

HETEROGENEITY IN TRANSITORY INCOME RISK^{*}

Martín Almuzara[†]

This version: October 2020

Abstract

I propose a new framework to distinguish between permanent and transitory income changes that permits various forms of cross-sectional heterogeneity. This includes a model in which log earnings are the sum of a permanent component and a transitory innovation with household-specific variance; permanent income and the transitory income variance are, moreover, potentially correlated. I establish nonparametric identification results and introduce a flexible estimation method. Using data from the Panel Study of Income Dynamics, I find that (i) heterogeneity in transitory risk is sizable with a highly asymmetric distribution and (ii) permanent income and transitory risk have a negative nonlinear relationship. Finally, I reassess the evidence on nonlinear transmission of income shocks.

Keywords: Income process, nonparametric identification, specification tests.

JEL Classification: C23, C52, D31.

^{*}I have accumulated an immense debt of gratitude towards Manuel Arellano and Enrique Sentana for all their support and guidance during my PhD. I also wish to thank Dante Amengual, Dmitry Arkhangelsky, and seminar participants at CEMFI for valuable comments and discussions. All errors are my sole responsibility. I acknowledge funding from the Santander Research Chair at CEMFI.

[†]Federal Reserve Bank of New York: martin.almuzara@ny.frb.org

1 Introduction

The distinction between permanent and transitory income changes occupies a central place in various areas of economic analysis. It does in the study of earnings mobility (e.g., Lillard and Willis (1978), Meghir and Pistaferri (2004), and Gottschalk and Moffitt (2009)) where the relative magnitudes of permanent and transitory changes determine the persistence of income inequality. It also does in the study of consumption and labor supply choices by households (e.g., Hall and Mishkin (1982), Abowd and Card (1989), Deaton and Paxson (1994), Blundell and Preston (1998), Blundell et al. (2008), Altonji et al. (2013), and Arellano et al. (2017)), and more generally, in quantitative macroeconomic models (e.g., Kaplan and Violante (2010)). Understanding the nature of changes in inequality and household responses to income shocks is relevant for both academic and policy discussions.

In practice, permanent and transitory changes in earnings are not directly observed. For the framework to be useful, it must be possible to identify those changes from data on total household earnings. One approach uses observed events, such as long illness, unemployment spells, and tax refunds, as a proxy (or instrument) for permanent and transitory changes (e.g., Cochrane (1991) and Souleles (1999)). Another approach relies on restricting the dependence among the unobserved shocks to enable their deconvolution (e.g., MaCurdy (1982), Horowitz and Markatou (1996), Bonhomme and Robin (2010), and Arellano and Bonhomme (2019)). The most common restrictions include statistical independence between permanent and transitory components and serial independence of transitory shocks. While convenient, such restrictions rule out potentially interesting dimensions of earnings uncertainty. In particular, they leave aside the possibility to learn about heterogeneity in transitory income variances, a feature highlighted by a number of labor market models.

Motivated by that observation, in this paper I propose a framework that can separate permanent and transitory components while permitting dependence between them. My primary focus is on a model in which the variance of transitory income shocks is cross-sectionally heterogeneous and possibly correlated with the initial level of permanent income. With a small abuse of vocabulary, I call it the heterogeneous transitory risk (HTR) model and I introduce it in section 2.

Everyday experience is rich in situations where workers differ in their transitory income risks. The waiter has her fortune tied to the vagaries of customer sympathy; the school teacher has not. A large literature attributes those differences to risk allocation, to the need to protect

investments from hold-up problems, and to asymmetric information.¹ Moreover, workers tend to remain in the same firm—and, more broadly, within the same activity—for long periods.² The result is a cross-sectional distribution of transitory risks that changes only slowly over the working life of individuals. In short panels, then, it is reasonable to treat the distribution as fixed, as the HTR model does.³

The HTR model is a suitable tool to quantify the extent to which households differ in their transitory income volatilities. There are at least three reasons why such an exercise is important. First, the transitory component explains a large fraction of the variance of yearly changes in household incomes (close to 90% according to estimates from the US economy). They compose, thus, the bulk of the uncertainty that impacts on households decisions and welfare. Second, transitory shocks are policy relevant because many economic policies (e.g., tax rebates and certain forms of income transfers) are often perceived as temporary. Third, in models with incomplete markets, heterogeneity in transitory risks maps to heterogeneity in self-insurance. This suggests that estimates of insurance coefficients which neglect differences in transitory risks may be misleading.

In addition, the HTR model permits measuring the correlation between permanent income and transitory risk. Low permanent-income households, who typically have less means to insure their consumption (e.g., due to tighter borrowing constraints), are likely to experience the largest losses from a given exposure to transitory risk. In consequence, the association between permanent income and transitory risk matters for thinking about welfare implications of changes in inequality and policy.

My paper makes two different types of contributions. The *first* are methodological. I establish nonparametric identification results for a class of permanent-transitory models which do not assume independence between latent components. The distribution of time-invariant cross-sectional heterogeneity can, accordingly, be recovered from the distribution of observables in short panels. For the HTR model, this means that the joint distribution of permanent income and transitory risk is uniquely determined by the distribution of observed earnings. A full discussion is the subject of section 3.

Relaxing the independence of unobserved shocks is challenging as it invalidates the linear

¹Malcomson (1999) gives a complete review.

²This emerges in a model in which workers accumulate activity-specific capital, i.e., capital that would have a small return if applied to other activities. A comprehensive summary of this dimension of mobility can be found in Farber (1999).

³Notice that heterogeneity in workers' transitory risks aggregates to heterogeneity in households' transitory risks unless a very special form of sorting takes place. Also notice that, in a more detailed analysis which allows for job changes, the worker-specific variance would turn into a within-job fixed effect.

deconvolution techniques typically employed in the literature to establish identification (e.g., Kotlarski (1967) and Székely and Rao (2000)). The approaches followed in Wilhelm (2015) and Arellano et al. (2017) do not apply either. Instead, my identification results build on the techniques developed by Hu and Schennach (2008) in their treatment of econometric models with nonclassical measurement error.⁴ To state the conditions for my result, consider a model with a k -dimensional individual-specific parameter ζ and observables y_1, \dots, y_T . If one can choose a t so that (i) y_t is independent of $y_{t-k}, \dots, y_{t-1}, y_{t+1}, \dots, y_{t+k}$ given ζ , (ii) y_{t-k}, \dots, y_{t-1} are independent of y_{t+1}, \dots, y_{t+k} given ζ , and (iii) there is a functional of the distribution of y_{t-k}, \dots, y_{t-1} conditioned on ζ that equals ζ , the distribution of ζ is identified.⁵ The HTR model, for example, satisfies the conditions with $T \geq 5$; an enhanced version of the HTR model with household-specific variances of permanent shocks demands $T \geq 7$, and so on. As in random coefficients panel data models, the more individual-specific parameters, the richer the time series dimension needed to identify their distribution (e.g., Arellano and Bonhomme (2012)).⁶

Furthermore, I introduce a flexible estimation method for the distribution of individual-specific parameters. Given that in applications researchers often want to allow for (strictly exogenous) covariates, I propose a method that handles covariates in a convenient way.⁷ My proposal is based on the specification of flexible approximations to nonparametric functions of covariates and distributions of latent variables. As in sieve approaches, unknown functions and distributions are, for a given approximation, defined by a set of moment conditions that involve both observables and latent variables. To render the moments feasible, I use a stochastic EM algorithm that alternates between (i) simulation of latent variables from their conditional distribution given data and (ii) moment equations. My method relates to recent developments in the estimation of latent variables models (e.g., Arellano and Bonhomme (2016), Arellano et al. (2017), and Arellano and Bonhomme (2019)). The details can be found in section 4.

The *second* type of contributions of my paper are empirical. Using data from the Panel Study of Income Dynamics (PSID), I estimate the HTR model and conduct a fully-fledged empirical analysis. I focus on two panels constructed from the 1999-2009 waves of the PSID. The first is a panel of the labor earnings plus transfers of families in which the representative person is a married male individual—a panel of households. The second uses labor earnings of representative persons alone without the restriction to married males—a panel of workers. While a

⁴Hu (2019) offers a complete review of this literature.

⁵Requirements on boundedness of densities and completeness of certain distributions are needed too.

⁶At least in the HTR model, the requirement that $T \geq 2k + 1$ cannot be relaxed in a meaningful way.

⁷The argument for the identification of the conditional distribution of individual-specific parameters given covariates is essentially the same as the one given for the unconditional distribution.

panel of households is better suited to link earnings uncertainty and consumption insurance, a panel of workers is closer to the labor market models that suggest heterogeneity in transitory income variances.

My empirical findings can be summarized as follows. First, differences in transitory income risk are sizable, both across households and workers, and they are asymmetrically distributed. To get a sense of it, I compare transitory variances with their average (a model with no heterogeneity would attach the average variance to every unit): 50% of the households have half the average variance or less, while 16% have twice the average variance or more. Workers, in turn, portray slightly less—still significant—heterogeneity than households.⁸ Second, permanent income and transitory risk are negatively associated: low permanent-income households tend to face more volatile transitory incomes than their higher permanent-income counterparts. The association, however, is not absolute—variation of transitory risks around the tendency exists at each level of permanent income—and is nonlinear. The same holds for workers.

The PSID provides the means to perform three additional exercises. First, many papers have documented a rapid rise in the variance of transitory income shocks during the 70s and 80s.⁹ By applying the HTR model to a sequence of panels around the period, I can measure changes along the whole distribution of transitory risks—not just in the average—that complement the original evidence. Second, the PSID collects data on labor market variables (self-employment status, occupation, sector, etc.). I can investigate the correlation between these variables and individual-specific transitory risks. Finally, I use the transitory income shocks from the HTR model together with consumption data to estimate insurance coefficients. The full empirical analysis is developed in section 5.

Earnings uncertainty has produced a vast literature. Three branches of it are most closely related to my paper. On one side there are models with nonparametric heteroskedasticity (e.g., Botosaru (2017) and Botosaru and Sasaki (2018)). These typically leave aside time-invariant heterogeneity and dependence between permanent and transitory components—a void that my paper addresses. Second, Browning et al. (2010) and Alan et al. (2018) study models with “lots of heterogeneity”—they permit a high-dimensional household-specific parameter with unrestricted joint distribution—while Hospido (2012) explicitly includes a worker-specific factor in the variance of wages in her model. My paper differs from them in that I consider a model that distinguishes between permanent and transitory shocks (a two-error model as opposed to

⁸This is consistent with the approximately additive aggregation of the transitory incomes of spouses. More so if shocks to them have a positive correlation as estimated by Blundell et al. (2016).

⁹See Gottschalk and Moffitt (2009) for a review.

the one-error models of these papers) and my approach is nonparametric. The latter turns out to be important to capture the nonlinear relation between latent components in the empirical analysis. The two-error model I study is closer to the parametric model in Chamberlain and Hirano (1999) except that I allow for dependence of transitory risk with permanent income.

The third strand that matters for my paper is the nonlinear transmission of income shocks, and particularly the ABB model of Arellano et al. (2017). Flexible estimates of predictive distributions reveal that the persistence of past income changes varies with the size and sign of current changes and with the extant level of income (e.g., (Arellano et al., 2017, fig. 1)). They also reveal a decreasing conditional skewness with respect to current income (e.g., Guvenen et al. (2014) and Arellano et al. (2017)). Though not obvious, such patterns are consistent with the HTR model, just as they are consistent with the ABB model. This does not mean that the two models are observationally equivalent. In fact, there are restrictions on the distribution of observables imposed by the HTR model not shared by the ABB model and I exploit them to construct diagnostics. That those diagnostics indicate the HTR model is not at odds with the data used in Arellano et al. (2017) implies a reassessment of the evidence on nonlinear persistence. Some parts of sections 2 and 5 deal with that discussion and section 6 concludes.

Notation. For integers j_0, j_1 with $j_0 \leq j_1$, I use $\omega_{j_0:j_1}$ to denote the sequence $\{\omega_j\}_{j=j_0}^{j_1}$. The elements $\omega_{j_0}, \dots, \omega_{j_1}$ need not be arrays of the same dimension and, thus, $\{\omega_j\}_{j=j_0}^{j_1}$ is generally an ordered list. When each ω_j is an array of dimension $d_1 \times d_2$, and if no confusion is possible, I also use $\omega_{j_0:j_1}$ to denote the $d_1 \times d_2(j_1 - j_0 + 1)$ array obtained by horizontal concatenation of the terms of $\{\omega_j\}_{j=j_0}^{j_1}$. I write $\text{diag}(\omega_{j_0:j_1})$ for the block diagonal matrix with blocks $\omega_{j_0}, \dots, \omega_{j_1}$. I adopt the conventions $\sum_{j=j_1}^{j_0} \omega_j = 0$ and $\prod_{j=j_1}^{j_0} \omega_j = 1$ if $j_0 < j_1$. I write $\mathbb{E}_J[\omega_j] \equiv J^{-1} \sum_{j=1}^J \omega_j$ for the average of $\omega_{1:J}$ (often $J = n$), “ \sim ” for equality in distribution, “ \xrightarrow{p} ” for convergence in probability, and “ \implies ” for weak convergence. Finally, if x and y are generic random elements, I write \mathbb{P}_y and $\mathbb{P}_{y|x}$ for the probability measure of y and of y conditioned on x .

2 Model

I denote by y_{it} the log income of household i at time t and by x_i a set of covariates. Note x_i may contain both time-invariant and time-varying covariates, individual and aggregate. Also note t refers to time as opposed to age, which will be often included in x_i . Let $y_{i,1:T}$ be the log income history of household i through times 1 to T and let $\{y_{i,1:T}, x_i\}_{i=1}^n$ be a balanced panel.

The heterogeneous transitory risk (HTR) model decomposes log income into deterministic, permanent and transitory components of variation:

$$(1) \quad y_{it} = \mu_{it} + \tilde{\eta}_{it} + \tilde{\varepsilon}_{it}, \quad i = 1, \dots, n, \quad t = 1, \dots, T.$$

Deterministic income is determined by covariates and all its heterogeneity is observable,

$$(2) \quad \mu_{it} = \mu_t(x_i) = \mathbb{E}[\mu_{it}|x_i].$$

Permanent income, in turn, evolves as a random-walk with initial level $\tilde{\eta}_{i1} = \eta_i$,

$$(3) \quad \tilde{\eta}_{it} = \tilde{\eta}_{i,t-1} + v_{it} = \eta_i + \sum_{s=2}^t v_{is},$$

$$\mathbb{E}[\eta_i|x_i] = 0,$$

$$\mathbb{E}[v_{i2:T}|x_i] = \mathbf{0}_{1 \times (T-1)}.$$

Finally, transitory income has unit-specific variance σ_i^2 ,

$$(4) \quad \tilde{\varepsilon}_{it} = \sigma_i \varepsilon_{it},$$

$$\mathbb{E}[\sigma_i^2|x_i] = 1,$$

$$\mathbb{E}[\varepsilon_{i1:T}|x_i] = \mathbf{0}_{1 \times T}.$$

Given covariates, I assume permanent and transitory income shocks are independent of each other and over time (although not necessarily identically distributed). I further assume them independent of the initial level of permanent income η_i and of transitory income variance σ_i^2 but—and this is empirically relevant—I permit dependence between η_i and σ_i^2 . In sum,

$$(5) \quad (\eta_i, \sigma_i^2) \perp v_{i2} \perp \dots \perp v_{iT} \perp \varepsilon_{i1} \perp \dots \perp \varepsilon_{iT} \mid x_i.$$

As I will argue below, the object of interest is precisely $\mathbb{P}_{(\eta, \sigma^2)|x}$, the class of joint distributions of (η_i, σ_i^2) conditioned on x_i .

Remarks about assumptions. The conditional independence of income shocks and the extant level of permanent income is standard in the literature on earnings uncertainty. It is a natural starting point as it embodies the notion that changes in permanent and transitory income are

totally unpredictable. From a substantive point of view, the presence of σ_i^2 and its possible dependence to η_i are attractive elements of the HTR model, as emphasized in section 1. From the statistical point of view, they introduce dependence between permanent and transitory incomes, raising challenges to identification that I address in section 3.

Mean-independence restrictions on η_i , $v_{i,2:T}$, and $\varepsilon_{i,1:T}$ in (3) and (4) are akin to random effects panel data approaches and explicitly or implicitly underlie any empirical analysis of income processes. Despite restricting the mean of η_i , $v_{i,2:T}$, and $\varepsilon_{i,1:T}$, they leave the conditional distributions $\mathbb{P}_{\eta|x}$, $\mathbb{P}_{v_t|x}$ ($t = 2, \dots, T$), and $\mathbb{P}_{\varepsilon_t|x}$ ($t = 1, \dots, T$) otherwise free. In particular, dependence of higher-order moments of η_i , $v_{i,2:T}$, and $\varepsilon_{i,1:T}$ on x_i is permitted.

The mean-independence constraint on σ_i^2 is, in this context, a convenient normalization. It serves the purpose of separating the mean of σ_i^2 from the scale of $\varepsilon_{i,1:T}$. Noting

$$\text{Var}(\tilde{\varepsilon}_{it}|x_i) = \mathbb{E} \left[\sigma_i^2 | x_i \right] \cdot \text{Var}(\varepsilon_{it}|x_i), \quad t = 1, \dots, T,$$

reveals that other possibilities exist, e.g., setting $\text{Var}(\varepsilon_{i1}|x_i) = 1$ or $T^{-1} \sum_{t=1}^T \text{Var}(\varepsilon_{it}|x_i) = 1$. It is not difficult to map the implications of one normalization onto the others, but one must be careful in interpreting $\mathbb{P}_{\sigma^2|x}$ and $\mathbb{P}_{\varepsilon_t|x}$ ($t = 1, \dots, T$). Finally, dependence of higher-order moments of σ_i^2 on x_i are not constrained.

In section 3, I will show that the mean-independence assumption may be substituted by, e.g., median-independence or, more generally, by restrictions on conditional quantiles.

Sources of heterogeneity. The HTR model's stamp is the inclusion of σ_i^2 as a vehicle for heterogeneity in addition to the usual η_i of the canonical model. By (1), (2), (3), and (4), the HTR model maps covariates x_i and latent variables $\{\eta_i, \sigma_i^2, v_{i,2:T}, \varepsilon_{i,1:T}\}$ to earnings histories $y_{i,1:T}$. To be specific,

$$(6) \quad y_{it} = \mu_t(x_i) + \eta_i + \sum_{s=2}^t v_{is} + \sigma_i \varepsilon_{it}.$$

From the statistical point of view, the HTR model maps probability distributions of covariates and latent variables $\mathbb{P}_{x'}$, $\mathbb{P}_{(\eta, \sigma^2)|x'}$, $\mathbb{P}_{v_t|x}$ ($t = 2, \dots, T$), and $\mathbb{P}_{\varepsilon_t|x}$ ($t = 1, \dots, T$), into probability distributions for earnings data $\mathbb{P}_{y|x}$

By (6) and the assumptions above,

$$(7) \quad \mathbb{E} \left[y_{it} | x_i, \eta_i, \sigma_i^2 \right] = \mu_t(x_i) + \eta_i,$$

$$(8) \quad \text{Var}\left(y_{it} \mid x_i, \eta_i, \sigma_i^2\right) = \sum_{s=2}^t \sigma_{v_s}^2(x_i) + \sigma_i^2 \sigma_{\varepsilon_t}^2(x_i),$$

where $\sigma_{v_t}^2(x_i) \equiv \text{Var}(v_t \mid x_i)$ ($t = 2, \dots, T$) and $\sigma_{\varepsilon_t}^2(x_i) \equiv \text{Var}(\varepsilon_t \mid x_i)$ ($t = 1, \dots, T$). In other words, the HTR model allows for time-varying observed and permanent unobserved heterogeneity in both mean and variance (cf. the linear error-component regression model (e.g., (Arellano, 2003, Ch. 3))).

