10ISSN 1183-1057

SIMON FRASER UNIVERSITY

Department of Economics

Working Papers

17-11

"Identifying Distributions in a Panel Model with Heteroskedasticity: An Application to Earning Volatility"

Irene Botosaru

July 2017

SFU DEPARTMENT OF ECONOMICS

Identifying Distributions in a Panel Model with Heteroskedasticity: An Application to Earnings Volatility*

Irene Botosaru†

July 24, 2017

Abstract

This paper considers a panel model with heteroskedasticity, where the parameter of interest is the probability density function of the heteroskedasticity. The nonparametric identification results are established sequentially via a deconvolution argument (in the first step) and solving a linear Fredholm integral equation of the first kind (in the second step). The identification results are constructive and give rise to nonparametric estimators. The model is relevant to the literature on earnings dynamics. Applied to data from the Panel Study of Income Dynamics (PSID), the method developed in this paper reveals a high degree of unobserved heterogeneity in earnings risk. In particular, the evolution over time of the quantiles of the conditional shock variance shows that it is those in the right tail of the distribution who experience the highest volatilities (particularly during recessions), with lower quantiles experiencing relatively constant volatilities during the business cycle. This type of heterogeneity may be relevant to the study of the cyclicality of income risk.

JEL codes: C14, C23, D31

Keywords: Earnings dynamics, panel data, deconvolution, integral equation.

*I would like to thank Bertille Antoine, Manuel Arellano, Richard Blundell, Jean-Pierre Florens, Eric Gautier, Hiro Kasahara, Nour Meddahi, Chris Muris, Stephen Shore, and participants at the Econometrics of Earnings Dynamics and Distributions Workshop (UCL 2012), University of British Columbia, and Touluose School of Economics for questions and suggestions.

†Email: ibotosar@sfu.ca. Address: Department of Economics, Simon Fraser University, Burnaby, BC, Canada.

1 Introduction

This paper considers a panel model with heteroskedasticity, where the parameter of interest is the distribution function of the heteroskedasticity. The model is relevant to the literature on earnings dynamics. Identifying the distribution of the heteroskedasticity of the shocks at a particular time period allows the study of the evolution (over the business cycle) of the quantiles of the conditional variance of both transitory and permanent shocks to earnings. The paper fits within the recent trend in the earnings dynamics literature that is concerned with distributional characteristics of the income process, see e.g. Arellano, Blundell, and Bonhomme (2017).

Applied to data from the Panel Study of Income Dynamics (PSID), the method developed in this paper reveals a high degree of unobserved heterogeneity in earnings risk. In particular, the evolution over time of the quantiles of the conditional shock variances shows that it is those in the right tail of the distribution who experience the highest volatilities (particularly during recessions), with lower quantiles experiencing relatively constant volatilities during the business cycle. This heterogeneity is neglected when focus is on the average shock volatilities, which is the current practice in the earnings dynamics literature. The type of heteroskedasticity found may be relevant to the study of the cyclicality of income risk.

The model considered in this paper is as follows. For individuals $j = 1, ..., n$, and time periods $t = 1, ..., T$, the observed outcome, Y_{jt}, is written as the sum of two latent random variables:

$$Y_{jt} = \tau_{jt} + p_{jt} \tag{1a}$$

$$p_{jt} = p_{jt-1} + \pi_{jt} \tag{1b}$$

where τ_{jt} and π_{jt} are independent shocks, and p_{jt} is modeled as a unit root. Letting Ω_{t^-} be the sigma field generated by $\{\tau_{jt}, \pi_{jt}\}$ for all j up to time t, the special feature considered in this paper is the following specification of the conditional variances of τ_{jt} and π_{jt} :

$$var\left(\tau_{jt}|\Omega_{t^-}\right) = z_{jt}^2,\ z_{jt} \overset{i.n.i.d.}{\sim} F_{z_t} \tag{2}$$

$$var\left(\pi_{jt}|\Omega_{t^-}\right) = s_{jt}^2,\ s_{jt} \overset{i.n.i.d.}{\sim} F_{s_t} \tag{3}$$

where F_{z_t} and F_{s_t} are cumulative distribution functions (CDF) that vary deterministically over time. All random variables are assumed to be absolutely continuous with respect to the Lebesgue measure. The parameters of interest are the probability density functions (pdf) f_{z_t} and f_{s_t}.

Without loss of generality, τ_{jt} and π_{jt} can be decomposed as:

$$\tau_{jt} = z_{jt}\eta_{jt},\ \eta_{jt} \overset{i.i.d.}{\sim} F_\eta,\ E\left(\eta_{jt}|z_{jt}\right) = 0,\ E\left(\eta_{jt}^2|z_{jt}\right) = 1 \tag{4a}$$

$$\pi_{jt} = s_{jt}\varepsilon_{jt},\ \varepsilon_{jt} \overset{i.i.d.}{\sim} F_\varepsilon,\ E\left(\varepsilon_{jt}|s_{jt}\right) = 0,\ E\left(\varepsilon_{jt}^2|s_{jt}\right) = 1 \tag{4b}$$

where η_{jt} and ε_{jt} are i.i.d. shocks drawn from F_η and F_ε, respectively.

The contribution of this paper is the nonparametric identification and estimation of the probability density functions $(f_{z_t}, f_{s_t}, f_\eta, f_\varepsilon)$. The identification results are established via a sequential identification strategy. In the first step, (f_η, f_ε) are identified via a deconvolution argument, while in the second step (f_{z_t}, f_{s_t}) are identified by solving a linear Fredholm integral equation of the first kind. Identification of these four density functions leads to identification of the distributions of τ_{jt} and π_{jt}. For example, the CDF of π_{jt} at time t, call it F_{π_t}, is given by:

$$F_{\pi_t}(c) = P(\pi_{jt} \leq c) = \int_0^\infty F_\varepsilon\left(\frac{c}{s}\right) f_{s_t}(s)\, ds.$$

In a similar set-up, Bonhomme and Robin (2010) identify F_{π_t} (and the CDF of τ_{jt} at time t) under the assumption that the conditional variances $\left\{z_{jt}^2, s_{jt}^2\right\}_{j=1,\ t=1}^{n,T} = \left\{z_t^2, s_t^2\right\}_{t=1}^T$ are known at each t.

The model is motivated by the literature on earnings dynamics. In this literature, Y_{jt} represents (residual) log-income of individual j at time t, which is decomposed into a transitory component, τ_{jt}, and a permanent component, p_{jt}. When τ_{jt} is assumed to be i.n.i.d., it is referred to as transitory *shock*. In the canonical model for the income process, p_{jt} is modeled as a unit root, with $p_{j0} = 0$, and π_{jt} represents the permanent shock. In this paper, the conditional variances z_{jt}^2 and s_{jt}^2 vary in a flexible way across both individuals and time, and are unpredictable with respect to Ω_{t^-}. The main object of interest is the pdf of the conditional variances of each shock, i.e. f_{z_t} and f_{s_t}.

From an empirical point of view, identifying the distribution functions of the shocks τ_{jt} and π_{jt} allows for the identification of different features of the distribution of income risk, such as skewness and kurtosis, symmetry and multimodality – all of which have recently been identified as important features of the earnings risk distribution, see e.g. Guvenen et. al. (2016), Arellano, Blundell, and Bonhomme (2017). These results have been derived under the assumption that the conditional variance of τ_{jt} and π_{jt} at time t is a degenerate random variable. The focus then has been on the average of the conditional variance at time t with the main finding being that the averages do not vary over time and, in particular, that they do not vary over the business cycle.

In this paper, I show that although the averages of the volatilites are fairly constant over the business cycles, the quantiles of the volatilities are not. In fact, when applied to data from the PSID, higher quantiles of the volatilities move substantially with the business cycle. The results of this paper then show that focusing on average volatilies ignores the high degree of heterogeneity that exists in the data. Additionally, the results of this paper may be used to test whether the volatilities are degenerate random variables (as assumed in most analyses of earnings dynamics) or whether specific restrictions on the volatilities hold across heterogeneous populations.[1]

[1] For similar ideas, see the motivation behind studying the identification of the joint distribution of random coefficient models, e.g. Hoderlein, Klemelä, and Mammen (2010), Breunig and Hoderlein (2016).

The identification strategy is sequential. First, (f_η, f_ε) are identified via a deconvolution argument for panel data. Second, (f_{z_t}, f_{s_t}) are identified by solving two resulting linear Fredholm integral equations of the first kind. The identification strategy is constructive, leading to consistent estimators for (f_η, f_ε) (as in Bonhomme and Robin (2010)), and for (f_{z_t}, f_{s_t}) as solutions to inverse problems solved by Tikhonov regularization via the method of Darolles, Fan, Florens, and Renault (2011).

Literature review. Many papers in the earnings dynamics literature have stressed the importance of allowing the shock variances to vary over individuals and time, see e.g. Meghir and Windmeijer (1999), Meghir and Pistaferri (2004), Hospido (2012), Jensen and Shore (2011), Jensen and Shore (2015). Most work specifies the variances as ARCH processes and focuses on estimating the cross-sectional average of the variances. For example, Meghir and Pistaferri (2004) specifies the conditional volatility of each shock as an ARCH process, where individual and time effects enter additively. The parameter of interest is the cross-sectional average of the conditional variance of each shock. Estimation of the ARCH process parameters relies critically on the linearity assumption, and the results are biased when the number of periods is small.

More recent work models the shock variances as time-varying functions that depend on lagged values of the permanent component in order to allow for a particular type of heteroskedasticity, see Arellano, Blundell, and Bonhomme (2017) and Botosaru and Sasaki (2017). None of these papers identifies the distributions of the conditional variances.

To the best of my knowledge, Jensen and Shore (2011) is the only other paper that estimates the distributions of the conditional variances. The authors allow the conditional variance of each shock to vary over individuals and time in a flexible way, and they assume that each volatility sequence is drawn from a Dirichlet process prior. The resulting (discrete) posterior distribution is estimated by Bayesian methods. The normality of the shocks τ and π is one of the many restrictions the authors impose,[2] assumption which is considered problematic in the earnings dynamics literature, cf. Bonhomme and Robin (2010), Guvenen et. al. (2016). Using data from the PSID, Jensen and Shore (2015) show that there is considerable latent heterogeneity in the distribution of the conditional variance. As opposed to Jensen and Shore (2011) and Jensen and Shore (2015), this paper presents a fully nonparametric analysis of the distribution of conditional variances. The empirical results parallel those of Jensen and Shore (2015).

