 (0.004)
Covariance of Slope Het before and after H	σ_{12}	-0.003 (0.0003)***	-0.001 (0.0002)***	-0.001 (0.0001)***	-0.002 (0.001)	0.001 (0.004)	0.001 (0.004)
Persistence of AR(1)	ρ	0.903 (0.005)***	0.901 (0.005)***	0.903 (0.005)***	0.885 (0.006)***	0.887 (0.007)***	0.886 (0.007)***
Number of Moments			**56,072**			**64,278**	
R^2		**0.968**	**0.968**	**0.968**	**0.862**	**0.862**	**0.862**

NOTES: This table shows results for the benchmark specification as described in equations (4.2) to (4.8) but with slope-heterogeneity allowed to vary over the life-cycle. The variance of returns to experience is assumed to be different across two stages in the life-cycle, where the split takes place at H quarters. Slopes at the different stages are allowed to be correlated. I estimate the model for different values of H: 20 quarters, 40 quarters, and 60 quarters. The "secondary degree group" includes individuals who have a high-school and a vocational degree, while the "dropout group" includes individuals without a formal degree. All specifications allow for factor loadings on the permanent and persistent component, all of which are found to be signficant on the 1%-level. *** 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.

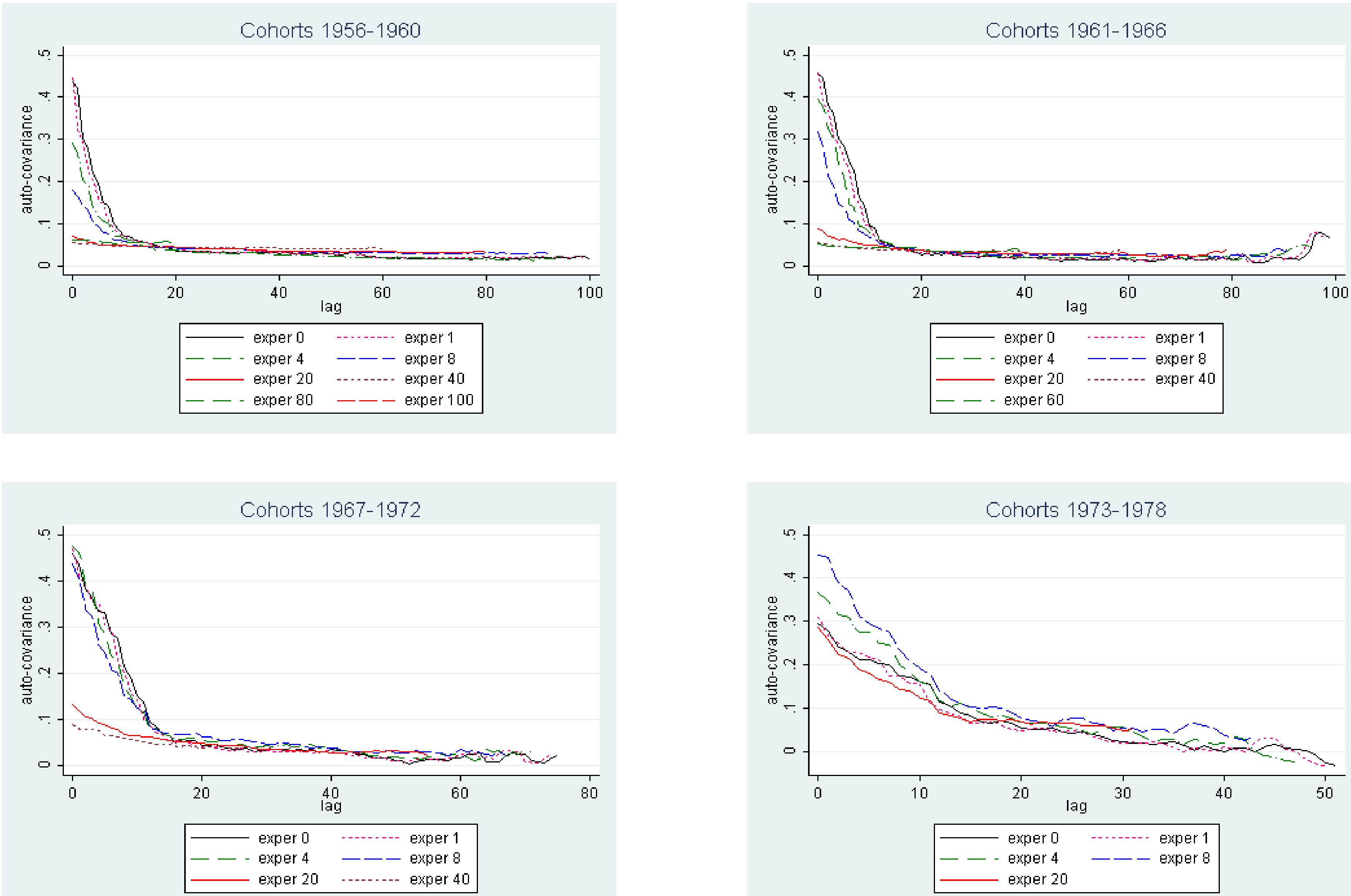
APPENDIX FIGURE 1 - LAG-PROFILES OF AUTO-COVARIANCES FOR DIFFERENT EXPERIENCE GROUPS, BY COHORTS

Sample: Secondary Degree Group

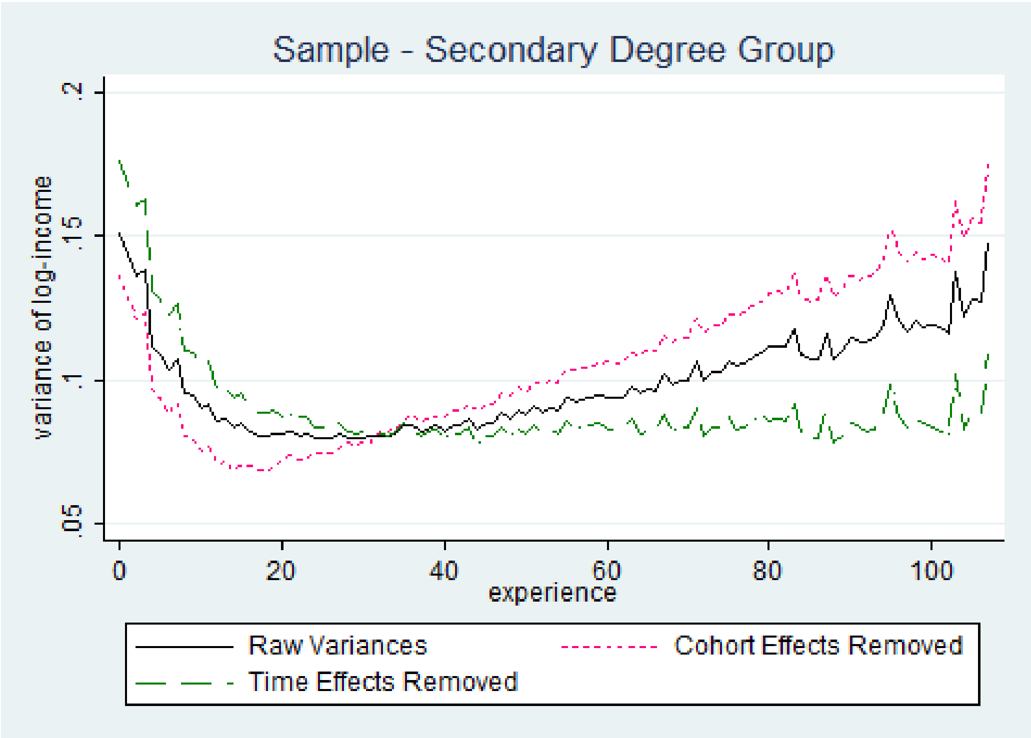


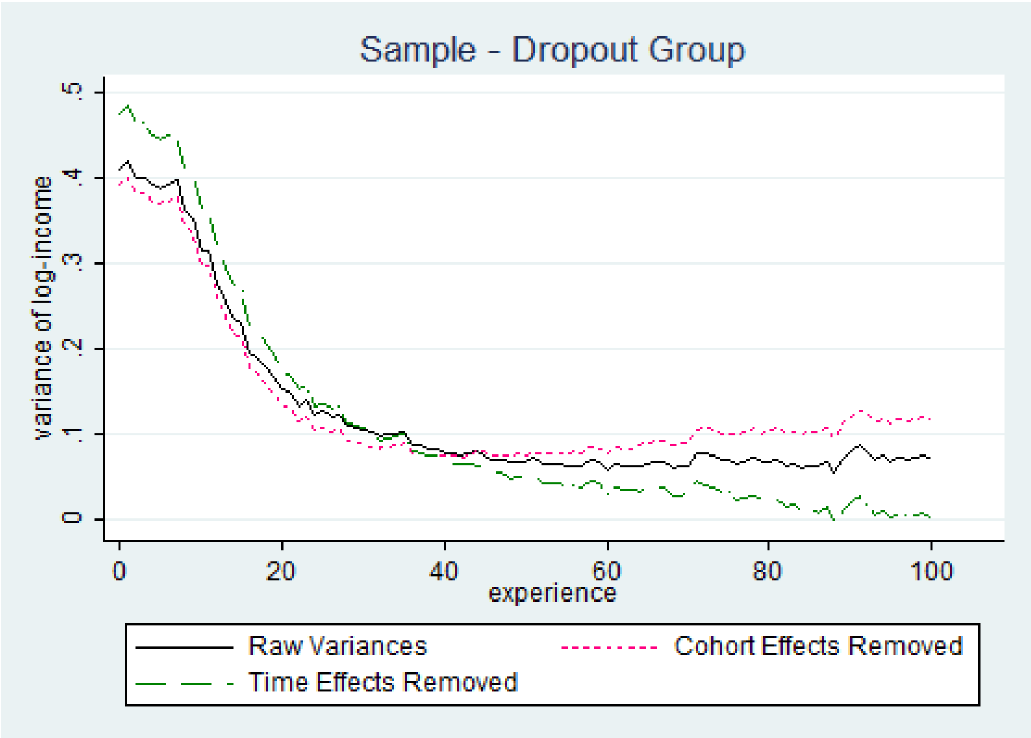
APPENDIX FIGURE 2 - LAG-PROFILES OF AUTO-COVARIANCES FOR DIFFERENT EXPERIENCE GROUPS, BY COHORTS

