Dynamic Unobservable Heterogeneity: Income Inequality and Job Polarization *

Silvia SARPIETRO †

November 14, 2020

PRELIMINARY AND INCOMPLETE

For the latest version click here

Abstract

I propose the use of state-space methods as a unified econometric framework for studying heterogeneity and dynamics in *micropanels* (large N, medium T), which are typical of administrative data. I formally study identification and inference in models with pervasive unobservable heterogeneity. I show how to consistently estimate the cross-sectional distributions of unobservables in the system and uncover how such heterogeneity has changed over time. A mild parametric assumption on the standardized error term offers key advantages for identification and estimation, and delivers a flexible and general approach. Armed with this framework, I study the relationship between job polarization and earnings inequality, using a novel dataset on UK earnings, the New Earnings Survey Panel Data (NESPD). I analyze how the distributions of unobservables in the earnings process differ across occupations and over time, and separate the role played on inequality by workers' skills, labor market instability, and other types of earnings shocks.

Keywords: Unobserved Heterogeneity; State-space Methods; Job Polarization; Income Inequality.

*Special thanks to my supervisors, Raffaella Giacomini, Toru Kitagawa, and Dennis Kristensen for their invaluable guidance, support, and patience. For very helpful comments, I would like to thank Irene Botosaru, Simon Lee, Uta Schoenberg, Martin Weidner, Daniel Wilhelm, Morten O. Ravn, Michela Tincani, Juan José Dolado, Juan Carlos Escanciano, Davide Melcangi, Martin Almuzara, Marta Lopes, Alan Crawford, Jan Stuhler, Felix Wellschmied, Warn Lekfuangfu, Arthur Taburet, Mimosa Distefano, Carlo Galli, Gonzalo Paz-Pardo, Julio Galvez, Richard Audoly, Riccardo D'Adamo, Matthew Read, Rubén Poblete-Cazenave, Alessandro Toppeta, Guillermo Uriz-Uharte, Jeff Rowley, David Zentler Munro, Riccardo Masolo, Ambrogio Cesa-Bianchi, as well as seminar participants at the Bank of England, at Carlos III, and at UCL for valuable comments and feedback. I gratefully acknowledge financial support from the ESRC and the Bank of England, and data access from UK Data Service/ONS.

†Department of Economics, UCL. Email: silvia.sarpietro.14@ucl.ac.uk.

1 Introduction

In recent years, administrative datasets have become increasingly available. While this wealth of data can be instrumental in answering several key questions in Economics, it also introduces modeling challenges. Most of the time, administrative data are *micropanels*, which are panel data where many units (N) are observed for a medium number of time periods (T), and thus provide rich information on individuals and firms over time. However, micropanels usually have few dimensions of observable heterogeneity: for instance, administrative data on earnings typically lack information on education, marital status, and health conditions, with demographical variables for each worker limited to age and gender. This drawback makes it crucial to model unobservable heterogeneity both over time and across individuals, and the medium time-series dimension requires careful modeling of the dynamics. Unobservable heterogeneity is not only interesting per se, but it also affects several other outcomes of interest.[1] Indeed, many important questions in the earnings literature, covering topics such as wage inequality or insurance against earnings shocks, require an understanding of the interplay between dynamics and heterogeneity.[2] In addition to this, a modeling framework that features pervasive unobservable heterogeneity and dynamics would be useful in addressing new empirical questions using administrative data.

In this paper, I propose the use of state-space methods as a unified econometric framework for the study of heterogeneity and dynamics in micropanels. I estimate unobservable heterogeneity and uncover how such heterogeneity has changed over time. As a key contribution, I formally study identification and inference in models with pervasive unobservable heterogeneity. Armed with this framework, I analyze how earnings dynamics of UK workers differ across occupations and over time, making use of a novel dataset on UK earnings, the New Earnings Survey Panel Data (NESPD). My approach and findings reconcile empirical evidence of an increase in the 50/10 wage gap (the ratio of median and low wages) and the documented phenomenon of job polarization (increase in employment in low- and high-skill occupations alongside a simultaneous decrease in middle-skill occupations).

Several econometric methods, often applied to the study of earnings dynamics, treat unobservable heterogeneity as nuisance parameters. Following Almuzara (2020) and Botosaru

[1]For instance, heterogeneity in earnings dynamics influences predicted mobility out of low earnings (Browning, Ejrnaes, and Alvarez, 2010); heterogeneity in income profiles conditional on parents' background is crucial to the study of intergenerational mobility (see Mello, Nybom, and Stuhler, 2020).

[2]The distinction between transitory and persistent shocks and the trade-off between heterogeneity and persistence are useful in explaining how individual earnings evolve over time and in decomposing residual earnings inequality into different variance components; the persistence of earnings affects the permanent or transitory nature of inequality (MaCurdy (1982), Lillard and Weiss (1979), Meghir and Pistaferri (2004)). The components of the stochastic earnings process drive much of the variation in consumption, savings, and labor supply decisions, (see Guvenen (2007), Guvenen (2009), Heathcote, Perri, and Violante (2010), Arellano, Blundell, and Bonhomme (2017)). Moreover, they play a crucial role for the determination of wealth inequality, and for the design of optimal taxation and optimal social insurance. Finally, separating permanent from transitory income shocks is relevant for income mobility studies and to test models of human capital accumulation.

(2020), I depart from the existing approach in the literature and explicitly treat unobservable heterogeneity as the main object of interest. Almuzara (2020) and Botosaru (2020) adopt a non-parametric approach for estimation of unobservable heterogeneity in earnings models. I consider comparatively richer heterogeneity and dynamics, while imposing a mild parametric assumption on the standardized error term. I assume that innovations are Gaussian, but this assumption can be relaxed, and several more flexible distributions, e.g. mixtures of normals, can be considered.[3] Moreover, the approach proposed in this paper lends itself to several generalizations, such as unbalanced panel data and measurement errors, and can be adapted to accommodate a treatment of heterogeneity as either fixed or random effects.

The paper's contribution is twofold: methodological and empirical.

My *first* contribution is to derive theoretical results on how to adapt state-space methods to the analysis of panel data featuring heterogeneous dynamic structures. The choice of using state-space methods with filtering and smoothing techniques is motivated by their usefulness for estimation and inference about unobservables in dynamic systems. As emphasized by Durbin and Koopman (2012) and Hamilton (1994), state-space methods include efficient computing algorithms that provide (smoothed) estimates of unobservables, while providing flexible and general modeling that can incorporate individual explanatory variables, macro shocks, trends, seasonality, and nonlinearities. Another main advantage is that these methods can be used in the presence of data irregularities, e.g. unbalanced panel data and measurement error. The models typically considered in the earnings literature, e.g. ARIMA, are a special case of state-space models but state-space methods include techniques for initialization, filtering, and smoothing. If the goal is to uncover the evolution of the state variables, state-space models are the most natural choice. Multivariate extensions with common parameters and time-varying parameters are much more easily handled in state-space modeling with respect to a pure ARIMA modeling context.

State-space methods have been mainly used in the context of time series models or with macropanels (panel data with few units observed over many time periods), but the unique structure of micropanels requires the development of new econometric tools for analysis. There is a lack of theoretical results on how to extend their use to micropanels for the analysis of heterogeneous dynamic structures.[4] Therefore, I adapt state-space methods to the analysis of unobserved heterogeneity in micropanels and formally study identification and inference in the context of these heterogeneous models. I show how to consistently estimate the cross-sectional distributions of unobservables in the system and uncover how such heterogeneity has changed over time.

[3]Note that, when errors are not normally distributed, results from Gaussian state-space analysis are still valid in terms of minimum variance linear unbiased estimation.

[4]Some notable exceptions are the Seemingly Unrelated Times Series Equations (SUTSE) by Commandeur and Koopman (2007), and Dynamic Hierarchical Linear models by Gamerman and Migon (1993) and by Petris and An (2010) but the focus of the analysis is rather different. I build on these models, discuss the differences, and provide theoretical results on how to recover the cross-sectional distribution of heterogeneous components.

A mild parametric assumption on the standardized error term offers substantial advantages for identification and estimation, and delivers a flexible and general approach. Following the literature on state-space methods, I propose an argument for identification based on a large-T approach. I also consider a fixed-T identification approach to establish a comparison with the existing non-parametric literature. I discuss the corresponding estimation procedures and further analyze the asymptotic properties of these distribution estimators. In the existing literature, properties of the distribution estimators for the individual parameter estimates obtained from state-space models are unknown. Moreover, it is computationally challenging to extend state-space analysis and filtering to heterogeneous micropanels, which feature large N.

As a first step of the analysis, I consider a simple state-space model and treat the history of each individual i as a separate time series. Identification of the parameters relies on a large T argument, while asymptotic properties of the distribution estimators are established under some ratio between N and T. Building on the work of Okui and Yanagi (2020) and Jochmans and Weidner (2018), I derive this ratio and propose a bias correction for small T.

In a second step, I introduce time-varying parameters in the state-space model and further consider extensions where these parameters are assumed to be common across groups of similar individuals. I discuss how the identification results change in this setting. To devise a tractable estimation strategy, I use stratification as a device to reduce the computational burden of a large cross-sectional dimension on filtering and smoothing algorithms. Once I estimate the parameters and state variables of interest, a larger cross-section is used to consistently estimate the distribution of heterogeneous unobservables.