Deconvolution techniques have been used in econometrics to deal with unobserved heterogeneity and measurement error, see e.g. Li and Vuong (1998), Li (2002), Schennach (2004a), Schennach (2004b), Bonhomme and Robin (2010), Evdokimov (2010), Botosaru and Sasaki (2017). Integral equations have been analyzed in many areas of econometrics, particularly in the literatures on instrumental regression estimation and random coefficients, see e.g. Carrasco,

[2] Other restrictions include: priors on the probability that an individual volatility value will change, on the number of unique values that an individual will experience during his or her lifetime, and on the number of unique values in the sample.

Florens, and Renault (2007), Darolles, Fan, Florens, and Renault (2011), Carrasco and Florens (2011), and Hoderlein, Nesheim, and Simoni (2016).

Organization. The outline of the paper is as follows: Section 2 presents the identification results, Section 3 provides the large sample theory, Section 4 presents the small sample properties of the estimators, Section 5 applies the method proposed to data from the PSID, and Section 6 concludes. All proofs are in the Appendix.

Notation. Let $L^p(\mathbb{R})$ be the space of real-valued functions that are $p-$integrable with respect to the Lebesgue measure, and endow the space with the standard L_p norm, i.e. $||g||_p = \left(\int_{\mathbb{R}} |g(\xi)|^p d\xi\right)^{1/p}$. Let $\lambda : \mathbb{R} \to \mathbb{R}_+$ be a non-negative weight function such that $\int_{\mathbb{R}} \lambda(\xi) d\xi < \infty$. Denote by $L^p_\lambda(\mathbb{R})$ the weighted space of real-valued functions that are p-integrable with respect to λ, i.e.

$$L^p_\lambda(\mathbb{R}) = \left\{ g : \mathbb{R} \to \mathbb{R} : ||g||_{\lambda,p} = \left(\int_{\mathbb{R}} |g(\xi)|^p \lambda(\xi) d\xi\right)^{1/p} < \infty \right\}$$

Each space, $(L^2(\mathbb{R}), ||.||_2)$ and $\left(L^2_\lambda(\mathbb{R}), ||.||_{\lambda,2}\right)$, is a Hilbert space.

Let W be a random variable with support $\mathcal{W}$ and with density function $f_W \in L^p(\mathbb{R})$, $1 \leq p \leq 2$. Letting $i = \sqrt{-1}$, define the characteristic function of W as:

$$\phi_W(\xi) = E\left(e^{i\xi W}\right) = \int_{\mathcal{W}} e^{i\xi w} f_W(w) dw, \ \xi \in \mathbb{R}$$

For every $\kappa \in \mathbb{N}$, define the κ^{th} moment of f_W as $m^{(\kappa)}_{f_W} \equiv \frac{d^\kappa}{d\xi^\kappa}\phi_W(\xi)\Big|_{\xi=0}$.

In what follows, I supress dependence on the j subscript.

2 Identification

The identification strategy is composed of two main steps. In the first step, I apply Kotlarski's lemma to deconvolve the densities of η, ε, $\tau_t = z_t\eta$, and $\pi_t = s_t\varepsilon$. In the second step, I show the existence and uniqueness of the densities of s_t and z_t by showing the injectivity of the integral operators associated to the resulting linear Fredholm integral equations for the densities of s_t and z_t. For identification of $(f_\eta,\ f_\varepsilon, f_{s_2}, f_{z_2})$ three time periods are sufficient. More generally, whenever T time periods are available, it is possible to identify $T-2$ volatility densities due to the nonstationarity of the volatility distributions.

Since three time periods are sufficient for identification, I let $T = 3$ in what follows. Define: $s \equiv (s_1, s_2, s_3) \in \mathbb{R}_+$, $z \equiv (z_1, z_2, z_3) \in \mathbb{R}_+$, $\eta \equiv (\eta_1, \eta_2, \eta_3) \in \mathbb{R}$, and $\varepsilon \equiv (\varepsilon_1, \varepsilon_2, \varepsilon_3) \in \mathbb{R}$.

Assumption 1K (Kotlarski)

(i) $s,\ z,\ \eta$, and ε are mutually independent; (ii) $E(\eta) = 0,\ E(\varepsilon) = 0$; (iii) $|\phi_\varepsilon(\xi)| > 0$ and $|\phi_\eta(\xi)| > 0$ for all $\xi \in \mathbb{R}$; (iv) $f_\varepsilon,\ f_\eta \in L^2(\mathbb{R})$.

Assumption 1K is typically made in the deconvolution literature that uses Kotlarski's lemma. The independence assumption in 1K(i) is usually made in panel models and factor models, although it is possible to weaken it, see e.g. Cuhna, Heckman, and Schennach (2010) and Botosaru and Sasaki (2017). Assumption 1K(ii) pins down the means of η and ε. Without this assumption, identification would be obtained up to a location shift. Assumption 1K(iii) is usually made in the deconvolution literature and it excludes e.g. distributions defined on bounded supports, see Hu and Ridder (2010). It is possible to weaken 1K(iii) to require that the zeros of the characteristic functions be isolated, see e.g. Carrasco and Florens (2011) and Evdokimov and White (2012). The integrability assumption in 1K(iv) is stronger than necessary for an application of Kotlarski's lemma, which requires only existence of ϕ_ε and ϕ_η. However, this assumption is needed for the identification of f_{z_t} and f_{s_t} for which square integrability of ϕ_ε and ϕ_η is required.

Assumption 1I (initialization)

$s_1 = 1 = z_1$.

Assumption 1I is restrictive, but it can be relaxed to $s_1 = c_s$ and $z_1 = c_z$, where c_s, c_z are known positive constants. It is possible to dispense with this assumption at the expense of obtaining identification of the moments (rather than the densities) of the conditional volatilities. I discuss the trade-offs in identification assumptions in Appendix A.2.

Assumption 1F (Fredholm)

(i) $f_{s_t},\ f_{z_t} \in L^2(\mathbb{R}_+)$ for each t; (ii) f_ε and f_η are strictly positive definite functions on $\mathbb{R}$; (iii) λ is such that $\int_{\mathbb{R}} \frac{1}{\xi}\lambda(\xi) < \infty$.

Assumptions 1F(i) and 1F(iii) guarantee the boundedness of the integral operators defined in the statement of Theorem 1 below. The positive definiteness assumption in 1F(ii) is sufficient for the characteristic functions ϕ_ε and ϕ_η to be positive everywhere. This assumption guarantees that the kernels of the resulting integral operators that have to be solved for f_{z_t} and f_{s_t} are strictly positive, resulting in the injectivity of the operators. Strictly positive definite functions are symmetric and attain their maximum value at zero. The standard normal density function is strictly positive definite, for example.

Theorem 1 *Let the distribution of $Y_t,\ t = 1, 2, 3$ be observed and assume that Y_t follows the model described by* (1a), (1b), (4a), *and* (4b). *Let assumptions 1I and 1K hold. Then the density* *functions f_η and f_ε are identified and given by:*

$$f_\eta(\xi) = \frac{1}{2\pi}\int_{\mathbb{R}} \exp(-i\xi w)\,\phi_\eta(w)\,dw,\ \xi\in\mathbb{R}, \tag{5}$$

$$f_\varepsilon(\xi) = \frac{1}{2\pi}\int_{\mathbb{R}} \exp(-i\xi w)\,\phi_\varepsilon(w)\,dw,\ \xi\in\mathbb{R}. \tag{6}$$

Further, letting $U,\widetilde{U} : L^2(\mathbb{R}_+) \to L^2_\lambda(\mathbb{R})$ be linear integral operators and letting assumption 1F hold, the density functions f_{s_2} and f_{z_2} are identified and given as the unique solutions to the following integral equations:

$$(Uf_{s_2})(\xi) = \int_{\mathbb{R}_+} \phi_\varepsilon(\xi w)\,f_{s_2}(w)\,dw,\ \xi\in\mathbb{R}, \tag{7}$$

$$\left(\widetilde{U}f_{z_2}\right)(\xi) = \int_{\mathbb{R}_+} \phi_\eta(\xi w)\,f_{z_2}(w)\,dw,\ \xi\in\mathbb{R}. \tag{8}$$

Proof. The proof is divided in two parts. In the first part, I apply Kotlarski's lemma twice. Kotlarski's lemma applied to the first two periods obtains identification of ϕ_ε and ϕ_η. By the inverse Fourier transform, this identifies f_ε and f_η. Kotlarski's lemma applied to the second and third time periods obtains identification of ϕ_{π_2} and ϕ_{τ_2}. Applying the law of iterated expectations to these latter two characteristic functions obtains integral equations (7) and (8), where f_{s_2} and f_{z_2} are the unknowns. Each equation can be solved uniquely by showing that the associated integral operator is injective with a dense range. The proof is in Appendix A.1. ■

3 Estimation

Following the identification strategy, I first estimate f_η, f_ε, ϕ_{π_2}, and ϕ_{τ_2}. The estimators of the characteristic functions become the kernels of two linear Fredholm equations for f_{s_2} and f_{z_2}. The integral operators are solved by Tikhonov regularization. The large sample theory for the deconvolution step follows closely Bonhomme and Robin (2010), where I compute upper bounds for the uniform rate of convergence of the estimators for f_η and f_ε, while the second step follows Darolles, Fan, Florens, and Renault (2011), where I compute L^2 rates of convergence for the estimators of f_{s_2} and f_{z_2}.