Sample: Dropout Group

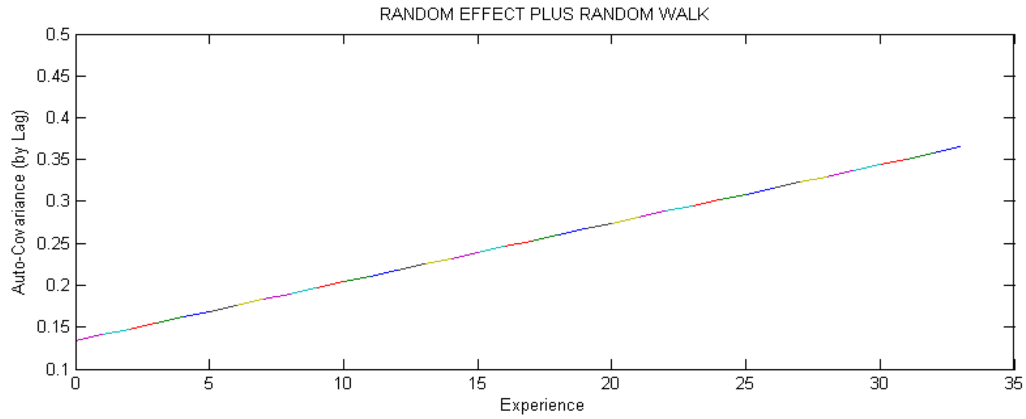


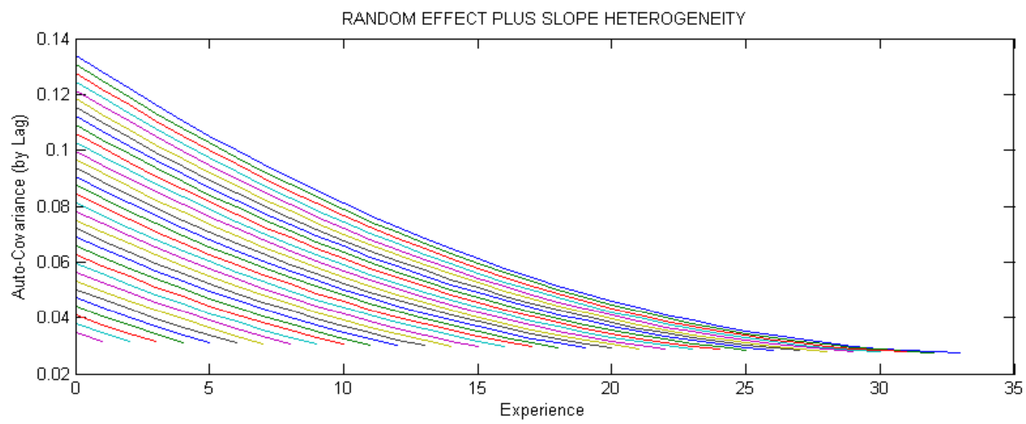
APPENDIX FIGURE 3 - EXPERIENCE-VARIANCE PROFILES OF LOG LABOR INCOME

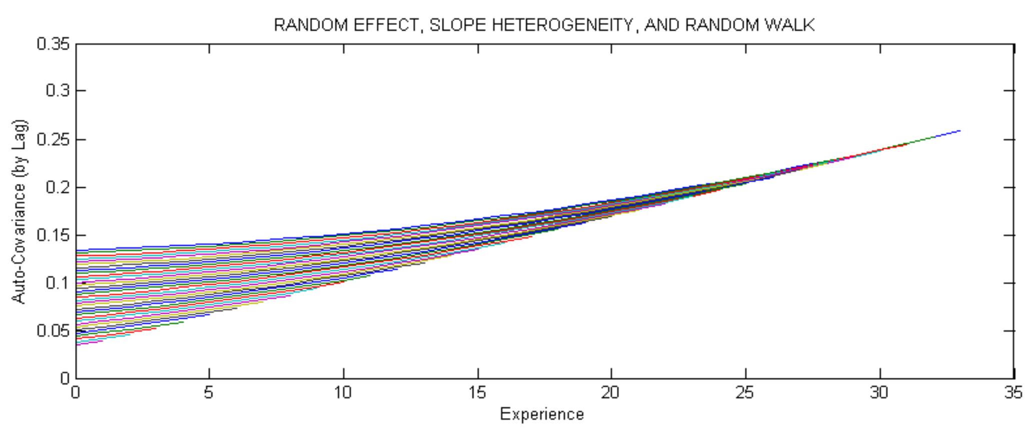


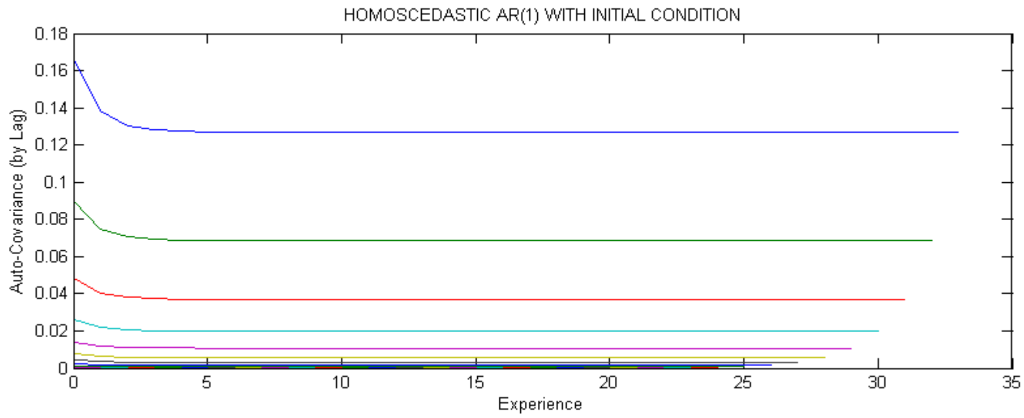


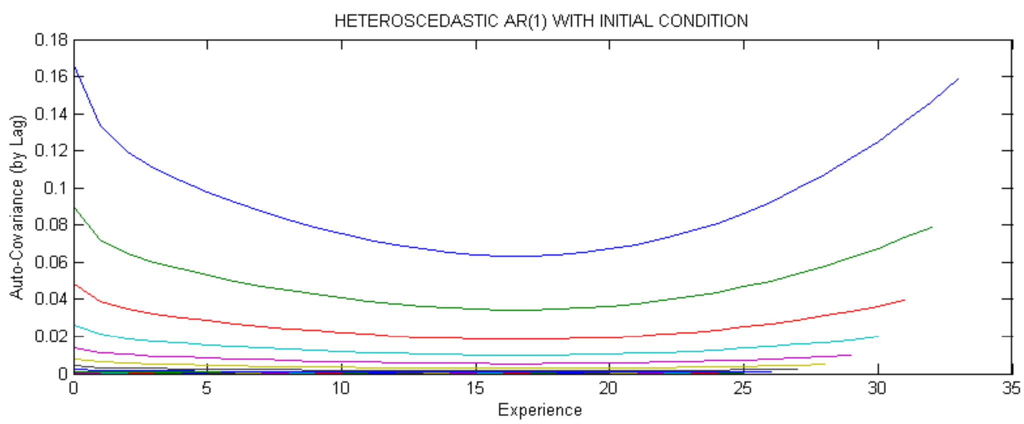
APPENDIX FIGURE 4 - VARIANCE COMPONENTS WITH BAKER-SOLON ESTIMATES, STATIONARY PART