Finally, in the last part of the theoretical analysis, I consider a fixed-T approach to explore the relationship to the current non-parametric approach, (see Almuzara, 2020, and Botosaru, 2020), which relies on a fixed-T argument for identification of the cross-sectional distribution of unobservables in the model. The main limitation of fixed-T approaches is that the condition for identification may be difficult or even impossible to verify and existing estimation techniques can be computationally expensive. I show how the parametric assumption on the error term can permit achieving identification with a short number of time periods, making the analysis feasible when richer heterogeneity is allowed in the model. I also discuss what the implications of a parametric assumption on error terms are for regular identification of the distribution of unobservables, following the work of Escanciano (2020).[5]

My *second* contribution is to provide new empirical evidence on the phenomenon of job polarization using a novel UK micropanel, the NESPD, and to study it within a dynamic framework. Analysis of job polarization in the literature is typically grounded on a static approach. The literature on job polarization, pioneered by Autor, Katz, and Kearney (2006), defines job polarization as a significant increase in employment shares in low-skill occupations and high-skill occupations, associated with a simultaneous decrease in employment shares in

[5]Regular identification of functionals of nonparametric unobserved heterogeneity means identification of these functionals with a finite efficiency bound.

middle-skill occupations, which is a pattern that has been been observed and documented in the US and UK over the last 40 years.[6] I use this novel dataset to test several hypotheses on the relation between job polarization and income inequality. The NESPD is a survey directed to the employer, running from 1975 to 2016, with large cross-sectional and time-series dimensions, which allow the earnings process to feature type dependence in a flexible way. Stratification by observables is possible and replaces the first-stage regression of earnings on covariates, which restricts the dependence of earnings on them. I analyze how the distributions of unobservables in earnings processes have evolved over time and across occupations, and separate the role that workers' skills, labor market instability, and other types of earnings shocks have played on inequality. I use the proposed modeling framework to test whether the distribution of individuals' skills among different occupations has evolved over time and by different age groups. Moreover, I investigate how the corresponding skill prices have changed, and how the distributions of permanent and transitory shocks have changed over time and by occupation.

This paper uses the answers to the above questions to reconcile the empirical evidence that an increase in the 50/10 wage gap (inequality between the low and median wages) has occurred despite the documented phenomenon of job polarization, which would predict the opposite if relative demand is rising the low-skill jobs relative to middle-skill jobs. The findings can provide key insights to inform policy decisions based on the dynamics of earnings and of their distributions over time, and are relevant to think about the evolution of labor markets and inequality, also during and after the COVID-19 pandemic. Another interesting empirical question is to uncover heterogeneity in firms' productivity and document how this has changed over time.

To conclude, I develop a state-space framework as a new tool for modelers, with several advantages for identification and estimation, which can be used to address questions on dynamic unobservable heterogeneity in many settings.

The outline of the paper is as follows: Section 2 presents an overview of the related literature. In Section 3, I establish the argument for identification, while the corresponding estimation procedures are discussed in Section 4. Section 5 provides a discussion of the Gaussian assumption and further extensions. In Section 5, I describe the dataset used for the empirical analysis. In Section 6, I present the empirical application and report empirical findings. Finally, the last Section concludes and discusses directions for future research.

2 Related literature

There is an extensive literature on state space methods for time series or macropanels, which are panel data with small N and large T (Durbin and Koopman (2012), Hamilton (1994)). However, the unique nature of micropanels requires the development of new econometric tools

[6]Following the literature, occupations are classified into the categories of low-, middle-, and high-skill jobs based on 1976 wage density percentiles.

to make use of state space methods. I contribute to this econometric literature on state-space by adapting existing methods to suit the characteristics of administrative data, i.e. micropanel data, which feature large N. In particular, I derive theoretical results on how to consistently estimate the cross-sectional distribution of unobservables estimated with state space models.

In order to establish the asymptotic properties of (and make inference on) the estimators of the cross-sectional distribution of unobservables, I rely on the literature on heterogeneous dynamic panel data (Okui and Yanagi (2020), Jochmans and Weidner (2018), Mavroeidis, Sasaki, and Welch (2015)). Okui and Yanagi (2020) propose a model-free approach, whereas Jochmans and Weidner (2018) consider a Gaussian assumption on error term but obtain similar results. Finally, Mavroeidis et al. (2015) consider heterogeneous AR(1) models with a fixed-T setting. I extend these existing approaches to investigate the asymptotic properties of the estimator of the cross sectional distribution of unobservables, which are estimated in a first-stage using a state-space model.

Panel data factor models, e.g. Bai (2009), are related to the analysis of panel data with state space methods since dynamic factor models are special cases of state-space models where the econometrician specifies dynamic properties for latent factors in the state equation. However, the state vector is small, and the goal of the analysis is to find commonalities in the covariance structure of a high dimensional dataset.

By developing the corresponding fixed-T approach, I explore the relation of my methodology with a recent literature on estimation of the cross-sectional distribution of unobservables with panel data for the analysis of earnings processes. Almuzara (2020) and Botosaru (2020) adapt the identification argument in Hu and Schennach (2008), with the aim of identifying the distribution of heterogeneous variance and permanent components in earning processes. I consider a more general process but impose a (flexible) parametric assumption on the error term: in particular, I focus on large dimensions of heterogeneity, with time-varying parameters, and I impose a mild parametric assumption on the standardized error term. Moreover, this approach lends itself to generalizations such as allowing for unbalanced panel data and measurement errors.

This paper also relates to the literature on earning dynamics. The literature on the analysis of earnings processes is large and can be distinguished into several strands: one strand focuses on the permanent-transitory decomposition of earnings residuals (Abowd and Card (1989), MaCurdy (1982), Lillard and Weiss (1979)); another strand introduces growth-rate heterogeneity, e.g. Baker (1997), Haider (2001), Guvenen (2009); a third strand considers income variance dynamics allowing for conditional heteroskedasticity in permanent and transitory shocks, e.g. Meghir and Pistaferri (2004), Botosaru et al. (2018)); finally, nonlinear models have recently been proposed by De Nardi, Fella, and Pardo (2016), Arellano et al. (2017). Guvenen, Karahan, Ozkan, and Song (2015) and Browning et al. (2010) introduce pervasive heterogeneity and are the closest to the present paper. However, Browning et al. (2010) do not consider a transitory-persistent decomposition of earnings shocks and both these papers do not propose arguments for identification and estimation of the cross-sectional distribution of unobservables.

Finally, I investigate the relationship between wage inequality and job polarization, which has only be analyzed using static approaches in the literature. The phenomenon of job polarization has been documented by Autor et al. (2006) for the US, and by Goos and Manning (2007) for the UK. The literature that supports the hypothesis of skill-biased technical change cannot explain the increase in employment in low- and high-skill occupations alongside a simultaneous decrease in medium-skill occupations (U-shape in figure 1) because it would only predict change in demand for unskilled vs skilled workers. The hypothesis of automation and routinization, advanced by Autor et al. (2006), can explain this U-shape, but contradicts the fact that wages in low-skill jobs have been falling relative to those in medium-skill jobs. Indeed, one would think that the opposite occurs if relative demand is rising in the low-skill jobs relative to middle-skill jobs. The modeling approach developed in my paper links the literature on earnings dynamics and wage inequality with the literature on job polarization and investigates this puzzle by testing different hypotheses on the equality of distributions of unobservables over time and across occupations.

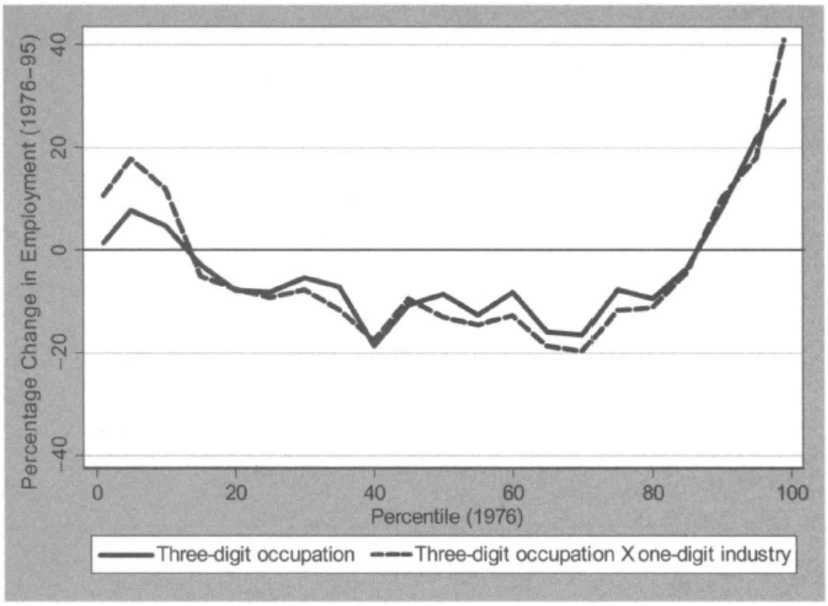

Figure 1: The graph is taken from Goos and Manning (2007). It shows the impact of job polarization on employment growth by wage percentile. Data are taken from NES using 3-digit SOC90 code. Employment changes are taken between 1976 and 1995. Percentiles are the 1976 wage density percentiles.

3 Identification

I start by describing a general state-space model and how a model of earning process can be written in terms of a state-space representation. I then discuss identification and present the main results on asymptotic properties of the distribution estimators of unobserved estimated from state-space models. I consider state-space models both with time-invariant and time-varying parameters. I discuss how the parametric assumption on the error terms helps to establish these results. Finally, I analyze the implications of this assumption and of a long-T approach for identification results in the existing literature.