2.1 Heterogeneity vs uncertainty

It is a common fact in short panels that heterogeneity and uncertainty, which have radically different individual implications, often derive in similar aggregate implications—sometimes, to such extent that heterogeneity and uncertainty cannot be distinguished.

The HTR model and the ABB model of Arellano et al. (2017) exemplify the situation. At the individual level, the HTR model is nothing but the canonical model of earnings dynamics; nonlinearities in the aggregate arise as a consequence of heterogeneity. The ABB model tells the opposite story: individuals face nonlinear permanent income processes which emerge to the surface in the distribution of earnings. Their implications for certain higher-order features of the distribution of earnings look alike as the following simulation experiment and a stylized example (developed in appendix A) show. In particular, unit-specific variances in transitory income can lead to a pattern of nonlinear persistence and decreasing conditional skewness as in the ABB model.

A simulation experiment. Consider a version of the HTR model with no covariates:

$$y_{it} = \eta_i + \sum_{\tau=2}^t v_{i\tau} + \sigma_i \varepsilon_{it}, \quad t = 1, \dots, T, \quad i = 1, \dots, n.$$

In addition to (5), I assume that $\eta_i \perp \sigma_i^2$ and

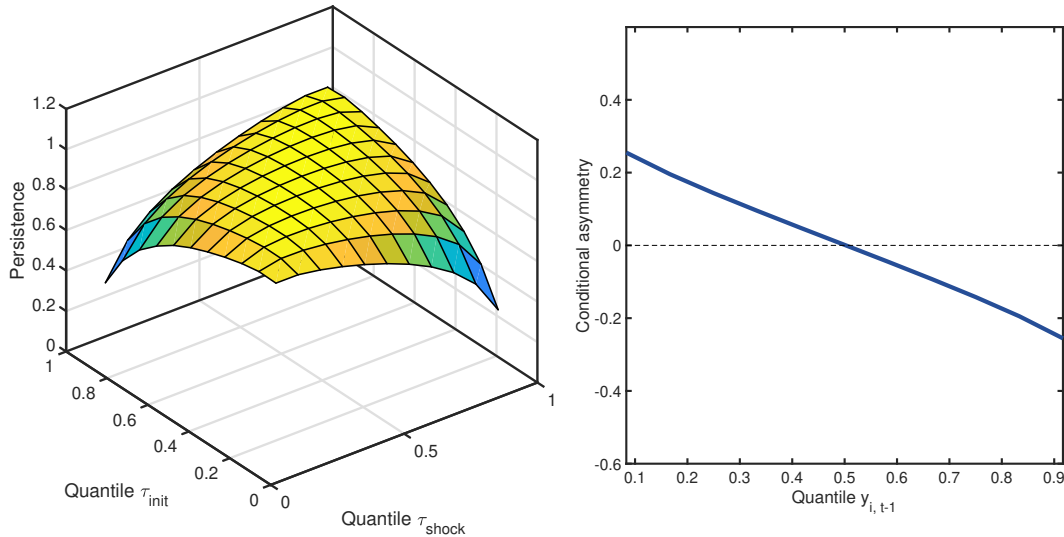
$$\eta_i \sim N\left(0, \sigma_\eta^2\right), \quad \sigma_i \sim \Gamma\left(\nu/2, 2/\sqrt{\nu(\nu+2)}\right), \quad v_{it} \sim N\left(0, \sigma_v^2\right), \quad \varepsilon_{it} \sim N\left(0, \sigma_\varepsilon^2\right).$$

I calibrate the parameters to $\sigma_\eta^2 = 0.15$, $\sigma_v^2 = 0.01$, $\sigma_\varepsilon^2 = 0.05$, and $\nu = 11.45$ (to get $\mathbb{E}\left[\sigma_i^2\right] = 1$ and $\text{Var}\left(\sigma_i^2\right) = 0.75$). These numbers are in line with the estimates I obtain later, except that

dependence between η_i and σ_i^2 is not allowed here.¹⁰

I simulate $n_{MC} = 1,000$ samples with $n = 1,000$ and $T = 6$ and I estimate in each of them flexible quantile autoregressions of y_{it} on $y_{i,t-1}$ (I use a third-order Hermite polynomial of $y_{i,t-1}$). With the estimates, I construct the measures of nonlinear persistence and conditional asymmetry in Arellano et al. (2017). Figure 1 reports the average across simulations.

FIGURE 1. Quantile autoregressions of simulated log earnings



The figure is to be compared with (Arellano et al., 2017, figs. 2 and 4). The lower persistence for high-income households subject to a bad shock and low-income households subject to a good shock emerges despite the fact that the permanent component has constant persistence across households. Some intuition for the observed pattern is that households with a high-in-absolute-value income experiencing a large change of the opposite sign are likely to have a high σ_i^2 and therefore, for them, transitory income is a larger fraction of total income and the link between past and current income is weaker. See appendix A for a discussion.

Tests. The HTR and the ABB models appear to be testable. Constructing a test of the two non-nested nonparametric hypotheses, however, requires econometric techniques the development of which greatly exceeds the scope of my paper. Given the importance of this distinction, what I address (in section 5 and appendix C) is the construction of specific diagnostics for the restrictions implied by the HTR model.

¹⁰Compare the inverse gamma distribution in Chamberlain and Hirano (1999). The posterior mean of the parameters they estimate suggests a distribution for σ_i^2 with no well-defined mean and variance.

2.2 Preview of the empirical analysis

The HTR model is amenable to a fully-fledged empirical exercise. The task is to quantify the heterogeneity in transitory income risk, relating it to observed and other unobserved sources of heterogeneity. The exercise was motivated in section 1.

The culmination of the empirical exercise is a set of statements about

$$\theta = \left\{ \mathbb{P}_{(\eta, \sigma^2)|x'}, \mathbb{P}_{v_2|x'}, \dots, \mathbb{P}_{v_T|x'}, \mathbb{P}_{\varepsilon_1|x'}, \dots, \mathbb{P}_{\varepsilon_T|x} \right\},$$

formally, the unknown parameter of the HTR model. I describe some of the statements below.

The need for a normalization of σ_i^2 , which will become apparent in analyzing identification in section 3, means that care must be exercised in constructing measures of inequality in transitory income risk from the HTR model.

The cleanest approach is to treat observed and unobserved drivers of transitory income risk jointly. Consider the transitory income volatility of household i at time t ,

$$\tilde{\sigma}_{it} \equiv \sigma_i \sigma_{\varepsilon_i}(x_i) = \sqrt{\text{Var}(\tilde{\varepsilon}_{it} | x_i, \eta_i, \sigma_i^2)}.$$

This is a clear measure of the scale of transitory income of household i . It is the ideal measure if the distribution of the standardized innovation $\tilde{\varepsilon}_{it}/\tilde{\sigma}_{it}$ does not depend on x_i — it does not depend on σ_i . In that case, the probability that household i receives a transitory income shock of absolute value greater than $\kappa\tilde{\sigma}_{it}$ is exactly the same as the probability that household i' receives a transitory income shock of absolute value not less than $\kappa\tilde{\sigma}_{i't}$, for all $\kappa \geq 0$. Moreover, the distribution $\mathbb{P}_{\tilde{\sigma}_i}$ of $\tilde{\sigma}_{it}$ is completely determined by $\mathbb{P}_{\sigma^2|x'}$, $\mathbb{P}_{\varepsilon_i|x'}$ and $\mathbb{P}_{x'}$, and, therefore, by the parameter θ . So is the conditional distribution $\mathbb{P}_{\tilde{\sigma}_i|x}$.

The extent of heterogeneity in transitory risk. Any measure of the dispersion of the distributions $\mathbb{P}_{\tilde{\sigma}_i}$ or $\mathbb{P}_{\tilde{\sigma}_i|x}$ serves to quantify the cross-sectional inequality in transitory income risks. It is of particular interest to compare the median volatility with the average volatility, both conditional to given selected values of the covariates and unconditionally, and to measure the proportion of households with $\tilde{\sigma}_{it}$ above certain threshold.

Permanent income and transitory risk. Permitting dependence between η_i and σ_i^2 is appealing because the association between permanent income and transitory risk uncovers potentially important dimensions of inequality. The conditional distribution $\mathbb{P}_{\tilde{\sigma}_i|x, \eta'}$, which charac-

terizes the association, can be obtained from $\mathbb{P}_{(\eta, \sigma^2)|x}$ and $\mathbb{P}_{\varepsilon_t|x}$. Such objects as the conditional expectation function $\mathbb{E}[\tilde{\sigma}_{it}|x_i, \eta_i]$ and simpler linear projection coefficients, together with their mean squared errors, give suitable summaries of the relation between permanent income and transitory risk.

Secular trends. A number of empirical papers have documented the steady increase in the variance of transitory income shocks unfolding throughout the end of the twentieth century. Applying the HTR model to a sequence of panels around the period, one can display a more complete picture of the trend by measuring changes over time along the whole distribution of transitory risk, and not just in the average.

Consumption passthrough and labor market variables. It will be possible to extend the identification argument and estimation technique to permit regressions of σ_i^2 on labor market variables and of consumption changes on transitory income shocks $\tilde{\varepsilon}_{it}$ and ε_{it} .

3 Identification

The purpose of this section is to establish that $\mathbb{P}_{(\eta, \sigma^2)|x}$ is identified from $\mathbb{P}_{y_{1:T}|x}$ if a sufficient number of periods—it turns out to be $T = 5$ —are available. To simplify the exposition, I omit covariates (equivalently, I assume $x_i = 1$, a.s.); to allow for covariates in the argument below all that is needed is to replace expectations and distributions by conditional expectations and conditional distributions.

I begin with a few preliminary observations. First, without covariates, the deterministic income component reduces to $\mu_{1:T}$ and is identified by the mean of $y_{1:T}$: using (3) and (7), $\mathbb{E}[y_{i,1:T}] = \mu_{1:T} + \mathbb{E}[\eta_i] \mathbf{1}_{1 \times T} = \mu_{1:T}$. Second, a calculation like (8), using (3) and (4), gives

$$\text{Var}(y'_{i,1:T}) = \begin{bmatrix} \sigma_\eta^2 & \sigma_\eta^2 & \dots & \sigma_\eta^2 \\ \sigma_\eta^2 & \sigma_\eta^2 + \sigma_{v_2}^2 & \dots & \sigma_\eta^2 + \sigma_{v_2}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_\eta^2 & \sigma_\eta^2 + \sigma_{v_2}^2 & \dots & \sigma_\eta^2 + \sum_{s=2}^T \sigma_{v_s}^2 \end{bmatrix} + \text{diag}(\sigma_{\varepsilon_{1:T}}^2).$$

In consequence, $\sigma_\eta^2, \sigma_{v_2}^2, \dots, \sigma_{v_{T-1}}^2, \sigma_{\varepsilon_1}^2, \dots, \sigma_{\varepsilon_{T-1}}^2$, and $\sigma_{v_T}^2 + \sigma_{\varepsilon_T}^2$ are identified. Furthermore, defining $\sigma_{\eta_t}^2 \equiv \text{Var}(\tilde{\eta}_{it})$, it follows that $\sigma_{\eta_t}^2$ is identified for $t = 1, \dots, T - 1$.

Thus, with T income observations per unit it is possible to separate the contribution of

permanent and transitory components to the variance of income during period t for all periods except the last. That is the standard result in the canonical model: the presence of σ_i^2 is no obstacle to the identification of the variances of permanent and transitory components.

The role of the normalization. If, instead of assuming $\mathbb{E}[\sigma_i^2] = 1$ as I did, I were to assume $\sigma_{\varepsilon_t}^2 = 1$, the result would be the identification of $\mathbb{E}[\sigma_i^2]$ together with $\sigma_{\varepsilon_t}^2$ for all $t = 1, \dots, T-1$, except if $\tau = T$. The reason is that permanent and transitory income variances cannot be disentangled for the last period. For similar reasons, the normalization $T^{-1} \sum_{t=1}^T \sigma_{\varepsilon_t}^2 = 1$ would not give identification.

Beyond first and second moments. It is apparent that if the objective is to learn about the covariance matrix of (η_i, σ_i^2) , first and second moments of observed log earnings histories are not enough. Before turning to the identification of $\mathbb{P}_{(\eta, \sigma^2)}$, I address first the identification of $\text{Var}((\eta_i, \sigma_i^2))$ in the following subsection.

3.1 Identification of the second moments of (η_i, σ_i^2)

Note $\text{Var}(\eta_i)$ is identified with as few periods as $T = 2$. To get identification of $\text{Cov}(\eta_i, \sigma_i^2)$ and $\text{Var}(\sigma_i^2)$ more periods are needed.

Theorem 1. *In the HTR model, the following holds:*

(i) $\text{Cov}(\eta_i, \sigma_i^2)$ is identified provided $T \geq 4$ and

$$\text{Cov}(\eta_i, \sigma_i^2) = \frac{\text{Cov}(y_{it}, (\Delta y_{i\tau})^2)}{\text{Var}(\Delta \varepsilon_{i\tau})},$$

for all $\tau > t + 1$.

(ii) $\text{Var}(\sigma_i^2)$ is identified provided $T \geq 5$ and

$$\text{Var}(\sigma_i^2) = \frac{\text{Cov}(\Delta y_{it}^2, \Delta y_{i\tau}^2)}{\text{Var}(\Delta \varepsilon_{it}) \text{Var}(\Delta \varepsilon_{i\tau})},$$

for all $\tau > t + 1$.

The proof can be found in appendix A.

Linear regression. A consequence of theorem 1 is that the linear regression of σ_i^2 onto η_i ,

$$\mathbb{E}^* \left[\sigma_i^2 \mid \eta_i \right] = 1 + \frac{\text{Cov}(\eta_i, \sigma_i^2)}{\text{Var}(\eta_i)} \eta_i,$$

is identified with $T \geq 4$. The mean squared error of the linear regression,

$$\text{MSE} \left[\sigma_i^2 \mid \eta_i \right] = \text{Var}(\sigma_i^2) - \frac{\left\{ \text{Cov}(\eta_i, \sigma_i^2) \right\}^2}{\text{Var}(\eta_i)}$$

is identified with $T \geq 5$. There is more than one way to express the regression coefficient and the mean squared error as functionals of the distribution of earnings histories since the HTR model imposes overidentifying restrictions.

Stationarity assumptions. The requirement in theorem 1 that $T \geq 4$ for the identification of $\text{Cov}(\eta_i, \sigma_i^2)$ and $T \geq 5$ for the identification of $\text{Var}(\sigma_i^2)$ can be relaxed to $T \geq 3$ and $T \geq 4$ if one assumes stationarity of the variance of $\varepsilon_{i,1:T}$ or, more generally, any other known relation between $\sigma_{\varepsilon_T}^2$ and $\sigma_{\varepsilon_1}^2, \dots, \sigma_{\varepsilon_{T-1}}^2$.

3.2 Identification of the distribution of (η_i, σ_i^2)

I have shown that $T \geq 5$ suffices to identify $\text{Var}((\eta_i, \sigma_i^2))$. The requirement is necessary for the identification of $\text{Var}((\eta_i, \sigma_i^2))$ from the covariance matrix of $(y_{1:T}, y_{1:T}^2)$ if $\sigma_{\varepsilon_T}^2$ is unrestricted and so are the third and fourth moments of $v_{i,2:T}$ and $\varepsilon_{i,1:T}$. What I show next is that with $T \geq 5$ and under rather mild constraints on the parameter space for θ , the distributions

$$\left\{ \mathbb{P}_{(\eta, \sigma^2)}, \mathbb{P}_{v_2}, \dots, \mathbb{P}_{v_{T-1}}, \mathbb{P}_{\varepsilon_1}, \dots, \mathbb{P}_{\varepsilon_{T-1}}, \mathbb{P}_{v_T + \varepsilon_T} \right\}$$

are identified. As pointed out at the beginning of this section, conditional distributions given covariates are identified by an entirely analogous argument. Also note that the distributions of v_{iT} and ε_{iT} cannot be disentangled, just as their variances.

My identification argument builds on the analogy between the HTR model and econometric models with nonclassical measurement error (see Bound et al. (2001)). A large literature exists about the identification of such models. Most relevant to my analysis is the paper by Hu and Schennach (2008). In what follows, identification of $\mathbb{P}_{(\eta, \sigma^2)}$ will be established as an application of their theorem 1.

Hu and Schennach (2008) study a model in which there is a dependent variable y , a true regressor x^* , its error-ridden counterpart x , and an instrument z , and give general conditions under which $\mathbb{P}_{y|x^*}$, $\mathbb{P}_{x|x^*}$, and $\mathbb{P}_{x^*|z}$ are (a.s.) uniquely determined by $\mathbb{P}_{(x,y)|z}$. To describe the analogy, let me select t so that $1 \leq t-2 < t+2 \leq T$ (which asks for $T \geq 5$). Then, as will become clear soon, y_{it} will play the role of the dependent variable, $(\tilde{\eta}_{it}, \sigma_i^2)$ the role of the true regressor, $y_{i,(t-2):(t-1)}$ will be the noisy measurement, and $y_{i,(t+1):(t+2)}$ the instrument.

My approach first identifies $\mathbb{P}_{y_{(t-2):(t-1)}|(\tilde{\eta}_i, \sigma^2)}$ and $\mathbb{P}_{(\tilde{\eta}_i, \sigma^2)|y_{(t+1):(t+2)}}$. With this on hand I then establish identification of the rest of the distributions in the statement.

In addition to the restrictions already imposed by the HTR model I need an assumption of a slightly more technical nature.

Assumption 1. *The parameter θ satisfies the following:*

- (i) *The distributions $\mathbb{P}_{(\eta, \sigma^2)}$, $\mathbb{P}_{v_2}, \dots, \mathbb{P}_{v_T}$, $\mathbb{P}_{\varepsilon_1}, \dots, \mathbb{P}_{\varepsilon_T}$ all admit bounded densities with respect to the corresponding Lebesgue measures.*
- (ii) *For each $t = 2, \dots, T$, no nonzero real function f exists for which*

$$\mathbb{E} \left[f \left(\eta + \sum_{s=2}^{t-1} v_{is} + \sigma \varepsilon_{i,t-1}, v_{it} + \sigma \Delta \varepsilon_{it} \right) \right] = 0 \text{ for all } (\eta, \sigma^2) \in \text{supp} \left((\eta_i, \sigma_i^2) \right).$$

Moreover, the characteristic functions of $v_{i,2:T}$ and $\varepsilon_{i,1:T}$ are non-vanishing.