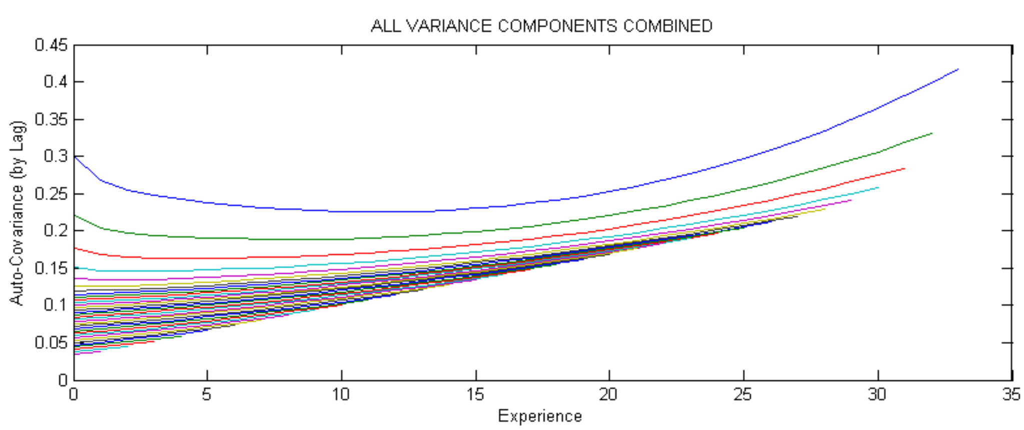




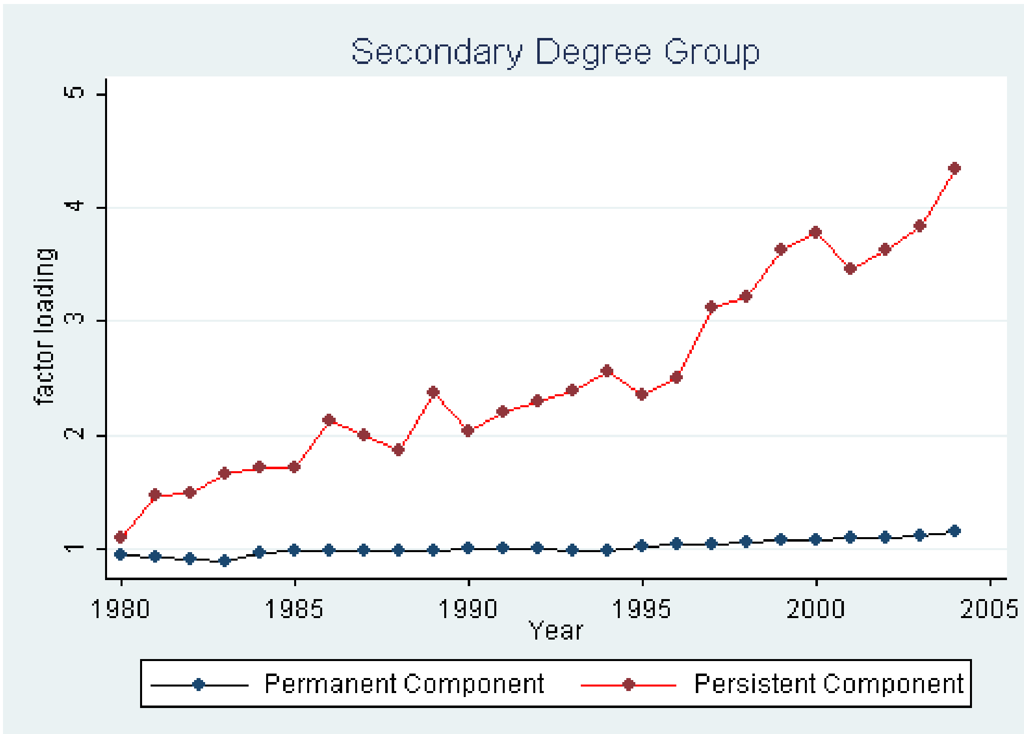


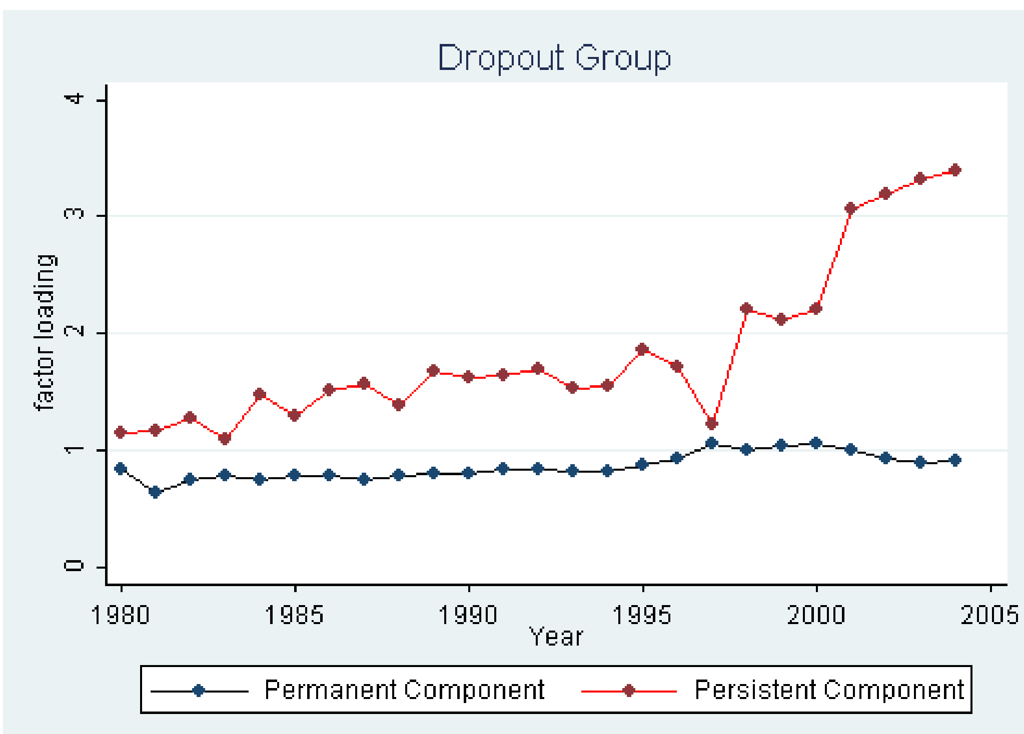

HOMOSCEDASTIC AR(1) WITH INITIAL CONDITION
Auto-Covariance (by Lag)
Experience


HETEROSCEDASTIC AR(1) WITH INITIAL CONDITION
Auto-Covariance (by Lag)
Experience

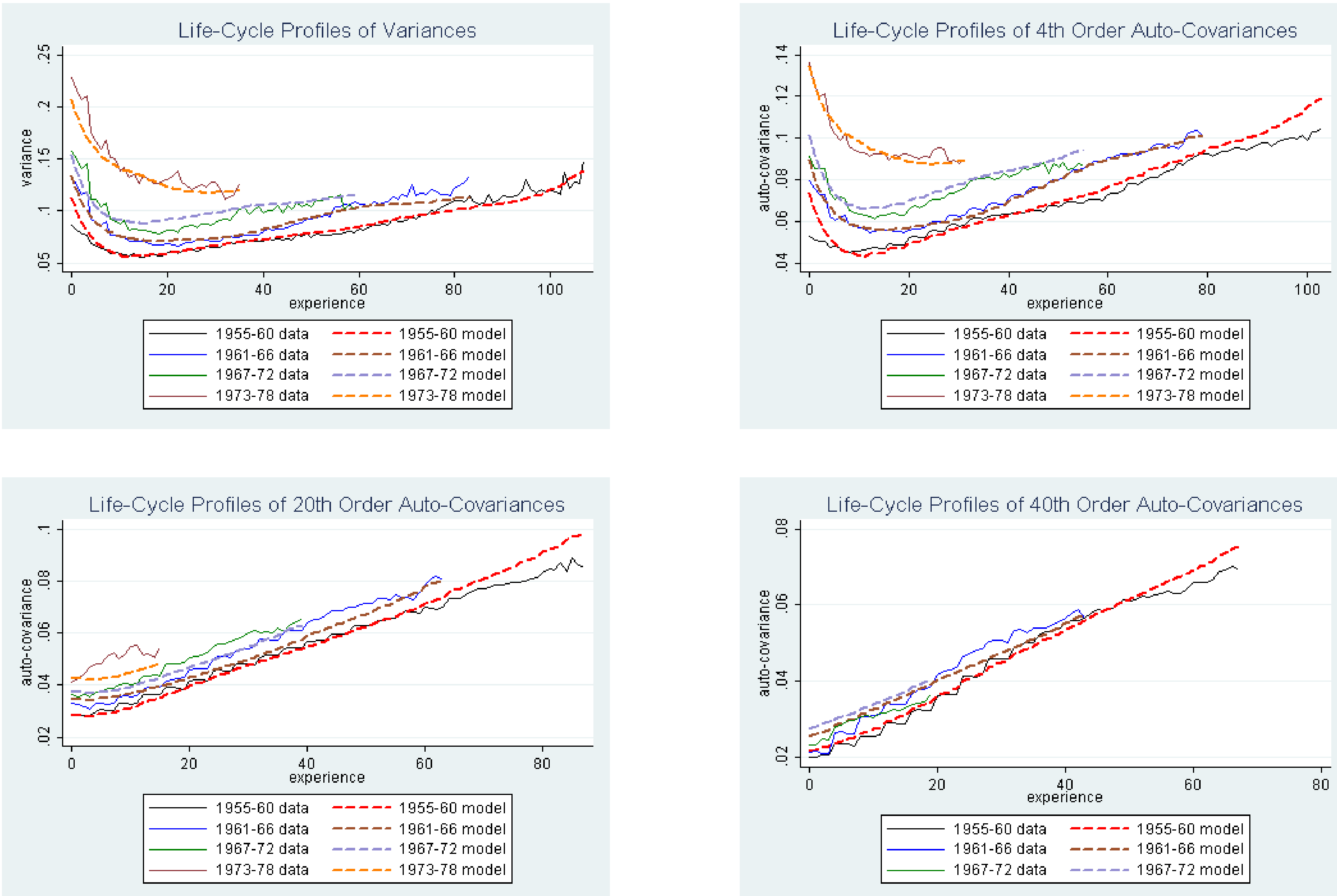