3.1 Model Setup

The state-space representation of a dynamic system is used to capture the dynamics of an observable variable, y_{it}, in terms of unobservables, known as the state variables for the system, z_{it}. Consider the following state-space representation to describe the dynamic behavior of y_{it}, for $i = 1, ..., N$, and $t = 1, ..., T$:

$$\begin{aligned} y_{it} &= A_{it} z_{it} + D_{it} x_{it} + \sigma_i \epsilon_{it} && \text{(observation equation)} \\ z_{it+1} &= T_{it} z_{it} + R_{it} \eta_{it} && \text{(state equation)} \end{aligned} \tag{1}$$

where I name $\tilde{\epsilon}_{it} \equiv \sigma_i \epsilon_{it}$ the raw errors and ϵ_{it} the standardized errors; $\epsilon_{it} \sim \mathcal{N}(0, H_t)$, $\eta_{it} \sim \mathcal{N}(0, S_{it})$; z_{it} denotes the state variables; $\tilde{\epsilon}_{it}$ and η_{it} are the errors. A vector of exogenous observed variables x_{it} can be added to the system. The state equation describes the dynamics of the state vector, while the observation equation relates the observed variables to the state vector. The unobservables of the model are the (potentially time-varying) parameters, the state variables, and the error terms. To complete the system and start the iteration via Kalman filter I further make the assumption that for each individual i, the initial value of the state vector, z_{i1} is drawn from a normal distribution with mean denoted by $\hat{z}_{i1|0}$ and variance $P_{i1|0}$.[7] Assuming the parameters are known, the Kalman filter recursively calculates the sequences of states $\{\hat{z}_{it+1|t}\}_{t=1}^T$ and $\{P_{it+1|t}\}_{t=1}^T$ where $\hat{z}_{it+1|t}$ is the optimal forecast of z_{it+1} given the set of all past observations $(y_{it}, ..., y_{i1}, x_{it}, ...x_{i1})$, and its mean squared forecast error is $P_{it+1|t}$. It does so by first getting the filtered values of the states $\{\hat{z}_{it|t}\}_{t=1}^T$ and variances $\{P_{it|t}\}_{t=1}^T$. When the interest is in the state vector per se, it is possible to improve inference on it by obtaining the smoothed estimates of the states, i.e. $\{\hat{z}_{it|T}\}_{t=1}^T$ and $\{P_{it|T}\}_{t=1}^T$, i.e. the expected value of the state when all information through the end of the sample, up to time T, is used, and its corresponding mean square error.[8] When parameters are unknown, maximum likelihood estimation is possible

[7] If the vector process z_{i1} is stationary, i.e. if the eigenvalues of T_{it} are all inside the unit circle, then $\hat{z}_{i1|0}$ and $P_{i1|0}$ would be the unconditional mean and variance of this process, respectively. If the system is not stationary or time-varying then they represent the initial guess for z_{i1} and the associated uncertainty.

[8] The general formulas used by the Kalman filter and smoother are provided in Appendix A.

but presupposes the model to be identified.[9]

The model for earnings y_{it}, of an individual i at time t, has the following state-space representation:

$$y_{it} = [p_t]\alpha_i + z_{it} + \sigma_i \epsilon_{it} \qquad \text{(observation equation)}$$
$$z_{it+1} = \rho_i z_{it} + \eta_{it} \qquad \text{(state equation)}$$

with $\epsilon_{it} \sim N(0, H_t)$ and $\eta_{it} \sim N(0, S_{it})$. In this specification, the individual specific component α_i enters the state vector, and the coefficient p_t enters the matrix of parameters A_{it} in the general model described in 1. The factor p_t might be included as a measure of skills price. Note that transitory shocks are assumed to be i.i.d. in these models. However, more general moving average representations, which are common in the earnings literature, can be accommodated by augmenting the state vector accordingly. An extension of this model to include a term $\beta_i t$ can account for an individual's ith specific income growth rate with cross-sectional variance σ^2_β (see HIP model in Guvenen (2009)). A model for earnings could further include job-specific effects γ_i, for job k, with $j_{ik} = \mathbb{1}(\mathrm{K}_i = k)$.

First, I consider a simpler model of earnings and treat the history of each individual i as a separate time series. I provide the argument for identification of the parameters and states, and of their cross-sectional distribution. The identification of the parameters relies on a large T argument, while the asymptotic properties of the estimator of the cross-sectional distribution of parameters and states are established under some ratio of N and T. I derive this ratio and propose a bias correction method to use when T is small. In the second step of the analysis, I introduce time-varying parameters in the state-space model. I discuss how the identification results change in this setting. Finally, in the last part of the analysis, I relate to the nonparametric existing approach, which relies on a fixed-T argument for identification of the unobservables in the model and of their cross-sectional distribution. I show how the parametric assumption on the error term can permit to achieve identification with a shorter number of time periods and discuss whether high-level assumptions for identification hold.

3.2 Benchmark Model

First, treat the earnings history of each individual i as a separate time series. In particular, assume that for each individual i, the time series is represented by the state-space model:

$$y_{it} = \alpha_i + z_{it} + \sigma_i \epsilon_{it} \qquad \text{(observation equation)}$$
$$z_{i,t+1} = \rho_i z_{it} + \eta_{it} \qquad \text{(state equation)}$$

[9] Details on the likelihood are provided in Appendix A.

with $\epsilon_{it} \sim N(0,1)$ and $\eta_{it} \sim N(0, \sigma^2_{i,\eta})$. This model decomposes earnings into a deterministic fixed effect and a stochastic term, which has a transitory and a persistent component. I first discuss how the model's parameters are identified and how it is possible to identify the cross-sectional distribution of the parameters and state variables.

A state-space model is identified when a change in any of the parameters of the state-space model would imply a different probability distribution for $\{y_{it}\}_{t=1}^{\infty}$. There exist several ways of checking for identification. Burmeister, Wall, and Hamilton (1986) provide a sufficient condition for identification: a state-space model is minimal if it is completely controllable with respect to the error term (and external variable directly affecting both the observed and the state variables) and completely observable. If the state-space is minimal, then it is identified.[10] An alternative way of checking identification of a state-space model is to rely on the exact relationship between the reduced form parameters of an ARIMA process and the structural parameters in the state-space model, and use the condition for identification of parameters in ARIMA models. The literature on linear systems has also extensively investigated the question of identification, see Gevers and Wertz (1984) and Wall (1987) for a survey of some of the approaches.

For the above state-space model, it is possible to verify that under stationarity the following holds, $\forall i$:

$$\rho_i = \frac{Cov(y_{it}, y_{it+2})}{Cov(y_{it}, y_{it+1})}$$

$$\sigma_i^2 = Var(y_{it}) - \frac{Cov(y_{it}, y_{it+1})}{\rho_i} = Var(y_{it}) - \frac{Cov(y_{it}, y_{it+1})}{\frac{Cov(y_{it}, y_{it+2})}{Cov(y_{it}, y_{it+1})}}$$

$$\sigma^2_{\eta i} = (Var(y_{it}) - \sigma_i^2)(1 - \rho_i^2)$$

$$\alpha_i = E(y_{it})$$

where the mean, variance, and covariances are moments of the distribution of y_{it} taken over time, for each individual i.

Once I establish identification of the model's parameters, which is based on properties of each individual's ith time series, I can exploit the cross-section of the time series to identify the

[10] The model considered above is observable as the observation matrix has rank equal to the number of state variable where the observation matrix is defined as: $O = [AATAT^2AT^3...AT^n]$ where n is the number of state variables. However, the model is not controllable because there are no observables entering additively into the state equation that one can use to change the direction of the states (to check for controllability test on full rank of controllability matrix). Require $\rho \neq 1$ for observability.

cross sectional distributions of the variables of interest (parameters and states), and analyze the asymptotic properties of these distribution estimators. In line with these results, I derive nonparametric bias correction via split panel Jackknife methods when T is small.

From the above state-space model, I collect all unknown parameters in a vector $\theta_i = \{\alpha_i, \rho_i, \sigma_i^2, \sigma_{\eta i}^2\}$. Let $\hat{\theta}_i$ be the MLE estimator for the vector of parameters θ_i, obtained as: $\hat{\theta}_i = \arg\max_{\theta_i} Q_T(\theta_i)$, where $Q_T(\theta_i) = T^{-1} \sum_{t=1}^{T} \log f(y_{it}; \theta_i) := m(w_{it}, \theta_i)$ and $f(y_{it}; \theta_i)$ is the likelihood from the state-space model as derived in Appendix A. Following a similar notation and argument as in Okui and Yanagi (2020), define $\mathbb{P}_N^{\hat{\theta}} := N^{-1} \sum_{i=1}^{N} \delta_{\hat{\theta}_i}$, as the empirical measure of $\hat{\theta}_i$, where $\delta_{\hat{\theta}_i}$ is the probability distribution degenerated at $\hat{\theta}_i$. Also, let P_0^{θ} be the probability measure of θ_i. Denote as $\mathbb{F}_N^{\hat{\theta}}$ the empirical distribution function, so $\mathbb{F}_N^{\hat{\theta}}(a) = \mathbb{P}_N^{\hat{\theta}} f$ for $f = \mathbb{1}_{(-\infty,a]}$, where $\mathbb{1}_{(-\infty,a]}(x) := \mathbb{1}(x \leq a)$ and the class of indicator functions is denoted as $\mathcal{F} := \{\mathbb{1}_{(-\infty,a]} : a \in \mathbb{R}\}$. Similarly, $\mathbb{F}_0^{\theta}(a) = P_0^{\theta} f$. Finally, denote as $P_T^{\hat{\theta}}$ the probability measure of $\hat{\theta}_i$. In the following, for simplicity of notation, I omit superscripts $\hat{\theta}$ and θ, so $\mathbb{P}_N = \mathbb{P}_N^{\hat{\theta}}$, $\mathbb{F}_N = \mathbb{F}_N^{\hat{\theta}}$, $P_0 = P_0^{\theta}$, $\mathbb{F}_0 = \mathbb{F}_0^{\theta}$, $P_T = P_T^{\hat{\theta}}$, $\mathbb{F}_T = \mathbb{F}_T^{\hat{\theta}}$.