3.1 Estimators for the Characteristic functions

For $\xi\in\mathbb{R}$, define

$$\widehat{\delta}_t(\xi) \equiv \int_0^\xi \frac{\sum_{j=1}^n iY_{j,t}\exp(iwY_{j,t+1})}{\sum_{j=1}^n \exp(iwY_{j,t+1})}dw$$

The sample analogues of the characteristic functions of η, ε, π_2, and τ_2 are given by, respectively:

$$\widehat{\phi}_{\varepsilon}(\xi) = \frac{1}{T-1}\sum_{t=1}^{T-1}\exp\left(\widehat{\delta}_t(\xi)\right) \tag{9}$$

$$\widehat{\phi}_{\eta}(\xi) = \frac{1}{\widehat{\phi}_{\varepsilon}(\xi)}\left[\frac{1}{n(T-1)}\sum_{t=1}^{T-1}\sum_{j=1}^{n}\exp\left(i\xi Y_{jt}\right)\right] \tag{10}$$

$$\widehat{\phi}_{\pi_2}(\xi) = \frac{1}{\widehat{\phi}_{\varepsilon}(\xi)}\left[\frac{1}{T-1}\sum_{t=1}^{T-1}\exp\left(\widehat{\delta}_{t+1}(\xi)\right)\right] \tag{11}$$

$$\widehat{\phi}_{\tau_2}(\xi) = \frac{1}{\sum_{t=1}^{T-1}\exp\left(\widehat{\delta}_{t+1}(\xi)\right)}\left[\frac{1}{n}\sum_{j=1}^{n}\sum_{t=1}^{T-1}\exp\left(i\xi Y_{jt+2}\right)\right] \tag{12}$$

3.2 Estimators of Density Functions

The density functions of η and ε can be obtained by inverting their corresponding Fourier transforms, as in (5) and (6). However replacing directly ϕ_η and ϕ_ε by their respective sample analogues results in estimators for the density functions that are not well-defined since the empirical characteristic functions are neither integrable nor square integrable, see e.g.Meister (2009). Numerically inverting Fourier transforms is an ill-posed inverse problem, see e.g. Horowitz (1998), so that a regularization method is needed for obtaining the estimators $\left(\widehat{f}_\varepsilon, \widehat{f}_\eta\right)$.

The most studied regularization method in the deconvolution literature is the deconvolution kernel density estimator proposed by Stefanski and Carroll (1990). The regularization method consists in using a kernel function, $K \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, whose characteristic function, $\phi_K \in [-1, 1]$, is compactly supported. This ensures that the problem is well-posed, and thus, that the estimator is well-defined with probability equal to one. Using the method proposed by Stefanski and Carroll (1990), the densities of ε and η can be estimated as:

$$\widehat{f}_{\varepsilon}(\xi) = \frac{1}{2\pi}\int_{-h_n}^{h_n} e^{-i\xi w}\phi_K\left(\frac{w}{h_n}\right)\widehat{\phi}_{\varepsilon}(w)\,dw \tag{13}$$

$$\widehat{f}_{\eta}(\xi) = \frac{1}{2\pi}\int_{-h_n}^{h_n} e^{-i\xi\tau}\phi_K\left(\frac{w}{h_n}\right)\widehat{\phi}_{\eta}(w)\,dw \tag{14}$$

Following Delaigle, Hall, and Meister (2008) and Bonhomme and Robin (2010), the characteristic function ϕ_K is given by:

$$\phi_K(\xi) = \left(1-\xi^2\right)^3 I\left(\xi \in [-1, 1]\right)$$

which corresponds to the second order kernel[3]

$$K(\xi) = \frac{48\cos(\xi)}{\pi\xi^4}\left(1 - \frac{15}{\xi^2}\right) - \frac{144\sin(\xi)}{\pi\xi^5}\left(2 - \frac{5}{\xi^2}\right)$$

Notice that $\widehat{\phi}_\eta(\xi)$ is obtained by division by $\widehat{\phi}_\varepsilon(\xi)$, running the risk of dividing by zero. I follow the suggestion of Delaigle, Hall, and Meister (2008) to replace the empirical characteristic function in the denominator by the following: Let $\widehat{\phi}_X$ be the empirical characteristic function of the density function f_X. Instead of dividing by $\widehat{\phi}_X$, they propose dividing by

$$\widetilde{\phi}_X(\xi) = \widehat{\phi}_X(\xi)\, I(\xi \in A) + \widehat{\phi}_P(\xi)\, I(\xi \notin A)$$

where A denotes the largest interval around zero where $\widehat{\phi}_X(\xi)$ does not oscillate and $\widehat{\phi}_P(\xi) = \left(1 + \alpha\xi^2\right)^{-\beta}$, where α and β are the second and the fourth moments of X. When the empirical moments are negative, I set $\alpha = \frac{1}{2}var(X)$ and $\beta = 1$. Likewise, following Bonhomme and Robin (2010), I set $\widehat{\phi}_X(\xi) = 0$ whenever $\left|\widehat{\phi}_X(\xi)\right| > 1$ in order to guarantee that $\widehat{\phi}_X(\xi)$ is a proper characteristic function.

The remaining density functions f_{s_2} and f_{z_2} are defined as solutions to linear Fredholm integral equations of the first kind. First, set $\Omega_s = \Omega_z = [0,1]$ and $f_{s_2} \in L^2(\Omega_s)$, $f_{z_2} \in L^2(\Omega_z)$. At the population level, the integral equations have unique solutions since the integral operators associated to them defined as, respectively:

$$\begin{aligned}(U_1 f_{s_2})(\xi) &= \int_0^1 \phi_\varepsilon(w\xi)\, f_{s_2}(w)\, dw,\ \xi \in \mathbb{R} \\ (U_2 f_{z_2})(\xi) &= \int_0^1 \phi_\eta(w\xi)\, f_{z_2}(w)\, dw,\ \xi \in \mathbb{R}\end{aligned} \tag{15}$$

are injective, see the proof of Theorem 1. At the sample level, the resulting equations may be ill-posed since the operators are usually estimated by discretization methods where one determines finitely many unknowns, so that the estimated operators are of finite rank.

The estimated operator $\widehat{U}_{1,n} : L^2([0,1]) \to L^2([-h_n, h_n])$ and its adjoint $\widehat{U}^*_{1,n} : L^2([-h_n, h_n]) \to L^2([0,1])$ are defined as, respectively:

$$\left(\widehat{U}_{1,n} f_{s_2}\right)(\xi) = \int_0^1 \widehat{\phi}_\varepsilon(w\xi)\, f_{s_2}(w)\, dw,\ \xi \in [-h_n, h_n] \tag{16}$$

$$\left(\widehat{U}^*_{1,n} f_{s_2}\right)(w) = \int_{-h_n}^{h_n} \widehat{\phi}_\varepsilon(w\xi)\, f_{s_2}(\xi)\, d\xi,\ w \in [0,1] \tag{17}$$

with $\widehat{U}_{2,n}$ and $\widehat{U}^*_{2,n}$ defined similarly. Thus it may happen that $\widehat{\phi}_{\pi_2}(\xi) \notin R\left(\widehat{U}_{1,n}\right)$ or $\widehat{\phi}_{\tau_2}(\xi) \notin$

[3] This kernel function has good performance in simulation studies as shown by Fan (1992) and Delaigle and Hall (2006).

$R\left(\widehat{U}_{2,n}\right)$ or that $\widehat{U}_{1,n}$ and $\widehat{U}_{2,n}$ may not be invertible, see e.g. Carrasco, Florens, and Renault (2007) and Kim (2004). As such, for consistent estimation, the sample counterparts of (7) and (8) have to be regularized. Then the estimators $\left(\widehat{f}_{s_2}, \widehat{f}_{z_2}\right)$ are defined as the stable solutions of regularized Fredholm integral equations of the first kind, where the chosen regularization scheme is Tikhonov regularization with regularization parameter α_n. That is, the estimators are defined as:

$$\widehat{f}_{s_2}(\xi) = \left(\alpha_n I + \widehat{U}_{1,n}^* \widehat{U}_{1,n}\right)^{-1} \widehat{U}_{1,n}^* \widehat{\phi}_{\pi_2}(\xi) \quad (18)$$

$$\widehat{f}_{z_2}(\xi) = \left(\alpha_n I + \widehat{U}_{2,n}^* \widehat{U}_{2,n}\right)^{-1} \widehat{U}_{2,n}^* \widehat{\phi}_{\pi_2}(\xi) \quad (19)$$

where $\xi \in [-h_n, h_n]$, $h_n \to \infty$, and $\alpha_n \to 0$ as $n \to \infty$.

4 Asymptotic Theory

The estimators of the two density functions, $\widehat{f}_\varepsilon$ and $\widehat{f}_\eta$, are shown to be uniformly consistent in Theorem 2, with upper bounds on the uniform rate of convergence derived, while in Theorem 3, I derive L^2 rates of convergence for the estimators $\widehat{f}_{s_2}$ and $\widehat{f}_{z_2}$ over the bounded sets, Ω_s and Ω_z, respectively.

Assumption 2

(i) Let $g_l : \mathbb{R}_+ \to [0,1]$, $l = \varepsilon, \eta$, be an integrable function such that $|\phi_l(\xi)| \leq g_l(|\xi|)$, for all $|\xi|$; (ii) Let $g_y : \mathbb{R}_+ \to [0,1]$ be an integrable, decreasing function, and let c be a constant such that for $|w| > c$, $|\phi_Y(w)| \geq g_y(|w|)$ with $\lim_{|w|\to\infty} g_y(|w|) = 0$; (iii) $E(Y) < \infty$; (iv) The moment generating functions of Y_t^2 and $|Y_t Y_{t+1}|$ exist in a neighborhood around zero.

Assumption 2(i) relates to the usual assumption made on the rate of decay of characteristic functions in the deconvolution literature, see the remark below for more details. Assumption 2(ii) controls the smoothness of the joint distribution function of the observable variables, ensuring that the tails of the characteristic function of Y do not approach zero too quickly. Similar assumptions appear in the literature on deconvoluting the density function of an unobservable variable from that of an error. In that literature, it is usually assumed that the distribution of the error or the signal is known and assumptions similar to 2(ii) on the characteristic function of the error are made, for example see Hu and Ridder (2010). Assumption 2(ii) is the analogue of this type of assumption when the error distribution is not known but when repeated measurements are observed. Assumption 2(iii) imposes that the mean of the dependent variable be finite. Assumption 2(iv) implies that the moments of the random variables Y_t^2 and $|Y_t Y_{t-1}|$ exist. This assumption may be relaxed at the expense of slower rates of convergence, see e.g. Bonhomme and Robin (2010).

