UNCERTAINTY, PERSISTENCE, AND HETEROGENEITY: A PANEL DATA PERSPECTIVE

Manuel Arellano
CEMFI, Madrid

Abstract
The purpose of this paper is to review newly developed identification and estimation tools that are relevant for the analysis of dynamic dependence structures of income risk. I present an application to nonlinear permanent–transitory models of household income using data from the Panel Study of Income Dynamics (PSID), but the empirical approach is more generally applicable. Household income processes are of interest because the size of shocks, the nature of their persistence, and cross-household heterogeneity are all important to understand how income inequality varies with age and cohort and how it translates into consumption inequality. I argue that going from an econometrics of autocovariances to an econometrics of flexible distributions is feasible and has the potential to reveal richer aspects of risk—for example, nonlinear persistence of unusual shocks. (JEL: C23, D31, D12)

1. Introduction

Using panel data to separate out permanent from transitory components of variation is the leading empirical approach in the analysis of individual earnings and the productivity of firms. Important economic questions have been addressed within this framework, including descriptive studies of inequality and mobility or structural links between life-cycle income and consumption.

Typically, the starting point is a matrix of covariances between one or more variables calculated at different points in time. A simple example is the error-components model. Classic examples are Hall and Mishkin (1982) on household income and consumption growth, and Abowd and Card (1989) on male earnings and hours of work. An influential recent study is Blundell, Pistaferri, and Preston (2008).

To some extent the focus on covariances is unfortunate; firstly, because the objects of economic interest are often probability distributions conditioned on past states

The editor in charge of this paper was Fabrizio Zilibotti.

Acknowledgments: This paper was originally written as a lecture for presentation as the Presidential Address to the 2013 Congress of the European Economic Association, Gothenburg, 26–30 August. It is based on the results of joint research with Richard Blundell and Stéphane Bonhomme. Many thanks to both of them. I also thank Olympia Bover, Josep Pijoan, Enrique Sentana and a referee for helpful comments, and Jose Pablo Vasquez Carvajal for able research assistance. I acknowledge research funding from the Ministerio de Economía y Competitividad, Grant No. ECO2011-26342.

E-mail: arellano@cemfi.es

Journal of the European Economic Association October 2014 12(5):1127–1153

DOI: 10.1111/jeea.12105

rather than autocorrelations—this is so when a predictive distribution of income is required to study the effect of income uncertainty on consumption decisions, or when transition probabilities are needed to measure the extent of earnings mobility; and secondly, because micro panels contain abundant information on the joint distribution of a sequence of outcomes for different periods—for example, the joint distribution of household earnings over five consecutive years.

The purpose of this paper is to review newly developed identification and estimation tools that are relevant for the analysis of dynamic dependence structures of income risk. I present an application to nonlinear permanent–transitory models of household income using data from the Panel Study of Income Dynamics (PSID), but the empirical approach is more generally applicable. Household income processes are of interest because the size of shocks, the nature of their persistence, and cross-household heterogeneity are all important to understand how income inequality varies with age and cohort and how it translates into consumption inequality.

The outline of the paper is as follows. Section 2 introduces permanent–transitory models of income risk, provides an overview of issues addressed in the literature, and presents the nonlinear dynamic model studied in Arellano, Blundell, and Bonhomme (2014, hereafter ABB). Section 3 discusses identification of two permanent–transitory models; one is a fixed-effect model and the other is the ABB Markov process model. Section 4 explains how to estimate the previous models using a simulation-based sequential calculation of quantiles and quantile regressions, which illustrate the estimation approach in Arellano and Bonhomme (2013). Section 5 reports some estimates of a nonlinear permanent–transitory model of household labor income using PSID data for the years 1998–2008. Finally, Section 6 presents the conclusions.

2. Income Processes

An income process is a central ingredient of quantitative macroeconomic models with incomplete markets and heterogeneous households. These models are a standard tool to quantify, for example, the welfare implications of changes in inequality or of changes in tax policy.[1]

In this context, an "income process" is a representation of the uncertainty about labor income in future periods that households face when deciding how much to spend and save. Faced with uncertainty, households will postpone consumption and accumulate assets to self-insure against future income shocks (precautionary savings). Moreover, the more persistent the shocks are the more difficult it will be for households to protect against those shocks.