Assumption 1 Assume that $\{\{\epsilon_{it}\}_{t=1}^{T}, \{\eta_{it}\}_{t=1}^{T}\}_{i=1}^{N}$ is i.i.d. across i and y_{it} is a scalar random variable.

Assumption 2 The true parameters θ_i must be continuously distributed.

Assumption 3 Further, assume that: $|\rho| < 1$; θ_i identified, and not on the boundary of parameter space.

Assumption 2 and 3 state standard and sufficient conditions that are required for the ML estimators of the unknown parameters in the time-invariant Gaussian state-space model to be consistent and asymptotically normal. In particular, Assumption 3 is required to establish convergence in probability of $\hat{\theta}_i$ to θ_{i0}, as $T \to \infty$. Note that even without normal distributions the quasi maximum likelihood estimates $\hat{\theta}_i$, obtained assuming Gaussian errors, is consistent and asymptotically normal under certain conditions, see White (1982).

Indeed, the above model is a Gaussian time-invariant state space model, which has a stationary underlying state process (ρ is assumed to be less than 1 in absolute value), and which has the smallest possible dimension (see Hannan and Deistler (2012)). Under these general and sufficient conditions, then the MLE estimator is consistent and asymptotically normal if the true parameters are identified and not at the boundary of the parameter space, see Douc, Moulines, and Stoffer (2014).

Assumption 4 The CDFs of θ_i is thrice boundedly differentiable. The CDFs of $\hat{\theta}_i$ is thrice boundedly differentiable uniformly over T.

Under these assumptions, it is possible to establish uniform consistency and asymptotic normality of the distribution estimator. In the following theorem, I show that the estimator for the distribution of the true individual parameters and states uniformly converges to their true population distribution and it converges in distribution at the rate $N^{3+\epsilon}/T^4$, where $\epsilon \in (0, 1/3)$, if the above assumptions hold.

Theorem 1 Under Assumption 1-4, when $N, T \to \infty$: (i) $\sup |\mathbb{P}_N f - P_0 f| \xrightarrow{as} 0$, where $\xrightarrow{as}$ signifies almost sure convergence. Moreover, (ii) when $N, T \to \infty$, with $N^{3+\epsilon}/T^4 \to 0$ and $\epsilon \in (0, 1/3)$: $\sqrt{N}(\mathbb{P}_N - P_0) \rightsquigarrow G_{P_0}$ in $1^\infty(\mathcal{F})$, where $\rightsquigarrow$ means weak convergence and $\mathbb{G}_{P_0}$ is a Gaussian process with zero mean and covariance function $F_0(a_i \wedge a_j) - F_0(a_i)F_0(a_j)$ with $f_i = \mathbb{1}(-\infty, a_i]$ and $f_j = \mathbb{1}(-\infty, a_j]$ for $a_i, a_j \in R$ and $a_i \wedge a_j$ is the minimum of a_i and a_j.

The key idea behind this result is that the asymptotic properties of the ML estimator $\hat{\theta}_i$ for each individual's i parameters guarantee that it is possible to bound the norm of the difference between the cross-sectional distribution of the ML estimators and the true distribution of the true parameters, i.e. the term $\sup |P_T f - P_0 f|$. See Appendix B for the proof.

Following Okui and Yanagi (2020) and Jochmans and Weidner (2018), when T is small I propose a nonparametric bias correction method via split-panel jackknife (HPJ). I divide the panel along the time series dimensions into two parts and obtain $\hat{F}^{HPJ} = 2\hat{F} - \bar{F}$, where $\hat{F}$ is the estimator obtained using the whole sample, while $\bar{F} = (\hat{F}^1 + \hat{F}^2)/2$ with $\hat{F}^j$ for $j = 1, 2$ being the estimators obtained when using each half of the panel.

3.3 Time-varying Model

When adding time-varying parameters in the state-space model for each i, the derivation of the Kalman filter and smoother is essentially the same as for the case of time-invariant matrices. Note that if the matrices are generic functions of the stochastic variable x_t, then, even if the error terms are normal, the unconditional distribution of the state variable and of the observation y_{it} is no longer normal, while normality can be established conditionally on the past observations and x_t.

Assumption 3 in Theorem 1 can be modified by using existing results that provide conditions on asymptotic properties of the ML estimator for time-varying state-space models. Indeed, assumption 3 can be relaxed along several dimensions: it is possible to rely on results in Chapter 7 of Jazwinski (1969) for a departure of the time-invariance assumption, and it is further possible to weaken the assumption that $\rho < 1$ for stability of the filter, as in Harvey (1990).

For time-varying parameters that are common across (groups of) individuals, I consider a multivariate version of the state-space model above. I consider stratification by observables and, within each group, I impose common time-varying parameters (e.g. price of skills) and individual-specific parameters. The main challenge is that the Kalman filter and smoother can be computationally intense or even infeasible when the cross-sectional dimension N is large. I give proposals on how to deal with these issues in the estimation section.

3.4 Relation to Non-Parametric Literature

Finally, I consider a fixed-T approach to establish a comparison with nonparametric estimation (Almuzara, 2020) and analyze how the results differ when I impose a parametric assumption on the error term. Consider the following simple process for log labor income of individual i at time t:

$$y_{it} = z_{it} + \sigma_i \epsilon_{i,t} \tag{2}$$

$$z_{it} = z_{it-1} + \eta_{it} \tag{3}$$

where z_{it} and $\epsilon_{i,t}$ are unobserved components; $E(\sigma_i^2) = 1$, and the initial level of the random walk is $z_{i1} = z_i$. But impose $\epsilon_{i,t} \sim N(0, \sigma_\epsilon^2)$; the distribution of the raw errors $\tilde{\epsilon}_{i,t} = \sigma_i \epsilon_{i,t}$ is quite flexible, depending on the distribution of heterogeneous variance. It is a special case of the general state-space model above. In the following, I show first identification of the moments of the cross-sectional distribution of (σ_i^2, z_i), and then identification of their joint distribution.

With stationarity only, need $T \geq 3$ for identification of $Cov(z_i, \sigma_i^2)$ and $T \geq 4$ for $Var(\sigma_i^2)$ (Almuzara, 2020).

$$Cov(y_{it}, y_{it+k}) = \begin{cases} \sigma_z^2 + \sigma_\epsilon^2 & \text{if } k = 0, \ \ t = 1 \\ \sigma_z^2 + \sum_{s=2}^{k} \sigma_\eta^2 + \sigma_\epsilon^2 & \text{if } k = 0, \ \ t > 1 \\ \sigma_z^2 & \text{if } k > 0, \ \ t = 1 \\ \sigma_z^2 + \sum_{s=2}^{k} \sigma_\eta^2 & \text{if } k > 0, \ \ t > 1 \end{cases}$$

$$Cov(z_i, \sigma_i^2) = \frac{Cov(y_{it}, (\Delta y_{i\tau+1})^2)}{2\sigma_\epsilon^2} \qquad \tau > t + 1$$

$$Var(\sigma_i^2) = \frac{Cov((\Delta y_{it})^2, (\Delta y_{i\tau+2})^2)}{4\sigma_\epsilon^4} \qquad \tau > t + 1$$

Can reduce T if assuming Gaussian shocks: need $T \geq 2$ for identification.

$$Cov(y_{it}, y_{it+k}) = \begin{cases} \sigma_z^2 + \sigma_\epsilon^2 & \text{if } k = 0, \ \ t = 1 \\ \sigma_z^2 + \sum_{s=2}^{k} \sigma_\eta^2 + \sigma_\epsilon^2 & \text{if } k = 0, \ \ t > 1 \\ \sigma_z^2 & \text{if } k > 0, \ \ t = 1 \\ \sigma_z^2 + \sum_{s=2}^{k} \sigma_\eta^2 & \text{if } k > 0, \ \ t > 1 \end{cases}$$

$$Cov(z_i, \sigma_i^2) = \frac{Cov(y_{it}, (\Delta y_{it+1})^2)}{2\sigma_\epsilon^2}$$

$$Var(\sigma_i^2) = \frac{Var((\Delta y_{it})^2) - 4(1-\sigma_\epsilon^4) + \sigma_\eta^2(\sigma_\eta^2 - 8\sigma_\epsilon^2)}{8 + 6\sigma_\epsilon^4}$$

For the latter use Gaussian nature of η but can relax this assumption using the moments $E[y_{it+1}^4] - E[y_{it}^4]$.