Theorem 2 *Let the data be i.i.d. and suppose that the assumptions of Theorem 1 and assumption 2 hold. Let K be a kernel function of even order $q \geq 2$ with its Fourier transform ϕ_K satisfying $\phi_K(s) = 0$ for $|s| > 1$. If $\varepsilon_n = \frac{\ln n}{\sqrt{n}}$ and $h_n = Cn^{\delta/2}$ for some $C, \delta > 0$, then there exist constants C_1 and C_2 such that*

$$\sup_z \left|\widehat{f}_\varepsilon(z) - f_\varepsilon(z)\right| \leq C_1 \frac{h_n^2}{g_y^2(h_n)} \varepsilon_n + C_2 \frac{1}{h_n^q} \int_{-h_n}^{h_n} \tau^q g_\varepsilon(|\tau|)\, d\tau + 2 \int_{h_n}^{\infty} g_\varepsilon(|\tau|)\, d\tau$$

and

$$\sup_z \left|\widehat{f}_\eta(z) - f_\eta(z)\right| \leq C_1 \frac{h_n^2}{g_y^2(h_n)} \varepsilon_n + C_2 \frac{1}{h_n^q} \int_{-h_n}^{h_n} \tau^q g_\eta(|\tau|)\, d\tau + 2 \int_{h_n}^{\infty} g_\eta(|\tau|)\, d\tau$$

Proof. The proof uses Lemma 1 of Bonhomme and Robin (2010) and it is similar to the proofs of Theorems 1 and 2 in Bonhomme and Robin (2010). See Appendix A.3. ■

Let $\{g_i\}_{i\geq 0}$ be an orthonormal sequence of $L^2([0,1])$ and let $\{\lambda_i\}_{i\geq 0}$ be a sequence of non-negative real numbers such that $\lambda_0 = 1 \geq \lambda_1 \geq ....$ The two sequences enter the singular value decomposition of the operators in (15), where $\{\lambda_i\}$ are the eigenvalues of the integral operators. Additionally, define the space

$$\Phi_\beta = \left\{ g \in L^2([0,1]) : \sum_{i\geq 0} \frac{\langle g, g_i\rangle^2}{\lambda_i^{2\beta}} < \infty \right\}$$

Letting Q be the following operator

$$(Qg)(\xi) = \int_0^1 \phi(\xi w)\, g(w)\, dw, \ \xi \in \mathbb{R}$$

and Q^* its adjoint, consider Proposition 3.2 in Darolles, Fan, Florens, and Renault (2011) included below for convenience:

Proposition 1 *If $g \in \Phi_\beta$ for some $\beta > 0$ and $g_\alpha = (\alpha I + Q^* Q)^{-1} Q^* Q g$, then $||g - g_\alpha||_{L^2}^2 = O\left(\alpha^{\beta \wedge 2}\right)$ when $\alpha \to 0$.*

Consider the operators defined in (16) and (17) and define $\widehat{A}_{\alpha,n} = \left(\alpha_n I + \widehat{U}_n^* \widehat{U}_n\right)^{-1}$. To derive rates of convergence for the two remaining estimators, I assume bounded support of the volatility parameters s and z, i.e. $\Omega_s = \Omega_z = [0,1]$.

Assumption 3

(i) Let $f_s,\ f_z \in L^2([0,1])$ be bounded with $\sup_s f(s) = M_1$ and $\sup_z f(z) = M_2$; (ii) Let $f_s, f_z \in \Phi_\beta$; (iii) Let $\phi_\varepsilon(s\xi)$ and $\phi_\eta(z\xi)$ be continuously differentiable in the interior of $[0,1] \times \mathbb{R}$; (iv) Let h_n be such that $h_n \to \infty$, $\frac{h_n \varepsilon_n}{g_y(h_n)} = o(1)$, $\frac{h_n^2 \varepsilon_n}{g_y^2(h_n)} = o(1)$, and $\frac{h_n^3 \varepsilon_n^2}{g_y^4(h_n)} = o(1)$; (v) Let α_n be such that $\alpha_n \to 0$, $\frac{h_n^3 \varepsilon_n^2}{\alpha_n g_y^4(h_n)} \to 0$, and either $\beta \geq 1$ or $\alpha_n^{1-\beta} g_y^4(h_n) \to \infty$.

Assumption 3(i) is usually made in the nonparametric instrumental variable literature that uses Tikhonov regularization in order to derive L^2 rates of convergence. Assumption 3(ii) is known in the inverse problem literature as a source condition, see Engl, Hanke, and Neubauer (2000). It is needed in order to control the regularization bias of the Tikhonov regularized parameter. I follow Darolles, Fan, Florens, and Renault (2011) in making assumption 3(iii), which allows me to use Proposition 1.

Theorem 3 *In addition to the assumptions required for Theorem 2, let Assumption 3 hold. Then:*

$$\begin{aligned}\left\|\widehat{f}_{s_2}(z) - f_{s_2}(z)\right\|_{L^2}^2 &= O_p\left(h_n^3 \frac{\varepsilon_n^2}{g_y^4(h_n)}\left(\alpha_n^{-1} + \alpha_n^{\beta-1\wedge 1} + \alpha_n^{\beta-1\wedge 0}\right) + \alpha_n^{\beta\wedge 2}\right) \\ \left\|\widehat{f}_{z_2}(z) - f_{z_2}(z)\right\|_{L^2}^2 &= O_p\left(h_n^3 \frac{\varepsilon_n^2}{g_y^4(h_n)}\left(\alpha_n^{-1} + \alpha_n^{\beta-1\wedge 1} + \alpha_n^{\beta-1\wedge 0}\right) + \alpha_n^{\beta\wedge 2}\right)\end{aligned}$$

Proof. The proof uses Proposition 3.2 in Darolles, Fan, Florens, and Renault (2011) and is similar to the proof of Theorem 4.1 in the same paper. See Appendix A.4. ■

5 Monte Carlo Simulations

In this section, I present results from a Monte Carlo study. The DGP is as follows:

$$\begin{aligned}Y_{j1} &= \varepsilon_{j1} + \eta_{j1} \\ Y_{j2} &= \varepsilon_{j1} + s_{j2}\varepsilon_{j2} + z_{j2}\eta_{j2} \\ Y_{j3} &= \varepsilon_{j1} + s_{j2}\varepsilon_{j2} + s_{j3}\varepsilon_{j3} + z_{j3}\eta_{j3}\end{aligned}$$

with $\varepsilon_{jt} \sim N(0,1)$, $\eta_{jt} \sim N(0,5)$ for $t = 1,2,3$, and $s_{j2},\ s_{j3} \sim \exp(2)$, $z_{j2},\ z_{j3} \sim \exp(0.5)$, where $\exp(\lambda)$ denotes the exponential distribution with parameter λ.

Using the proposed estimation method, I estimate f_ε, f_η, f_{s_2}, and f_{z_2} using $(Y_{j1}, Y_{j2}, Y_{j3})_{j=1}^{N=1000}$ generated as above. The parameters that I vary in the simulations are the bandwidth parameter h_n and the Tikhonov regulation parameter α_n.

Figures 1 to 4 show simulation results based on 500 Monte Carlo replications. The results are shown for $h_n \in \left\{\frac{1}{4}, \frac{1}{6}\right\}$ and $\alpha_n \in \{0.5, 0.7\}$.[4] In each display, the solid curve represents the true function, the dark dashed curve draws the average estimator across simulations, and the dashed curves draw MC percentiles at $50 \pm 33\%$. The inter-quantile range captures the true function well, with the average following the true function closely.

[4] In the simulations, the bandwidth parameter $h_n \in \left\{\frac{1}{4}, \frac{1}{6}\right\}$ and the regulation parameter $\alpha_n \in \{0.1, ..., 0.9\}$. The estimators are fairly robust across different spefications of h_n and α_n, and across different combinations of the parameters.

Simulation results, $h_n = \frac{1}{6}$ and $\alpha_n = 0.5$

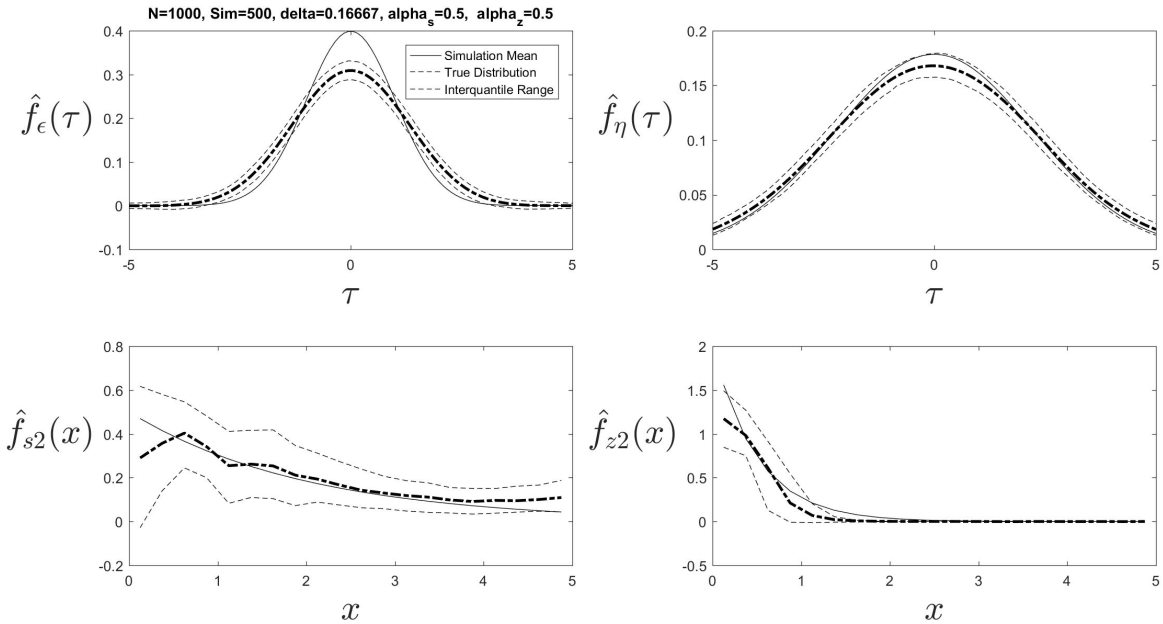

Figure 1: The figure shows the true function as a straight line, the average estimator across simulations as a bold dashed line, and the interquantile range as light dashed lines.