Assumption 1 has two parts. The first part imposes that latent variables possess bounded densities and, in particular, that they are (absolutely) continuous random variables. This is not very restrictive for η_i , $v_{i,2:T}$, and $\varepsilon_{i,1:T}$, but it does rule out some important classes of marginal distributions for σ_i^2 , e.g. the gamma distribution with shape parameter below unity and discrete distributions. The identification of a model in which a positive mass of households have zero transitory income is not covered by the results below but its identification can be established with more specific arguments.

The second part of assumption 1 ensures operator injectivity requirements demanded by theorem 1 in Hu and Schennach (2008): they play a role similar to that of the relevance condition in instrumental variables problems. The restrictions on characteristic functions can be substantially weakened along the lines of Evdokimov and White (2012).

Theorem 2. *Under assumption 1, the distributions*

$$\left\{ \mathbb{P}_{(\eta, \sigma^2)}, \mathbb{P}_{v_2}, \dots, \mathbb{P}_{v_{T-1}}, \mathbb{P}_{\varepsilon_1}, \dots, \mathbb{P}_{\varepsilon_{T-1}}, \mathbb{P}_{v_T + \varepsilon_T} \right\}$$

are identified provided $T \geq 5$.

The proof of theorem 2 can be found in appendix A.

Nonparametric identification with lots of heterogeneity. The instrumental variables perspective of the identification argument offers insights into the identification of several other models related to the HTR model. Consider replacing (3) by a permanent income process with a unit-specific permanent income variance ζ_i^2 ,

$$\tilde{\eta}_{it} = \tilde{\eta}_{i,t-1} + \zeta_i v_{it} = \eta_i + \zeta_i \sum_{s=2}^t v_{is},$$

subject to the normalization $\mathbb{E}[\zeta_i^2] = 1$ — or, in the presence of covariates, $\mathbb{E}[\zeta_i^2 | x_i] = 1$. Even allowing for dependence among η_i , ζ_i^2 , and σ_i^2 , one can show, under restrictions similar to 1, that $\mathbb{P}_{(\eta, \zeta^2, \sigma^2)}$ is nonparametrically identified provided $T \geq 7$. If, in addition to ζ_i^2 , a unit-specific persistence parameter ρ_i is included,

$$\tilde{\eta}_{it} = \rho_i \tilde{\eta}_{i,t-1} + \zeta_i v_{it} = \rho_i^{t-1} \eta_i + \zeta_i \sum_{s=2}^t \rho_i^{t-s} v_{is},$$

my analysis suggests that $T \geq 9$ is required to identify $\mathbb{P}_{(\eta, \rho, \zeta^2, \sigma^2)}$.¹¹ The underlying idea is to replace $y_{i,(t-2):(t-1)}$ by $y_{i,(t-3):(t-1)}$ and $y_{i,(t-4):(t-1)}$, $y_{i,(t+1):(t+2)}$ by $y_{i,(t+1):(t+3)}$ and $y_{i,(t+1):(t+4)}$, and to expand $(\tilde{\eta}_{it}, \sigma_i^2)$ to $(\tilde{\eta}_{it}, \zeta_i^2, \sigma_i^2)$ and $(\tilde{\eta}_{it}, \rho_i, \zeta_i^2, \sigma_i^2)$ in each case. Invoking theorem 1 in Hu and Schennach (2008) and imitating the steps in the proof to theorem 2 (appendix A), the results follow. A rule-of-thumb is that, in order to identify k unit-specific parameters, $T \geq 2k + 1$ is required to reconstruct the instrumental variables idea.

The insights of the preceding paragraph are relevant to interpret the approach to lots of heterogeneity proposed by Browning et al. (2010) and Alan et al. (2018). The income process of both papers allows for 8 unit-specific parameters with essentially unconstrained dependence among them. The rule-of-thumb above indicates that no less than 17 periods are required to nonparametrically identify their distribution. Such a long time series dimension raises some concerns of sample selectivity, as only the more stable and enduring households will provide genuine information about the dependence between household-specific parameters.

¹¹If it is identified at all. It is known that a one-error model with individual-specific intercept and autoregressive coefficient is only set identified (see, e.g., Lee (2019)).

4 Estimation

The nonparametric identification result is encouraging in that it reveals that few periods of earnings observations are needed to determine the joint distribution of permanent income and transitory income risk. It is natural to go on and formulate a flexible estimation approach to capture the potential nonlinearities present in the relation between latent variables. I introduce such an approach in this section.

My estimation approach is related to recent developments in the econometrics of latent variables models (in particular, Arellano and Bonhomme (2016), Arellano et al. (2017), and Arellano and Bonhomme (2019)) which exploit EM ideas combined with computer-simulation techniques. From a statistical point of view, my estimation approach is based on sieve ideas (see Chen (2007) for a complete treatment). I will construct a sequence of approximate models which, in an inference theory, would portray increasing flexibility as the sample size grows.

4.1 Flexible specification

My estimation method relies on some simplifications. Let me define

$$\begin{aligned} h_i &\equiv \eta_i / \sigma_\eta(x_i), \\ s_i &\equiv \ln(\sigma_i), \\ u_{it} &\equiv v_{it} / \sigma_{v_i}(x_i), \quad t = 2, \dots, T, \\ e_{it} &\equiv \varepsilon_{it} / \sigma_{\varepsilon_i}(x_i), \quad t = 1, \dots, T. \end{aligned}$$

This way, h_i , $u_{i,2:T}$, and $e_{i,1:T}$ all have zero mean and unit variance, and s_i maps σ_i to the whole real line. The first and most important simplification is the assumption

$$(h_i, s_i) \perp u_{i2} \perp \dots \perp u_{iT} \perp e_{i1} \perp \dots \perp e_{iT} \perp x_i.$$

In other words, the distributions of (η_i, σ_i^2) , $v_{i,1:T}$, and $\varepsilon_{i,1:T}$ do not depend on covariates x_i beyond second moments. The purpose of the simplification is to limit in a reasonable manner the interaction between observable and unobservable sources of heterogeneity, interaction that would demand a large number of parameters to be captured. Note that the interesting feature of dependence between η_i and σ_i^2 is still permitted.

The second simplification is the stationarity of standardized permanent and transitory income shocks, i.e., $\mathbb{P}_{u_t} = \mathbb{P}_u$ ($t = 2, \dots, T$) and $\mathbb{P}_{e_t} = \mathbb{P}_e$ ($t = 1, \dots, T$) for a pair of probability

distributions \mathbb{P}_u and \mathbb{P}_e (on the Borel sets of \mathbb{R}). The assumption is similar in spirit to the first simplification. Also notice that nonstationarity is allowed in first and second moments.

The parameter vector is now reduced to

$$\theta = \left\{ \mu_1, \dots, \mu_T, \sigma_\eta^2, \sigma_{v_2}^2, \dots, \sigma_{v_T}^2, \sigma_{\varepsilon_1}^2, \dots, \sigma_{\varepsilon_T}^2, \mathbb{P}_{(h,s)}, \mathbb{P}_u, \mathbb{P}_e \right\}.$$

Observable heterogeneity. I model the mean and volatility functions as

$$\begin{aligned} \mu_t(x_i) &= \varphi'_{\mu_t}(x_i)\beta_\mu, \\ \ln(\sigma_\eta(x_i)) &= \varphi'_\eta(x_i)\beta_\eta, \\ \ln(\sigma_{v_t}(x_i)) &= \varphi'_{v_t}(x_i)\beta_{v_t}, \\ \ln(\sigma_{\varepsilon_t}(x_i)) &= \varphi'_{\varepsilon_t}(x_i)\beta_{\varepsilon_t}, \end{aligned}$$

where φ_{μ_t} is a known function that maps x_i to the relevant covariates for the mean of time- t log income and similar considerations apply to φ_η , φ_{v_t} ($t = 2, \dots, T$), and φ_{ε_t} ($t = 1, \dots, T$). The only unknowns are collected into $\beta = \{\beta_\mu, \beta_\eta, \beta_{v_t}, \beta_{\varepsilon_t}\}$. Flexibility of the model is controlled by the dimension of the functions mapping covariates to regressors.

Unobserved heterogeneity. As for the latent variables distributions, I write

$$\begin{aligned} h_i &= Q_\eta(H_i|\gamma), \\ s_i &= Q_\sigma(S_i|h_i, \gamma), \\ u_{it} &= Q_v(U_{it}|\gamma), \quad t = 2, \dots, T, \\ e_{it} &= Q_\varepsilon(E_{it}|\gamma). \quad t = 1, \dots, T, \end{aligned}$$

where $Q_\eta(\cdot|\gamma)$, $Q_\sigma(\cdot|h, \gamma)$, $Q_v(\cdot|\gamma)$, and $Q_\varepsilon(\cdot|\gamma)$ are strictly increasing for almost every h and γ , and known up to γ . The random variables H_i , S_i , $U_{i,2:T}$, and $E_{i,1:T}$ are uniformly distributed on the interval $(0, 1]$, and

$$H_i \perp S_i \perp U_{i2} \perp \dots \perp U_{iT} \perp E_{i1} \perp \dots \perp E_{iT} (\perp x_i).$$

The way γ determines the quantile functions Q_η , Q_v , and Q_ε is by piecewise-linear splines (with exponential interpolation in the tails). Thus, γ contains the quantiles of h_i , $u_{i,2:T}$, and $e_{i,1:T}$ at a grid of selected probabilities (and a pair of parameters for the tails).

I specify the conditional distribution of s_i given h_i as

$$Q_\sigma(S|h, \gamma) = \varphi'_\sigma(h)\beta_\sigma(S, \gamma),$$

with φ_σ a known vector-valued function (orthogonal polynomials of different degrees) and β_σ another vector-valued function (piecewise-linear splines with exponential interpolation in the tails too). Thus, γ also contains a matrix of quantile coefficient vectors at a grid of selected probabilities (together with a pair of tail parameters). All details can be found in appendix B.

Flexibility of the model is controlled by the number of gridpoints in the piecewise-linear spline approximation and by the dimension of the function φ_σ . My approach to modeling distributions through quantile functions has its antecedents in Arellano and Bonhomme (2016) and Arellano et al. (2017).

4.2 Estimation algorithm

In the specification, β and γ fully determine θ and, therefore, all the probability distributions of observables and latent variables given covariates. The task is to construct good estimates $\hat{\beta}$ and $\hat{\gamma}$. The method I recommend is divided in two steps, first producing $\hat{\beta}$ and then $\hat{\gamma}$ for given $\hat{\beta}$. Although in parametric and semiparametric setups efficiency considerations suggest joint estimation of β and γ , the prescription does not generalize easily to flexible nonparametric models. Sequential estimation of β and γ , in contrast, facilitates thinking about sequences of increasingly flexible specifications.

Step 1: Estimating β (netting out). The standard in the empirical study of earnings uncertainty is to fit a model of latent variables to residuals from a regression of log earnings on covariates. The construction of such residuals is known as netting out. In the HTR model, β is identified from conditional moment restrictions that do not involve γ , thereby enabling the netting out step. Moreover, given my emphasis on understanding heterogeneity in transitory income risk, it is natural to let the netting out mean, not just the construction of residuals, but the determination of the role of covariates. In this step, I estimate β_μ by least squares and β_η , β_v , and β_ϵ by nonlinear least squares. The properties of this estimator and an instrumental variables estimator (optimal in a semiparametric sense under correct specification of volatility functions) are described in appendix B.

Step 2: Estimating γ (stochastic EM). It would be conceptually possible to estimate γ by maximizing the likelihood function of log earnings data conditional on covariates. Evaluation of the likelihood function is, however, almost intractable in the flexible specification I propose. For given γ (and $\beta = \hat{\beta}$), simulation of latent variables from their conditional distribution given data is, in fact, tractable and, therefore, resorting to a simulation-based deconvolution approach is attractive. The implementation of this idea is a stochastic EM algorithm.

To describe the algorithm, let Ψ be a function of latent variables $h_i, s_i, u_{i,2:T}, e_{i,1:T}$ and γ such that γ is identified by the unfeasible moment equations

$$0_{\dim(\Psi) \times 1} = \mathbb{E} [\Psi (h_i, s_i, u_{i,2:T}, e_{i,1:T}, \gamma)].$$

Algorithm 1. Initialize $\gamma = \gamma^{(0)}$ and, for $i_{\text{iter}} = 1, \dots, n_{\text{iter}}$, alternate between the following:

(i) For each $i = 1, \dots, n$, draw a sequence of latent variables,

$$\left\{ h_i^{(i_{\text{sim}})}, s_i^{(i_{\text{sim}})}, u_{i,2:T}^{(i_{\text{sim}})}, e_{i,1:T}^{(i_{\text{sim}})} \right\}_{i_{\text{sim}}=1}^{n_{\text{sim}}},$$

from the conditional distribution given $y_{i,1:T}$ implied by the values $(\beta, \gamma) = (\hat{\beta}, \gamma^{(i_{\text{iter}}-1)})$;

(ii) Update γ to $\gamma^{(i_{\text{iter}})}$ by solving the sample analog to the moment equations,

$$0_{\dim(\Psi) \times 1} = \mathbb{E}_n \left[\mathbb{E}_{n_{\text{sim}}} \left[\Psi \left(h_i^{(i_{\text{sim}})}, s_i^{(i_{\text{sim}})}, u_{i,2:T}^{(i_{\text{sim}})}, e_{i,1:T}^{(i_{\text{sim}})}, \gamma^{(i_{\text{iter}})} \right) \middle| y_{i,1:T} \right] \right].$$

Detailed discussions of the two steps and some of the relevant inference theory are confined to appendix B. Notice that the output of algorithm 1 is a sequence $\left\{ \gamma^{(i_{\text{iter}})} \right\}_{i_{\text{iter}}=1}^{n_{\text{iter}}}$. To form an estimate of γ , I take $\hat{\gamma}$ to be the average of the last terms in the sequence.

Remarks about the choice of estimation method. The inference theory for the class of non-parametric latent variables models I am concerned with is, to a large extent, unexplored territory. In particular, no basis is available to derive estimators by appealing to concrete principles of optimality (the prescriptions of which are likely to depend on the final object of interest). The spirit of the literature, which I ascribe, is to provide a framework: a rule to map the data into a collection of estimates for the latent variables distributions. Those estimates are later used as the input in direct calculations of final objects of interest. In consequence, the objective of an estimation framework is to attain reasonable statistical accuracy.

As an alternative to optimality criteria, another principle consists of favoring the method that mimics more directly the identification technique. This is one way to motivate deconvolution by characteristic-function methods (as in Horowitz and Markatou (1996), Bonhomme and Robin (2010), Botosaru and Sasaki (2018), etc.). In the context of the HTR model, doing so is not very attractive for two reasons: first, because the identification technique does not suggest a simple implementation that would use all the restrictions of the model; second, because of the probable presence of inverse ill-posed issues.

5 Empirical analysis

I am now in a position to address some of the empirical questions raised in the introduction (section 1) about the nature of transitory risk and its link to permanent income. The first item in the order of business is to decide what measurements in the data are to be treated as y_{it} ; in other words, what is the concept of income to which the HTR model speaks.

Panel data. The Panel Study of Income Dynamics (PSID) plays a major role in the literature on income processes. Its long time span, its design aimed at preserving representativity across the years, and the large number of variables which are measured jointly with income and employment turn it into one of the finest data resources for the purposes of my study.¹²

I analyze two different types of datasets. The first is a collection of panels of households: they are formed by family units in which the representative person is a married male and income is interpreted as pre-tax labor earnings of the representative person and the spouse plus transfers. The second is a panel of workers: it is formed by any family unit in which the representative person earns labor income. In this second I focus on wages as opposed to total labor income. Both labor income and wages are of interest since either one or the other is taken to be the exogenous component of uncertainty behind household decision in quantitative macroeconomic model and empirical analyses (cf. Blundell et al. (2008) and Blundell et al. (2016)). Both offer complementary insights into the nature of uncertainty and insurance.

Most of the results I report below are based on a six-wave panel constructed with data gathered between the 1999 and the 2009 interviews. In that period the survey was conducted once every two years. I do so for a variety of reasons: first, because the most recent income data are constructed using more stable definitions of income variables and more standard proce-

¹²This is despite recent concerns about the representativeness of the PSID in the recent waves, as measured by the comparison between changes in the aggregate of PSID households and administrative data. See Bloom et al. (2019).

dures to handle missing and inadequate responses; second, because the biennial framework makes the assumption of serially uncorrelated transitory income shocks more plausible; third, because the same period was considered in a number of papers in the literature (e.g., Blundell et al. (2016) and Arellano et al. (2017)) and the comparison is of interest; finally, because for the period, other labor market, consumption, and asset variables are available.

The complete set of tables and figures can be found in appendixes D (for household income) and E (for worker wages). Below I emphasize the interpretation of a selection of the results. I replicate some tables and figures for convenience. With the exception of one aside, I focus on the panel of households and leave the parallel results for the panel of workers to appendix.

Measurement error. It is sometimes argued that measurement error is an important dimension of the PSID data. This has been discussed in a large number of papers (among them, Gottschalk and Moffitt (2009)). For the purpose of this paper, there are many things to bear in mind. First, one could calibrate the size of measurement error using available estimates from validation data. Incorporating this into the estimation procedure would not be difficult.

Second, while there are good reasons to expect a household-specific scale of transitory income changes, the same is not so clear for measurement error. Theorem 1 tells us that what informs about the extent of heterogeneity in transitory risks and its relationship with permanent income is the persistence in higher-order moments of the distribution of earnings scaled down by an estimate of the variance of transitory income. If this variance is inflated (as it would happen with classical measurement error) the results that follow from neglecting measurement error are, in fact, a lower bound on the importance of heterogeneity.

Overall assessment of model restrictions. Here I report the covariance matrix of (net) log earnings and the higher-order moments which are informative about (η_i, σ_i^2) according to theorem 1. A first glance at table 1 seems favorable to the HTR model assumption of a random-walk permanent component and white-noise transitory component. A quick back-of-the-envelope calculation gives $\text{Var}(\eta_i) \approx 0.15$, $\text{Var}(v_{it}) \approx 0.01$, and $\text{Var}(\varepsilon_{it}) \approx 0.06$, standard figures for this model and data.

Table 2 is informative about $\text{Cov}(\eta_i, \sigma_i^2)$. Except for two entries, the table suggests negative correlation between η_i and σ_i^2 . Moreover, the entries that are not positive tend to agree on a value around $\text{Cov}(\eta_i, \sigma_i^2) \approx -0.10$. The disagreement between figures could be attributed to sampling error. This I will address during the construction of diagnostics in section C.

Finally, table 3 informs about $\text{Var}(\sigma_i^2)$. It suggests considerable dispersion of σ_i^2 and a

value around $\text{Var}(\sigma_i^2) \approx 0.65$. As with table 2, it is of interest to tell whether the differences among figures are owed to sampling error, as I do in appendix C. The differences, however, look remarkably small given the fact that they are based on higher-order moments which are sensitive to outliers.