ALL VARIANCE COMPONENTS COMBINED
Auto-Covariance (by Lag)
Experience

APPENDIX FIGURE 5 - ESTIMATED FACTOR LOADINGS FOR THE FULL MODEL

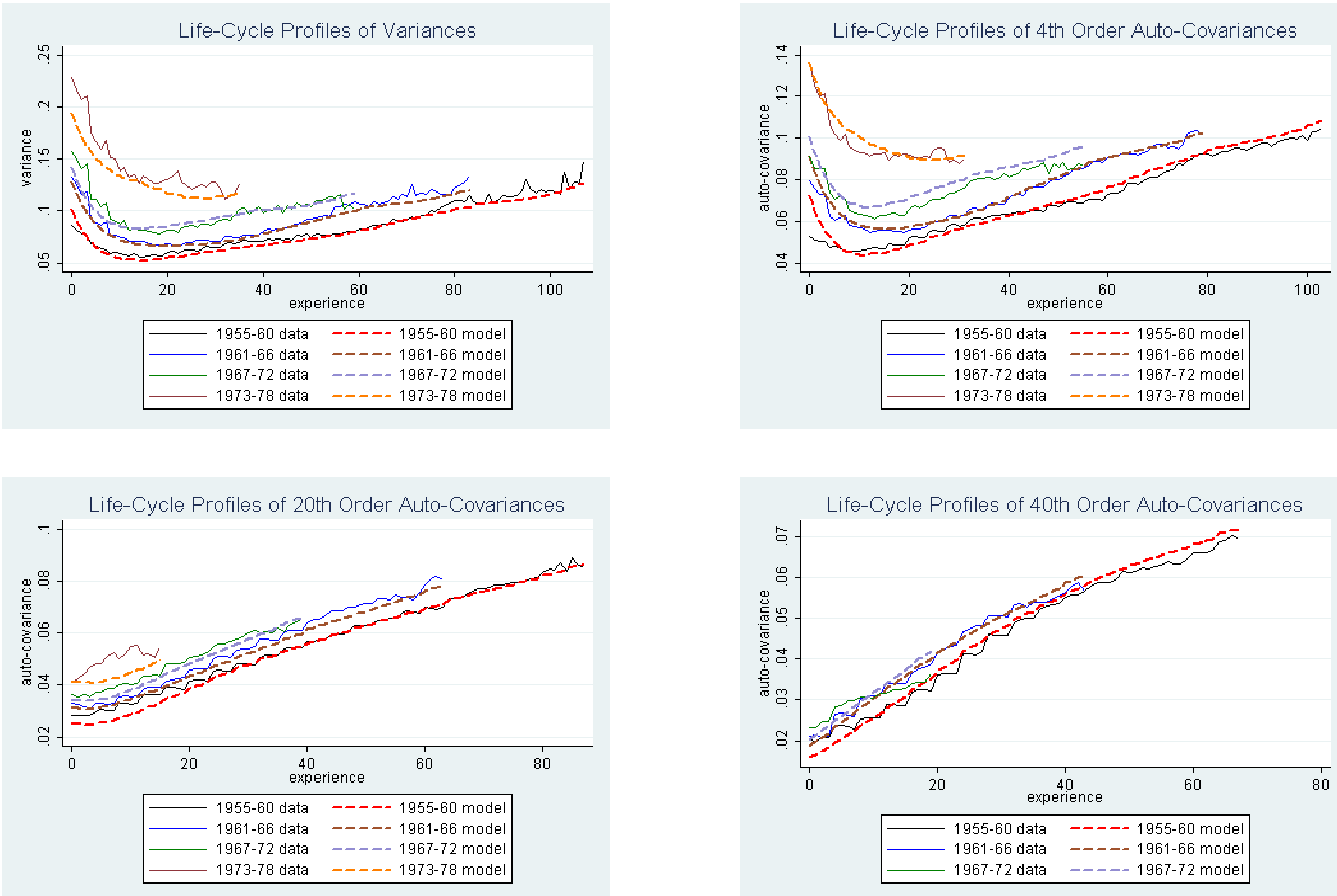




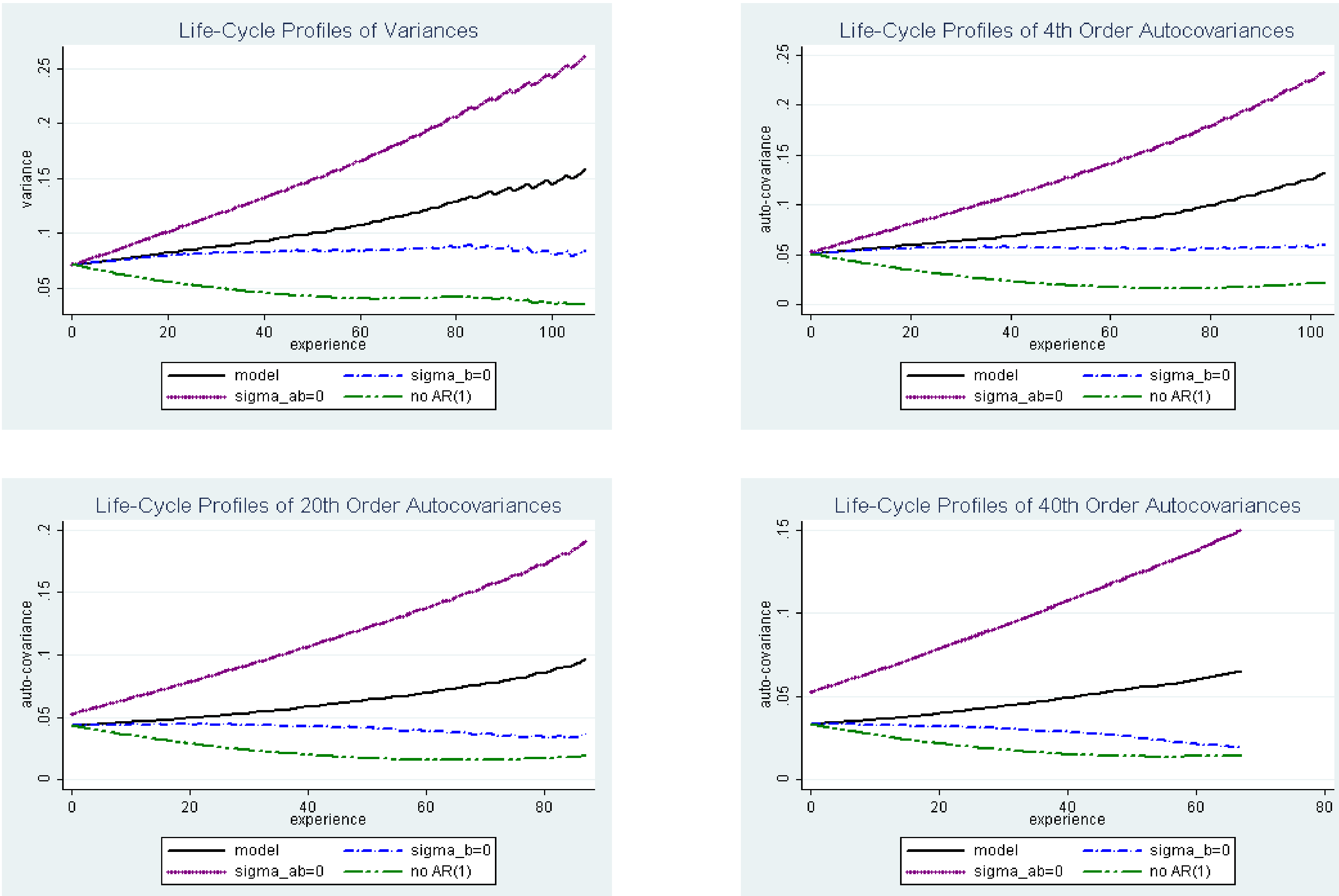
APPENDIX FIGURE 6 - FIT OF MODEL WITH HOMOSCEDASTIC UNIT ROOTS: SECONDARY DEGREE GROUP