As for identification of the cross sectional distribution of the unobservables (σ_i^2, z_i) under Gaussian error, the argument in Hu and Schennach (2008) would simplify here as there is no need for instruments. Let's denote by y earnings, by x lagged earnings, and by x^* the unobservables of interest (σ_i^2, z_i).

$$f(y,x) = \int f(y|x^*)f(x|x^*)f(x^*)dx^*$$

Note that $f(y|x^*)$ and $f(x|x^*)$ are known up to parameters. Then, it is possible to identify the unobserved distribution of interest $f(x^*)$ with just (y,x), no need for additional z, by solving the above for $f(x^*)$ in terms of known objects. Identifiability requires the integral operator to be invertible, this is a completeness condition. If I define y to be two-dimensional I do not need x and identification of $f(x^*)$ is obtained as follows:

$$f(y) = \int f(y|x^*)f(x^*)dx^*$$

Without the parametric assumption on the error term, I need to introduce the variable z, which is further lags or leads of y, i.e. more time periods are required (5 time periods for this simple model, see argument in Almuzara (2020)). Note the analogy with the logic of Mavroeidis et al. (2015), which is based on a fixed-T setting and require a parametric assumption on the distribution of error term. Consider again the simple state-space model:

$$y_{it} = z_{it} + \sigma_i \epsilon_{i,t} \tag{4}$$

$$z_{it} = z_{it-1} + \eta_{it} \tag{5}$$

Identification relies on the equality:

$$f_{Y_T,...,Y_2|Y_1}(y_T,...,y_2|y_1) = \int\int\int f_{\zeta,\sigma_\epsilon,\sigma_\eta|Y_1}(z,s_\epsilon,s_\eta|y_1)$$
$$f_{Y_T,...,Y_2|\zeta,\sigma_\epsilon,\sigma_\eta,Y_1}(y_T,...,y_2|z,s_\epsilon,s_\eta,y_1)dz ds_\epsilon ds_\eta$$

Provided that the solution exists, one can recover the unknown primitive $f_{\zeta,\sigma|Y_1=y_1}$ by solving the linear equation:

$$f_{\zeta,\sigma_\epsilon,\sigma_\eta|Y_1}(z, s_\epsilon, s_\eta|Y_1 = y_1) = L^{-1} f_{Y_T,...,Y_2|Y_1=y_1}$$

where L is the linear integral operator:

$$\begin{aligned} & L(\xi)(Y_T, ..., Y_2) \\ = & \int\int\int \xi(z, s_\epsilon, s_\eta) \\ & f_{Y_T,...,Y_2|\zeta,\sigma_\epsilon,\sigma_\eta}(y_T, ..., y_2|z, s) dz ds_\epsilon ds_\eta \end{aligned}$$

For identification, need the linear operator $L : \mathcal{L}^2(F_{\zeta,\sigma|Y1=y1}) \to \mathcal{L}^2(F_{Y_T,...,T_2|Y1=y1})$ to be complete, i.e. $Lf = 0$ in $\mathcal{L}^2(F_{Y_T,...,T_2|Y1=y1})$ implies $f = 0$ in $\mathcal{L}^2(F_{\zeta,\sigma_\epsilon,\sigma_\eta|Y1=y1})$.

[On the conditions for identification, the L^2-completeness conditions can be very difficult or impossible to test.[11] The paper of Andrews (2011) proposes a class of distributions satisfying this conditions but it doesn't extend to multivariate case. Characterization of completeness via characteristic function as in D'Haultfoeuille (2011) may extend to multivariate cases. See also paper of Seely on Completeness for a Family of Multivariate Normal Distributions, given that both ϵ and η are normally distributed.] It is possible to use the argument in Newey and Powell (2003) to this case given the assumption of normality in the univariate case. Extension to the multivariate case can be established using the results in Lemma 7 of Hu and Schennach, which reduce a multivariate completeness problem to a single variate one, under some independence assumptions on the endogenous variables. Gaussian likelihood introduces irregular identification (Escanciano, 2020), one way of dealing with this is to employ sieve methods with incomplete sieve basis.

4 Estimation

In this Section I provide some details on the estimation procedure, starting from the long-T approach, which I adopt in the empirical application, and then considering the alternative fixed-T estimation procedure.

4.1 Main estimation

State-space estimation and filtering with heterogeneous dynamic panel data pose econometric challenges. Estimation of the distribution of unobservables is performed in 2 stages: a first step of estimation is performed via state-space methods; then, in a second step, I obtain the

[11] Canay, Santos, and Shaikh (2013) conclude that no nontrivial tests for testing completeness conditions in nonparametric models with endogeneity involving mean independence restrictions exist.

empirical cross-sectional distribution of unobservables estimated in the first step.
In the first step, estimation of model' s parameters is based on maximum likelihood.[12] I employ the kalman filtering and smoothing algorithm to get smoothed estimates of state variables and error terms.

The econometric challenge in this first step of estimation is on how to deal with state-space models for a dataset featuring a large cross-section N: given recursive nature of filter, at each period inversion of $F_t = Var(v_t|y_{t-1})$, where $v_t = y_t - A_t E[z_t|Y_{t-1}]$ is the innovation, can be problematic, see Durbin and Koopman (2012) (F_t has size N x N, computationally costly with large N). In the models I consider, H_t is diagonal, hence, it is possible to adopt matrix identity for inverse of F_t. Moreover, I perform stratification as a way to avoid intractability while also addressing the issue of not restricting the dependence of earnings on covariates. When introducing time-varying parameters, I impose that within each group some parameters are common and time-varying parameters (e.g. price of skills), while others are individual-specific (e.g. the standard deviation of the shocks as reported in the matrix R_{it} in model **??**).
For starting the recursions, I implement diffuse initialization as in De Jong et al. (1991), i.e. the uncertainty around initial states is represented in the model with an arbitrarily large covariance matrix for the initial state distribution.[13]

Once (smoothed) estimates of unobservables are obtained, I obtain the empirical cross-sectional distribution of the unobserved components estimated from the state-space models in the second step of the estimation strategy. Note that dimensionality of vector y_t can vary over time. Thus, the methodology can be easily extended to deal with unbalanced panel data.

4.2 Fixed-T estimation

For fixed-T, the corresponding estimation approach is based on sieve nonparametric maximum likelihood (see Mavroeidis et al. (2015)):

$$max_{\theta \in \Theta_{k(N)}} \sum_{i=1}^{N} log \int \int \int f_{\zeta,\sigma_\epsilon,\sigma_\eta,Y_1:\theta}(z, s_\epsilon, s_\eta, y_{i1})$$
$$f_{Y_T,...,Y_2|\zeta,\sigma_\epsilon,\sigma_\eta,Y_1}(y_T, ..., y_2|z, s_\epsilon, s_\eta, y_1)dzds_\epsilon ds_\eta$$

where $\Theta_{k(N)}$ denotes a sieve space whose dimension $k(N)$ increases with the sample size N; and $\Theta \subset \mathcal{L}^1(F_{\zeta,\sigma_\epsilon,\sigma_\eta,Y1=y1})$.

[12] Details on the likelihood are provided in Appendix A.
[13] Durbin and Koopman (2012) show that initialization of the Kalman filter is not affected the choice of representing the initial state as a random variable with infinite variance as opposed to assuming that it is fixed, unknown and estimated from observations at t=1.

5 Discussion on Gaussian Error and Extension

One might be worried that the parametric assumption is quite restrictive. Horowitz and Markatou (1996) provide empirical evidence that the normal distribution can approximate well the distribution of the permanent component of the income process. However, there might be concerns that the parametric assumption is restrictive for the transitory component of earnings shocks. Indeed, there is empirical evidence that the cross-sectional distribution of transitory shocks features negative skewness and high kurtosis. These stylized facts have been documented, among others, by Arellano et al. (2017) as relevant features of the earnings process.

First and importantly, note that when errors are not normally distributed, results from Gaussian state-space analysis are still valid in terms of Minimum Variance Linear Unbiased Estimation (MVLUE): Kalman Filter estimates are not necessarily optimal, but they will have the smallest mean squared errors with respect to all other estimates based on a linear function of the observed variables $(y_{it}, y_{it-1}, ..., y_{i1}, x_{it}, x_{it-1}, ..., x_{i1})$, see Anderson and Moore (1989).[14]

Second, the homogeneity assumptions may explain some of these stylized facts: once allowing for rich heterogeneity, it is unclear whether the residuals will still display the same features. One interesting empirical question is to test to what extent these features are still present when allowing for rich heterogeneity and time-varying parameters. Assuming individual Gaussian shocks with heterogeneous variances allows obtaining flexible cross-sectional distributions and, depending on the cross-sectional distributions of the heterogeneous variances, the resulting cross-sectional distribution might display the above key features.

Finally, extensions to different distributions are feasible within a state-space framework. Alternative assumptions on error terms can be considered by non-Gaussian state-space models; for instance, the error term can be assumed to follow a Mixture of Normals distribution. It would be interesting to see how much the goodness of fit improves when the assumption on Gaussian shocks is relaxed.

6 Data

The dataset used for the empirical application is a novel confidential dataset for the UK, the New Earnings Survey Panel Data (NESPD). It is an annual panel, running from 1975 to 2016. All individuals whose National Insurance Number ends in a given pair of digits are included in the survey, making it representative of the UK workforce.[15] It surveys around 1% of the UK workforce.

[14] The sketch of the argument is provided in appendix C.

[15] It might under-sample part-time workers if their weekly earnings falls below the threshold for paying National Insurance and those that moved jobs recently. Thus, the following categories are likely to be under-sampled: self-employed, some groups of seasonal workers, and those working only few hours irregularly. To address these concerns, one could perform robustness checks using the Labour Force Survey data, which, however, has a much smaller sample size.