Simulation results, $h_n = \frac{1}{6}$ and α_n = **0.7**

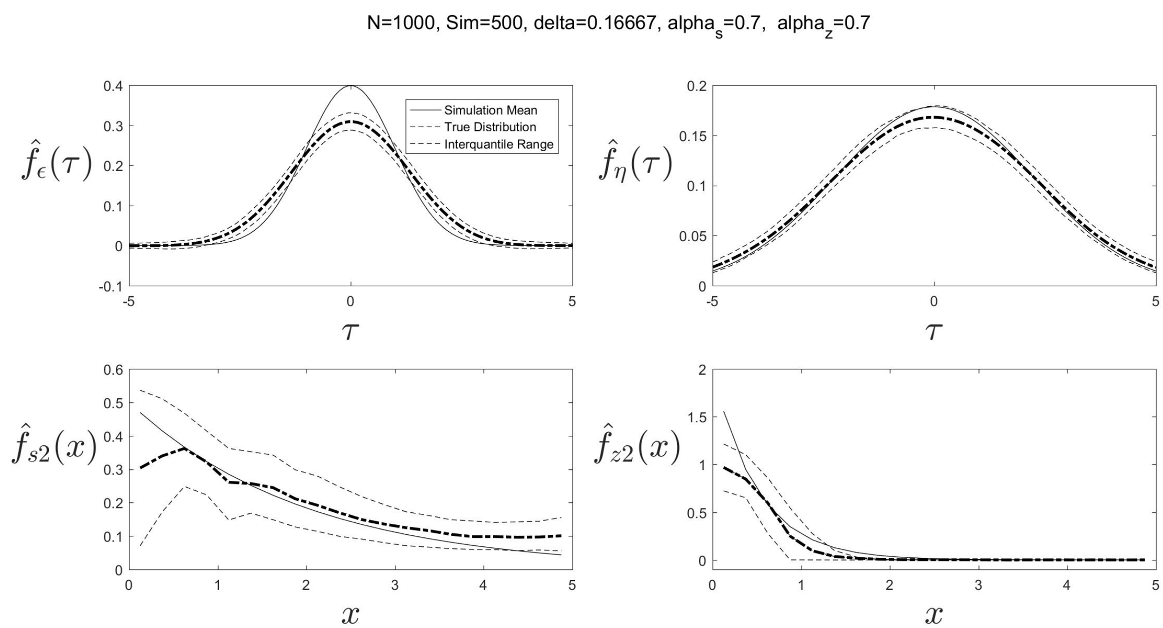

Figure 2: The figure shows the true function as a straight line, the average estimator across simulations as a bold dashed line, and the interquantile range as light dashed lines.

Simulation results, $h_n = \frac{1}{4}$ and $\alpha_n = 0.5$

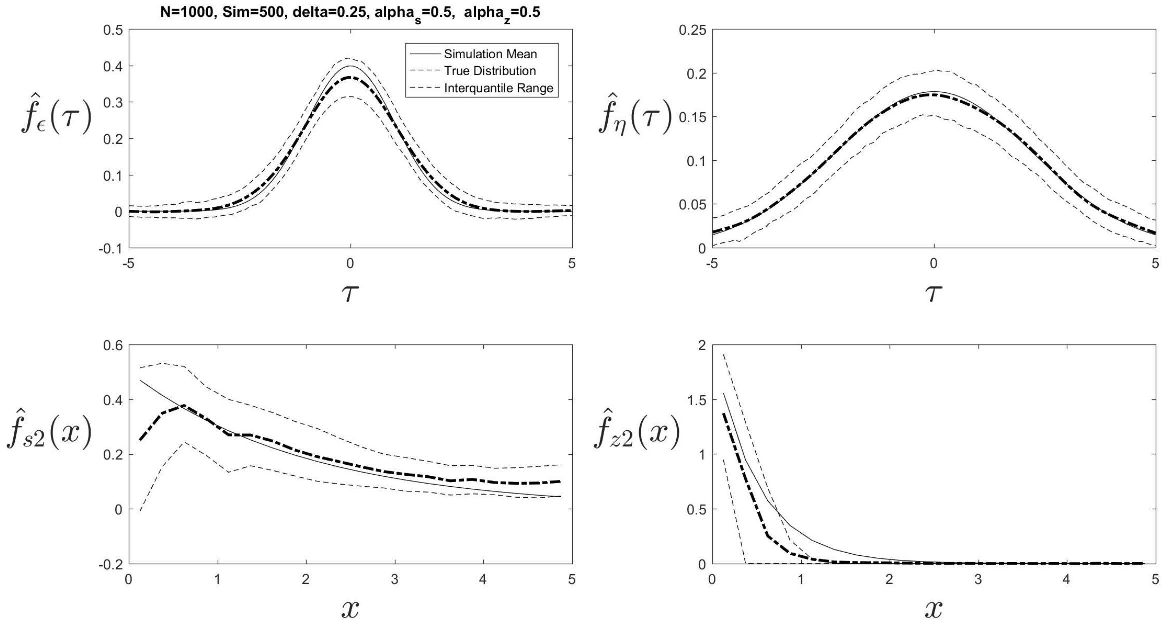

Figure 3: The figure shows the true function as a straight line, the average estimator across simulations as a bold dashed line, and the interquantile range as light dashed lines.

Simulation results, $h_n = \frac{1}{4}$ and $\alpha_n = 0.7$

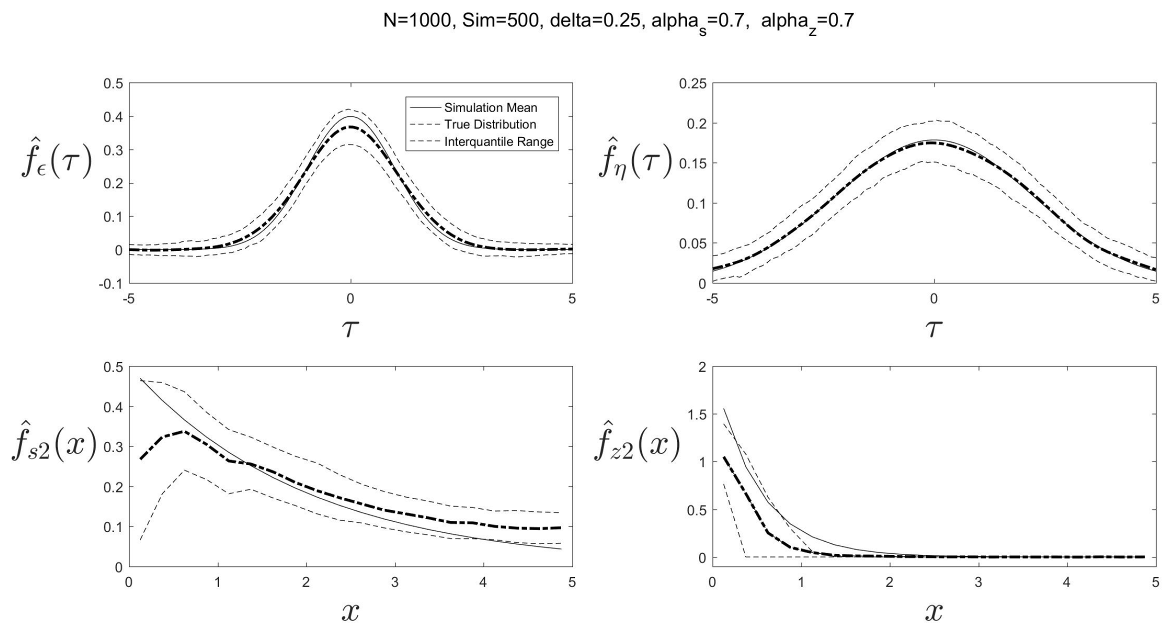

Figure 4: The figure shows the true function as a straight line, the average estimator across simulations as a bold dashed line, and the interquantile range as light dashed lines.

6 Empirical Illustration

In this section, I apply the estimation method to the data used by Bonhomme and Robin (2010). The data is a balanced panel from the PSID for the years $1977 - 1989$, with a cross-sectional sample size of 659. Following Bonhomme and Robin (2010), the measured component of earnings is the OLS residual of log wages on education, age, geographic characteristics, and year dummies.

Using the method proposed, I estimate the densities of the volatilities for each of the eleven years $1978 - 1988$. I use $h_n = \frac{1}{2}$ and $\alpha_n = 0.7$. As in the estimation, different other parameters have been considered but the results obtained were qualitatively the same as the ones presented here. Figure 5 presents the evolution over time for the different quantiles of the volatilities. The results are similar to those in Jensen and Shore (2015). Figure 5 shows that there is considerable heterogeneity in the distribution of the volatilities, with those in the right tail of the distribution experiencing much higher levels of volatility than those below the 75^{th} percentile. It is interesting to notice that the permanent volatilities of those in the 90^{th} percentile have been very volatile during the sample studied, while those of the individuals in the 95^{th} volatile have been relatively constant but much higher than the other percentiles. It is also interesting to notice that during the $1982-1983$ recession, it was the volatilities of those in the 90^{th} and 95^{th} percentiles that increased, while those of the median stayed relatively constant. Hence, those who are in the right tail tend to suffer more from business cycle variations, which shows that the answer to the debate about whether or not volatilies are procyclical may be quite nuanced (Salgado, Guvenen, and Bloom (2017)).

Simulation results, $h_n = \frac{1}{4}$ and $\alpha_n = 0.7$

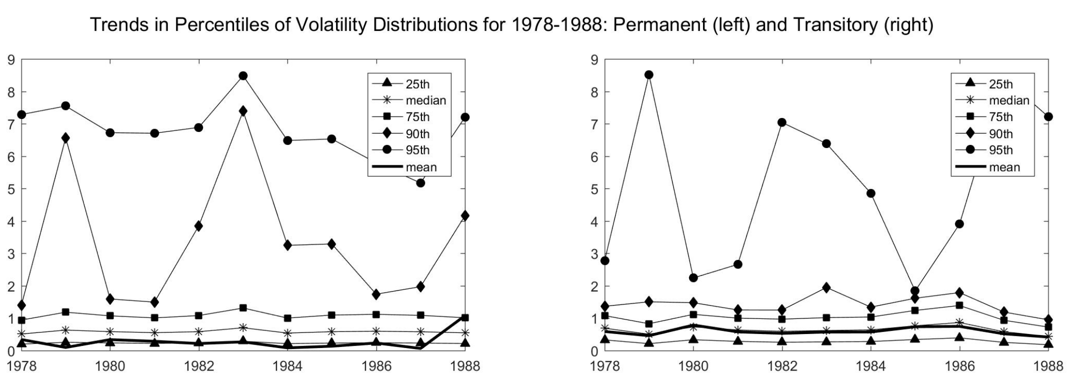

Figure 5: The figure shows the percentiles of the volatilities during 1978-1988.

7 Conclusion

The labor and macroeconomics literature has been dominated by income models in which the way income volatility varies over individuals and time is restricted. In this paper, I relax this assumption by indexing the volatility of income by both time and individual. I decompose the income risk into risk associated to permanent shocks and to transitory shocks. I show the nonparametric identification of the distributions of the conditional variances of the permanent and transitory shocks with a minimum of three time periods. I propose nonparametric estimators, which I then apply to data from the PSID, and I show that there is considerable heterogeneity in the distribution of the volatilities. The results show that it is those in the right tail of the volatility distribution who experienced high levels of volatility, particularly during recessions. Those in percentiles lower than the 75^{th} have experienced relatively constant levels of volatilities, which seem to be relatively constant over the business cycle.