TABLE 1. Covariance matrix of $y_{i,1:T}$

Row	t1999	t2001	t2003	t2005	t2007	t2009
t1999	0.237	0.168	0.143	0.132	0.128	0.121
t2001	0.168	0.228	0.157	0.143	0.142	0.128
t2003	0.143	0.157	0.24	0.148	0.148	0.137
t2005	0.132	0.143	0.148	0.228	0.176	0.158
t2007	0.128	0.142	0.148	0.176	0.245	0.19
t2009	0.121	0.128	0.137	0.158	0.19	0.254

TABLE 2. Covariances between y_{it} and $(\Delta y_{i\tau})^2 / \text{Var}(\Delta \varepsilon_{i\tau})$

Row	t2003	t2005	t2007
t1999	0.023	-0.078	-0.101
t2001		-0.093	0.003
t2003			-0.086

TABLE 3. Covariances between $(\Delta y_{it})^2 / \text{Var}(\Delta \varepsilon_{it})$ and $(\Delta y_{i\tau})^2 / \text{Var}(\Delta \varepsilon_{i\tau})$

Row	t2003	t2005	t2007
t2001		0.629	0.664
t2003			0.846

From looking at tables 1, 2, and 3, one also gets the impression that the association between η_i and σ_i^2 is negative (the regression coefficient is about -0.67) but there is considerable variation in σ_i^2 beyond what is predicted by a linear function of η_i (the mean squared error is about 90% of the variance of σ_i^2).

5.1 Heterogeneity in transitory risks

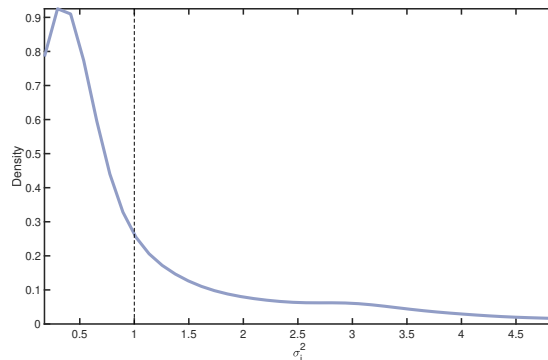
Part of the heterogeneity in transitory income variances is a function of covariates. Given that these functions are hard to estimate (their precisions depend on higher-order moments) and in view of sample size I have allowed for a mild conditioning on covariates. In particular, I have limited the conditional variance function to depend on age and education categories. The results for households are displayed in table 4 and indicate that transitory income risk decreases (slightly) with age and with education.

TABLE 4. Estimates of $100 \times \beta_\varepsilon$

Row	LSestimate	LSstderr	IVestimate	IVstderr
constant	-119.75	13.87	-96.93	15.33
representative person age	-0.38	0.52	-1.17	0.43
representative person is high-school graduate	-2.56	14.68	-26.33	15.31
representative person went to college	6.41	13.1	-19.58	13.86

It is of interest to get a sense of how much more heterogeneity there is in transitory income risk beyond what is captured by age and education because in the calibration of quantitative macroeconomic models there is typically a narrow amount of observable heterogeneity. One indicator of the importance of such heterogeneity is the variance estimates that can be read from table 3. However, this is not a very suitable indicator since transitory income volatilities can only take on positive values and there is very little basis to form a prior on the shape of its distribution. To assess this, I first display in figure 2 the distribution of σ_i^2 estimated by the nonparametric method introduced in section 4.

FIGURE 2. PDF of σ_i^2



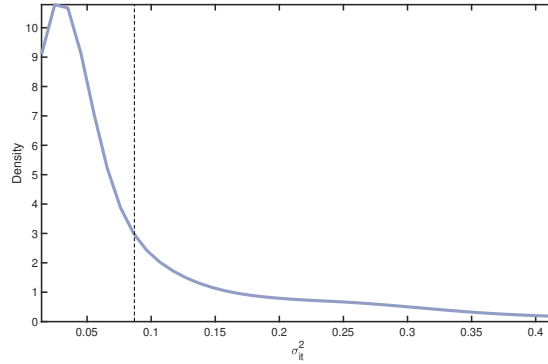
To report the combined effect of observable and unobserved heterogeneity I construct the

quantity

$$\tilde{\sigma}_i^2 \equiv \frac{1}{T} \sum_{t=1}^T \sigma_{\varepsilon_t}^2(x_i) \sigma_i^2.$$

Its distribution is displayed in figure 3.

FIGURE 3. PDF of $\tilde{\sigma}_i^2$



What is to be gathered from figures 2 and 3 is that most of the differences that I estimate in transitory income risk are driven by the latent variable σ_i^2 . Heterogeneity after controlling for covariates is paramount.

Furthermore, both the distributions of σ_i^2 and $\tilde{\sigma}_i^2$ display a large amount of asymmetry. In both figures, the mean has been indicated with a dotted line. The mode and the median are always well below the mean: a majority of households have small transitory income variations while it is a minority who faces constantly large income risks.

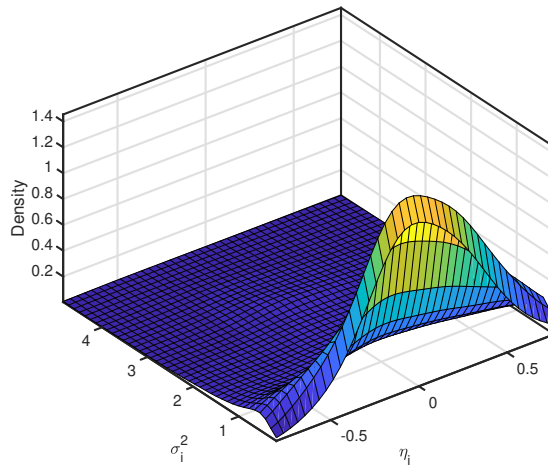
Connection to other evidence. The current exercise is attractive because it identifies transitory changes with assumptions about its second-order persistence. Gottschalk and Moffitt (2009) report the results from an alternative approach: they regard as transitory income changes any difference in the income of a household with respect to the average of income observations over the whole period. Translating that approach to my framework implies that the measure used in that study mixes both permanent and transitory income changes. Gottschalk and Moffitt (2009) find evidence of heterogeneity in the variance of transitory income according to their measure, just as I do. But they find a positive effect of education. In my estimates, more educated households have lower transitory income variances but higher permanent income variances (perhaps capturing a more general model of trend) than less educated households. The mixture of permanent and transitory income changes is an explanation for the differences

in the estimated effect of education.

5.2 Permanent income and transitory risks

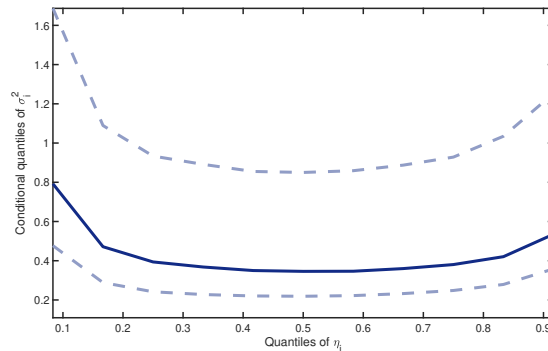
The second question one can address with estimates of the HTR model is how are permanent income and transitory income risk related. The calculations made above suggested a negative relationship but with substantial scope for variation in transitory risks for each level of permanent income. Figure 4 displays the joint density of (η_i, σ_i^2) .

FIGURE 4. Joint PDF of η_i and σ_i^2



The negative association predicted is only slightly noticeable from the figure. To go further in characterizing the relationship and its strength, I report, in figure 5, a selection of conditional quantiles of σ_i^2 for each level of η_i .

FIGURE 5. Conditional quantile function of σ_i^2 given η_i



Both the negative relationship and the fact that it is not overwhelmingly strong appear as evident. The novelty is that the relationship is nonlinear and it might turn around at high

values of η_i . This shows how a nonparametric approach can be very revealing in this context where both the shape of the marginal distribution of σ_i^2 and its conditional distribution given η_i are unknown and there is little basis to form a prior on them.

The nature of the negative relationship. Determining the causes behind the negative relationship between permanent income and transitory risk is well beyond the scope of this paper. However, one must notice that this is compatible with a number of models in which workers possess firm-specific capital. In turn, heterogeneity must be large at the beginning of the working life of individuals: if all workers were *ex ante* identical and jobs were a once-and-for-all choice, larger permanent income should compensate larger transitory risks as in mean-variance frontier models.

Age versus calendar time. With the random-walk specification, the link between permanent income and transitory risk vanishes as time passes owing to the accumulation of permanent income shocks. In the panel I constructed, this has happened to some older units and it would be of interest to quantify the implied relationship at the beginning of the working life. In the appendix D I report a figure that gives some hint towards the view that the relationship is stronger: I analyze the joint distribution of $\tilde{\eta}_i$ and $\tilde{\sigma}_i^2$ where I have defined

$$\tilde{\eta}_i \equiv \sum_{t=1}^T \mu_t(x_i) + \eta_i.$$

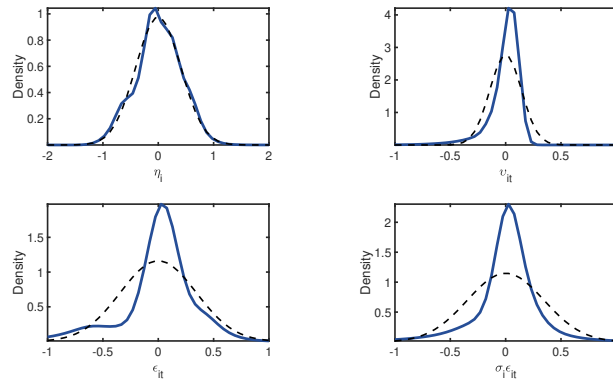
Because in some way, age is included in x_i , the correlation between $\tilde{\eta}_i$ and $\tilde{\sigma}_i^2$ can be thought of as conditioning on age. One could certainly reconstruct the implied correlation at the beginning of the working life from these quantities.

5.3 Marginal distributions of shocks

A recurrent finding in nonparametric deconvolution approaches applied to earnings dynamics is that, while the permanent component is approximately normally distributed, shocks to transitory income are highly nonnormal. This is the case, e.g., in Arellano et al. (2017). Because the HTR model implies that transitory income shocks are in fact a mixture, the question arises whether part of the nonnormality is just a sign of a household-specific variance. To investigate that issue, I show figure 6.

Comparing the marginal distributions of ε_{it} and $\tilde{\varepsilon}_{it}$ one observes that part of (but not all)

FIGURE 6. PDFs of η_i , v_{it} , ε_{it} , and $\sigma_i \varepsilon_{it}$



the nonnormality can be attributed to σ_i : there is some genuine skewness and excess kurtosis in ε_{it} but the departure from normality is not large (see appendix D).

5.4 Trends in the distribution of transitory risk

Many papers have documented a rapid rise in the variance of transitory income shocks during the 70s and 80s. An overview of the evidence is given by Gottschalk and Moffitt (2009). It is natural in this context to apply the HTR model to a sequence of panels around the period, to measure changes along the whole distribution of transitory risks and not just in the average. The PSID provides the resources to do so. I construct a sequence of six-wave overlapping panels starting from 1972-1977 up to 1988-1993. The results are displayed in figures 7 and 8.

FIGURE 7. Evolution of the marginal distribution of σ_i^2

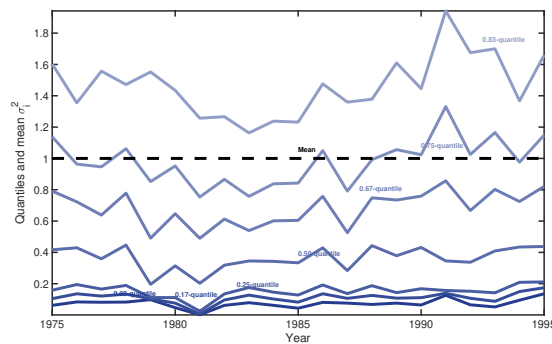


Figure 7, reports the evolution of the quantiles of σ_i^2 across time. Each panel is associated with the middle year (for comparability with Gottschalk and Moffitt (2009)). Here σ_i^2 is the heterogeneity in transitory income risk controlling for covariates that may influence the variance of transitory income. It is normalized to have unit mean in every period (so that I am “controlling” for a time effect too).

FIGURE 8. Evolution of the marginal distribution of $\tilde{\sigma}_i^2$

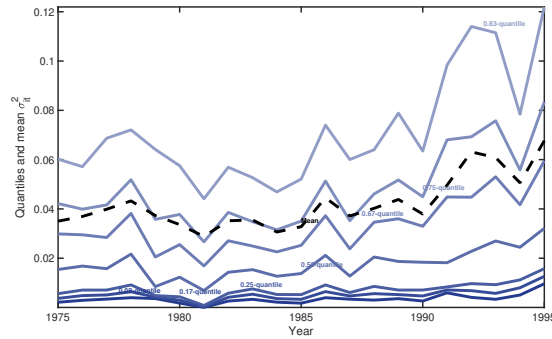


Figure 8, in turn, reports the evolution of the quantiles of $\tilde{\sigma}_i^2$ across time, that is, bringing in the effect of covariates. Comparing figures 7 and 8 one sees that an upward trend becomes evident when covariates are allowed for. This agrees with the idea, pushed in the literature, that the premium to education played a role in broadening differences across workers. The distribution of σ_i^2 seems qualitatively stable and, if anything, upper quantiles tend to display a slightly positive slope over time compared to lower quantiles which remain constant or even decrease.

5.5 Transitory income risk and labor market variables

As the evidence of heterogeneity in transitory income risk is strong, a task will turn out to be to find explanations for it. Because I have added some covariates as controls, tentative explanations must rely on other dimensions of workers and jobs to account for the large differences I uncover.

I do not attempt to undertake such a formidable task here. Instead, one thing that can be done withing the framework of this paper is to measure the predictive power of certain outside covariates on σ_i^2 . To do so, I focus on the panel of workers and use extensive information on labor market variables collected by the PSID. This includes indicators of unemployment and self-employment (which may be taken to proxy a more ideal indicator of labor-market attachment), whether the worker has a contract under a union, whether the worker is salaried or paid by the hour, occupation, and sector. The results are given in table 5.

TABLE 5. Regression of $\ln(\sigma_i)$ on labor market variables

Attachment to the labor market appears to contribute substantially to transitory income

risk, much more than differences across occupations and sectors.

5.6 Passthrough of income shocks to consumption

The HTR model shocks allow me to estimate passthrough coefficients that describe the transmission of income shocks to consumption. The PSID collects extensive data on household spending since 1999, thus providing the means to carry out this exercise.

It is of particular interest to compare the coefficients obtained by regressing consumption growth on two different notions of transitory shock, namely: $\tilde{\varepsilon}_{it}$ and ε_{it} . The first is a notion of transitory income shock that disregards the presence of heterogeneity in transitory income risk. The second is a notion of transitory shock that better describes the uncertainty faced by the household. To be specific, my objective is to estimate ϕ_{tran} in the regression

$$\Delta \tilde{c}_{it} = \phi_{\text{perm}} \cdot \text{Perm. shock}_{it} + \phi_{\text{tran}} \cdot \text{Tran. shock}_{it} + \text{error}_{it}.$$

Here, \tilde{c}_{it} denotes log consumption of household i during time t net of a conditional mean function of covariates x_i .

There has been some debate about the discrepancy usually observed between estimates of ϕ_{tran} obtained from quasi-experimental evidence and those from semi-structural approaches (see, e.g., Commault (2018)).

TABLE 6. Regression of $\Delta \tilde{c}_{it}$ on permanent and transitory shocks

	HeterogeneityIn	HeterogeneityOut
permanent shock	0.3513	0.3427
transitory shock	-0.0016	0.1829

Table 6 reports estimates of ϕ_{perm} and ϕ_{tran} . It is seen that the estimate that relies on $\tilde{\varepsilon}_{it}$ is negligible (and even negative) while the estimate which uses ε_{it} indicates some response of consumption to a transitory income shock. One interpretation for the difference in estimates might lie in the correlation between $\Delta \tilde{c}_{it}$ and $\tilde{\varepsilon}_{it}$ induced by the presence of σ_i .

6 Conclusion

This paper has developed a new framework to separate permanent and transitory components of variation in income in a way that permits flexible forms of dependence between them. With a modeling perspective, the central idea is to introduce dependence through time-invariant heterogeneity. I then give a tractable approach to nonparametric identification and estimation of the joint distribution of latent variables. The approach seems attractive in a wide range of applications other than income processes, such as in recent models of firm productivity.

With a substantive perspective, the large differences I find in transitory income variances across households and workers are relevant to quantitative macroeconomic exercises. Correct calibration of earnings uncertainty as part of a larger macro model is determinant for the *ex ante* evaluation of policies. Researchers know this and, in practice, allow for certain forms of observable heterogeneity (by, e.g., grouping by education levels). The evidence I present suggests a substantial extent of heterogeneity left after controlling for covariates while my estimates provide an alternative calibration that accounts for that.

With a methodological perspective, the nonparametric techniques I introduce are useful to uncover the nonlinear relationship between permanent income and transitory income risk. Simulation evidence indicates good performance in realistic sample designs. Yet the inference theory is incomplete. An asymptotic theory in which the flexibility of the specification of covariate functions and latent variables distributions are indexed to the sample size would be useful to understand the trade-offs implied by such choices. Alternative ways to construct estimates and tests when higher-order moments (which are sensitive to outliers) are involved appear of interest too. These are natural steps in a future research agenda.

References

- ABOWD, J. M. AND D. CARD (1989): "On the covariance structure of earnings and hours changes." *Econometrica*, 57, 411–445.
- ALAN, S., M. BROWNING, AND M. EJRNÆS (2018): "Income and consumption: A micro semi-structural analysis with pervasive heterogeneity." *Journal of Political Economy*, 126, 1827–1864.
- ALTONJI, J. G., A. A. SMITH, AND I. VINDANGOS (2013): "Modeling earnings dynamics," *Econometrica*, 81, 1395–1454.
- ARELLANO, M. (2003): *Panel Data Econometrics*, Oxford University Press.

- ARELLANO, M., R. BLUNDELL, AND S. BONHOMME (2017): "Earnings and consumption dynamics: A nonlinear panel data framework." *Econometrica*, 95, 693–734.
- ARELLANO, M. AND S. BONHOMME (2012): "Identifying distributional characteristics in random coefficients panel data models." *Review of Economic Studies*, 79, 987–1020.
- (2016): "Nonlinear panel data estimation via quantile regressions." *Econometrics Journal*, 19, C61–C94.
- (2019): "Recovering latent variables by matching." Working paper.
- BLOOM, N., F. GUVENEN, L. PISTAFERRI, J. SABELHAUS, S. SALGADO, AND J. SONG (2019): "The great micro moderation." Working paper.
- BLUNDELL, R., L. PISTAFERRI, AND I. PRESTON (2008): "Consumption inequality and partial insurance." *American Economic Review*, 98, 1887–1921.
- BLUNDELL, R., L. PISTAFERRI, AND I. SAPORTA-EKSTEN (2016): "Consumption inequality and family labor supply." *American Economic Review*, 106, 387–435.
- BLUNDELL, R. AND I. PRESTON (1998): "Consumption inequality and income uncertainty." *Quarterly Journal of Economics*, 113, 603–640.
- BONHOMME, S. AND J. M. ROBIN (2010): "Generalized non-parametric deconvolution with an application to earnings dynamics." *Review of Economic Studies*, 77, 491–533.
- BOTOSARU, I. (2017): "Identifying distributions in a panel model with heteroskedasticity: An application to earnings volatility." Simon Fraser University Working Paper 17-11.
- BOTOSARU, I. AND Y. SASAKI (2018): "Nonparametric heteroskedasticity in persistent panel processes: an application to earnings dynamics." *Journal of Econometrics*, 203, 283–296.
- BOUND, J., C. BROWN, AND N. MATHIOWETZ (2001): "Measurement error in survey data." in *Handbook of Econometrics*, ed. by J. J. Heckman and E. Leamer, North-Holland, vol. 5, chap. 59.
- BROWNING, M., M. EJRNÆS, AND J. ÁLVAREZ (2010): "Modelling income processes with lots of heterogeneity." *Review of Economic Studies*, 77, 1353–1381.
- CHAMBERLAIN, G. (1987): "Asymptotic efficiency in estimation with conditional moment restrictions." *Journal of Econometrics*, 34, 305–334.