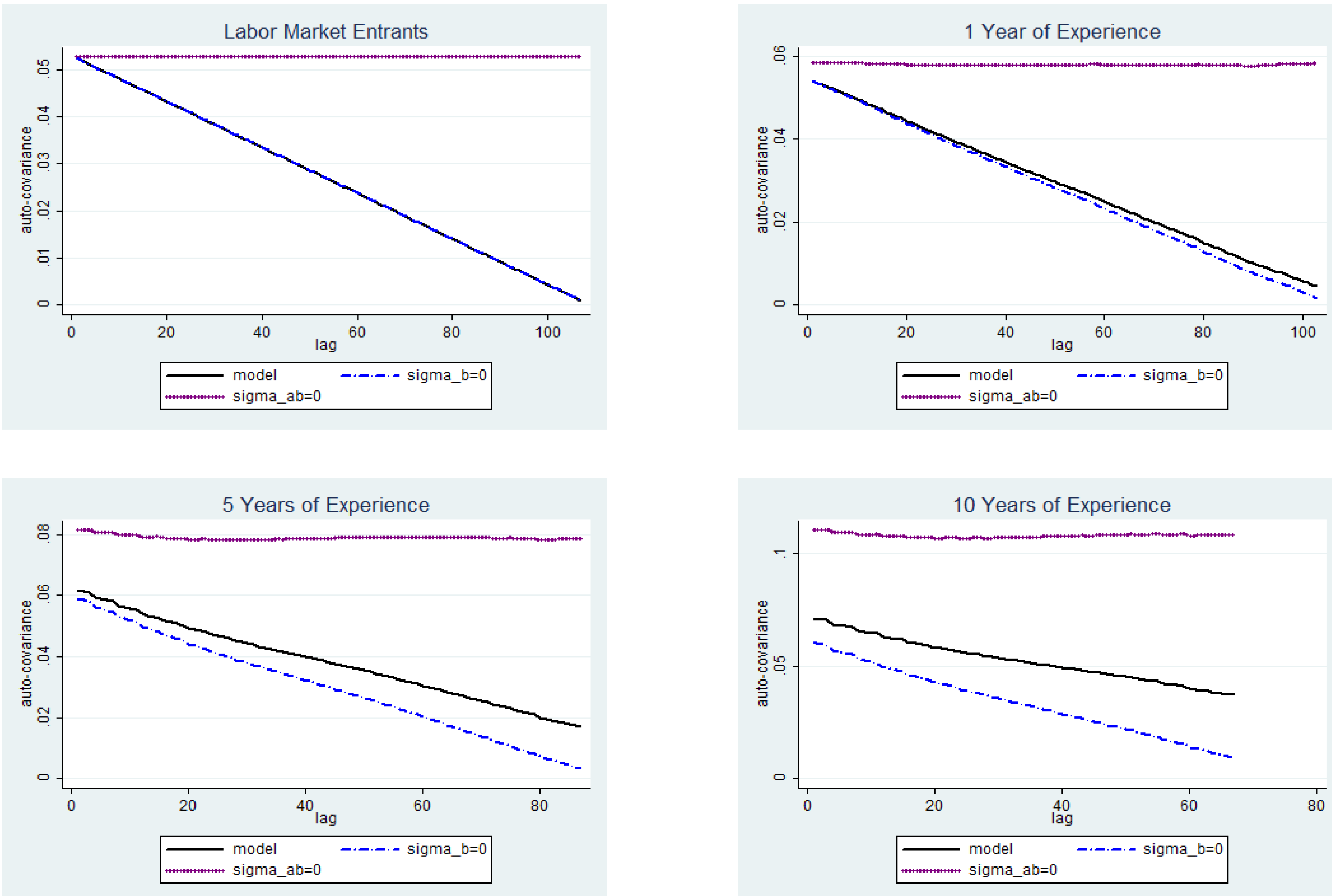
APPENDIX FIGURE 7 - FIT OF MODEL WITHOUT SLOPE HETEROGENEITY: SECONDARY DEGREE GROUP



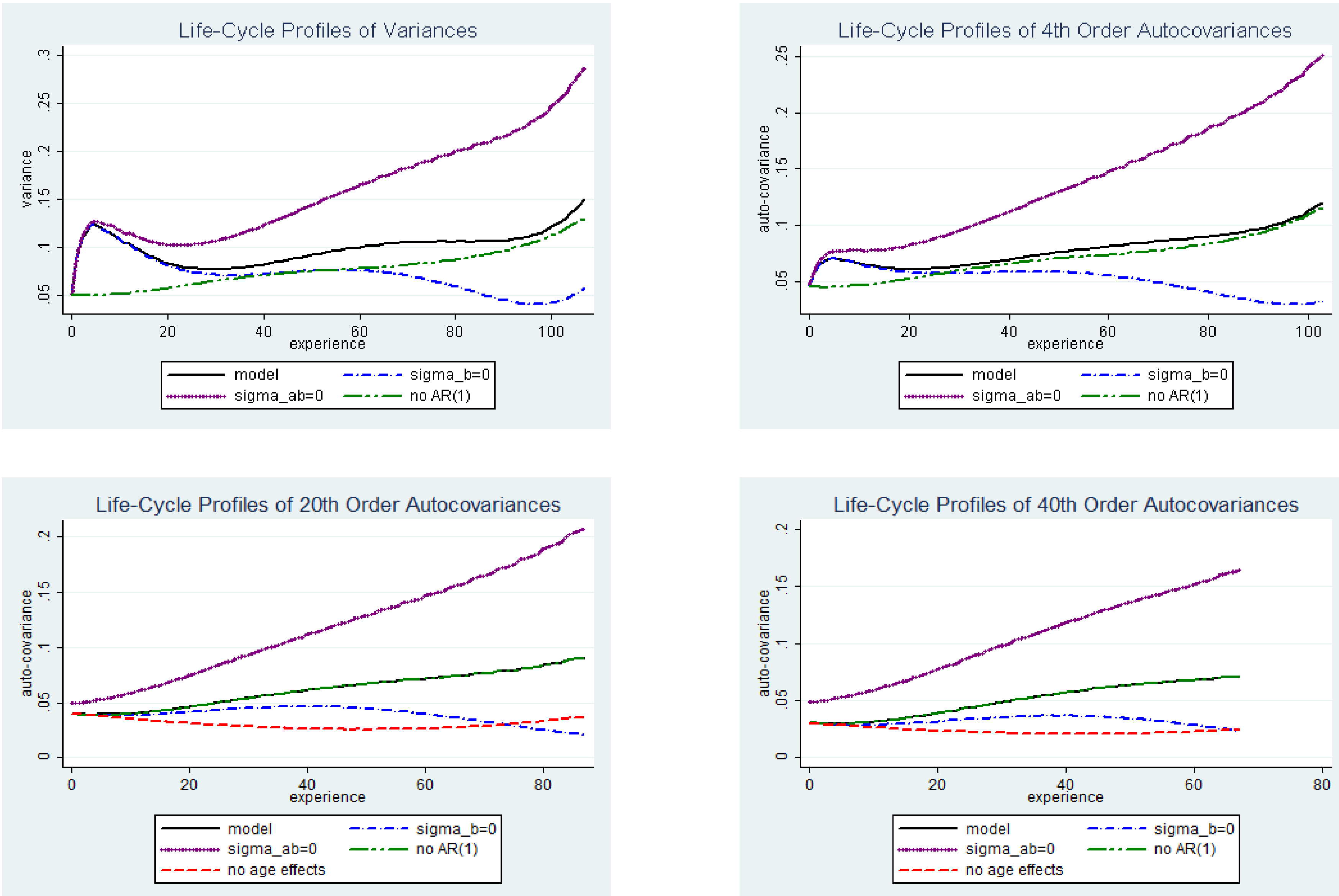
APPENDIX FIGURE 8 - COUNTERFACTUAL EXPERIENCE-PROFILES IN GUVENEN'S MODEL: SECONDARY DEGREE GROUP



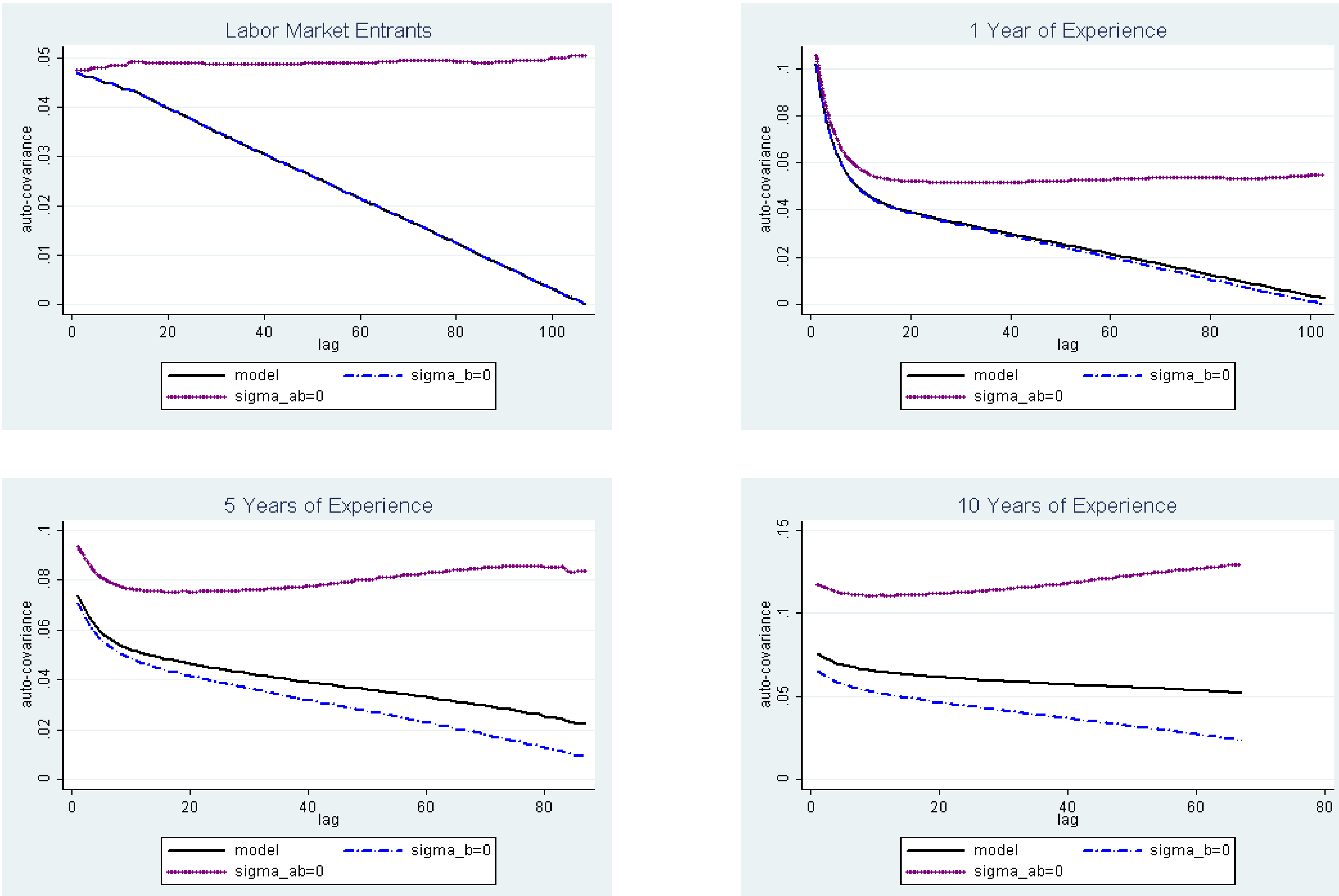
APPENDIX FIGURE 9 - COUNTERFACTUAL LAG-PROFILES IN GUVENEN'S MODEL: SECONDARY DEGREE GROUP