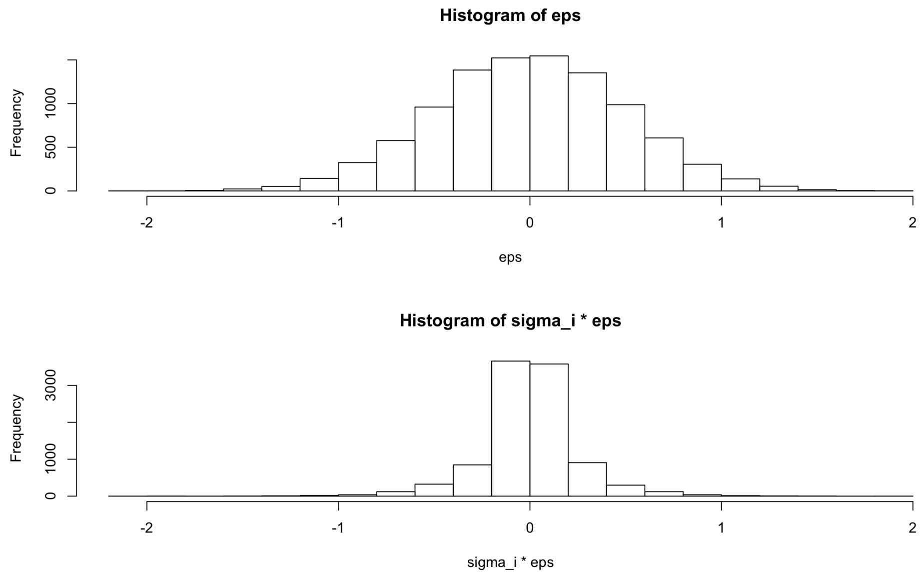

Figure 2: Monte Carlo simulation showing that, despite the assumption of Gaussian errors at individual level, the cross sectional distribution of raw errors can display very high kurtosis (and potentially also skewness) depending on the cross-sectional distribution of heterogeneous variances, σ_i^2.

The questionnaire is directed to the employer, who completes it based on payroll records for the employee; the survey contains information on earnings, hours of work, occupation, industry, gender, age, working area, firms' number of employers, and unionization. This information relates to a specified week in April of each year: the data sample is taken on the 1st of April of each calendar year and concerns complete employee records only. As a result of being directed to the employer, NESPD has a low non-classical measurement error and attrition rate. Descriptive statistics are reported in Appendix D.

For both Standard Industry Classification (SIC) and Standard Occupation Classification (SOC) codes, different classifications have been used over time. I report SIC and SOC codes to the same classification using conversion documents provided by ONS on their website: I map all SIC codes into the SIC07 Division (2-digit); for those divisions where there are multiple correspondences, I use the information on whether the individual has stayed in the same job in the last twelve months to identify the mapping; I proceed in an analogous way for SOC codes.

I rank occupations by percentiles of the median wage distribution in the starting year and separate them accordingly into three groups: low-skill, medium-skill, and high-skill occupations.[16]

Given large dimensions, this dataset is particularly suited to obtain a flexible treatment of

[16] Another classification might be based on routine task intensity of occupations since one of the main hypotheses put forward to explain job polarization is the bias of recent technological change towards replacing labor in routine tasks (this is called routine-biased technological change, RBTC, by Goos, Manning, and Salomons (2014)).

covariates by stratification. I stratify by observables instead of running a first-stage regression on covariates which restricts the dependence of earnings on them. Stratification allows considering a specification for the earning process that features type dependence in a flexible way. In particular, I perform stratification by occupations: high-skill occupations, medium-skill occupations, and low-skill occupations; by age groups; and by gender.

7 Empirical Application

Over the last 40 years, in the US and UK, there has been a significant increase in employment shares in low-skill occupations and high-skill occupations, and a simultaneous decrease in employment shares in middle-skill occupations. Goos and Manning (2007) document that this phenomenon, known as job polarization, has occurred in the UK since 1975. A likely explanation for it is the automation of some types of jobs only, the middle-skill jobs, which require precision and are easy to be replaced by machines.

In the following figures, the phenomenon of job polarization results in the characteristic U-shape with much a negative change in employment share for middle-skill occupations. Note that this pattern is observed over the whole period, and is not driven by a change in the gender composition of the workforce.[17]

As a result of job polarization, one would expect an increase in wages for both low-skill and high-skill occupations, while a decrease in wages for medium-skill occupations. Indeed, job polarization would predict a rising relative demand in the low-skill relative to middle-skill jobs. However, this has not been the case: on the contrary, earnings inequality also between low and median wages has increased over time. Part of the increase in wage inequality might be justified by the fact that wage growth is monotonically positively related to the quality of jobs. If one includes more controls, the within job inequality significantly reduces. Once one controls for job-specific effects, there should only be between job inequality, not within. However, as suggested by Goos and Manning (2007) the findings that wages in low-skill jobs are falling relative to those in middle-skill jobs presents something of a problem for the routinization hypothesis, as one might expect the opposite if relative demand is rising in the low-skill jobs relative to middle-skill jobs.

The methodology proposed in the paper is used to shed light on the relation between job polarization and earnings inequality, which is relevant to think about the evolution of labor markets and inequality, also during and after the COVID-19 pandemic. The goal of the empirical analysis is to relate the components and dynamics of the earnings process to the phenomenon of job polarization, which is usually investigated only with a static approach. To this aim, I am going to test different hypotheses on the degree of heterogeneity of the distributions of unobservables, by observables and over time, to shed light on this puzzling empirical evidence.

[17]Results are robust to the chosen level of disaggregation by occupation.

More specifically, first I am going to consider the time-invariant model used as benchmark model in the analysis. In a second step of the analysis, I am going to introduce time varying parameters, in the form of a time-varying price of skills (p_t) in the model above, and by allowing the variances of the shocks to be time-varying. For both models, for each group obtained by stratification by observables, I use state-space analysis to obtain (smoothed) estimates of unobservables. Finally, I estimate the cross-sectional distribution of the unobservables, potentially for aggregated strata in order to recover a larger cross-sectional dimension needed for inference on distributions. I compare these distributions via tests of the null hypothesis of equal distributions by Kolmogorov-Smirnov test to test for different degrees of heterogeneity.[18]

7.1 Toy Model

The following toy model is used to motivate things and illustrate some of the underlying mechanisms that I would like to test.

Consider a model with two types of individuals, $i \in \{LG, HG\}$, where LG stands for low growth type and HG for high growth type. Further, assume that there are 3 types of occupations, $k \in \{LS, MS, HS\}$, i.e. low-skill, medium-sill, and high-skill occupations. The price of the skills in occupation k, at time t, is $\pi_{k,t}$, and $\pi_{LS,t} \leq \pi_{MS,t} \leq \pi_{HS,t}$. The individual i's earnings at time t from occupation k is:

$$y_{i,k,t} = \pi_{k,t}(\alpha_{i,k} + \beta_{i,k}h_{i,t})$$

where α_{ik} and β_{ik} are the heterogeneous level and slope, which are time-invariant. The individual's problem at time s is:

$$max_k \sum_{t=s}^{T} E(y_{i,k,t})\beta_d^t$$

where β_d is the discount factor. In this scenario one moves from MS to LS occupation if either displaced with probability δ_i or if $\pi_{MS,t}(\alpha_{i,MS}+\beta_{i,MS}h_{i,t}) < \pi_{LS,t}(\alpha_{i,LS}+\beta_{i,LS}h_{i,t})$. Analogously from HS to MS.

I model routinization as a negative demand shock in MS occupation, i.e. $\pi_{MS,t}$ decreases. After this shock, all HG type move from MS occupations to HS occupations, or stay in MS occupations. Vice versa all LG type move from MS occupations to LS occupations, or stay in MS occupations. Assume that, after the shock, for $i = HG$, $\pi_{MS,t}(\alpha_{HG,MS} + \beta_{HG,MS}h_{HG,t}) \leq \pi_{HS,t}(\alpha_{HG,HS} + \beta_{HG,HS}h_{HG,t})$ and $\pi_{LS,t}(\alpha_{LG,LS} + \beta_{LG,LS}h_{LG,t}) \geq \pi_{MS,t}(\alpha_{LG,MS} + \beta_{LG,MS}h_{LG,t})$. Assuming that there is a nonzero outflow of people from MS occupation, the overall effect would be an increase in inequality.

[18] There might be a problem of independence if aggregate time effects are taken into account.

Now, let's consider a more realistic earnings process by adding the stochastic persistent and transitory components $z_{i,t} + \epsilon_{i,t}$:

$$y_{i,k,t} = \pi_{k,t}(\alpha_{i,k} + \beta_{i,k}h_{i,t}) + z_{i,t} + \epsilon_{i,t}$$

An increase in inequality might occur also if the variances of the stochastic components significantly changed over time and by different type of occupation. This might happen as a result of changes in institutions that have lead to a decline in wages at the bottom of the distribution. In UK there has been a marked decline of both unionization and minimum wage over time.

Several hypotheses can be tested to investigate this phenomenon:

H1: Change in prices of skills by occupation and over time. (?)

H2: Distribution of skills changed over time as result of job mobility/displacement. (?)

H3: Alternative explanation: higher skills for those in middling occupations. Also in line with literature on displaced workers.

H4: Distribution of variance of transitory shocks more concentrated depending on the evolution of unionization and minimum wage over time. I investigate this channel given that a possible explanation for increase in inequality might be that change in institutions have been in such a way to lead to a fall in wages at the bottom of the distribution.

8 Empirical Findings

The empirical findings provide evidence that earnings dynamics feature considerable unobservable heterogeneity. First, I uncover the amount of unobservable heterogeneity using the simple time-invariant model considered as benchmark model in the theoretical section. I document that workers in middle-skill occupations display significantly different earnings dynamics with respect to workers in other occupations. In particular, as shown in the table below, persistence to earnings shocks for workers in middle-skill jobs is on average smaller, over the entire time period. The distribution of persistence has the largest dispersion for workers in low-skill occupations. Moreover, empirical evidence suggests a relatively higher correlation between the skills of workers in middle-skill occupations and the dispersion of earnings shocks they face.

To test the hypotheses presented in the above section, I introduce time-varying parameters in the state-space model. I obtain that there has been a pattern of increase in the prices of skills for workers in low- and high-skill occupations, while the change over time of the skill prices for workers in middle-skill occupations has been unstable as shown in the figure below.