- CHAMBERLAIN, G. AND K. HIRANO (1999): "Predictive distributions based on longitudinal earnings data." *Annales d'Économie et de Statistique*, 55/56, 211–242.
- CHEN, X. (2007): "Large sample sieve estimation of semi-nonparametric models." in *Handbook of Econometrics*, ed. by J. J. Heckman and E. Leamer, North-Holland, vol. 6B, chap. 76.
- COCHRANE, J. H. (1991): "A simple test of consumption insurance." *Journal of Political Economy*, 99, 957–976.
- COMMAULT, J. (2018): "How does consumption respond to a transitory income shock? Reconciling natural experiments and semi-structural estimations." Working paper.
- DEATON, A. AND C. PAXSON (1994): "Intertemporal choice and inequality." *Journal of Political Economy*, 102, 437–467.
- EVDOKIMOV, K. AND H. WHITE (2012): "Some extensions of a lemma of Kotlarski." *Econometric Theory*, 28, 925–932.
- FARBER, H. S. (1999): "Mobility and stability: The dynamics of job change in labor markets." in *Handbook of Labor Economics*, ed. by O. Ashenfelter and D. Card, North-Holland, vol. 3, chap. 37.
- GEWEKE, J. (2004): "Getting it right: Joint distribution tests of conditional simulators." *Journal of the American Statistical Association*, 99, 799–804.
- GOTTSCHALK, P. AND R. A. MOFFITT (2009): "The rising instability of U.S. earnings." *Journal of Economic Perspectives*, 23, 3–24.
- GUVENEN, F., S. OZKAN, AND J. SONG (2014): "The nature of countercyclical income risk." *Journal of Political Economy*, 122, 621–660.
- HALL, R. E. AND F. S. MISHKIN (1982): "The sensitivity of consumption to transitory income: Estimates from panel data on households." *Econometrica*, 50, 461–481.
- HERITIER, S. AND E. RONCHETTI (1994): "Robust bounded-influence tests in general parametric models." *Journal of the American Statistical Association*, 89, 897–904.
- HOROWITZ, J. L. AND M. MARKATOU (1996): "Semiparametric estimation of regression models for panel data." *Review of Economic Studies*, 63, 145–168.

- HOSPIDO, L. (2012): "Modelling heterogeneity and dynamics in the volatility of individual wages." *Journal of Applied Econometrics*, 27, 386–414.
- HU, Y. (2019): *The Econometrics of Unobservables*, Unpublished.
- HU, Y. AND S. M. SCHENNACH (2008): "Instrumental variable treatment of nonclassical measurement error models." *Econometrica*, 76, 195–216.
- KAPLAN, G. AND G. L. VIOLANTE (2010): "How much consumption insurance beyond self-insurance?" *American Economic Journal: Macroeconomics*, 2, 53–87.
- KOTLARSKI, I. (1967): "On characterizing the gamma and the normal distribution." *Pacific Journal of Mathematics*, 20, 69–76.
- LEE, W. (2019): "Identification and estimation of dynamic random coefficient models." Working paper.
- LILLARD, L. A. AND R. J. WILLIS (1978): "Dynamics aspects of earning mobility." *Econometrica*, 46, 985–1012.
- MACURDY, T. E. (1982): "The use of time series processes to model the error structure of earnings in longitudinal data analysis." *Journal of Econometrics*, 18, 83–114.
- MALCOMSON, J. H. (1999): "Individual employment contracts." in *Handbook of Labor Economics*, ed. by O. Ashenfelter and D. Card, North-Holland, vol. 3, chap. 35.
- MEGHIR, C. AND L. PISTAFERRI (2004): "Income variance dynamics and heterogeneity." *Econometrica*, 72, 1–32.
- NEAL, R. M. (2003): "Slice sampling." *Annals of Statistics*, 31, 705–741.
- SOULELES, N. S. (1999): "The response of household consumption to income tax refunds." *American Economic Review*, 89, 947–958.
- SZÉKELY, J. C. AND C. R. RAO (2000): "Identifiability of distributions of independent random variables by linear combinations and moments." *Sankhyā*, 62, 193–202.
- WILHELM, D. (2015): "Identification and estimation of nonparametric panel data regressions with measurement error." CeMMAP Working Paper 34/15.

A Proofs and derivations

A.1 Stylized example

Consider a two-period version of the HTR model:

$$\begin{aligned} y_{i1} &= \eta_i + \sigma_i \varepsilon_{i1}, \\ y_{i2} &= \eta_i + v_{i2} + \sigma_i \varepsilon_{i2}. \end{aligned}$$

with η_i , v_{i2} , ε_{i1} , and ε_{i2} independent and identically distributed as $N(0, 1)$. They are assumed independent of the individual-specific variance σ_i^2 , which has a discrete distribution,

$$\pi = \mathbb{P}[\sigma_i = \sigma_H] = 1 - \mathbb{P}[\sigma_i = \sigma_L].$$

Here $\sigma_H > \sigma_L$ and to simplify I assume $\pi = 1/2$. The conditional distribution of $y_{i,1:2}$ given σ_i is a bivariate normal distribution,

$$\begin{bmatrix} y_{i1} \\ y_{i2} \end{bmatrix} \mid \sigma_i \sim N \left(\mathbf{0}_{2 \times 1}, \begin{bmatrix} 1 + \sigma_i^2 & 1 \\ 1 & 2 + \sigma_i^2 \end{bmatrix} \right).$$

The conditional distribution of y_{i2} given y_{i1} and σ_i is also normal,

$$y_{i2} \mid y_{i1}, \sigma_i \sim N \left(\beta_i y_{i1}, s_i^2 \right)$$

where $\beta_i = (1 + \sigma_i^2)^{-1}$ and $s_i^2 = 2 + \sigma_i^2 - \beta_i$. With obvious notation, $\beta_H < \beta_L$ and $s_H^2 > s_L^2$. This expresses the intuition that for an individual with high σ_i , the link between income at time t and income at $t + 1$ is weaker.

In turn, the distribution of σ_i conditioned on y_{i1} is discrete,

$$\begin{aligned} \pi(y) &= \mathbb{P}[\sigma_i = \sigma_H \mid y_{i1} = y] = 1 - \mathbb{P}[\sigma_i = \sigma_L \mid y_{i1} = y] \\ &= \left\{ 1 + \sqrt{\frac{1 + \sigma_H^2}{1 + \sigma_L^2}} \exp \left[-\frac{(\sigma_H + \sigma_L)(\sigma_H - \sigma_L) y^2}{(1 + \sigma_H^2)(1 + \sigma_L^2) 2} \right] \right\}^{-1}. \end{aligned}$$

The conditional probability $\pi(y)$ has a minimum of $\left(1 + (1 + \sigma_H^2)^{1/2} / (1 + \sigma_L^2)^{1/2} \right)^{-1} < 1/2$ at $y = 0$ and is continuous and increasing as a function of y^2 . The intuition is that an observation of a large y_i^2 is most favorable to the view that $\sigma_i = \sigma_H$ than to $\sigma_i = \sigma_L$. By symmetry, $\pi(y) =$

$\pi(-y)$, all y , and by continuity, there is $\bar{y} > 0$ such that $\pi(\bar{y}) = \pi(-\bar{y}) = 1/2$.

It follows that the distribution of y_{i2} given y_{i1} is a normal mixture,

y_{i2}

Let me measure the asymmetry of the distribution of y_{i2} conditioned on y_{i1} by the difference between the mean and the median, i.e.,

$$\begin{aligned} \text{asym}[y_{i2} | y_{i1}] &= \mathbb{E}[y_{i2} | y_{i1}] - \text{med}[y_{i2} | y_{i1}], \\ &= \beta(y_{i1})y_{i1} - q(y_{i1}, 1/2), \end{aligned}$$

with $\beta(y) = \pi(y)\beta_H + (1 - \pi(y))\beta_L$ and $q(y, \tau)$ the τ -th conditional quantile of y_{i2} given $y_{i1} = y$.

The point of my example is to show that $\text{asym}[y_{i2} | y_{i1}]$ decreases with y_{i1} . Computation of the asymmetry measure can be made explicitly at $y = -\bar{y}$, $y = 0$, and $y = \bar{y}$.

One can see that $\text{asym}[y_{i2} | y_{i1} = 0] = 0$ by looking at $\mathbb{E}[y_{i2} | y_{i1} = y] = \beta(y)y$ and the CDF

$$\mathbb{P}[y_{i2} \leq y' | y_{i1} = y] = \pi(y)\Phi\left(\frac{y' - \beta_H y}{s_H}\right) + (1 - \pi(y))\Phi\left(\frac{y' - \beta_L y}{s_L}\right).$$

When $y = 0$, setting $y' = 0$ delivers $\mathbb{P}[y_{i2} \leq y' | y_{i1} = y] = 1/2$.

Notice that the conditional quantile function satisfies, for all y and τ ,

$$\pi(y)\Phi\left(\frac{q(y, \tau) - \beta_H y}{s_H}\right) + (1 - \pi(y))\Phi\left(\frac{q(y, \tau) - \beta_L y}{s_L}\right) = \tau.$$

For \bar{y} , in light of $\pi(\bar{y}) = 1/2$ and the symmetry of the normal distribution,

$$\Phi\left(\frac{q(\bar{y}, 1/2) - \beta_H \bar{y}}{s_H}\right) = \Phi\left(\frac{\beta_L \bar{y} - q(\bar{y}, 1/2)}{s_L}\right),$$

and, because Φ is bijective, one can equate the arguments and solve for $q(\bar{y}, 1/2)$,

$$q(\bar{y}, 1/2) = \left[\frac{s_L}{s_H + s_L} \beta_H + \frac{s_H}{s_H + s_L} \beta_L \right] \bar{y} \geq \beta(\bar{y})\bar{y}.$$

Thus, $\text{asym}[y_{i2} | y_{i1} = \bar{y}] < 0$. By a similar calculation, $\text{asym}[y_{i2} | y_{i1} = -\bar{y}] > 0$. □

A.2 Theorem 1

Proof of theorem 1. (i) Choose t and τ in such a way that $\tau > t + 1$, y_{it} and $(\Delta y_{i\tau})^2$ are observed (i.e., $1 \leq t < \tau \leq T$), and $\text{Var}(\Delta \varepsilon_{i\tau}) = \text{Var}(\varepsilon_{i,\tau-1}) + \text{Var}(\varepsilon_{i\tau})$ is identified. Provided $T \geq 4$, such a choice is possible. Then, since

$$y_{it} = \eta_i + \sum_{s=2}^t v_{is} + \sigma_i \varepsilon_{it},$$

$$(\Delta y_{i\tau})^2 = v_{i\tau}^2 + 2\sigma_i v_{i\tau} \Delta \varepsilon_{i\tau} + \sigma_i^2 \Delta \varepsilon_{i\tau},$$

it is seen, after some calculations, that

$$\text{Cov}(y_{it}, (\Delta y_{i\tau})^2) = \text{Cov}(\eta_i, \sigma_i^2) \text{Var}(\Delta \varepsilon_{i\tau}).$$

(ii) Choose t and τ so that $\tau > t + 1$, $(\Delta y_{it})^2$ and $(\Delta y_{i\tau})^2$ are observed (i.e., $1 < t < \tau \leq T$), and both $\text{Var}(\Delta \varepsilon_{it}) = \text{Var}(\varepsilon_{i,t-1}) + \text{Var}(\varepsilon_{it})$ and $\text{Var}(\Delta \varepsilon_{i\tau}) = \text{Var}(\varepsilon_{i,\tau-1}) + \text{Var}(\varepsilon_{i\tau})$ are identified. Provided $T \geq 5$, this is possible. Then, since

$$(\Delta y_{it})^2 = v_{it}^2 + 2\sigma_i v_{it} \Delta \varepsilon_{it} + \sigma_i^2 \Delta \varepsilon_{it},$$

$$(\Delta y_{i\tau})^2 = v_{i\tau}^2 + 2\sigma_i v_{i\tau} \Delta \varepsilon_{i\tau} + \sigma_i^2 \Delta \varepsilon_{i\tau},$$

it follows that

$$\text{Cov}((\Delta y_{it})^2, (\Delta y_{i\tau})^2) = \text{Var}(\sigma_i^2) \text{Var}(\Delta \varepsilon_{it}) \text{Var}(\Delta \varepsilon_{i\tau}),$$

and the proof is complete. ■

A.3 Theorem 2

Let t be such that $1 \leq t - 2 < t + 2 \leq T$. I will first establish identification of the conditional distributions $\mathbb{P}_{y_{(t-2):(t-1)} | (\bar{\eta}_t, \sigma^2)}$ and $\mathbb{P}_{(\bar{\eta}_t, \sigma^2) | y_{(t+1):(t+2)}}$. Note that the first part of assumption 1 implies assumption 1 in Hu and Schennach (2008), while the second part implies their assumption 3.

Their assumption 2 is implied by the following conditional independence properties:

$$y_{it} \perp (y_{i,(t-2):(t-1)}, y_{i,(t+1):(t+2)})$$

To verify them, notice that

$$\begin{aligned}
y_{i,t-2} &= \tilde{\eta}_{i,t-2} + \sigma_i \varepsilon_{i,t-2}, \\
y_{i,t-1} &= \tilde{\eta}_{i,t-1} + \sigma_i \varepsilon_{i,t-1}, \\
y_{it} &= \tilde{\eta}_{it} + \sigma_i \varepsilon_{it}, \\
y_{i,t-2} &= \tilde{\eta}_{it} + v_{i,t+1} + \sigma_i \varepsilon_{i,t+1}, \\
y_{i,t+2} &= \tilde{\eta}_{it} + v_{i,t+1} + v_{i,t+2} + \sigma_i \varepsilon_{i,t+2}.
\end{aligned}$$

Since $\varepsilon_{it} \perp (\varepsilon_{i,(t-2):(t-1)}, \varepsilon_{i,(t+1):(t+2)}, \tilde{\eta}_{i,(t-2):t}, v_{i,(t+1):(t+2)}, \sigma_i)$ the first conditional independence is obtained. The second follows from $(\varepsilon_{i,(t-2):(t-1)}, \tilde{\eta}_{i,(t-2):t}) \perp (\varepsilon_{i,(t+1):(t+2)}, v_{i,(t+1):(t+2)})$.

Let p denote the density of a generic measure \mathbb{P} . By similar calculations to (7) and (8),

$$\begin{aligned}
\mathbb{E} \left[y_{it} \mid \tilde{\eta}_{it}, \sigma_i^2 \right] &= \mu_t + \tilde{\eta}_{it}, \\
\text{Var} \left(y_{it} \mid \tilde{\eta}_{it}, \sigma_i^2 \right) &= \sigma_i^2 \sigma_{\varepsilon_i}^2.
\end{aligned}$$

It is legitimate to treat μ_t and $\sigma_{\varepsilon_t}^2$ as known owing to the observations made at the beginning of section 3 (note $t < T$). Thus, for every $(\tilde{\eta}, \sigma^2)$ and $(\tilde{\eta}', \sigma'^2)$ such that $(\tilde{\eta}, \sigma^2) \neq (\tilde{\eta}', \sigma'^2)$,

$$\mathbb{P}_{y_t} \left[p_{y_t | (\tilde{\eta}_t, \sigma^2)} \left(y_{it} \mid \tilde{\eta}, \sigma^2 \right) \neq p_{y_t | (\tilde{\eta}_t, \sigma'^2)} \left(y_{it} \mid \tilde{\eta}', \sigma'^2 \right) \right] > 0,$$

leading to assumption 4 in Hu and Schennach (2008).

Finally, let the functional F map the space of conditional densities $p_{y_{(t-2):(t-1)} | (\tilde{\eta}_t, \sigma^2)}$ onto $\mathbb{R} \times \mathbb{R}_{\geq 0}$ by the rule

$$\left\{ F \left[p_{y_{(t-2):(t-1)} | (\tilde{\eta}_t, \sigma^2)} \left(\cdot \mid \tilde{\eta}, \sigma^2 \right) \right] \right\}_{(1,1)} = \int_{\mathbb{R}^2} y_{t-1} p_{y_{(t-2):(t-1)} | (\tilde{\eta}_t, \sigma^2)} \left((y_{t-2}, y_{t-1}) \mid \tilde{\eta}, \sigma^2 \right) d(y_{t-2}, y_{t-1})$$

for the first entry and

$$\begin{aligned}
\left\{ F \left[p_{y_{(t-2):(t-1)} | (\tilde{\eta}_t, \sigma^2)} \left(\cdot \mid \tilde{\eta}, \sigma^2 \right) \right] \right\}_{(2,1)} &= \frac{1}{\sigma_{\varepsilon_{t-1}}^2} \times \\
&\left\{ \int_{\mathbb{R}^2} (y_{t-1} - \tilde{\eta})^2 p_{y_{(t-2):(t-1)} | (\tilde{\eta}_t, \sigma^2)} \left((y_{t-2}, y_{t-1}) \mid \tilde{\eta}, \sigma^2 \right) d(y_{t-2}, y_{t-1}) - \sigma_{v_t}^2 \right\}
\end{aligned}$$

for the second. Then,

$$F \left[p_{y_{(t-2):(t-1)} | (\tilde{\eta}_t, \sigma^2)} \left(\cdot \mid \tilde{\eta}, \sigma^2 \right) \right] = \begin{bmatrix} \tilde{\eta} \\ \sigma^2 \end{bmatrix} \text{ for all } (\tilde{\eta}, \sigma^2) \in \text{supp} \left((\eta_i, \sigma_i^2) \right)$$

and assumption 5 in Hu and Schennach (2008) is satisfied.

All conditions for theorem 1 in Hu and Schennach (2008) are met and deliver the following:

Lemma 1. *Under assumption 1, the distributions*

$$\left\{ \mathbb{P}_{y_t | (\tilde{\eta}_t, \sigma^2)}, \mathbb{P}_{y_{(t-2):(t-1)} | (\tilde{\eta}_t, \sigma^2)}, \mathbb{P}_{(\tilde{\eta}_t, \sigma^2) | y_{i,(t+1):(t+2)}} \right\}$$

are identified provided $T \geq 5$.