A Appendix

A.1 Proof of Theorem 1

For $L \in \{1, 2\}$, let

$$y_L = M + U_L \tag{20}$$

$$y_{L+1} = M + U_{L+1} \tag{21}$$

Using assumption 1I, for the first two time periods (i.e. $L = 1$), $M = \varepsilon_1,\ U_1 = \eta_1,\ U_2 = \tau_2 + \pi_2$, while for the second and third time periods (i.e. $L = 2$), $M = \varepsilon_1 + \pi_2,\ U_2 = \tau_2,\ U_3 = \tau_3 + \pi_3$.

The proof of Kotlarski's lemma can be found in Rao (1992). I show below how to obtain the characteristic functions of $M,\ U_L$, and U_{L+1} by applying Kotlarski's lemma to (20) and (21) with $L = 1$ and $L = 2$, respectively.

Let $\phi_Y(\xi_1, \xi_2),\ \xi_1, \xi_2 \in \mathbb{R}$ be the characteristic function of (y_L, y_{L+1}). By assumption 1K(i), $M, U_L, U_{L+1},$ are independent for $L = 1, 2$, so that:

$$\phi_Y(\xi_1, \xi_2) = \phi_M(\xi_1 + \xi_2)\, \phi_{U_L}(\xi_1)\, \phi_{U_{L+1}}(\xi_2)$$

Notice that

$$\frac{1}{\phi_Y(0, \xi_2)} \frac{\partial}{\partial \xi_1} \phi_Y(\xi_1, \xi_2)\bigg|_{\xi_1 = 0} = \frac{d}{d\xi_2} \log \phi_M(\xi_2) + \phi'_{U_L}(0)$$

so that, letting $\xi \in \mathbb{R}$:

$$\begin{aligned}
\phi_M(\xi) &= \exp \int_0^{\xi} \left(\frac{\partial}{\partial \xi_1} \log \phi_Y(\xi_1, s) \Big|_{\xi_1 = 0} - \phi'_{U_L}(0) \right) ds & (22) \\
\phi_{U_L}(\xi) &= \frac{\phi_Y(\xi, 0)}{\phi_M(\xi)} & (23) \\
\phi_{U_{L+1}}(\xi) &= \frac{\phi_Y(0, \xi)}{\phi_M(\xi)} & (24)
\end{aligned}$$

By assumptions 1K(i) and 1K(ii), $\phi'_{U_1}(0) = \phi'_{\eta_1}(0) = 0$ and $\phi'_{U_2}(0) = \phi'_{\tau_2}(0) = 0$. Then applying expressions (22) to (24) to the first two time periods obtains:

$$\begin{aligned}
\phi_{\varepsilon}(\xi) &= \exp \left[\int_0^{\xi} \left[\frac{E[iY_1 \exp(isY_2)]}{E[\exp(isY_2)]} \right] ds \right] & (25) \\
\phi_{\eta}(\xi) &= \frac{1}{\phi_{\varepsilon}(\xi)} E[\exp(i\xi Y_1)] & (26) \\
\phi_{\pi_2 + \tau_2}(\xi) &= \frac{1}{\phi_{\varepsilon}(\xi)} E[\exp(i\xi Y_2)] & (27)
\end{aligned}$$

By the inverse Fourier transform and by assumption 1K(iv), expressions (25) and (26) can be solved for the density functions of ε and η, respectively, obtaining expressions (5) and (6) in the main text.

Applying expressions (22), (23) , and (24) to the second and third time periods obtains:

$$\begin{aligned}
\phi_{\varepsilon + \pi_2}(\xi) &= \exp \left[\int_0^{\xi} \left[\frac{E[iY_2 \exp(isY_3)]}{E[\exp(isY_3)]} \right] ds \right] & (28) \\
\phi_{\tau_2}(\xi) &= \frac{1}{\phi_{\varepsilon + \pi_2}(\xi)} E[\exp(i\xi Y_2)] & (29) \\
\phi_{\pi_3 + \tau_3}(\xi) &= \frac{1}{\phi_{\varepsilon + \pi_2}(\xi)} E[\exp(i\xi Y_3)] & (30)
\end{aligned}$$

Consider then (27), (28), and (30). Notice that (27) and (29) are multicolinear. By assumption 1K(i):

$$\begin{aligned}
\phi_{\tau_2 + \pi_2}(\xi) &= \phi_{\tau_2}(\xi) \phi_{\pi_2}(\xi) & (31) \\
\phi_{\varepsilon + \pi_2}(\xi) &= \phi_{\varepsilon}(\xi) \phi_{\pi_2}(\xi) & (32) \\
\phi_{\tau_3 + \pi_3}(\xi) &= \phi_{\tau_3}(\xi) \phi_{\pi_3}(\xi) & (33)
\end{aligned}$$

Then (25) and (32) identify:

$$\phi_{\pi_2}(\xi) = \frac{\phi_{\varepsilon + \pi_2}(\xi)}{\phi_{\varepsilon}(\xi)} \qquad (34)$$

The second step identifies f_{z_2} and f_{s_2} by showing that the integral operators associated with

(29) and (34) are injective.

Consider first (34). By the law of iterated expectations, the characteristic function $\phi_{\pi_2}(\xi)$ can be written as:

$$\phi_{\pi_2}(\xi) = E\left[E\left(\exp\left(i\xi S_2 \varepsilon\right)|S_2 = s\right)\right] = \int_{\mathbb{R}_+} \phi_\varepsilon(\xi s) f_{s_2}(s)\, ds \tag{35}$$

Since ϕ_ε has been identified, (35) is a linear Fredholm integral equation of the first kind for f_{s_2}. Let $U : L^2(\mathbb{R}_+) \to L^2_\lambda(\mathbb{R})$ be the integral operator associated to it, defined as:

$$(Uf_{s_2})(\xi) = \int_{\mathbb{R}_+} \phi_\varepsilon(\xi s) f_{s_2}(s)\, ds$$

Equation (35) has a unique solution for f_{s_2} if U is injective and $\phi_{\pi_2} \in \mathcal{R}(U)$, where $\mathcal{R}$ is the range of the operator U. First, I show that $\phi_{\pi_2} \in \mathcal{R}(U)$. Consider:

$$\begin{aligned} ||\phi_{\pi_2}||^2_{\lambda,2} &= \int_{\mathbb{R}} \left|\phi_{\pi_2}(\xi)\right|^2 \lambda(\xi)\, d\xi \\ &\leq \int_{\mathbb{R}} \left(\int_{\mathbb{R}_+} \left|\phi_\varepsilon(\xi s)\right|^2 ds\right)\left(\int_{\mathbb{R}_+} f^2_{s_2}(s)\, ds\right)\lambda(\xi)\, d\xi \\ &< ||f_{s_2}||^2_2 \left(\int_{\mathbb{R}} \frac{1}{\xi}\lambda(\xi)\, d\xi\right)\left(\int_{\mathbb{R}} \phi^2_\varepsilon(x)\, dx\right) < \infty \end{aligned} \tag{36}$$

where the first inequality follows by the Cauchy-Schwarz inequality and by assumption 1F(ii) which implies that $\phi_\varepsilon(\xi s) > 0$ for all ξ, s, while (36) follows by a change of variables $x = \xi s$, and by assumptions 1K(iv) (which implies square integrability of ϕ_ε) and 1F(iii).

Second, I show that U is an injective operator. Consider two density functions, $f_{s_2}, f'_{s_2} \in L^2(\mathbb{R}_+)$ such that $(Uf_{s_2})(\xi) = \left(Uf'_{s_2}\right)(\xi)$. Then since $\phi_\varepsilon(\xi s) > 0$ by assumption 1F(ii),

$$\int_{\mathbb{R}_+} \phi_\varepsilon(\xi s)\left(f_{s_2}(s) - f'_{s_2}(s)\right) ds = 0 \Rightarrow f_{s_2} = f'_{s_2} \tag{37}$$

Similarly, f_{η_2} is identified by showing that the linear Fredholm integral operator $\widetilde{U}$ below is injective:

$$\phi_{\tau_2}(\xi) = \int_{\mathbb{R}_+} \phi_\eta(\xi s) f_{z_2}(s)\, ds \equiv \left(\widetilde{U} f_{z_2}\right)(\xi)$$

Since the proof of identification is similar to the above, it is not repeated here.

A.2 Identification without assumption 1I

Consider two time periods.

Assumption R

(i) $\pi_1,\ \tau_1,\ \pi_2,\ \tau_2$ are mutually independent; (ii) $E(\tau_1) = 0$; (iii) $f_{\pi_1}, f_{\tau_1} \in L_1(\mathbb{R})$; (iv) The first $K \geq 1$ moments of π_1 and τ_1 exist, $E\left(|\pi_1|^k\right) < \infty$ and $E\left(|\tau_1|^k\right) < \infty$ a.s. $k = 1, ..., K$; (v) The first $K \geq 1$ moments of ε_1 and η_1 exist, are non-zero, and are known.

Theorem 4 *Let the distribution of* (Y_1, Y_2) *be observed and assume that it follows the model described by* (1*a*), (1*b*), (4*a*), *and* (4*b*). *Let assumption R hold. Then* $\phi_{\pi_1}(\xi)$ *and* $\phi_{\tau_1}(\xi)$ *are identified by Kotlarski's lemma and the first* $K \geq 1$ *moments of* s_1 *and* z_1 *are identified and given by*

$$m_{s_1}^{(k)} = \frac{\phi_{\pi_1}^{(k)}(0)}{\phi_{\varepsilon_1}^{(k)}(0)} \text{ and } m_{z_1}^{(k)} = \frac{\phi_{\tau_1}^{(k)}(0)}{\phi_{\eta_1}^{(k)}(0)},\ k = 1, 2, ..., K$$

Proof. Consider the first two time periods:

$$\begin{aligned} Y_1 &= \pi_1 + \tau_1 \\ Y_2 &= \pi_1 + \pi_2 + \tau_2 \end{aligned}$$

By Kotlarski's lemma, the characteristic functions ϕ_{π_1} and $\phi_{\tau_1}$are identified. Then, as in the proof of Theorem 1, the law of iterated expectations obtains:

$$\phi_{\pi_1}(\xi) = \int_{\mathbb{R}_+} \phi_{\varepsilon_1}(\xi s) f_{s_1}(s)\, ds$$

where ϕ_{ε_1} and f_{s_1} are unknown. By assumptions R(iv) and (v), it is possible to differentiate the expression above $k = 1, 2, ..., K$ times with respect to ξ and evaluate the resulting expression at $\xi = 0$. This obtains

$$\phi_{\pi_1}^{(k)}(0) = \phi_{\varepsilon_1}^{(k)}(0)\, m_{s_1}^{(k)},\ k = 1, 2, ..., K \tag{38}$$

By assumption R(v), $\phi_{\varepsilon_1}^{(k)}(0) \neq 0$ and since it is known for all $k = 1, ..., K$, (38) can be solved for the sequence of moments $\left\{m_{s_1}^{(k)}\right\}_{k=1}^{K}$, obtaining the expression in Theorem 4.