2.1. Permanent–Transitory Models

The standard approach decomposes uncertain future income into a permanent component and a transitory component. Permanent shocks come to stay while transitory

1. See Heathcote, Storesletten, and Violante (2009, 2013) for recent surveys.

shocks only last one period. Households are assumed to be able to distinguish one type of shock from the other and to form joint probabilities about their future occurrence. The canonical permanent–transitory income process takes the form

$$Y_{it} = \eta_{it} + \varepsilon_{it} \tag{1}$$

$$\eta_{it} = \eta_{i,t-1} + v_{it}, \tag{2}$$

where Y_{it} denotes log income of household i at age t, net of cohort effects (and other demographic characteristics). The permanent component η_{it} is a random walk, while the transitory component ε_{it} is independent over time and independent of the permanent shock v_{it}.[2]

From a statistical point of view this is a standard unobserved components model with a long tradition in time series and panel data analysis. Attractive aspects of this formulation are its ability to distinguish between two sources of uncertainty, and the fact that it can be easily calibrated using panel data on household income.

In a standard calibration the variance of permanent shocks is 0.01, the variance of transitory shocks is 0.05, and the initial variance of the permanent component is 0.15 (Kaplan and Violante 2010). These figures imply that around 90% of the variation in earnings changes is due to the transitory component.

Various extensions of the canonical model have been used in the macro literature. For example, Storesletten, Telmer, and Yaron (2004) add a fixed effect to (1) and allow an autoregressive coefficient that can be less than unity in (2).[3] Guvenen (2007) adds another fixed effect interacted with an age trend and assumes that households learn about their individual-specific income profile over time. Other extensions include transitory shocks that last more than one period (a moving-average process) and variances of shocks that vary with age.

There is a large literature on longitudinal earnings in labor economics, even if some of these papers did not have an explicit objective to provide a representation of income uncertainty. For example, Moffitt and Gottschalk (1995) documented the changes in earnings inequality over the 1970s and 1980s by focusing on changes in the covariance structure of earnings. Early contributions include Hause (1977), Lillard and Willis (1978), and MaCurdy (1982), while the more recent ones include Meghir and Pistaferri (2004), Browning, Ejrnæs, and Alvarez (2010), and Altonji, Smith, and Vidangos (2013).

One- or two-error formulations? The income process (1)–(2) is a two-error model, but is well known to have an equivalent representation as a single-error integrated moving-average model. The moving-average parameter is negative and can be determined from the permanent–transitory variance ratio.[4] However, despite their

2. In the simplest form of the model η_{i1}, ε_{it}, and v_{it} are normally distributed with age-invariant standard deviations.

3. Note that (1)–(2) already includes an additive fixed effect given by the initial condition: $Y_{it} = \eta_{i1} + \left(v_{i2} + \cdots + v_{it}\right) + \varepsilon_{it}$.

4. See for example Arellano (2003, Section 5.5).

observational equivalence on income data, the two models are different representations of household uncertainty with different implications for the response of consumption to income (as noted in Quah 1990). In particular, a single income shock of moderate persistence may have quite different implications for life-cycle consumption than two shocks, one of which is highly persistent.[5]

Empirics of Consumption–Income Links. A strand of work that is closely connected to both earnings dynamics and macro consumption has sought to establish the links between income shocks and consumption choices from the joint covariance structure of earnings and expenditures. Amongst others, the papers by Hall and Mishkin (1982), Deaton and Paxson (1994), Blundell and Preston (1998), and Blundell, Pistaferri, and Preston (2008) belong to this line of research.[6] Blundell et al. (2008) interpret the degree of transmission of income shocks to consumption growth as a measure of the extent of insurance opportunities (partial insurance) available to households. A major lesson from this literature is that the degree of persistence in income shocks is key to understanding the differences in the patterns of income and consumption inequality.

The empirical work on the consumption-income links using latent covariance structures occupies a (quasi-structural) middle ground between quasi-experimental approaches and structural model-based macro approaches. In a quasi-experimental approach one seeks to estimate the response of consumption to an observable income shock whose characteristics are clear cut.[7] If successful, the quasi-experimental approach has the attraction of a strong context-specific causal claim. In contrast, a latent-structure approach seeks to construct economy-wide measures of income risk, and to understand the channels through which it affects consumption inequality.