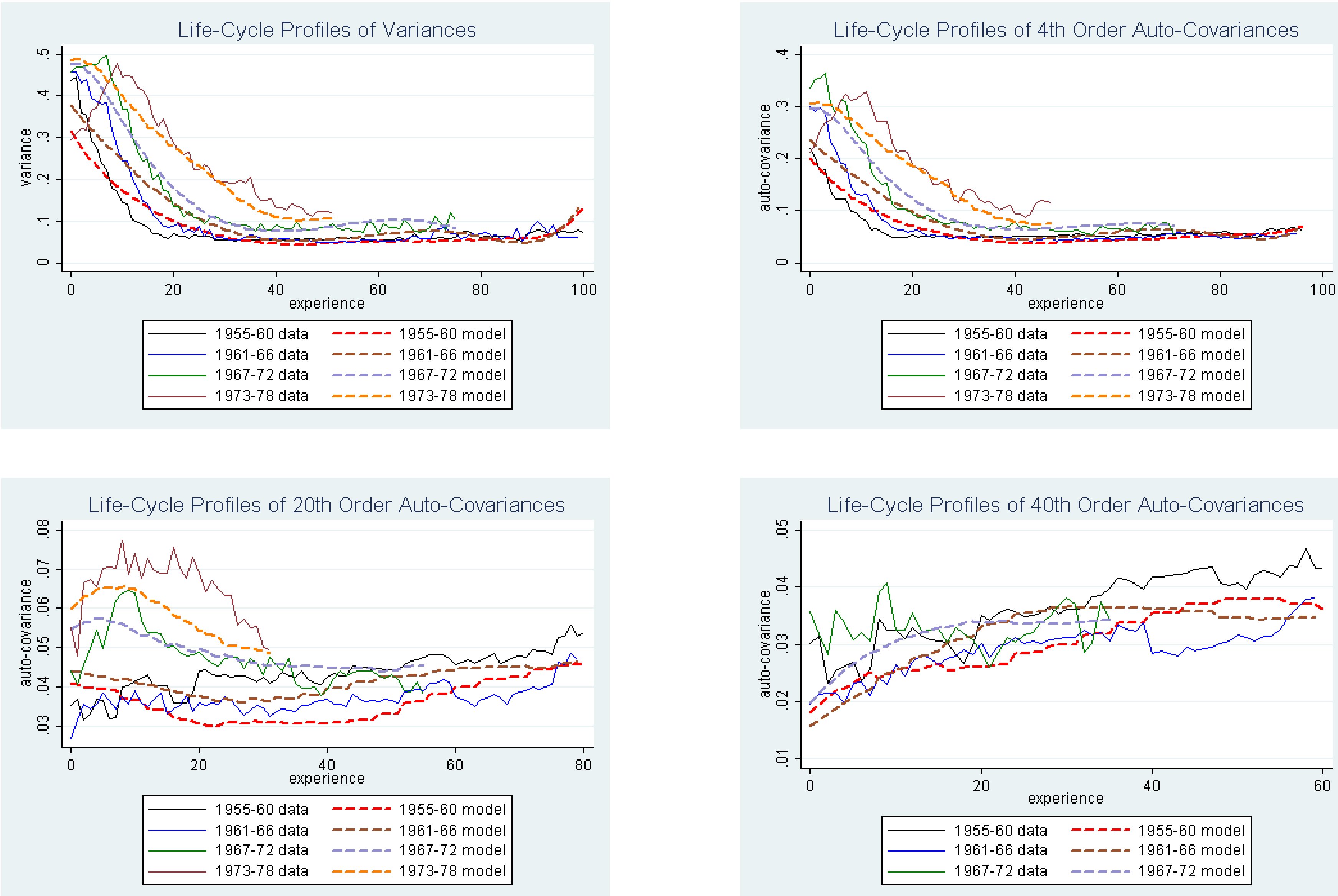
APPENDIX FIGURE 10 - COUNTERFACTUAL EXPERIENCE-PROFILES IN THE MODEL WITHOUT PERSISTENT INITIAL CONDITION: SECONDARY DEGREE GROUP



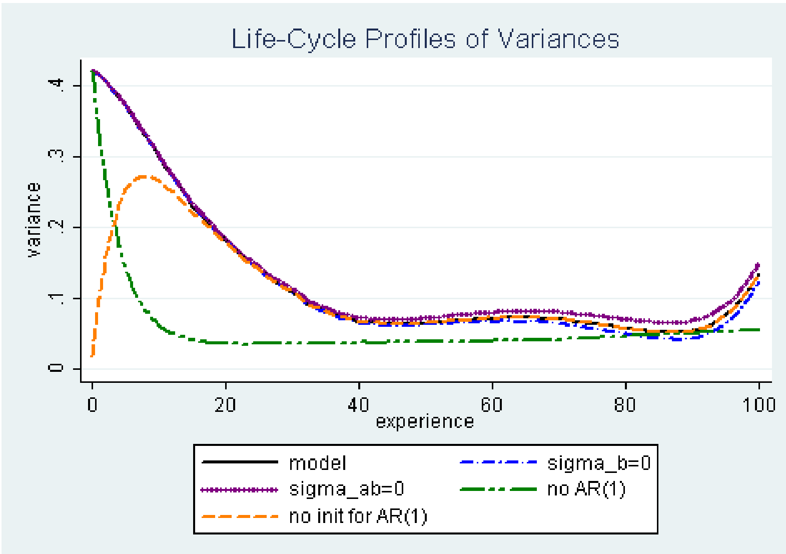
APPENDIX FIGURE 11 - COUNTERFACTUAL LAG-PROFILES IN THE MODEL WITHOUT PERSISTENT INITIAL CONDITION: SECONDARY DEGREE GROUP

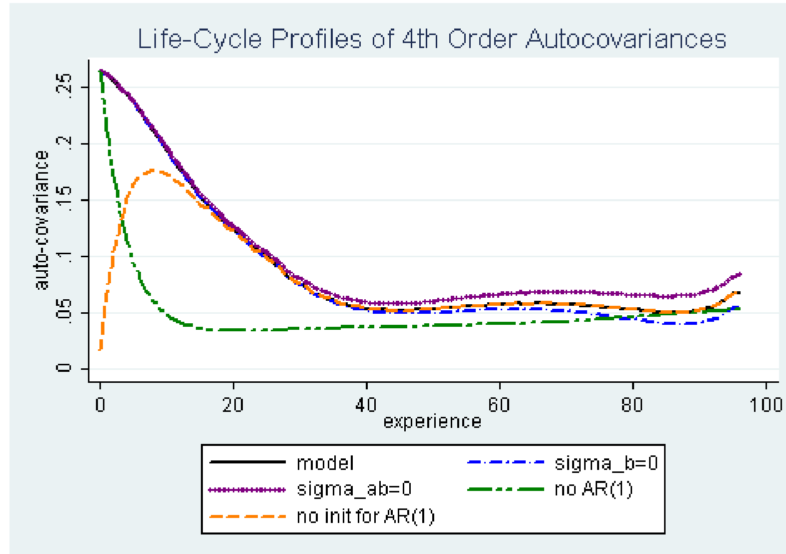


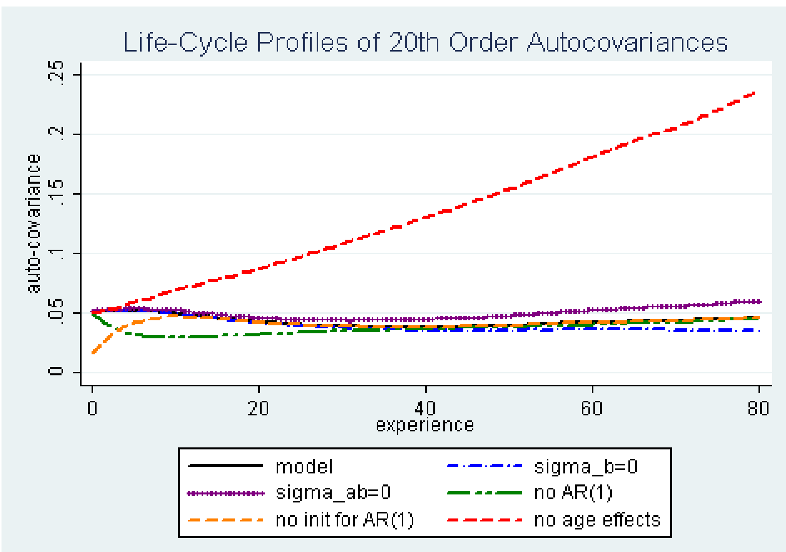
APPENDIX FIGURE 12 - FIT OF FULL MODEL: DROPOUT GROUP