A similar analysis applies to show the identification of the sequence of moments $\left\{m_{z_1}^{(k)}\right\}_{k=1}^{K}$. ■

A.3 Proof of Theorem 2

The estimation error of $\widehat{f}_\varepsilon(z)$ is decomposed into a stochastic error component and a deterministic error part, as follows:

$$\begin{aligned} \widehat{f}_\varepsilon(z) - f_\varepsilon(z) &= \frac{1}{2\pi}\int \exp(-iz\xi)\, \phi_K\left(\frac{\xi}{h_n}\right)\left[\widehat{\phi}_\varepsilon(\xi) - \phi_\varepsilon(\xi)\right] d\xi & (39) \\ &+ \frac{1}{2\pi}\int \exp(-iz\xi)\, \phi_\varepsilon(\xi)\left[\phi_K\left(\frac{\xi}{h_n}\right) - 1\right] d\xi & (40) \end{aligned}$$

where $\phi_\varepsilon(z)$ is given by (25) and $\widehat{\phi}_\varepsilon(z)$ is its sample counterpart.

Consider first the stochastic error part in (39). For n large enough:

$$\begin{aligned}&\sup_z\left|\frac{1}{2\pi}\int_{-h_n}^{h_n}\exp(-iz\xi)\,\phi_K\left(\frac{\xi}{h_n}\right)\left[\widehat{\phi}_\varepsilon(\xi)-\phi_\varepsilon(\xi)\right]d\xi\right|\\ \leq\ &\frac{h_n}{\pi}\sup_{\xi\in[-h_n,h_n]}\left|\widehat{\phi}_\varepsilon(\xi)-\phi_\varepsilon(\xi)\right|\end{aligned}$$

By using assumption 2(i), the deterministic part appearing in (40) can be bounded as:

$$\begin{aligned}&\sup_z\left|\frac{1}{2\pi}\int_{-h_n}^{h_n}\exp(-iz\xi)\,\phi_\varepsilon(\xi)\left[\phi_K\left(\frac{\xi}{h_n}\right)-1\right]d\xi\right|\\ \leq\ &\frac{1}{2\pi}\int_{-h_n}^{h_n}g_\varepsilon(|\xi|)\left|\phi_K\left(\frac{\xi}{h_n}\right)-1\right|d\xi\end{aligned}$$

Then for n large enough

$$\begin{aligned}\sup_z\left|\widehat{f}_\varepsilon(z)-f_\varepsilon(z)\right| &\leq \frac{h_n}{\pi}\sup_{\xi\in[-h_n,h_n]}\left|\widehat{\phi}_\varepsilon(\xi)-\phi_\varepsilon(\xi)\right|\\ &\quad+\frac{1}{2\pi}\int_{-h_n}^{h_n}g_\varepsilon(|\xi|)\left|\phi_K\left(\frac{\xi}{h_n}\right)-1\right|d\xi \qquad (41)\end{aligned}$$

Define

$$a(s)=E\left[iY_{il}\exp(isY_{ik})\right] \text{ and } b(s)=E\left[\exp(isY_{ik})\right]$$

and let $\widehat{a}(s)$ and $\widehat{b}(s)$ be the sample counterparts of $a(s)$ and $b(s)$, respectively. Consider:

$$\begin{aligned}\sup_{\xi\in[-h_n,h_n]}\left|\widehat{\phi}_\varepsilon(\xi)-\phi_\varepsilon(\xi)\right| &\leq \sup_{\xi\in[-h_n,h_n]}|\phi_\varepsilon(\xi)|\left|\exp\left[\int_0^\xi\frac{\widehat{a}(s)}{\widehat{b}(s)}ds-\int_0^\xi\frac{a(s)}{b(s)}ds\right]\right|\\ &\leq \sup_{\xi\in[-h_n,h_n]}\int_0^\xi\left|\frac{\widehat{a}(s)}{\widehat{b}(s)}-\frac{a(s)}{b(s)}\right|ds \qquad (42)\\ &\leq h_n\frac{C\varepsilon_n}{g_y^2(h_n)} \qquad (43)\end{aligned}$$

The last inequality follows by the arguments below. Consider the term inside the integral in (42). Applying the same type of arguments as Bonhomme and Robin (2010) we can write

$$\frac{\widehat{a}(s)}{\widehat{b}(s)}-\frac{a(s)}{b(s)}=\frac{1}{b(s)}\left(\widehat{a}(s)-a(s)\right)-\frac{\widehat{a}(s)}{b(s)}\frac{\frac{\widehat{b}(s)-b(s)}{b(s)}}{\frac{\widehat{b}(s)-b(s)}{b(s)}+1}$$

Now we bound each term appearing in the expression above. Consider first:

$$\left|\frac{1}{b(s)}\right| = \left|\frac{1}{E\exp(isY)}\right| \leq \frac{1}{\inf_{|s|>c}|\phi_Y(s)|} \leq \frac{1}{g_y(h_n)} \tag{44}$$

Inequality (44) above follows since by assumption 2(ii) we have that

$$\inf_{|s|\leq h_n}|\phi_Y(s)| = \begin{cases} g_y(h_n) \text{ if } c < |s| \leq h_n \\ \inf_{|s|\leq c}|\phi_Y(s)| \text{ if } |s| \leq c < h_n \end{cases}$$

By Lemma 1 in Bonhomme and Robin (2010),

$$|\widehat{a}(s) - a(s)| \leq \sup_{|s|\leq h_n}|\widehat{a}(s) - a(s)| < C\varepsilon_n \tag{45}$$

Then by (44) and (45),

$$\left|\frac{\widehat{b}(s) - b(s)}{b(s)}\right| \leq \sup_{|s|\leq h_n}\left|\frac{\widehat{b}(s) - b(s)}{b(s)}\right| < \frac{C\varepsilon_n}{g_y(h_n)}$$

and by assumption 2(iii),

$$|\widehat{a}(s)| \leq \sup_{|s|\leq h_n}|\widehat{a}(s)| \leq \sup_{|s|\leq h_n}|\widehat{a}(s) - a(s)| + E|Y_{il}| = O(1)$$

Then we have that

$$\left|\frac{\widehat{a}(s)}{\widehat{b}(s)} - \frac{a(s)}{b(s)}\right| \leq \frac{O(\varepsilon_n)}{g_y^2(h_n)}$$

Plugging this into (42) obtains (43).

Then (41) becomes

$$\sup_z\left|\widehat{f}_\varepsilon(z) - f_\varepsilon(z)\right| \leq \frac{h_n^2}{\pi}\frac{O(\varepsilon_n)}{g_y^2(h_n)} + \frac{1}{2\pi}\int_{-h_n}^{h_n} g_\varepsilon(|\xi|)\left|\phi_K\left(\frac{\xi}{h_n}\right) - 1\right| d\xi$$

and by the same arguments as in the proof of Theorem 2 of Bonhomme and Robin (2010), we have that

$$\sup_z\left|\widehat{f}_\varepsilon(z) - f_\varepsilon(z)\right| \leq \frac{h_n^2}{\pi}\frac{O(\varepsilon_n)}{g_y^2(h_n)} + \sup_{\xi\in[-1,1]}|m(\xi)|\frac{1}{h_n^q}\int_{-h_n}^{h_n}\tau^q g_\varepsilon(|\xi|)\,d\xi + 2\int_{h_n}^{\infty} g_\varepsilon(|\xi|)\,d\xi$$

A.4 Proof of Theorem 3

We prove the theorem only for the density function f_{s_2}. The proof relies heavily on Darolles, Fan, Florens, and Renault (2011). The proof for f_{z_2} is similar.

As explained in the main text, the (Tikhonov) regularized solution for $\widehat{f}_{s_2}$ is given by

$$\widehat{f}_{s_2}(\tau) = \left(\alpha_n I + \widehat{U}_n^* \widehat{U}_n\right)^{-1} \widehat{U}_n^* \widehat{\phi}_{s_2\varepsilon}(\tau) = \widehat{A}_{\alpha,n} \widehat{U}_n^* \widehat{\phi}_{s_2\varepsilon}(\tau)$$

where $\widehat{A}_{\alpha,n} = \left(\alpha_n I + \widehat{U}_n^* \widehat{U}_n\right)^{-1}$. Letting $A_{\alpha,n} = (\alpha_n I + U^* U)^{-1}$, we decompose the estimation error as follows:

$$\begin{aligned} \widehat{f}_{s_2}(\tau) - f_{s_2}(\tau) &= \widehat{A}_{\alpha,n} \widehat{U}_n^* \widehat{\phi}_{s_2\varepsilon}(\tau) - \widehat{A}_{\alpha,n} \widehat{U}_n^* \phi_{s_2\varepsilon}(\tau) \\ &\quad + \widehat{A}_{\alpha,n} \widehat{U}_n^* \phi_{s_2\varepsilon}(\tau) - A_{\alpha,n} U^* \phi_{s_2\varepsilon}(\tau) \\ &\quad + A_{\alpha,n} U^* \phi_{s_2\varepsilon}(\tau) - f_{s_2}(\tau) \end{aligned} \quad (46)$$

where the first term in (46) is the error associated with estimating $\widehat{\phi}_{s_2\varepsilon}$, the second term is the error associated with estimation of U, and the last terms is the regularization error.