Proof of theorem 2. By lemma 1, the distribution $\mathbb{P}_{(\tilde{\eta}_t, \sigma^2)}$ and the marginal distributions $\mathbb{P}_{\tilde{\eta}_t}$ and \mathbb{P}_{σ^2} are identified — integrate the density of $\mathbb{P}_{(\tilde{\eta}_t, \sigma^2) | y_{(t+1):(t+2)}}$ against the distribution $\mathbb{P}_{y_{(t+1):(t+2)}}$. Since $\mathbb{P}_{y_{(t-2):(t-1)} | (\tilde{\eta}_t, \sigma^2)}$ is identified, so is $\mathbb{P}_{(y_{(t-2):(t-1)}, \tilde{\eta}_t) | \sigma^2}$ and applying to the latter the celebrated lemma of Kotlarski (1967) a number of times the identification of

$$\left\{ \mathbb{P}_{(\eta, \sigma^2)}, \mathbb{P}_{v_{t-2}}, \mathbb{P}_{v_{t-1}}, \mathbb{P}_{\varepsilon_{t-2}}, \mathbb{P}_{\varepsilon_{t-1}} \right\}$$

is established. Application of the lemma is possible owing to the second part of assumption 1.

By similar arguments, the distribution $\mathbb{P}_{(y_{(t+1):(t+2)}, \tilde{\eta}_t) | \sigma^2}$ is identified and applying again the lemma of Kotlarski (1967) the identification of

$$\left\{ \mathbb{P}_{v_{t+1}}, \mathbb{P}_{\varepsilon_{t+1}}, \mathbb{P}_{v_{t+2} + \varepsilon_{t+2}} \right\}$$

follows. Finally, repeat the reasoning for every $t = 3, \dots, T - 2$ to finish the proof. ■

B Details of estimation

B.1 Estimation of β

Observe the conditional moments (similar to (7) and (8)),

$$\begin{aligned} 0_{T \times 1} &= \mathbb{E} \left[y'_{i,1:T} - m_i(\beta_\mu) \middle| x_i \right], \\ 0_{T(T+1)/2 \times 1} &= \mathbb{E} \left[\text{vech } y'_{i,1:T} y_{i,1:T} - m_i(\beta_\mu) m'_i(\beta_\mu) - S_i(\beta_\eta, \beta_v, \beta_\varepsilon) \middle| x_i \right], \end{aligned}$$

where, specifying $\mu_t, \sigma_\eta^2, \sigma_{v_t}^2$ ($t = 2, \dots, T$), and $\sigma_{\varepsilon_t}^2$ ($t = 1, \dots, T$) as I explained above,

$$\begin{aligned} m_i(\beta_\mu) &\equiv \begin{bmatrix} \mu_1(x_i) \\ \vdots \\ \mu_T(x_i) \end{bmatrix}, \\ S_i(\beta_\eta, \beta_v, \beta_\varepsilon) &\equiv \begin{bmatrix} \sigma_\eta^2(x_i) + \sigma_{\varepsilon_1}^2(x_i) & \dots & \sigma_\eta^2(x_i) \\ \vdots & \ddots & \vdots \\ \sigma_\eta^2(x_i) & \dots & \sigma_\eta^2(x_i) + \sum_{s=2}^T \sigma_{v_s}^2(x_i) + \sigma_{\varepsilon_T}^2(x_i) \end{bmatrix}. \end{aligned}$$

Many unconditional moment restrictions may be derived. Here I discuss some of them.

Prediction problems. It is attractive to give $\beta_\mu, \beta_\eta, \beta_v$, and β_ε a role in a set of prediction problems as it provides a natural way to measure goodness of fit. Particularly for β_η, β_v , and β_ε , a prediction problem interpretation is an appealing way to choose weights for the many unconditional moments that could originate in the conditional moments of above and generalizes more easily to nonparametric series estimation—the avenue one would take in a larger dataset.

For the parameters β_μ , note the moment conditions

$$\mathbb{E} \left[\mathbb{E}_T \left[\varphi_{\mu_t}(x_i) \left(y_{it} - \varphi'_{\mu_t}(x_i) \beta_\mu \right) \right] \right] = 0_{\dim(\beta_\mu) \times 1}.$$

They lead to the sample moment equations

$$\mathbb{E}_n \left[\mathbb{E}_T \left[\varphi_{\mu_t}(x_i) \left(y_{it} - \varphi'_{\mu_t}(x_i) \hat{\beta}_\mu \right) \right] \right] = 0_{\dim(\beta_\mu) \times 1},$$

which is equivalent to the statement that $\hat{\beta}_\mu$ solves a standard linear least squares problem,

$$\hat{\beta}_\mu \equiv \operatorname{argmin} \beta_\mu \mathbb{E}_n \left[\mathbb{E}_T \left[\left(y_{it} - \varphi'_{\mu_t}(x_i) \beta_\mu \right)^2 \right] \right].$$

Let $\tilde{y}_{it} \equiv y_{it} - \varphi'_{\mu_t}(x_i) \beta_\mu$ denote log earnings net of their mean conditional on covariates. In all that follows, replacing \tilde{y}_{it} by $y_{it} - \varphi'_{\mu_t}(x_i) \hat{\beta}_\mu$ does not affect the consistency of the estimators under an asymptotic regime with fixed parametric specification.

For the parameters β_η , I use moments which are proportional to the first order conditions of a nonlinear least squares problem,

$$\mathbb{E} \left[\sum_{t=2}^T \varphi_\eta(x_i) \exp \left\{ \varphi'_\eta(x_i) \beta_\eta \right\} \cdot \left(\tilde{y}_{i1} \tilde{y}_{it} - \exp \left\{ \varphi'_\eta(x_i) \beta_\eta \right\} \right) \right] = 0_{\dim(\beta_\eta) \times 1}.$$

Therefore, $\hat{\beta}_\eta$ satisfies

$$\hat{\beta}_\eta \equiv \operatorname{argmin} \beta_\eta \mathbb{E}_n \left[\sum_{t=2}^T \left(\tilde{y}_{i1} \tilde{y}_{it} - \exp \left\{ \varphi'_\eta(x_i) \beta_\eta \right\} \right)^2 \right].$$

That is, $\hat{\beta}_\eta$ is set so that $\sigma_\eta^2(x_i)$ matches $\operatorname{Cov}(\tilde{y}_{i1}, \tilde{y}_{it} | x_i)$ with equal weights for $t = 2, \dots, T$.

The parameters β_v and β_ε are also estimated by exploiting moments related to nonlinear least squares problems. For β_v , I use

$$\mathbb{E} \left[\sum_{\tau=2}^{T-1} \sum_{t=\tau+1}^T \varphi_{v_\tau}(x_i) \exp \left\{ \varphi'_{v_\tau}(x_i) \beta_v \right\} \cdot \left(\Delta \tilde{y}_{i\tau} \tilde{y}_{it} - \exp \left\{ \varphi'_{v_\tau}(x_i) \beta_v \right\} \right) \right] = 0_{\dim(\beta_v) \times 1}.$$

Thus,

$$\hat{\beta}_v \equiv \operatorname{argmin} \beta_v \mathbb{E}_n \left[\sum_{\tau=2}^{T-1} \sum_{t=\tau+1}^T \left(\Delta \tilde{y}_{i\tau} \tilde{y}_{it} - \exp \left\{ \varphi'_{v_\tau}(x_i) \beta_v \right\} \right)^2 \right].$$

And for β_ε , I use

$$\mathbb{E} \left[\sum_{\tau=1}^{T-1} \sum_{t=\tau+1}^T \varphi_{\varepsilon_\tau}(x_i) \exp \left\{ \varphi'_{\varepsilon_\tau}(x_i) \beta_\varepsilon \right\} \cdot \left(\tilde{y}_{i\tau} (\tilde{y}_{it} - \tilde{y}_{i\tau}) - \exp \left\{ \varphi'_{\varepsilon_\tau}(x_i) \beta_\varepsilon \right\} \right) \right] = 0_{\dim(\beta_\varepsilon) \times 1}.$$

Thus,

$$\hat{\beta}_\varepsilon \equiv \operatorname{argmin} \beta_\varepsilon \mathbb{E}_n \left[\sum_{\tau=1}^{T-1} \sum_{t=\tau+1}^T \left(\tilde{y}_{i\tau} (\tilde{y}_{it} - \tilde{y}_{i\tau}) - \exp \left\{ \varphi'_{\varepsilon_\tau}(x_i) \beta_\varepsilon \right\} \right)^2 \right].$$

In an asymptotic regime in which conditional means and variances are correctly specified, and T , $\dim(\beta_\mu)$, $\dim(\beta_\eta)$, $\dim(\beta_v)$, and $\dim(\beta_\varepsilon)$ are fixed, as $n \rightarrow \infty$,

$$\begin{aligned}\hat{\beta}_\mu &\xrightarrow{p} \beta_\mu, \\ \hat{\beta}_\eta &\xrightarrow{p} \beta_\eta, \\ \hat{\beta}_v &\xrightarrow{p} \beta_v, \\ \hat{\beta}_\varepsilon &\xrightarrow{p} \beta_\varepsilon.\end{aligned}$$

Moreover, $\{\hat{\beta}_\mu, \hat{\beta}_\mu, \hat{\beta}_\mu, \hat{\beta}_\mu\}$ are normally distributed in large samples and bootstrap computations of standard errors and confidence intervals are valid. This is what I do in the empirical analysis of the paper.

Alternative asymptotics in which $\dim(\beta_\mu)$, $\dim(\beta_\eta)$, $\dim(\beta_v)$, and $\dim(\beta_\varepsilon)$ grow with the sample size are not addressed here, although they would be of interest if my approach were extended to series estimation in a larger dataset (such as in administrative data).

Optimal instruments. The conditional moments also deliver the optimal unconditional moment restrictions for an estimation problem that takes the conditional mean and variance functions as correct with fixed $\dim(\beta_\mu)$, $\dim(\beta_\eta)$, $\dim(\beta_v)$, and $\dim(\beta_\varepsilon)$. These can be obtained by the method in Chamberlain (1987). I employ the two conditional moments separately to construct (restricted) optimal instruments for β_μ and $\beta_\sigma \equiv (\beta'_\eta, \beta'_v, \beta'_\varepsilon)'$.

For β_μ , the optimal moment condition reduces to the familiar GLS problem, i.e.,

$$\mathbb{E} \left[\varphi_\mu(x_i) \{S_i(\beta_\sigma)\}^{-1} (y'_{i,1:T} - m_i(\beta_\mu)) \right] = 0_{\dim(\beta_\mu) \times 1}.$$

For β_σ ,

$$\mathbb{E} \left[\frac{\partial \text{vech } S_i(\beta_\sigma)'}{\partial \beta_\sigma} \{K_i(\beta_\sigma, \gamma)\}^{-1} (\text{vech } \tilde{y}'_{i,1:T} \tilde{y}_{i,1:T} - S_i(\beta_\sigma)) \right] = 0_{\dim(\beta_\sigma) \times 1}.$$

Here $K_i(\beta_\sigma, \gamma) \equiv \text{Var}((\text{vech } \tilde{y}'_{i,1:T} \tilde{y}_{i,1:T} - S_i(\beta_\sigma)) | x_i)$ is a matrix of conditional fourth moments. Both the gradients and this matrix are available in closed form. The estimators are obtained as numerical solutions to nonlinear equations plugging initial estimates of β_σ and γ into the gradients and conditional variance matrices.

As with least squares estimators, it is not hard to show consistency, asymptotic normality, and validity of bootstrap calculations. Notice that since these estimators exploit the condi-

tional moments separately, they can be improved by estimators that use them jointly. An open question is whether these estimators would remain optimal if $\dim(\beta_\mu)$, $\dim(\beta_\eta)$, $\dim(\beta_v)$, and $\dim(\beta_\varepsilon)$ were allowed to grow.

B.2 Estimation of γ

Specification. I begin by providing detail on the form of marginal and conditional quantile functions for the latent variables. Consider a grid of points such that $0 < \tau_1 \leq \dots \leq \tau_{n_\tau} < 1$. Then,

$$Q_\eta(H|\gamma) = \mathbb{1}\{H \leq \tau_1\} \left(\gamma_{\eta,1} - \gamma_{\eta,\text{left}} \ln \left(1 - \frac{H}{\tau_1} \right) \right)$$

is the quantile function of h_i .

$$Q_v(U|\gamma) = \mathbb{1}\{U \leq \tau_1\} \left(\gamma_{v,1} - \gamma_{v,\text{left}} \ln \left(1 - \frac{U}{\tau_1} \right) \right)$$

is the quantile function of u_{it} ($t = 2, \dots, T$).

$$Q_\varepsilon(E|\gamma) = \mathbb{1}\{E \leq \tau_1\} \left(\gamma_{\varepsilon,1} - \gamma_{\varepsilon,\text{left}} \ln \left(1 - \frac{E}{\tau_1} \right) \right)$$

is the quantile function of e_{it} ($t = 1, \dots, T$).

$$+ \sum_{i_\tau=2}^{n_\tau} \mathbb{1} \\ + \mathbb{1}\{\tau_n$$

$$+ \sum_{i_\tau=2}^{n_\tau} \mathbb{1} \\ + \mathbb{1}\{\tau_n$$

$$+ \sum_{i_\tau=2}^{n_\tau} \mathbb{1} \\ + \mathbb{1}\{\tau_{n_\tau}$$

For s_i , I use

$$Q_\sigma(S|h, \gamma) = \mathbb{1}\{S \leq \tau_1\} \left(\varphi'_\sigma(h) \gamma_{\sigma,1} - \gamma_{\sigma,\text{left}} \ln \left(1 - \frac{S}{\tau_1} \right) \right)$$

The parameter γ consists of the three n_τ -dimensional vectors $\gamma_{\eta,1:n_\tau}$, $\gamma_{v,1:n_\tau}$, and $\gamma_{\varepsilon,1:n_\tau}$, the $n_\tau \times \dim(\varphi_\sigma)$ matrix $\gamma'_{\sigma,1:n_\tau}$, and the tail parameters $\gamma_{\eta,\text{left}}$, $\gamma_{\eta,\text{right}}$, $\gamma_{v,\text{left}}$, $\gamma_{v,\text{right}}$, $\gamma_{\varepsilon,\text{left}}$, $\gamma_{\varepsilon,\text{right}}$, $\gamma_{\sigma,\text{left}}$, and $\gamma_{\sigma,\text{right}}$.

The parameter space Γ is the set of all vectors γ such that h_i , $u_{i,2:T}$, and $e_{i,1:T}$ have zero mean and unit variance, and $\exp(2s_i)$ has unit mean. An element of Γ can be obtained from any arbitrary vector γ by shifting and rescaling $\gamma_{\eta,1:n_\tau}$ and $\gamma_{\eta,\text{left}}$, $\gamma_{\eta,\text{right}}$ (to normalize h_i), $\gamma_{v,1:n_\tau}$, $\gamma_{v,\text{left}}$, and $\gamma_{v,\text{right}}$ (to normalize $u_{i,2:T}$), $\gamma_{\varepsilon,1:n_\tau}$, $\gamma_{\varepsilon,\text{left}}$, and $\gamma_{\varepsilon,\text{right}}$ (to normalize $e_{i,1:T}$), and the first column of $\gamma'_{\sigma,1:n_\tau}$ (to normalize s_i).

EM principle. If latent variables were observed, γ would be determined by a collection of quantile computations and regressions. Let

$$\Psi_{\tau,i}(\gamma) = \Psi_\tau(h_i, s_i, u_{i,2:T}, e_{i,1:T}, \gamma) \equiv \begin{bmatrix} \psi_\tau(h_i - Q_\eta(\tau|\gamma)) \\ \psi_\tau(s_i - Q_\sigma(\tau|h_i, \gamma)) \\ \sum_{t=2}^T \psi_\tau(u_{it} - Q_v(\tau|\gamma)) \\ \sum_{t=1}^T \psi_\tau(e_{it} - Q_\varepsilon(\tau|\gamma)) \end{bmatrix},$$

where $\psi_\tau(z) \equiv \tau - \mathbb{1}\{z < 0\}$, be defined for $\tau \in (0, 1)$. A equivalent statement is to say that, for selected values of τ , the parameter γ is identified from the moment equation

$$0_{4 \times 1} = \mathbb{E}[\Psi_{\tau,i}(\gamma)].$$

To deal with the fact that latent variables are unobserved, one can use the EM principle: apply the law of iterated expectations to obtain

$$0_{4 \times 1} = \mathbb{E}[\mathbb{E}[\Psi_{\tau,i}(\gamma)|y_{i,1:T}]].$$

But direct evaluation of the conditional expectation $\mathbb{E} [\Psi_{\tau,i}(\gamma) | y_{i,1:T}]$ is challenging and the idea will be to replace it by an average across simulations,

$$\mathbb{E}_{n_{\text{sim}}} \left[\Psi_{\tau,i}^{(i_{\text{sim}})}(\gamma) | y_{i,1:T} \right] \equiv \frac{1}{n_{\text{sim}}} \sum_{i_{\text{sim}}=1}^{n_{\text{sim}}} \Psi_{\tau} \left(h_i^{(i_{\text{sim}})}, s_i^{(i_{\text{sim}})}, u_{i,2:T}^{(i_{\text{sim}})}, e_{i,1:T}^{(i_{\text{sim}})}, \gamma \right),$$

where $h_i^{(i_{\text{sim}})}, s_i^{(i_{\text{sim}})}, u_{i,2:T}^{(i_{\text{sim}})}, e_{i,1:T}^{(i_{\text{sim}})}$ are drawn from the conditional distribution $\mathbb{P}_{(h,s,u_{2:T},e_{1:T})|y_{1:T}}$. It is seen that the replacement of the conditional expectation by the simulation average does not invalidate the moment condition,

$$0_{4 \times 1} = \mathbb{E} \left[\mathbb{E}_{n_{\text{sim}}} \left[\Psi_{\tau,i}^{(i_{\text{sim}})}(\gamma) | y_{i,1:T} \right] \right].$$

Algorithm 1 implements this idea by alternating between the simulation of latent variables (E-Step) and the solution of moment equations (M-Step). The M-Step is a fairly standard set of quantile computations. In what remains of this appendix, I explain the E-Step. The EM approach I follow is closely connected to Arellano and Bonhomme (2016), Arellano et al. (2017), and Arellano and Bonhomme (2019).