These preliminary findings can be interpreted as suggestive of a pattern of negative demand shocks in MS occupations over the considered time period. Moreover, there has been a shift in the distribution of skills for individual in MS-occupations due to a compositional change in the UK workforce. Finally, the dispersion of the variance of transitory shocks has increased over

		α_i	ρ_i	σ_i^2	α_i	ρ_i	σ_i^2
		1975-1999			2000-2005		
LS	Mean	-0.1909	0.5424	0.0442	-0.3085	0.5028	0.0671
	St. Dev.	0.3112	0.5391	0.0639	0.3534	0.5607	0.1278
	IQR	0.3993	0.7996	0.0422	0.4107	0.8280	0.0724
MS	Mean	-0.1140	0.4620	0.0354	-0.1054	0.4731	0.0368
	St. Dev.	0.2909	0.5526	0.0408	0.3342	0.5440	0.0626
	IQR	0.3719	0.7686	0.0354	0.4760	0.8267	0.0329
HS	Mean	0.1527	0.5095	0.0278	0.2507	0.5926	0.0340
	St. Dev.	0.2873	0.5260	0.0416	0.3879	0.5366	0.0750
	IQR	0.3718	0.7501	0.0259	0.4536	0.7926	0.0293

Table 1: The table reports the means, standard deviation, and interquartile range (IQR) of the cross-sectional distributions of α_i, ρ_i, and σ_i^2, for workers in LS occupations, MS-occupations, and HS-occupations, for two time windows: 1975-1999, 2000-2005. Split-panel jackknife (HPJ) is used for bias correction.

time, comparatively more for workers in LS-occupations.

9 Conclusion

In this paper, I propose a formal econometric framework for studying identification and estimation of unobservable heterogeneity and its dynamics. I adapt state-space methods to the analysis of heterogeneous dynamic structures with micropanels. The framework proposed in this paper allows for rich heterogeneity and dynamics in models, while a mild parametric yet flexible assumption on the distribution of the shocks provides several advantages for identification and estimation.

The framework in this paper will enable empirical researchers to answer a variety of new empirical questions using administrative data. Moreover, it naturally lends itself to important and useful generalizations such as allowing for common unobservable macro shocks, trends, seasonality, and nonlinearities.

In the empirical application, I use a novel dataset on UK workers, the NESPD, to uncover unobserved heterogeneity in earnings processes and investigate how this is related to the phenomenon of job polarization.

A natural next step in the analysis is to combine the information on UK workers provided by the NESPD with information about the supply side as reported in another novel UK dataset, the Business Structure Database (BSD), which can be merged with NESPD to get a matched employee-employer dataset for UK.

Appendices

A Kalman Filter and Smoother

The Kalman filter is an algorithm that recursively calculates $\{\hat{z}_{it+1|t}\}_{t=1}^{T}$ and $\{P_{it+1|t}\}_{t=1}^{T}$ and given the initial $\hat{z}_{i1|0}$ and $P_{i1|0}$, it is implemented by iterating on the following two equations:

$$\hat{z}_{it+1|t} = T_{it}\hat{z}_{it|t-1} + T_{it}P_{it|t-1}A_t(A_tP_{it|t-1}A_t + \sigma_iH_t)^{-1}(y_{it} - A_t\hat{z}_{it|t-1} - D_tx_{it})$$
$$P_{it+1|t} = T_{it}P_{it|t-1}T'_{it} - T_{it}P_{it|t-1}A_t(A_tP_{it|t-1}A_t + \sigma_iH_t)^{-1}A'_tP_{it|t-1}T'_{it} + S_{it}$$

given that:

$$\hat{z}_{it|t} = \hat{z}_{it|t-1} + P_{it|t-1}A_t(A_tP_{it|t-1}A_t + \sigma_iH_t)^{-1}(y_{it} - A_t\hat{z}_{it|t-1} - D_tx_{it})$$
$$P_{it|t} = P_{it|t-1} - P_{it|t-1}A_t(A_tP_{it|t-1}A_t + \sigma_iH_t)^{-1}A'_tP_{it|t-1}$$

and

$$\hat{z}_{it+1|t} = T_{it}\hat{z}_{it|t}$$
$$P_{it+1|t} = T_{it}P_{it|t}T'_{it} + S_{it}$$

Once I run the Kalman filter and get the sequences $\{\hat{z}_{it+1|t}\}_{t=1}^{T}$ and $\{P_{it+1|t}\}_{t=1}^{T}$, and $\{\hat{z}_{it|t}\}_{t=1}^{T}$ and $\{P_{it|t}\}_{t=1}^{T}$, it is possible to proceed in reverse order in order to calculate the sequence of smoothed estimates $\{\hat{z}_{it|T}\}_{t=1}^{T}$ and their corresponding mean squared errors $\{P_{it|T}\}_{t=1}^{T}$, as follows:

$$\hat{z}_{it|T} = \hat{z}_{it|t} + P_{it|t}T'_{it}P^{-1}_{it+1|t}(\hat{z}_{it+1|T} - \hat{z}_{it+1|t})$$
$$P_{it|T} = P_{it|t} + P_{it|t}T'_{it}P^{-1}_{it+1|t}(P_{it+1|T} - P_{it+1|t})P^{-1}_{it+1|t}T_{it}P_{it|t}$$

for $t = T-1, T-2, ..., 1$, while $\hat{z}_{iT|T}$ and $P_{iT|T}$ are set equal to the terminal state of the sequence obtained with the Kalman filter and associated variance.

The above recursions are made assuming that the matrices of parameters are known. However, typically parameters are unknown. Denote by θ_i the vector containing all the unknown elements in these matrices for individual i. When one needs to estimate the parameter vector θ_i, one builds the likelihood for the observations y_{it} given its past values and the observables $x_{it}, x_{it-1}, ..., x_{i1}$, for an initial arbitrary guess on θ_i, θ_{i0}. In particular, $y_{it}|x_{it}, ..., x_{i1}, y_{it-1}, ..., y_{i1}; \theta_{i0} \sim \mathcal{N}(\mu_{it}(\theta_{i0}), \Sigma_{it}(\theta_{i0}))$, where $\mu_{it}(\theta_{i0} = D_{it}(\theta_{i0})x_{it} + A_{it}(\theta_{i0})\hat{z}_{it|t-1}(\theta_{i0})$ and $\Sigma_{it}(\theta_{i0}) = A_{it}(\theta_{i0})P_{t|t-1}(\theta_{i0})A_{it}(\theta_{i0})' + \sigma_i(\theta_{i0})H_t(\theta_{i0})$. Based on this, the value of the log-likelihood is:

$$\sum_{t=1}^{T} log f(y_{it}|x_{it},...,x_{i1},y_{it-1},...,y_{i1};\theta_{i0}) =$$

$$k - \frac{1}{2}\sum_{t=1}^{T} log|\Sigma_{it}(\theta_{i0})| - \frac{1}{2}\sum_{t=1}^{T}[y_{it}-\mu_{it}(\theta_{i0})]'\Sigma(\theta_{i0})^{-1}[y_{it}-\mu_{it}(\theta_{i0})]$$

where k is a constant, and the likelihood is evaluated at the initial guess for the unknown parameters. For alternative guesses one proceed to maximize the value of the log-likelihood by numerical method and find the Maximum Likelihood estimates of θ_{i0}. Many alternative optimization techniques exist, one attractive option is the EM algorithm of Watson and Engle (1983).

B Proof of Theorem 1

The proof of Theorem 1 has two parts, one for uniform convergence (i) and the other for convergence in distribution (ii): (i) As in Okui and Yanagi (2020), the proof for uniform convergence starts from the following triangle inequality:
$\sup_{f\in\mathcal{F}}|\mathbb{P}_N f - P_0 f| \leq \sup_{f\in\mathcal{F}}|\mathbb{P}_N f - P_T f| + \sup_{f\in\mathcal{F}}|P_T f - P_0 f|$. The goal is to show that the term on the left-hand side is bounded by 0. This proof is composed of two steps: in a first step, I bound the second term on the right-hand side by using the convergence in distribution of the MLE estimator. In a second step, I follow Okui and Yanagi (2020) and bound the first term using a modification of the steps in the Glivenko-Cantelli theorem that accounts for the fact that the true distribution of $\hat{\theta}_i$ changes as T increases. In particular, in the first step, I use Assumption 3 to ensure that $\hat{\theta}_i$ converges to θ_i in distribution. Moreover, given that θ_i is continuously distributed by Assumption 2, then Lemma 2.11 in van der Vaart (1998) implies that $\sup_{f\in\mathcal{F}}|P_T f - P_0 f| \to 0$. The second part of the proof is exactly as in Okui and Yanagi (2020) to show that the first term almost surely converges to 0. The assumptions required for this step are assumption 1, condition 1.5 in Hu et al. (1989) and Condition 1.6 in Hu et al. (1989) when I set $X = 2$ in Condition 1.6, which are both satisfied here.

(ii) The proof for convergence in distribution follows a similar logic. □

C Properties of Kalman Filter Estimator

The Kalman filter estimator obtained under the Gaussian assumption is the Minimum Variance Linear Unbiased Estimator (MVLUE) even when true errors are not normally distributed. To see this consider estimation of x when x is unknown and y is known.

$$E\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \qquad Var\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{xy} & \Sigma_{yy} \end{bmatrix} \tag{6}$$

$$\hat{x} = E[x|y] = \mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y)$$
$$Var(\hat{x} - x) = \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{xy}$$

then $\hat{x}$ is MVLUE regardless normality of (x, y).
Consider an unbiased linear estimator $\bar{x} = \beta + \gamma y$, and see that for $\gamma = \Sigma_{xy}\Sigma_{yy}^{-1}$ then $\bar{x} = \hat{x}$, so it is unbiased and it can be shown to have minimum variance. When matrices Z_t and T_t do not depend on previous y_t's, then under appropriate assumptions the values of the states given by the Kalman Filter minimize the variance matrices of the estimates of $z_{t|t}$ and z_{t+1} given y_t.