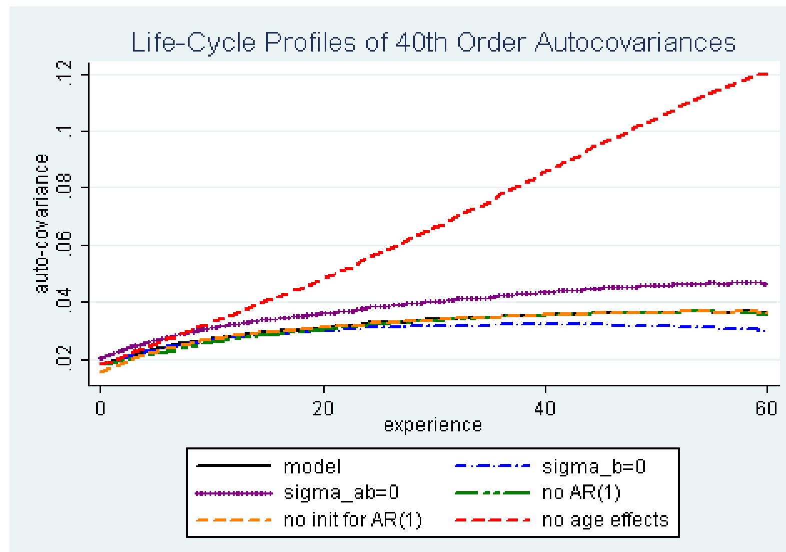


APPENDIX FIGURE 13 - COUNTERFACTUAL EXPERIENCE-PROFILES IN FULL MODEL: DROPOUT GROUP

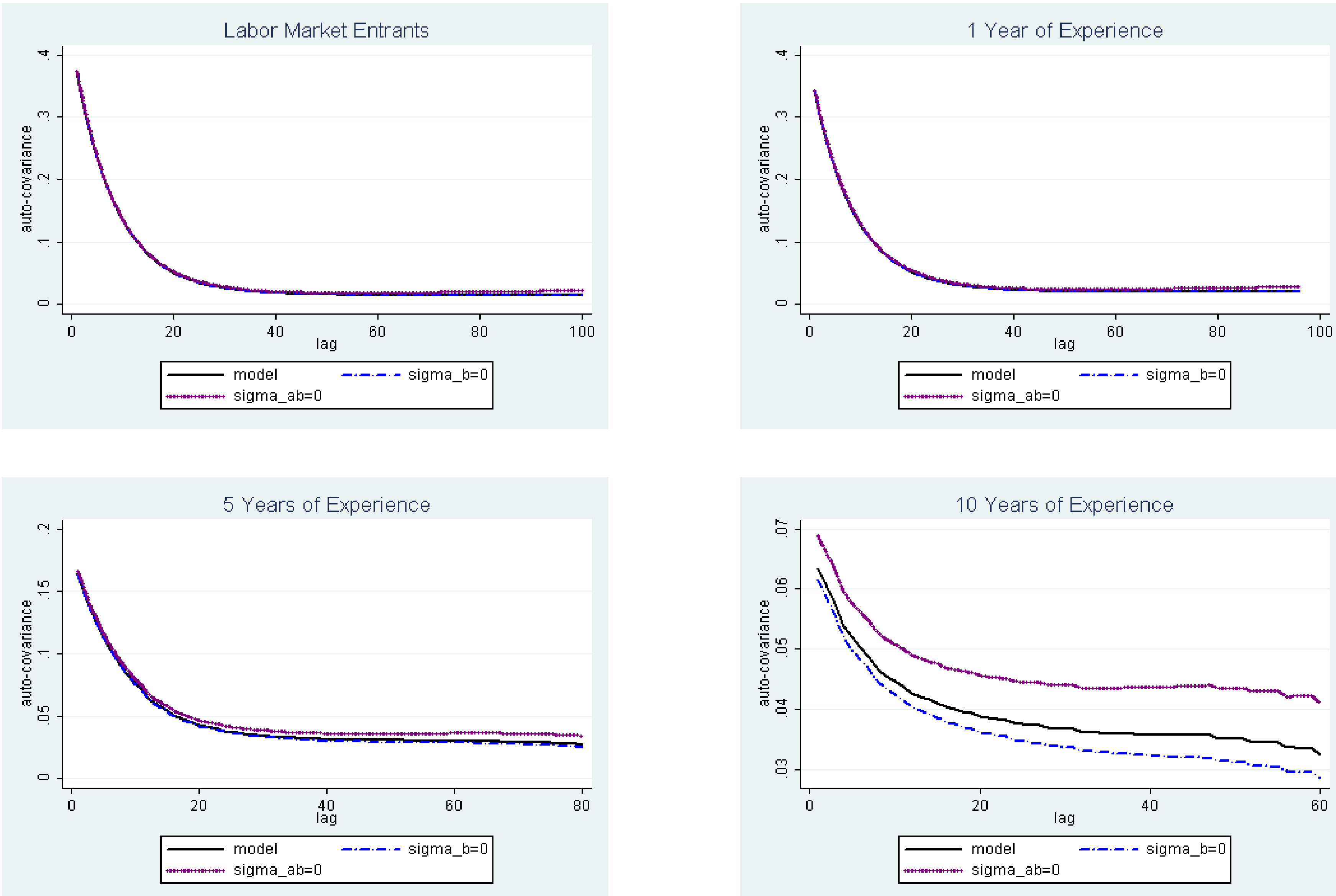




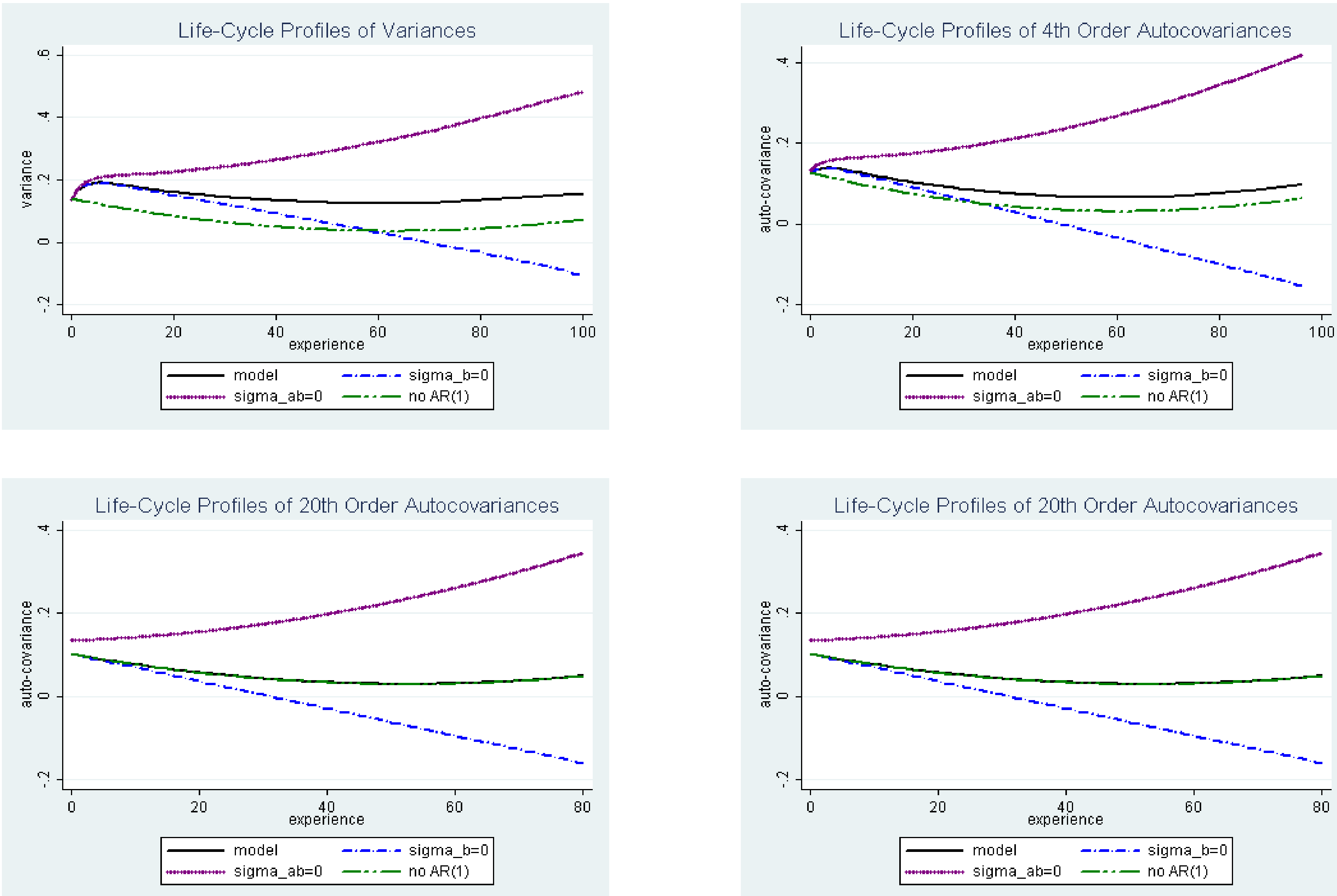




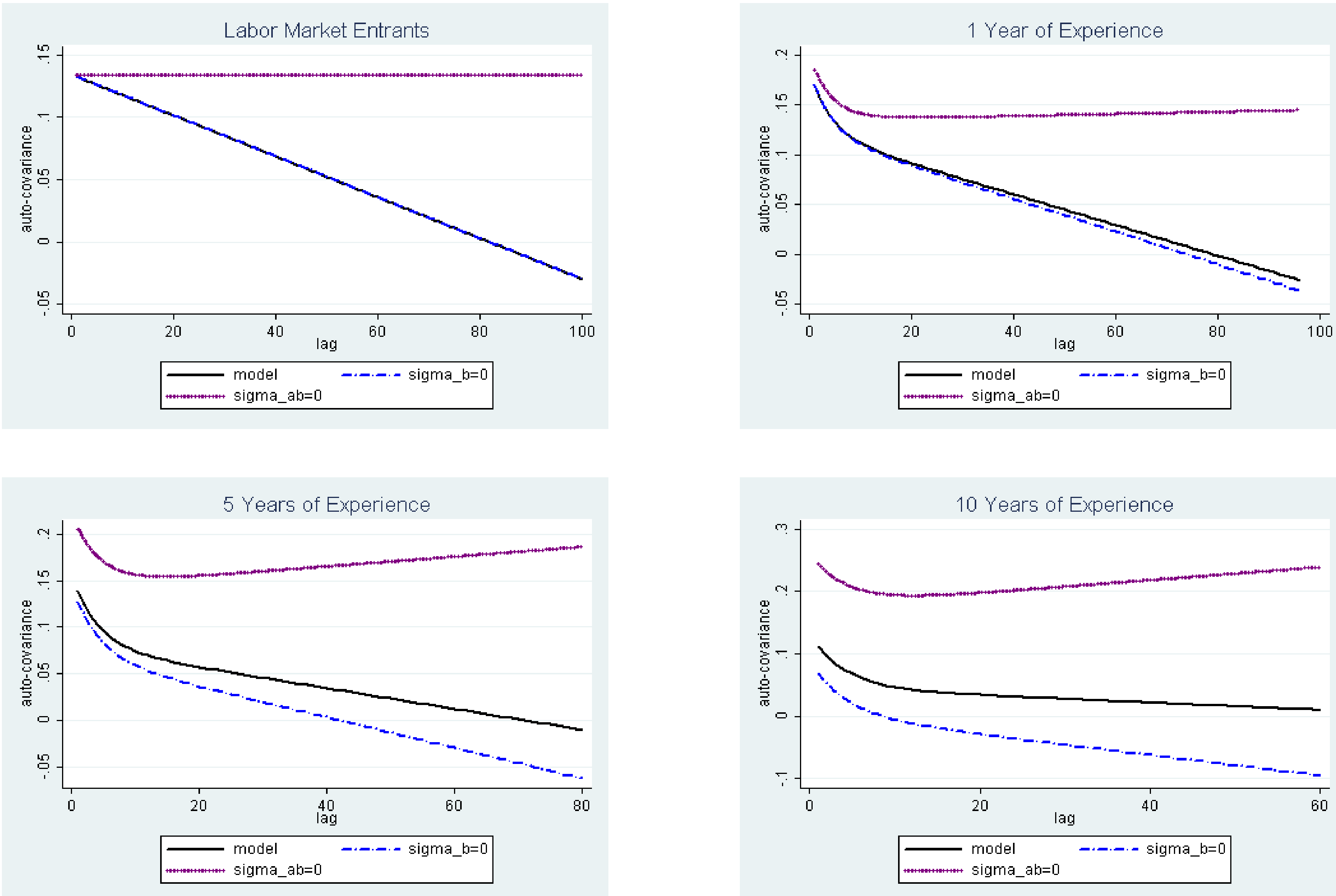
APPENDIX FIGURE 14 - COUNTERFACTUAL LAG-PROFILES IN FULL MODEL: DROPOUT GROUP



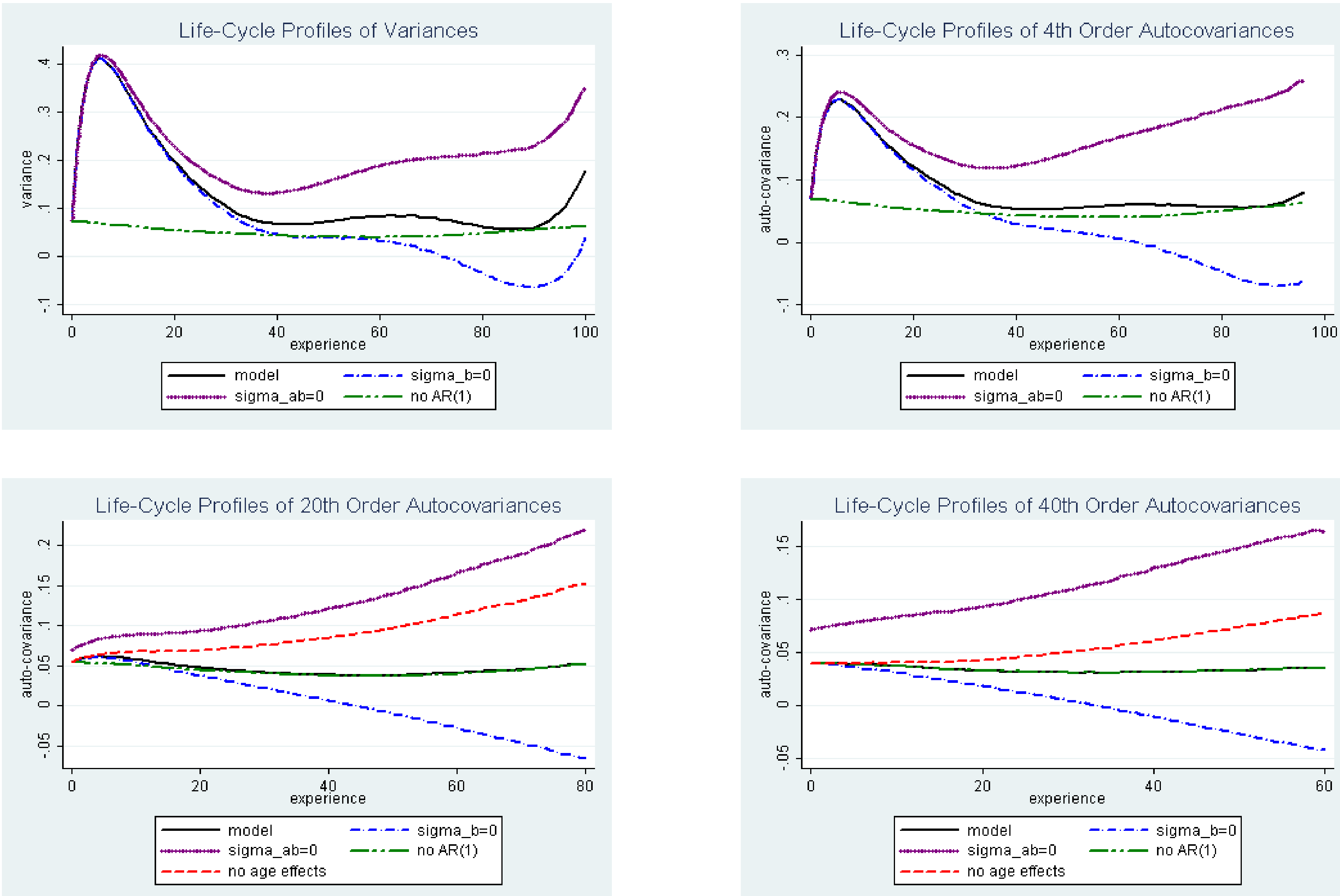
APPENDIX FIGURE 15 - COUNTERFACTUAL EXPERIENCE-PROFILES IN GUVENEN'S MODEL: DROPOUT GROUP



APPENDIX FIGURE 16 - COUNTERFACTUAL LAG-PROFILES IN GUVENEN'S MODEL: DROPOUT GROUP



APPENDIX FIGURE 17 - COUNTERFACTUAL EXPERIENCE-PROFILES IN THE MODEL WITHOUT PERSISTENT INITIAL CONDITION: DROPOUT GROUP



APPENDIX FIGURE 18 - COUNTERFACTUAL LAG-PROFILES IN THE MODEL WITHOUT PERSISTENT INITIAL CONDITION: DROPOUT GROUP