Quasi-experiments and quasi-structures are complementary, since finding instances of context-specific causality and finding ways of measuring and interpreting broad economic concepts are both important. This is not to deny that measuring household income risk is plagued with difficulties. Some of them are reviewed next.

Heterogeneity and Advance Information. A major difficulty of estimating uncertainty from earnings data is that we cannot fully rule out the possibility of overstating uncertainty because of unaccounted heterogeneity (omitted fixed effects) or because what the model labels as a shock has been foreseen by the consumer (advance or superior information).

The canonical model (1)–(2) includes a fixed effect given by the initial permanent component, which in the Kaplan–Violante calibration accounts for 50% of the variation in log earnings residuals ten years after entering the sample. However, a model with heterogeneous growth and less persistent shocks is hard to distinguish from the

5. See for example Heathcote, Storesletten, and Violante (2009).

6. Meghir and Pistaferri (2011) contains an excellent survey of this literature.

7. For example, the response of consumption to tax rebates exploiting randomized timing in receiving tax rebate checks by households (Johnson, Parker, and Souleles 2006). See also references in Meghir and Pistaferri (2011, Section 4.4.2).

canonical model, yet it implies a very different description of the uncertainty faced by households. In fact, a fully unstructured distinction between unobserved heterogeneity and individual dynamics in a finite-horizon panel is not possible.[8] Some papers have used structural models of schooling, labor supply or consumption choices to try to distinguish between predetermined heterogeneity and shocks taking place over the working life. The list of papers includes those in the income–consumption covariance-structure literature, and also Keane and Wolpin (1997), Cunha, Heckman, and Navarro (2005) and Guvenen and Smith (2010), among others.

Similarly, earnings data alone cannot hope to conclusively identify to what extent (if any) agents are able to anticipate the stochastic variation in the estimated process. Even in joint income and consumption data, advance information and partial insurance will be observationally equivalent in general (Kaufmann and Pistaferri 2009). In effect, if agents partly anticipate the statistical innovations in the income model, an attenuated response to those shocks will be estimated, similar to what would follow under suitable insurance opportunities against those shocks.

Problems with the Quality of Earnings Data Available. First, there is an issue of right tail error. The sample design of panel data on incomes such as the PSID makes it difficult to capture variation in inequality driven by changes in the right tail of the income distribution.[9] Surveys of household wealth that oversample the rich, such as the Survey of Consumer Finances (SCF), are able to produce better measures of cross-sectional inequality but typically lack a longitudinal dimension.

More generally, there has been a long-standing concern with measurement error in survey data on earnings.[10] Unfortunately, there are not many validation studies.[11] A two-wave validation study of PSID was conducted in 1983 and 1987, which compared the survey responses of workers in a particular manufacturing plant with the payroll records of the same workers collected by the firm. It emerged that PSID earnings tended to exaggerate the actual fluctuations in earnings by between 20% and 45% depending on the year.

However, poor reliability ratios may still leave permanent–transitory decompositions largely unaffected. This possibility was suggested in Pischke (1995), whose analysis of the PSID validation data was consistent with offsetting effects of additional noise and underreporting of transitory earnings, both brought about by misreporting.

Univariate versus Multivariate Models of Earnings. Risk measures constructed from earnings histories alone are necessarily coarse as they abstract from income variation due to endogenous labor market choices rather than exogenous shocks. If households

8. See Arellano (2003, chapter 5) for a related discussion.

9. A point emphasized in Castañeda, Díaz-Giménez, and Ríos-Rull (2003).

10. See Bound, Brown, and Mathiowetz (2001) for a survey.

11. Kapteyn and Ypma (2007) is a recent one.

change hours worked or jobs in response to wage or unemployment shocks, we would like to use this information in quantifying the uncertainty they face. Given data on employment and wage histories of household members, one could consider a joint model with a more detailed set of shocks than the canonical model. In an early contribution, Abowd and Card (1989) examined the joint covariance structure of wages and hours of work of male PSID employees.