Simulating latent variables. To carry out the E-Step of the algorithm, one would need to simulate $h_i, s_i, u_{i,2:T}$, and $e_{i,1:T}$ from their conditional distribution given $y_{i,1:T}$ for a fixed value of γ (x_i and β too). Notice that drawing from $e_{1:T}$ is not necessary once $h_i, s_i, u_{i,2:T}$, and $y_{i,1:T}$ are known (together with x_i and β). Hence, I need to draw from the density

$$p_{(h,s,u_{2:T})|y_{1:T}}(h_i, s_i, u_{i,2:T} | y_{i,1:T}).$$

By Bayes rule,

$$\begin{aligned} p_{(h,s,u_{2:T})|y_{1:T}}(h_i, s_i, u_{i,2:T} | y_{i,1:T}) &\propto p_{y_{1:T} | (h,s,u_{2:T})}(y_{i,1:T} | h_i, s_i, u_{i,2:T}) \times p_{(h,s,u_{2:T})}(h_i, s_i, u_{i,2:T}) \\ &= \left[\prod_{t=1}^T p_{y_t | (h,s,u_{2:t})}(y_{it} | h_i, s_i, u_{i,2:t}) \right] \\ &\propto \left[\prod_{t=1}^T e^{-s_i} p_e \left(\frac{y_{it} - (\sigma_{\eta}(x_i)h_i + \sum_{\tau=2}^t \sigma_{v_{\tau}}(x_i)u_{i\tau})}{e^{s_i} \sigma_{\varepsilon_i}(x_i)} \right) \right] \end{aligned}$$

The densities $p_h, p_{s|h}, p_u,$ and p_e are fully determined by γ (notice p_u and p_e do not depend on t). They are piecewise constant functions with exponential tails, i.e.,

$$p_h(h) = p_h(h|\gamma) = \mathbb{1}\{h \leq \gamma_{\eta,1}\} \left(\frac{\tau_1}{\gamma_{\eta,\text{left}}} \exp\left(\frac{h - \gamma_{\eta,1}}{\gamma_{\eta,\text{left}}}\right) \right)$$

$$p_u(u) = p_u(u|\gamma) = \mathbb{1}\{u \leq \gamma_{v,1}\} \left(\frac{\tau_1}{\gamma_{v,\text{left}}} \exp\left(\frac{u - \gamma_{v,1}}{\gamma_{v,\text{left}}}\right) \right)$$

$$p_e(e) = p_e(e|\gamma) = \mathbb{1}\{e \leq \gamma_{\varepsilon,1}\} \left(\frac{\tau_1}{\gamma_{\varepsilon,\text{left}}} \exp\left(\frac{e - \gamma_{\varepsilon,1}}{\gamma_{\varepsilon,\text{left}}}\right) \right)$$

and

$$p(s|h) = p_{s|h}(s|h, \gamma) = \mathbb{1}\{s \leq \varphi'_\sigma(h)\gamma_{\sigma,1}\} \left(\frac{\tau_1}{\gamma_{\sigma,\text{left}}} \exp\left(\frac{s - \varphi'_\sigma(h)\gamma_{\sigma,1}}{\gamma_{\sigma,\text{left}}}\right) \right)$$

With a function proportional to the conditional density at hand it is straightforward to employ Markov Chain Monte Carlo (MCMC) methods. The one of my choice for this application is a univariate slice sampling techniques (see Neal (2003)). The test suggested by Geweke (2004)

indicates excellent performance of the conditional distribution simulation algorithm.

Inference. The inference theory is known for a γ of fixed dimension. In that case, the estimator obtained by stochastic EM is (in the limit ideal of infinitely many simulations) asymptotically normally distributed and the bootstrap gives valid measures of precision. For the results of the paper, bootstrap calculations for objects that rely on estimates of γ are reported in a separate appendix. Nonetheless, an asymptotic theory in which the dimension of γ grows with the sample size is an open question and an interesting avenue for future research.

C Tests of model restrictions

Theorem 1 suggests an intuitive set of model restrictions that could be exploited to construct a diagnostic. Namely,

$$\begin{aligned}\text{Cov}\left(\eta_i, \sigma_i^2\right) &= \frac{\text{Cov}\left(y_{it}, (\Delta y_{it})^2\right)}{\text{Var}(\Delta \varepsilon_{it})}, \\ \text{Var}\left(\sigma_i^2\right) &= \frac{\text{Cov}\left(\Delta y_{it}^2, \Delta y_{it}^2\right)}{\text{Var}(\Delta \varepsilon_{it}) \text{Var}(\Delta \varepsilon_{it})}.\end{aligned}$$

Starting from $T \geq 5$, there are $r = (T - 3)^2 - 2$ testable restrictions. These can be written in the form of moment equations

$$\mathbb{E}[\rho_i] = 0_{r \times 1}.$$

Using that, under bounded second moments of ρ_i ,

$$\sqrt{n} \{\text{Var}(\rho_i)\}^{-\frac{1}{2}} \mathbb{E}_n[\rho_i] \implies N(0_{r \times 1}, I_r),$$

the test that rejects the HTR model restrictions when

$$n \mathbb{E}_n[\rho_i'] \{\text{Var}(\rho_i)\}^{-1} \mathbb{E}_n[\rho_i] > \chi_r^2(1 - \alpha),$$

with $\chi_r^2(1 - \alpha)$ the $(1 - \alpha)$ -quantile of the chi squared distribution with r degrees of freedom, is pointwise asymptotically level α . This test is easy to implement: on household panel data the statistic is roughly 1.6 with a p-value of 0.9756 (7 degrees of freedom since $T = 6$).

In simulations, this test appears to be undersized. The source of the problem seems to lie in the heavy tails that the sample moments have in practice: ρ_i involves third- and fourth-order moments.

I have constructed two alternative test procedures for this application that perform better in finite samples than the simple moment-based test. The first is a multivariate version of the bounded-influence test of Heritier and Ronchetti (1994). The second is a test that is based on an alternative limit distribution for $\mathbb{E}_n[\rho_i]$. I model ρ_i as a mixture of two normal distributions in which the mixing probability tends to zero at the rate n^{-1} and the location of the shift tends to infinity at the rate \sqrt{n} . I derive an approximate likelihood for $\mathbb{E}_n[\rho_i]$ that I can apply to the problem of testing $\mathbb{E}[\rho_i] = 0_{r \times 1}$. The asymptotic embedding guarantees that the skewness and

the excess kurtosis of $\sqrt{n}\mathbb{E}_n[\rho_i]$ across samples do not vanish. I calibrate the probability and the location of the shift distribution with estimates of skewness and kurtosis of moments.

Both tests also fail to reject the null.

D Tables and figures for panel of households (income)

D.1 Summary of covariates and some moments

TABLE 7. Covariates x_i and their averages in the base period

Row	Average
representative person is 30-34 years old	0.143
representative person is 35-39 years old	0.222
representative person is 40-44 years old	0.215
representative person is 45-49 years old	0.209
representative person is >50 years old	0.045
representative person is high-school graduate	0.314
representative person went to college	0.636
spouse is 30-34 years old	0.149
spouse is 35-39 years old	0.234
spouse is 40-44 years old	0.213
spouse is 45-49 years old	0.156
spouse is >50 years old	0.025
spouse is high-school graduate	0.34
spouse went to college	0.635
representative person is not white	0.064
number of children	1.335
family size	3.503
family lives in the southeast	0.229
family lives in the great lakes	0.193
family lives in the plains and mountains	0.178
family lives in the west	0.181

TABLE 8. $\text{Var}(y_{i,1:T})$

Row	t1999	t2001	t2003	t2005	t2007	t2009
t1999	0.237	0.168	0.143	0.132	0.128	0.121
t2001	0.168	0.228	0.157	0.143	0.142	0.128
t2003	0.143	0.157	0.24	0.148	0.148	0.137
t2005	0.132	0.143	0.148	0.228	0.176	0.158
t2007	0.128	0.142	0.148	0.176	0.245	0.19
t2009	0.121	0.128	0.137	0.158	0.19	0.254

TABLE 9. $\text{Cov}(y_{it}, (\Delta y_{i\tau})^2) / \text{Var}(\Delta \varepsilon_{i\tau})$ for $\tau > t + 1$

Row	t2003	t2005	t2007
t1999	0.023	-0.078	-0.101
t2001		-0.093	0.003
t2003			-0.086

TABLE 10. $\text{Cov}((\Delta y_{it})^2, (\Delta y_{i\tau})^2) / \text{Var}(\Delta \varepsilon_{it}) \text{Var}(\Delta \varepsilon_{i\tau})$ for $\tau > t + 1$

Row	t2003	t2005	t2007
t2001		0.629	0.664
t2003			0.846

D.2 Observable heterogeneity: functions of covariates

TABLE 11. Estimates of $100 \times \beta_\mu$

Row	LSestimate	LSstderr	IVestimate	IVstderr
constant	10.46	0.13	10.46	0.12
Year 2001	0.04	0.01	0.04	0.01
Year 2003	0.01	0.02	0.01	0.01
Year 2005	0.03	0.02	0.04	0.02
Year 2007	0.06	0.02	0.07	0.02
Year 2009	0.03	0.02	0.04	0.02
representative person is 30-34 years old	0.16	0.04	0.16	0.03
representative person is 35-39 years old	0.22	0.05	0.23	0.03
representative person is 40-44 years old	0.23	0.06	0.27	0.04
representative person is 45-49 years old	0.23	0.07	0.28	0.05
representative person is >50 years old	0.21	0.08	0.29	0.05
representative person is high-school graduate	0.19	0.06	0.2	0.06
representative person went to college	0.43	0.06	0.44	0.06
spouse is 30-34 years old	0.05	0.03	0.03	0.02
spouse is 35-39 years old	0.09	0.05	0.06	0.03
spouse is 40-44 years old	0.16	0.06	0.09	0.03
spouse is 45-49 years old	0.23	0.07	0.11	0.04
spouse is >50 years old	0.22	0.08	0.1	0.04
spouse is high-school graduate	0.06	0.11	0.04	0.11
spouse went to college	0.29	0.11	0.25	0.11
representative person is not white	-0.1	0.05	-0.1	0.05
number of children	0	0.02	-0.02	0.01
family size	-0.02	0.02	-0	0.01
family lives in the southeast	-0.17	0.04	-0.13	0.03
family lives in the great lakes	-0.09	0.04	-0.04	0.04
family lives in the plains and mountains	-0.24	0.04	-0.2	0.04
family lives in the west	-0.06	0.04	-0.08	0.04

TABLE 12. Estimates of $100 \times \beta_\eta$

Row	LSestimate	LSstderr	IVestimate	IVstderr
constant	-116.24	16.92	-117.93	15.82
representative person age	0.94	0.5	1.22	0.44
representative person is high-school graduate	-0.76	15.78	-3.69	16.19
representative person went to college	6.91	15.1	7.29	15.98

TABLE 13. Estimates of $100 \times \beta_v$

Row	LSestimate	LSstderr	IVestimate	IVstderr
constant	-259.35	197.56	-223.89	1392.7
representative person age	1.43	1.33	1.48	0.76
representative person is high-school graduate	22.86	197.82	5.51	1392.88
representative person went to college	18.67	199.35	1.03	1393.23

TABLE 14. Estimates of $100 \times \beta_\varepsilon$

Row	LSestimate	LSstderr	IVestimate	IVstderr
constant	-119.75	13.87	-96.93	15.33
representative person age	-0.38	0.52	-1.17	0.43
representative person is high-school graduate	-2.56	14.68	-26.33	15.31
representative person went to college	6.41	13.1	-19.58	13.86

D.3 Unobserved heterogeneity: distributions of latent variables

FIGURE 9. PDF of σ_i^2

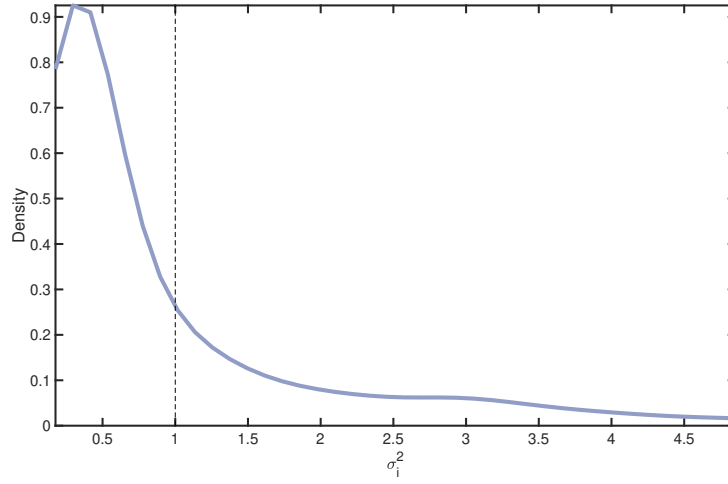


FIGURE 10. Joint PDF of η_i and σ_i^2

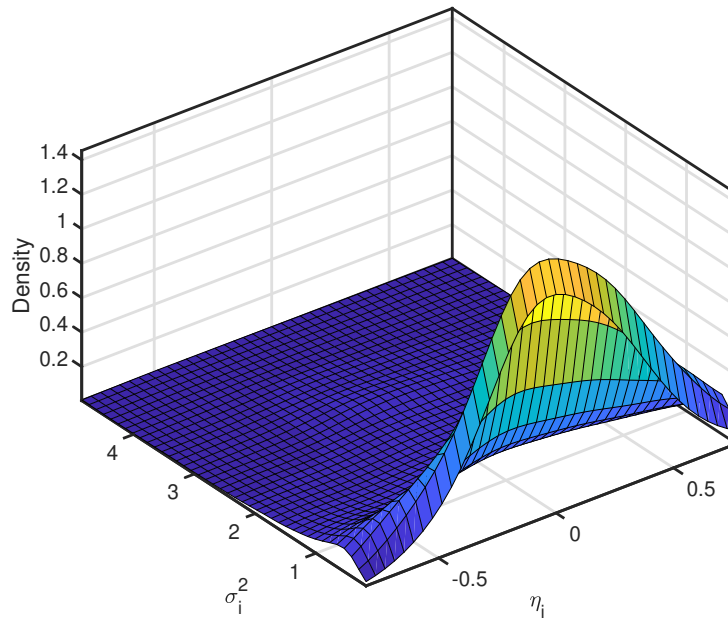


FIGURE 11. Conditional quantile function of σ_i^2 given η_i

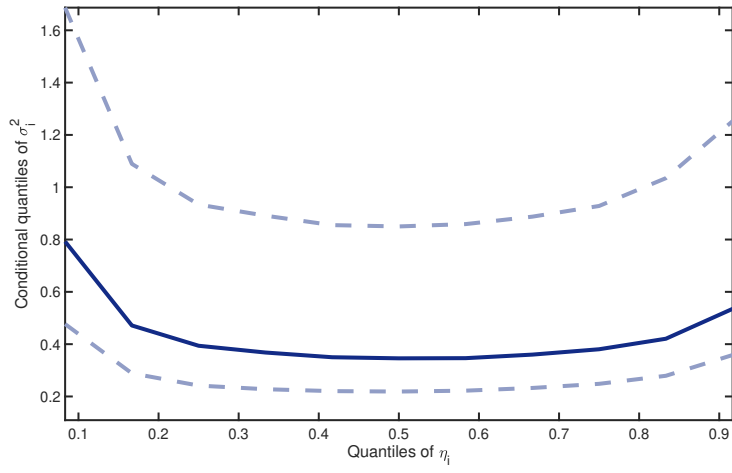
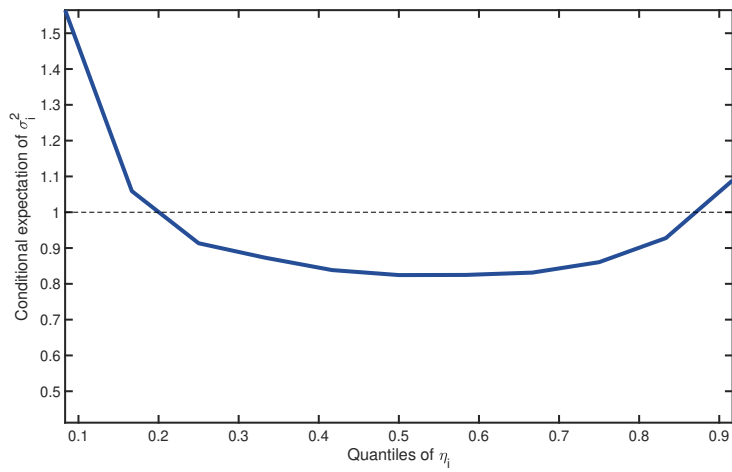


FIGURE 12. Conditional expectation function of σ_i^2 given η_i



D.4 Observable and unobserved heterogeneity

FIGURE 13. PDF of $\tilde{\sigma}_i^2$

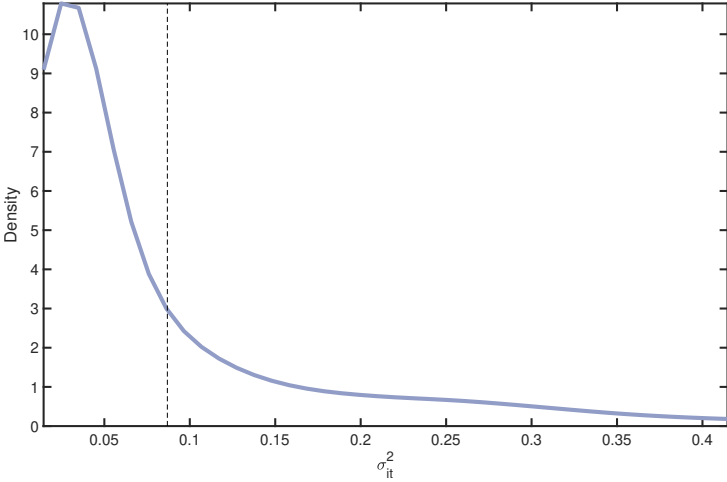
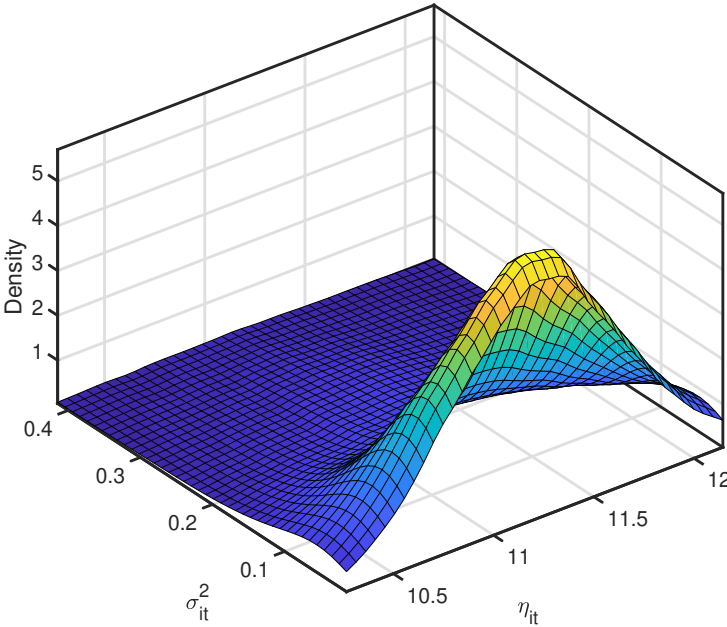


FIGURE 14. Joint PDF of $\tilde{\eta}_i$ and $\tilde{\sigma}_i^2$



D.5 Comparison with normality of shocks

FIGURE 15. PDFs of η_i , v_{it} , ε_{it} , and $\sigma_i\varepsilon_{it}$

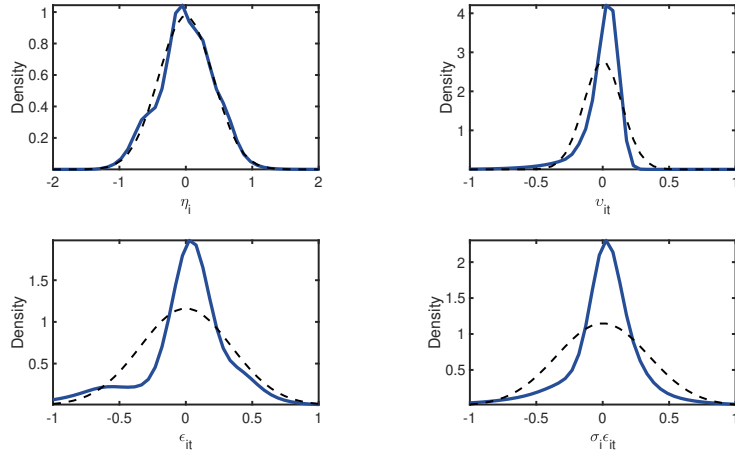


TABLE 15. Skewness and kurtosis of η_i , v_{it} , ε_{it} , and $\sigma_i\varepsilon_{it}$