D Descriptive Statistics

In the following table I report some descriptive statistics.

References

Abowd, J. M. and D. Card (1989): "On the Covariance Structure of Earnings and Hours Changes," *Econometrica*, 57, 411–445.

Acemoglu, D. and J. A. Robinson (2014): "The rise and fall of general laws of capitalism," *Unpublished Paper, MIT, Department of Economics*.

Anderson, B. and J. Moore (1989): "Optimal control-linear optimal control," *Prcntice—Hall, Englewood Cliffs*.

Arellano, M., R. Blundell, and S. Bonhomme (2017): "Earnings and consumption dynamics: a nonlinear panel data framework," *Econometrica*, 85, 693–734.

Autor, D. H., L. F. Katz, and M. S. Kearney (2006): "The polarization of the US labor market," *The American economic review*, 96, 189–194.

Bai, J. (2009): "Panel data models with interactive fixed effects," *Econometrica*, 77, 1229–1279.

Baker, M. (1997): "Growth-rate heterogeneity and the covariance structure of life-cycle earnings," *Journal of labor Economics*, 338–375.

Baltagi, B. H. (2008): "Forecasting with panel data," *Journal of Forecasting*, 27, 153–173.

Blundell, R., L. Pistaferri, and I. Preston (2008): "Consumption inequality and partial insurance," *The American Economic Review*, 98, 1887–1921.

Bonhomme, S. and J.-M. Robin (2010): "Generalized non-parametric deconvolution with an application to earnings dynamics," *The Review of Economic Studies*, 77, 491–533.

Browning, M., M. Ejrnaes, and J. Alvarez (2010): "Modelling income processes with lots of heterogeneity," *The Review of Economic Studies*, 77, 1353–1381.

Burmeister, E., K. D. Wall, and J. D. Hamilton (1986): "Estimation of unobserved expected monthly inflation using Kalman filtering," *Journal of Business & Economic Statistics*, 4, 147–160.

Canay, I. A., A. Santos, and A. M. Shaikh (2013): "On the testability of identification in some nonparametric models with endogeneity," *Econometrica*, 81, 2535–2559.

Chamberlain, G. (1984): "Panel data," *Handbook of econometrics*, 2, 1247–1318.

Chamberlain, G. and K. Hirano (1999): "Predictive distributions based on longitudinal earnings data," *Annales d'Economie et de Statistique*, 211–242.

Clark, T. E. and M. W. McCracken (2001): "Tests of equal forecast accuracy and encompassing for nested models," *Journal of econometrics*, 105, 85–110.

Commandeur, J. J. and S. J. Koopman (2007): *An introduction to state space time series analysis*, Oxford University Press.

Daly, M., D. Hryshko, and I. Manovskii (2014): "Reconciling estimates of earnings processes in growth rates and levels," .

De Jong, P. et al. (1991): "The diffuse Kalman filter," *The Annals of Statistics*, 19, 1073–1083.

De Nardi, M., G. Fella, and G. P. Pardo (2016): "The implications of richer earnings dynamics for consumption, wealth, and welfare," Tech. rep., National Bureau of Economic Research.

DEATON, A. (1991): "SAVING AND LIQUIDITY CONSTRAINTS," *Econometrica*, 59, 1221–1248.

Douc, R., E. Moulines, and D. Stoffer (2014): *Nonlinear time series: Theory, methods and applications with R examples*, CRC press.

Durbin, J. and S. J. Koopman (2012): *Time series analysis by state space methods*, Oxford university press.

Ejrnæs, M. and M. Browning (2014): "The persistent–transitory representation for earnings processes," *Quantitative Economics*, 5, 555–581.

Fields, G. S. and E. A. Ok (1996): "The meaning and measurement of income mobility," *Journal of Economic Theory*, 71, 349–377.

——— (1999): "The measurement of income mobility: an introduction to the literature," in *Handbook of income inequality measurement*, Springer, 557–598.

Fröhwirth-Schnatter, S. and S. Kaufmann (2008): "Model-based clustering of multiple time series," *Journal of Business & Economic Statistics*, 26, 78–89.

Gabaix, X., J.-M. Lasry, P.-L. Lions, and B. Moll (2016): "The dynamics of inequality," *Econometrica*, 84, 2071–2111.

Gamerman, D. and H. S. Migon (1993): "Dynamic hierarchical models," *Journal of the Royal Statistical Society: Series B (Methodological)*, 55, 629–642.

Giacomini, R. and H. White (2006): "Tests of conditional predictive ability," *Econometrica*, 74, 1545–1578.

Goos, M. and A. Manning (2007): "Lousy and lovely jobs: The rising polarization of work in Britain," *The review of economics and statistics*, 89, 118–133.

Goos, M., A. Manning, and A. Salomons (2014): "Explaining job polarization: Routine-biased technological change and offshoring," *American economic review*, 104, 2509–26.

Gu, J. and R. Koenker (2015): "Unobserved heterogeneity in income dynamics: an empirical Bayes perspective," *Journal of Business & Economic Statistics*.

Guvenen, F. (2007): "Learning your earning: Are labor income shocks really very persistent?" *The American economic review*, 687–712.

——— (2009): "An empirical investigation of labor income processes," *Review of Economic dynamics*, 12, 58–79.

Guvenen, F., F. Karahan, S. Ozkan, and J. Song (2015): "What do data on millions of US workers reveal about life-cycle earnings risk?" Tech. rep., National Bureau of Economic Research.

Guvenen, F. and A. A. Smith (2014): "Inferring Labor Income Risk and Partial Insurance From Economic Choices," *Econometrica*, 82, 2085–2129.

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

Hamilton, J. D. (1985): "Uncovering financial market expectations of inflation," *Journal of Political Economy*, 93, 1224–1241.

——— (1994): "State-space models," *Handbook of econometrics*, 4, 3039–3080.

Hannan, E. J. and M. Deistler (2012): *The statistical theory of linear systems*, SIAM.

Hansen, P. R., A. Lunde, and J. M. Nason (2011): "The model confidence set," *Econometrica*, 79, 453–497.

Harvey, A. C. (1990): *Forecasting, structural time series models and the Kalman filter*, Cambridge university press.

Heathcote, J., F. Perri, and G. L. Violante (2010): "Unequal we stand: An empirical analysis of economic inequality in the United States, 1967–2006," *Review of Economic dynamics*, 13, 15–51.

Hoffmann, F. (????): "HIP, RIP and the Robustness of Empirical Earnings Processes," .

Horowitz, J. L. and M. Markatou (1996): "Semiparametric estimation of regression models for panel data," *The Review of Economic Studies*, 63, 145–168.

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

Hu, Y. and S. M. Schennach (2008): "Instrumental variable treatment of nonclassical measurement error models," *Econometrica*, 76, 195–216.

Jantti, M. and S. P. Jenkins (2013): "Income mobility," .

Jazwinski, A. H. (1969): "Adaptive filtering," *Automatica*, 5, 475–485.

Jochmans, K. and M. Weidner (2018): "Inference on a distribution from noisy draws," *arXiv preprint arXiv:1803.04991*.

Karahan, F. and S. Ozkan (2013): "On the persistence of income shocks over the life cycle: Evidence, theory, and implications," *Review of Economic Dynamics*, 16, 452–476.

Kleinberg, J., J. Ludwig, S. Mullainathan, and Z. Obermeyer (2015): "Prediction policy problems," *The American Economic Review*, 105, 491–495.

Korobilis, D. (2016): "Prior selection for panel vector autoregressions," *Computational Statistics & Data Analysis*, 101, 110–120.

Lillard, L. A. and Y. Weiss (1979): "Components of variation in panel earnings data: American scientists 1960-70," *Econometrica: Journal of the Econometric Society*, 437–454.

Liu, L. (2017): "Density Forecasts in Panel Data Models: A Semiparametric Bayesian Perspective," Tech. rep., Working paper, University of Pennsylvania.

Liu, L., H. R. Moon, and F. Schorfheide (2016): "Forecasting with Dynamic Panel Data Models," .

MaCurdy, T. E. (1982): "The use of time series processes to model the error structure of earnings in a longitudinal data analysis," *Journal of econometrics*, 18, 83–114.

Mavroeidis, S., Y. Sasaki, and I. Welch (2015): "Estimation of heterogeneous autoregressive parameters with short panel data," *Journal of Econometrics*, 188, 219–235.

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

Okui, R. and T. Yanagi (2020): "Kernel estimation for panel data with heterogeneous dynamics," *The Econometrics Journal*, 23, 156–175.

Petris, G. and R. An (2010): "An R package for dynamic linear models," *Journal of Statistical Software*, 36, 1–16.

Postel-Vinay, F. and H. Turon (2010): "On-the-job search, productivity shocks, and the individual earnings process," *International Economic Review*, 51, 599–629.

Primiceri, G. E. and T. Van Rens (2009): "Heterogeneous life-cycle profiles, income risk and consumption inequality," *Journal of monetary Economics*, 56, 20–39.

Shorrocks, A. F. (1978): "The measurement of mobility," *Econometrica: Journal of the Econometric Society*, 1013–1024.

Topel, R. H. and M. P. Ward (1992): "Job Mobility and the Careers of Young Men," *The Quarterly Journal of Economics*, 439–479.