Consider the L^2-norm of the first term in (46):

$$\begin{aligned} \left\|\widehat{A}_{\alpha,n} \widehat{U}_n^* \left(\widehat{\phi}_{s_2\varepsilon}(\tau) - \phi_{s_2\varepsilon}(\tau)\right)\right\|_{L^2}^2 &\leq \frac{1}{\alpha_n} \left\|\widehat{\phi}_{s_2\varepsilon}(\tau) - \phi_{s_2\varepsilon}(\tau)\right\|_{L^2}^2 \\ &\leq 2\frac{C^2}{\alpha_n} h_n^3 \frac{\varepsilon_n^2}{g_y^4(h_n)} \end{aligned}$$

where we used that the operator norm $\left\|\widehat{A}_{\alpha,n} \widehat{U}_n^*\right\| \leq \frac{1}{\sqrt{\alpha_n}}$ (see Groetsch (1984)) and that

$$\begin{aligned} \left\|\widehat{\phi}_{s_2\varepsilon}(\tau) - \phi_{s_2\varepsilon}(\tau)\right\|_{L^2}^2 &\leq \|f_{s_2}\|_{L^2}^2 \int_{-h_n}^{h_n} \sup_s \left|\widehat{\phi}_\varepsilon(s\tau) - \phi_\varepsilon(s\tau)\right|^2 d\tau \\ &\leq 2h_n M_1^2 \sup_{s,\tau} \left|\widehat{\phi}_\varepsilon(s\tau) - \phi_\varepsilon(s\tau)\right|^2 \\ &\leq (2h_n) M_1^2 \left(h_n \frac{C\varepsilon_n}{g_y^2(h_n)}\right)^2 \end{aligned} \quad (47)$$

The second term in (46) can be written as follows

$$\begin{aligned} &\widehat{A}_{\alpha,n} \widehat{U}_n^* \widehat{U}_n f_{s_2}(\tau) - A_{\alpha,n} U^* U f_{s_2}(\tau) = \alpha_n \left[\widehat{A}_{\alpha,n} - A_{\alpha,n}\right] f_{s_2}(\tau) \\ = &-\left[\alpha_n \widehat{A}_{\alpha,n} \widehat{U}_n^* \left(\widehat{U}_n - U\right) A_{\alpha,n} f_{s_2}(\tau) + \alpha_n \widehat{A}_{\alpha,n} \left(\widehat{U}_n^* - U^*\right) U A_{\alpha,n} f_{s_2}(\tau)\right] \end{aligned} \quad (48)$$

where we used the identity $M^{-1} - N^{-1} = M^{-1}(N - M) N^{-1}$ with $M^{-1} = \widehat{A}_n$ and $N^{-1} = A$. Consider the terms appearing in (48). By the results in Groetsch (1984) and by Proposition 1 we have that:

$$\begin{aligned}\left|\left|\widehat{A}_{\alpha,n}\right|\right|^2 &\leq \frac{1}{\alpha_n^2},\ \left|\left|\widehat{A}_{\alpha,n}\widehat{U}_n^*\right|\right| \leq \frac{1}{\alpha_n},\\ ||\alpha_n A_{\alpha,n} f||_{L^2}^2 &= O_p\left(\alpha_n^{\beta\wedge 2}\right),\ ||\alpha_n U A_{\alpha,n} f||_{L^2}^2 = O_p\left(\alpha_n^{\beta+1\wedge 2}\right).\end{aligned}$$

Additionally, by using the definition of the operator norm and the inequality in (47) we have that:

$$\begin{aligned}\left|\left|\widehat{U}_n - U\right|\right| &\leq \sup_{||f||_{L^2}\neq 0} \frac{\left|\left|\left(\widehat{U}_n - U\right) f\right|\right|_{L^2}}{||f||_{L^2}} \leq \sqrt{2}\left(h_n^{3/2}\frac{C\varepsilon_n}{g_y^2\left(h_n\right)}\right)\\ \left|\left|\widehat{U}_n^* - U^*\right|\right| &\leq \sup_{||f||_{L^2}\neq 0} \frac{\left|\left|\left(\widehat{U}_n - U\right) f\right|\right|_{L^2}}{||f||_{L^2}} \leq \sqrt{2}\left(h_n^{3/2}\frac{C\varepsilon_n}{g_y^2\left(h_n\right)}\right)\end{aligned}$$

Then (48) obtains that

$$\left|\left|\widehat{A}_{\alpha,n}\widehat{U}_n^*\widehat{U}_n f_{s_2}\left(\tau\right) - A_{\alpha,n}U^*U f_{s_2}\left(\tau\right)\right|\right|_{L^2}^2 = O_p\left(2\left(h_n^{3/2}\frac{C\varepsilon_n}{g_y^2\left(h_n\right)}\right)^2\left(\alpha_n^{\beta-1\wedge 1} + \alpha_n^{\beta-1\wedge 0}\right)\right)$$

Consider now the last term entering (46). By Proposition 1:

$$||A_{\alpha,n}U^*U f_{s_2}\left(\tau\right) - f_{s_2}\left(\tau\right)||_{L^2}^2 = O\left(\alpha_n^{\beta\wedge 2}\right)$$

References

Arellano, M. Blundell, R. and Bonhomme, S. (2017) Earnings and Consumption Dynamics: A Nonlinear Panel Data Framework. Econometrica, 85(3): 693-734.

Bonhomme, S. and Robin, J-M. (2010) Generalized Nonparametric Deconvolution with an Application to Earnings Dynamics. *Review of Economic Studies,* 77(2): 491-533.

Botosaru, I., and Sasaki, Y. (2017) Nonparametric Heteroskedasticity in Persistent Panel Processes: An Application to Earnings Dynamics. Working paper.

Breunig, C. and Hoderlein, S. (2016) Specification Testing in Random Coefficient Models. SFB 649 Discussion Paper 2015-053.

Carrasco, M., Florens, J-P., and Renault, E. (2007) Linear Inverse Problems in Structural Econometrics: Estimation Based on Spectral Decomposition and Regularization. *Handbook of*

Econometrics, Volume 6B, ed. by J. Heckman and E. Leamer. Elsevier/North Holland, 5633-5751.

Carrasco, M. and Florens, J.P. (2011) A Spectral Method for Deconvolving a Density. Econometric Theory, 27: 546-581.

Cuhna, F., Heckman, J.J., and Schennach, S. (2010) Estimating the Technology of Cognitive and Noncognitive Skill Formation. Econometrica 78(3): 883-931.

Darolles, S., Fan, Y., Florens, J-P., and Renault, E. (2011) Nonparametric Instrumental Regression. *Econometrica,* 79(5): 1541-1565.

Delaigle, A., Hall, P., and Meister, A. (2008) On Deconvolution with Repeated Measurements. *The Annals of Statistics,* 36(2): 665-685.

Delaigle, A., and Hall, P. (2006) On Optimal Kernel Choice for Deconvolution. *Statistics and Probability Letters,* 76: 1594-1602.

Engl, H.W., Hanke, M., and Neubauer, A. (2000) *Regularization of Inverse Problems.* Netherlads: Kluver Academic.

Evdokimov, K. (2010) Identification and Estimation of a Nonparametric Panel Data Model with Unobserved Heterogeneity. Working paper.

Evdokimov, K. and White, H. (2012) Some Extensions of a Lemma of Kotlarski. *Econometric Theory*, 28(4): 925-932.

Fan, J. (1992) Deconvolution for Supersmooth Distributions. *Canadian Journal of Statistics,* 20: 155-169.

Groetsch, C. (1984). The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind, Pitman, London.

Guvenen, F., Ozkan, S., and Song, J. (2014) The Nature of Countercyclical Income Risk. *Journal of Political Economy*, 122(3): 621-660.

Guvenen, F., Karahan, F., Ozkan, S., and Song, J. (2016) What Do Data on Millions of U.S. Workers Reveal about Life-Cycle Earnings Dynamics? Working paper.

Hoderlein, S., Klemelä, J., and Mammen, E. (2010) Analyzing the random coefficient model nonparametrically. *Econometric Theory*, 26(3): 804-837.

Hoderlein, S., Nesheim, L. and Simoni, A. (2016) Semiparametric Estimation of Random Coefficients in Structural Economic Models. *Econometric Theory*, 1-41. doi:10.1017/S0266466616000396.

Horowitz, J. (1998) Semiparametric Methods in Econometrics. Springer-Verlag, New York.

Hospido, L. (2012) Modelling Heterogeneity and Dynamics in the Volatility of Individual Wages. *Journal of Applied Econometrics*, 27: 386-414.

Hu, Y. and Ridder, G. (2010) On Deconvolution as a First Stage Nonparametric Estimator. *Econometric Reviews,* 29(4): 365-396.

Jensen, S.T., and Shore, S.H. (2011) Semi-Parametric Bayesian Modelling of Income Volatility Heterogeneity. *Journal of the American Statistical Association*, 106(496): 1280-1290.

Jensen, S.T., and Shore, S.H. (2015) Changes in the Distribution of Earnings Volatility. *Journal of Human Resources*, 50(3): 811-836.

Kaplan, G., and Violante, G.L. (2010) How Much Consumption Insurance beyond Self-Insurance? *American Economic Journal: Macroeconomics*, 2(4), 53-87.

Kim, W. (2004) Identification and Estimation of Nonparametric Structural Models by Instrumental Variables Method. Manuscript, Humbolt University.

Li, T. (2002) Robust and Consistent Estimation of Nonlinear Errors-in-Variables Models. *Journal of Econometrics*, 110(1): 1-26.

Li, T. and Vuong, Q. (1998) Nonparametric Estimation of the Measurement Error Model Using Multiple Indicators. *Journal of Multivariate Analysis*, 65(2): 139-165.

Meghir, C. and Pistaferri, L. (2004) Income Variance Dynamics and Heterogeneity. *Econometrica,* 72(1): 1-32.

Meghir, C. and Windmeijer, F. (1999) Moment Conditions for Dynamic Panel Data Models with Multiplicative Individual Effects in the Conditional Variance. *Annales d'Économie et de Statistique*, 55/56: 317-330.

Meister, A. (2009) Deconvolution Problems in Nonparametric Statistics. Lectures Notes in Statistics Editors: P. Bickel, P.J. Diggle, S. Fienberg, U. Gather, I. Olkin, and S. Zeger. Springer-Verlang Berlin Heidelberg.

Rao, P.B.L.S. (1983) Nonparametric Functional Estimation. Academic Press, London.

Rao, P.B.L.S. (1992) Identifiability in Stochastic Models. Academic Press, Boston.

Salgado, S., Guvenen, F. and Bloom, N. (2017) Skewed Business Cycles. Working paper.

Schennach, S.M. (2004a) Estimation of Nonlinear Models with Measurement Error. *Econometrica*, 72(1): 33-75.

Schennach, S.M. (2004b) Nonparametric Regression in the Presence of Measurement Error. *Econometric Theory*, 20(6): 1046-1093.

Stefanski, L., and Carroll, R.J. (1990) Deconvolution Kernel Density Estimators. *Statistics,* 2: 169-184.