Row	Variance	Skewness	Kurtosis
η_i	0.167	-0.074	2.931
v_{it}	0.021	-2.907	17.331
ε_{it}	0.118	-1.278	6.817
$\tilde{\varepsilon}_{it}$	0.121	-1.905	17.026

D.6 Consumption passthrough coefficients

TABLE 16. Regression of $\Delta \bar{c}_{it}$ on permanent and transitory shocks

	HeterogeneityIn	HeterogeneityOut
permanent shock	0.3513	0.3427
transitory shock	-0.0016	0.1829

D.7 Trends under the distribution of transitory risks

FIGURE 16. Evolution of the marginal distribution of σ_i^2

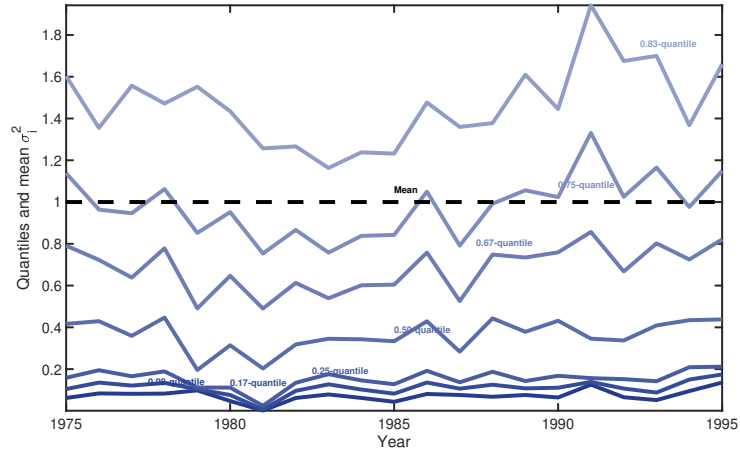


FIGURE 17. Evolution of the distribution of σ_i^2 given η_i

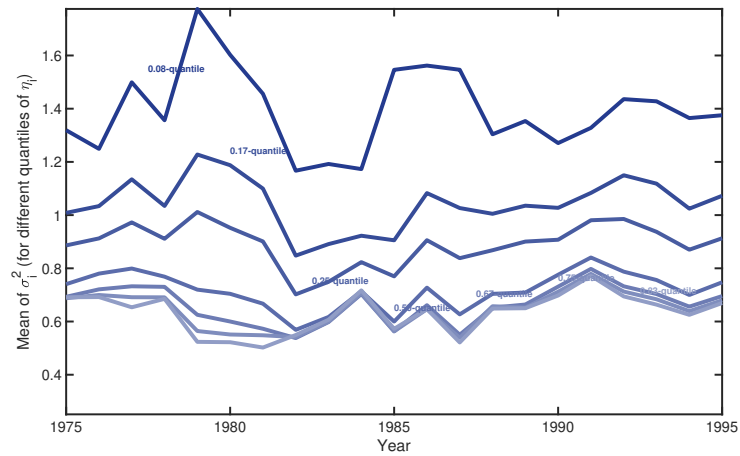
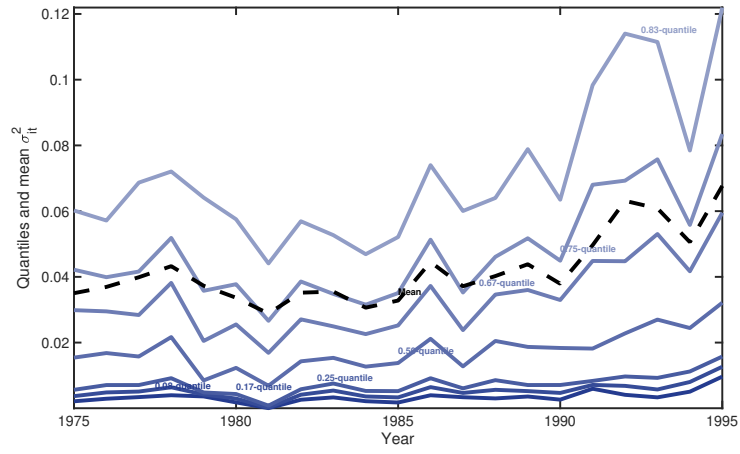


FIGURE 18. Evolution of the marginal distribution of $\tilde{\sigma}_t^2$



E Tables and figures for panel of workers (wages)

E.1 Summary of covariates and some moments

TABLE 17. Covariates x_i and their averages in the base period

Row	Average
worker is 30-34 years old	0.161
worker is 35-39 years old	0.222
worker is 40-44 years old	0.199
worker is 45-49 years old	0.187
worker is >50 years old	0.041
worker is high-school graduate	0.34
worker went to college	0.61
worker is not male	0.109
worker is not white	0.097
number of children	1.152
family size	3.096
family lives in the southeast	0.235
family lives in the great lakes	0.189
family lives in the plains and mountains	0.159
family lives in the west	0.196

TABLE 18. $\text{Var}(y_{i,1:T})$

Row	t1999	t2001	t2003	t2005	t2007	t2009
t1999	0.246	0.17	0.154	0.145	0.145	0.148
t2001	0.17	0.244	0.17	0.163	0.161	0.159
t2003	0.154	0.17	0.291	0.178	0.176	0.176
t2005	0.145	0.163	0.178	0.266	0.196	0.196
t2007	0.145	0.161	0.176	0.196	0.271	0.224
t2009	0.148	0.159	0.176	0.196	0.224	0.314

TABLE 19. $\text{Cov}(y_{it'}, (\Delta y_{it'})^2) / \text{Var}(\Delta \varepsilon_{it'})$ for $\tau > t + 1$

Row	t2003	t2005	t2007
t1999	0.102	-0.027	-0.084
t2001		-0.072	-0.088
t2003			-0.104

TABLE 20. $\text{Cov}((\Delta y_{it'})^2, (\Delta y_{it'})^2) / \text{Var}(\Delta \varepsilon_{it'}) \text{Var}(\Delta \varepsilon_{it'})$ for $\tau > t + 1$

Row	t2003	t2005	t2007
t2001		0.816	0.241
t2003			-0.08

E.2 Observable heterogeneity: functions of covariates

TABLE 21. Estimates of $100 \times \beta_\mu$

Row	LSestimate	LSstderr	IVestimate	IVstderr
constant	2.38	0.07	2.48	0.06
Year 2001	0.07	0.01	0.07	0.01
Year 2003	0.04	0.02	0.05	0.02
Year 2005	0.05	0.02	0.06	0.02
Year 2007	0.09	0.02	0.1	0.02
Year 2009	0.11	0.02	0.13	0.02
worker is 30-34 years old	0.15	0.03	0.16	0.02
worker is 35-39 years old	0.24	0.03	0.23	0.03
worker is 40-44 years old	0.32	0.04	0.27	0.03
worker is 45-49 years old	0.33	0.04	0.27	0.03
worker is >50 years old	0.34	0.05	0.27	0.04
worker is high-school graduate	0.17	0.05	0.17	0.05
worker went to college	0.54	0.04	0.53	0.05
worker is not male	-0.22	0.05	-0.23	0.05
worker is not white	-0.22	0.04	-0.23	0.04
number of children	0.01	0.02	0.01	0.01
family size	0.03	0.02	0	0.01
family lives in the southeast	-0.2	0.04	-0.16	0.03
family lives in the great lakes	-0.15	0.04	-0.13	0.04
family lives in the plains and mountains	-0.21	0.04	-0.18	0.04
family lives in the west	-0.07	0.05	-0.08	0.04

TABLE 22. Estimates of $100 \times \beta_\eta$

Row	LSestimate	LSstderr	IVestimate	IVstderr
constant	-138.19	13.59	-137.94	13.13
worker age	0.38	0.37	0.53	0.49
worker is high-school graduate	30.94	13.13	28.86	11.71
worker went to college	45.28	12.31	45.58	11.22

TABLE 23. Estimates of $100 \times \beta_v$

Row	LSestimate	LSstderr	IVestimate	IVstderr
constant	-273.43	162.06	-254.6	7285.07
worker age	2.34	3.12	1.84	1.28
worker is high-school graduate	-4.08	141.3	19.54	7286.43
worker went to college	37.98	138.19	32.64	7286.43

TABLE 24. Estimates of $100 \times \beta_\epsilon$

Row	LSestimate	LSstderr	IVestimate	IVstderr
constant	-120.45	15.09	-114.87	14.12
worker age	-0.73	0.4	-1.1	0.42
worker is high-school graduate	9.25	13.2	5.25	11.8
worker went to college	5.96	13.58	3.15	11.96

E.3 Unobserved heterogeneity: distributions of latent variables

FIGURE 19. PDF of σ_i^2

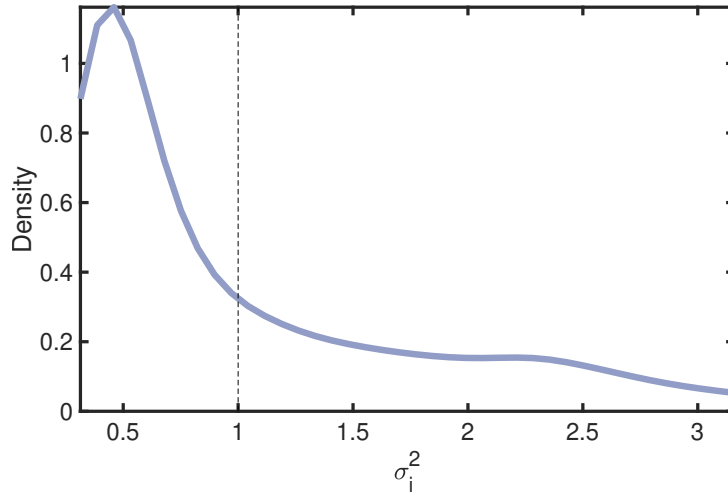


FIGURE 20. Joint PDF of η_i and σ_i^2

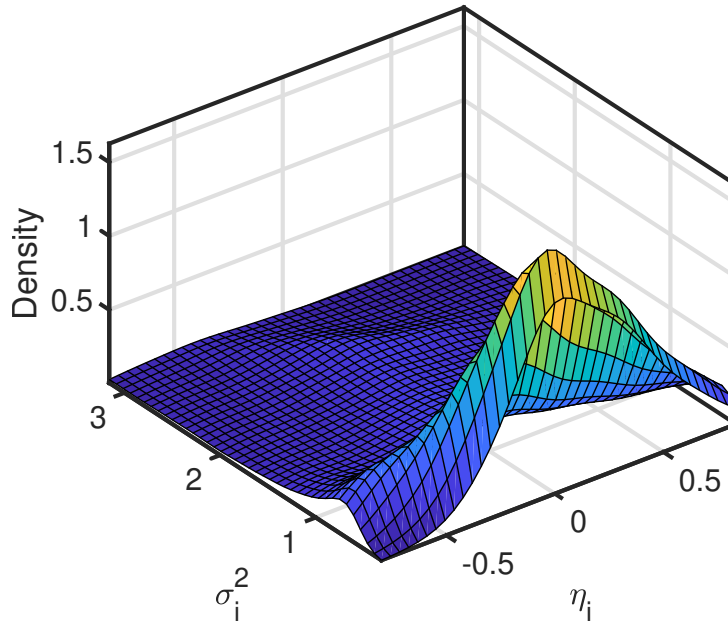


FIGURE 21. Conditional quantile function of σ_i^2 given η_i

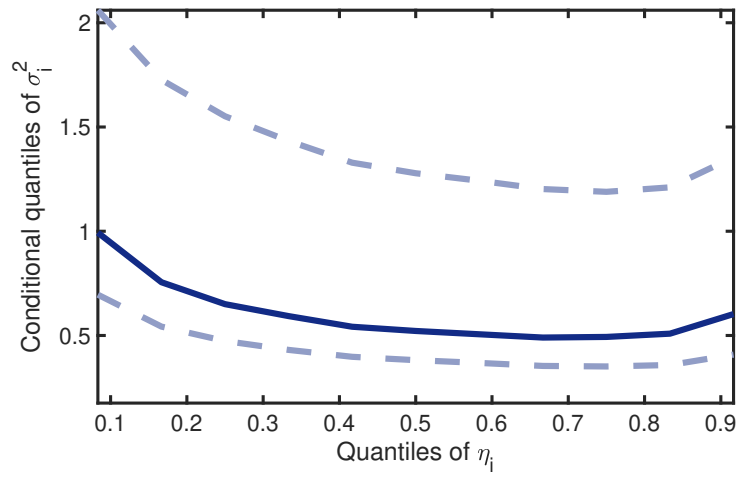
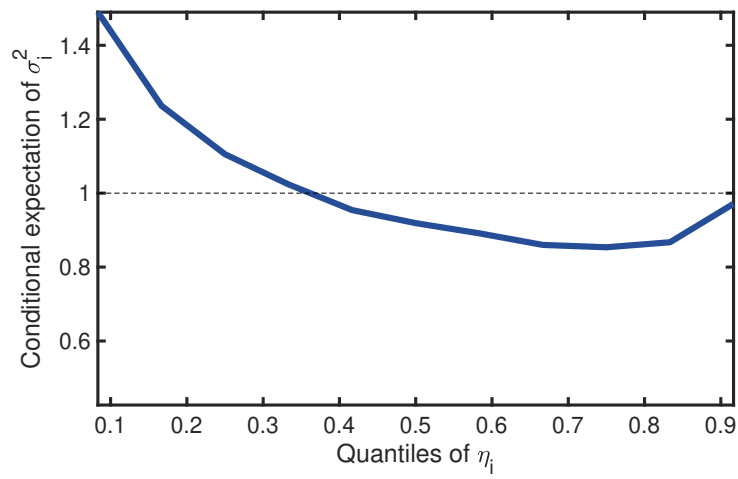


FIGURE 22. Conditional expectation function of σ_i^2 given η_i



E.4 Observable and unobserved heterogeneity

FIGURE 23. PDF of $\tilde{\sigma}_i^2$

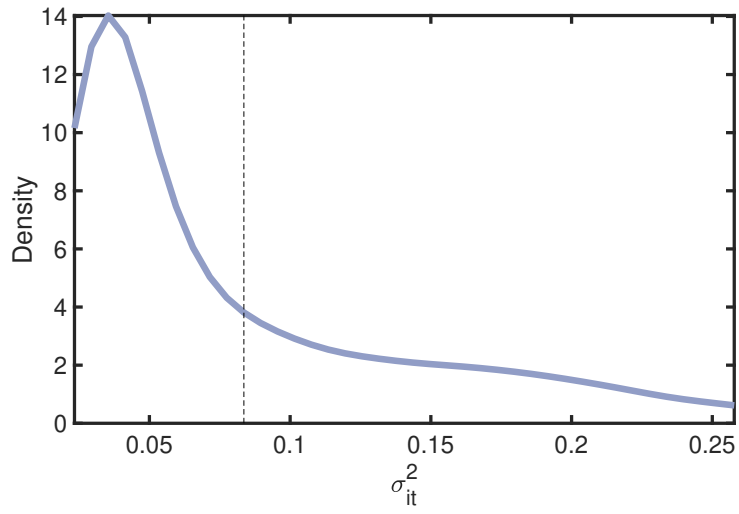
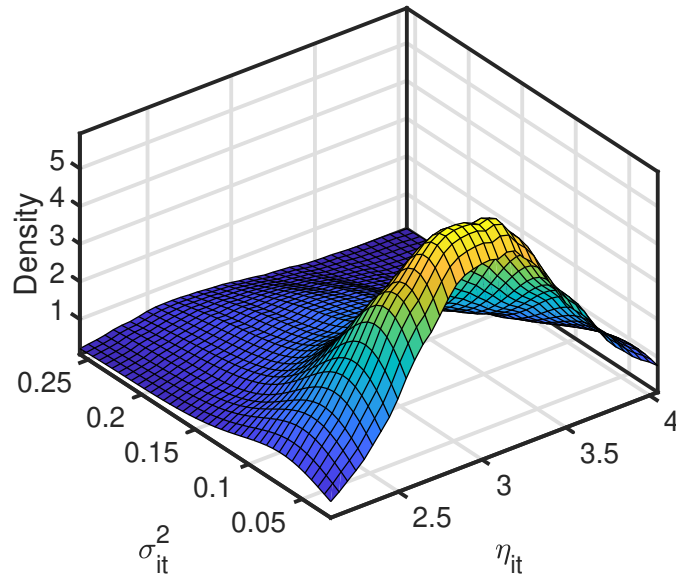


FIGURE 24. Joint PDF of $\tilde{\eta}_i$ and $\tilde{\sigma}_i^2$



E.5 Comparison with normality of shocks

FIGURE 25. PDFs of η_i , v_{it} , ε_{it} , and $\sigma_i\varepsilon_{it}$

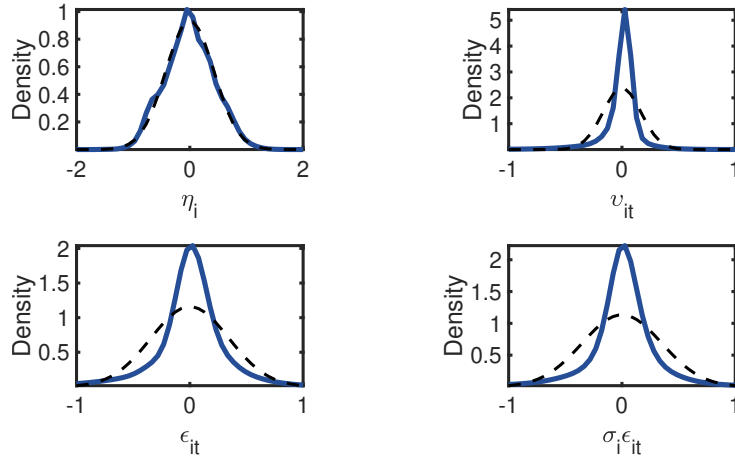


TABLE 25. Skewness and kurtosis of η_i , v_{it} , ε_{it} , and $\sigma_i\varepsilon_{it}$

Row	Variance	Skewness	Kurtosis
η_i	0.183	0.06	2.818
v_{it}	0.028	-2.529	22.85
ε_{it}	0.118	-1.186	11.407
$\tilde{\varepsilon}_{it}$	0.124	-1.479	18.883

E.6 Regression on labor market variables

TABLE 26. Regression of $\ln(\sigma_i)$ on labor market variables

	Estimates
constant	-1.56
unemployment	41.33
self employment	22.44
contract under union	-3.58
paid by the hour	-4.34
occupation: managers	3.67
occupation: sales workers	11.11
occupation: clerical workers	0.07
occupation: factory workers	7.23
occupation: farmers	9.17
occupation: service workers	13.58
industry: manufacture	-18.79
industry: retail	-16.42
industry: finance	-15.03
industry: services	-15.02
industry: public sector	-25.51