Progress in this area has been held back partly because additional labor market variables are discrete events, which are hard to combine with continuous variables in a convenient statistical framework. Thus, it has proved difficult to formulate generalized models of shock transmission that could replace standard univariate representations of labor income uncertainty. More fundamentally, it is difficult to examine a detailed menu of shocks without simultaneous consideration of a comparable menu of sequential household choices, as future exposure to a particular type of shock may be contingent on previous choices. Significant progress in this direction has been made in recent approaches developed in Low, Meghir, and Pistaferri (2010) and Altonji, Smith, and Vidangos (2013).

Recent papers that look at the transmission of wage shocks to consumption by considering the joint covariance structure of consumption, hours of work and earnings are Blundell, Pistaferri, and Saporta-Eksten (2012) and Heathcote, Storesletten, and Violante (2012). In these models labor supply is endogenous, so that wages instead of income become the exogenous source of uncertainty faced by households. In the following sections, I will discuss distributional dependence structures in the context of univariate models of earnings for simplicity, but the main ideas generalize to multivariate contexts.

Using Subjective Probabilistic Questions to Measure Income Risk. Usually we indirectly inferred risk from data on income realizations. An alternative is to directly ask survey respondents about their subjective income expectations. That is, to ask a question of the form: “What do you think is the percent chance that your total household income, before taxes, will be less than y over the next 12 months?” Dominitz and Manski (1997) report results from a survey that asked this question for four different thresholds y, together with actual income and other variables. They used responses to the probability questions to fit respondent-specific parametric distributions, which they compared with those implied from the income processes used in Hall and Mishkin (1982) and other papers in the literature. They found that subjective interquartile ranges were neither constant across households nor proportional to subjective medians. This finding is at odds with the properties of the canonical income process but not with those of the distributional models that I consider in what follows.

Expectations data from subjective probability questions can help make decisive progress on some of the key difficulties in measuring risk that we have discussed. The available evidence is that individuals are willing and able to respond to probabilistic questions about variables that are meaningful to them (Manski 2004). Much progress has been made in understanding the implications of different methods of eliciting expectations. Unfortunately, probabilistic questions are still absent from most major household income surveys.[12]

Data on subjective expectations and data on realizations can play complementary roles in constructing more credible measures of risk. Expectations data can be used to assess the assumptions made in models of expectations based on realizations. In turn those models should help fit probabilistic models to data from probabilistic questions. In particular, the combination of subjective expectations and realizations facilitates the estimation of robust measures of shock persistence (Attanasio and Augsburg 2012). A case in point is Kaufmann and Pistaferri (2009), who combine both types of data to disentangle information from insurance using the canonical income process.

Progress in this area is a complex path that requires not only the input of researchers but also of data producing agencies.

Beyond Covariance Restrictions. The mainstream approach to earnings dynamics fits the canonical income process or one of its variants to autocovariances of household income panel data. For all its limitations, the income history of a household and other similar households should play a prominent role in measuring uncertainty and persistence.

An advantage of the canonical model is its tractability. The model implies simple restrictions that can be used for identification and estimation. Moreover, linearity leads to simple approximations to consumption responses. However, a focus on covariance restrictions is overly restrictive. For example, the canonical model rules out asymmetric persistence and nonlinear transmission of shocks.

Some papers have considered income processes with heteroskedastic shocks, and/or non-normal and mixing distributions, thereby going beyond covariance analysis in significant ways. Papers in this category include Horowitz and Markatou (1996), Chamberlain and Hirano (1999), Geweke and Keane (2000), Alvarez and Arellano (2004), Meghir and Pistaferri (2004), Bonhomme and Robin (2010), Browning, Ejrnæs, and Alvarez (2010), and Hospido (2012), among others.

Next I turn to a measuring framework that shifts the focus from covariances and linear models to distributions and nonlinear transmission of income shocks.

2.2. Nonlinear Dynamics

ABB retain the permanent–transitory decomposition

$$Y_{it} = \eta_{it} + \varepsilon_{it},$$

12. The Survey of Household Income and Wealth from the Bank of Italy has been one of the few exceptions (Guiso, Jappelli, and Terlizzese 1992). Recent progress on subjective expectation data in developing countries is reviewed in Attanasio (2009).

but model the law of motion of η_{it} as an unrestricted Markovian process:

$$\eta_{it} = Q_t(\eta_{i,t-1}, V_{it}), \quad t = 2, \dots, T, \tag{3}$$

where $Q_t(\eta, v)$ is strictly monotonic increasing in v. The innovation V_{it} given $\eta_{i,t-1}, \dots, \eta_{i1}$ is uniformly distributed on (0, 1). In this setting $Q_t(\eta_{i,t-1}, \tau)$ is the τth conditional quantile of η_{it} given $\eta_{i,t-1}$.

The distribution of the initial value η_{i1} is left unrestricted. Moreover, the function $Q_t(\eta, v)$ may be nonstationary (period-specific). This allows for age-specific uncertainty and persistence, but also for aggregate shocks.

The transitory error ε_{it} is assumed to be zero-mean, independent over time. While this fits the application to the biannual PSID data well, generalizing identification and estimation to serially correlated (e.g., moving-average) transitory shocks seems important.

Another important generalization is to a richer specification of unobserved heterogeneity. The only fixed effect in (3) is the initial condition η_{i1}. I will focus on this model for simplicity, although the identification arguments in ABB allow for a nonlinear fixed effect in (3).

Although separate realizations of the components of income are of course unobservable, a nonparametric Markov process as well as the transitory shock and initial condition densities are identified from (short) panel data on total income. Identification critically hinges on the differential persistence between the components. I will return to formal identification arguments later. For now note that an implication is that a household can learn ex-ante about permanent and transitory risk not only from its own earnings history but also from those of similar households. Moreover, an economic statistician can hope to measure separate risks of different persistence based on earnings histories without data on household choices.

Nonlinear Persistence. A parameter of interest is the function

$$\rho_t(\tau) = E\left[\frac{\partial Q_t(\eta_{i,t-1}, \tau)}{\partial \eta_{i,t-1}}\right].$$

The average derivative effect $\rho_t(\tau)$ measures the persistence of $\eta_{i,t-1}$ when the process is hit by a shock V_{it} that has rank τ. The average is taken with respect to the distribution of $\eta_{i,t-1}$. Besides the mean, other characteristics of the distribution of the derivative effect are also of interest. For example, for a given shock, the persistence of positive and negative $\eta_{i,t-1}$ may differ.

In the canonical random walk model with normal shocks,

$$Q_t(\eta_{i,t-1}, \tau) = \eta_{i,t-1} + \sigma_v \Phi^{-1}(\tau),$$

so $\rho_t(\tau) = 1$, independent of τ. In contrast, in the current setup the persistence of η_{it} may depend on the magnitude and direction of the income shock V_{it}. In particular, this setting allows for stochastic volatility and asymmetric persistence. Under asymmetric persistence, the weight of the history of the process may differ according to the rank of the current shock (Koenker and Xiao 2006).[13]

A Measure of Uncertainty. Another parameter of interest is the function

$$\sigma_t(\tau) = E\left[Q_t(\eta_{i,t-1}, \tau) - Q_t(\eta_{i,t-1}, 1-\tau)\right].$$

The quantity $\sigma_t(\tau)$ measures the uncertainty generated by the presence of shocks to the persistent component of income. For example, $\sigma_t(0.75)$ is an average interquartile range.

In the canonical random walk model with $v_{it} \sim \mathcal{N}(0, \sigma_v^2)$,

$$\sigma_t(\tau) = 2\sigma_v \Phi^{-1}(\tau).$$

An analogous measure of the uncertainty generated by the transitory shocks is

$$\sigma_{\varepsilon_t}(\tau) = F_{\varepsilon_t}^{-1}(\tau) - F_{\varepsilon_t}^{-1}(1-\tau).$$

Conditional Skewness and Kurtosis. In a similar vein, the present framework leads naturally to consideration of quantile-based measures of skewness and kurtosis, along the lines of those discussed in Kim and White (2004). A measure of conditional skewness is

$$sk_t(\eta_{i,t-1}, \tau) = \frac{Q_t(\eta_{i,t-1}, \tau) + Q_t(\eta_{i,t-1}, 1-\tau) - 2Q_t(\eta_{i,t-1}, 0.5)}{Q_t(\eta_{i,t-1}, \tau) - Q_t(\eta_{i,t-1}, 1-\tau)},$$

whereas a measure of conditional kurtosis is given by

$$kr_t(\eta_{i,t-1}) = \frac{Q_t(\eta_{i,t-1}, 0.975) - Q_t(\eta_{i,t-1}, 0.025)}{Q_t(\eta_{i,t-1}, 0.75) - Q_t(\eta_{i,t-1}, 0.25)}.$$

Example: A Linear Quantile Model. A special case of model (3) is the following specification:

$$\eta_{it} = \alpha(V_{it}) + \beta(V_{it})h(\eta_{i,t-1}),$$

13. Regime-switching models that produce asymmetric persistence are popular in the time series analysis of business cycles. See for example Evans and Wachtel (1993)'s model of inflation uncertainty, or Kuan, Huang, and Tsay (2005)'s model in which an unobserved-component innovation can have permanent and transitory effects in different periods. See also Teräsvirta (1994) on smooth-transition autoregressive models.

where V_{it} are i.i.d. uniform variables, independent of $\eta_{i,t-1}, \ldots, \eta_{i1}$. In this model, $\alpha(\cdot)$ and $\beta(\cdot)$ are nonparametric functions, while $h(\cdot)$ is some pre-specified function of the lagged permanent latent variable. For example, $h(\eta_{i,t-1}) = |\eta_{i,t-1}|$ and $h(\eta_{i,t-1}) = (\eta^{+}_{i,t-1}, \eta^{-}_{i,t-1})$ are panel data counterparts of the CAViaR models in Engle and Manganelli (2004).

Since the persistence measure is given by

$$\rho_t(\tau) = \beta(\tau) E[h'(\eta_{i,t-1})],$$

this model allows shocks to affect the persistence of $\eta_{i,t-1}$ in rather general ways.

This nonlinear persistence may be rich enough to capture empirically the effects of the disparate variety of shocks (job changes, promotions, bad health, etc.), which are aggregated into the latent permanent component of household earnings.

3. Identification

3.1. Fixed-Effect Model

Let us first consider a simpler model in which the permanent component is just a fixed effect and $T = 2$:

$$Y_{i1} = \eta_i + \varepsilon_{i1} \tag{4}$$

$$Y_{i2} = \eta_i + \varepsilon_{i2}. \tag{5}$$

This model leads to the covariance structure

$$\mathrm{Var}\begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix} = \begin{pmatrix} \sigma^2_\eta + \sigma^2_{\varepsilon_1} & \sigma^2_\eta \\ \sigma^2_\eta & \sigma^2_\eta + \sigma^2_{\varepsilon_2} \end{pmatrix}.$$

The covariance between Y_{i1} and Y_{i2} identifies the permanent-component variance, which subtracted from the data variances leads to identification of the transitory-shock variances.

A similar argument can be made for the identification of the densities under the assumption that the characteristic functions are nonvanishing (Kotlarski's lemma).[14] The relationship between the (log) characteristic function of (Y_{i1}, Y_{i2}) and those of η_i, ε_{i1}, and ε_{i2} is

$$\kappa(s_1, s_2) \equiv \ln E\left[e^{\mathrm{i}(s_1 Y_1 + s_2 Y_2)}\right] = \kappa_\eta(s_1 + s_2) + \kappa_{\varepsilon_1}(s_1) + \kappa_{\varepsilon_2}(s_2),$$

14. Kotlarski (1967). See also Li and Vuong (1998) or Bonhomme and Robin (2010). The assumption of nonvanishing characteristic functions can be relaxed somewhat or replaced by tail conditions on the densities of the shocks, which may be more easily interpretable (Evdokimov and White 2012). A nonvanishing characteristic function (*cf*) excludes uniform and triangular distributions but contains common parametric distributions such as the normal, Student's t, chi-squared, gamma, and double exponential.

where s_1, s_2 are the arguments in the characteristic function and i denotes the imaginary unit. Second partial derivatives are given by

$$\begin{pmatrix} \frac{\partial^2 \kappa(s_1,s_2)}{\partial s_1^2} & \frac{\partial^2 \kappa(s_1,s_2)}{\partial s_1 \partial s_2} \\ \frac{\partial^2 \kappa(s_1,s_2)}{\partial s_1 \partial s_2} & \frac{\partial^2 \kappa(s_1,s_2)}{\partial s_2^2} \end{pmatrix} = \begin{pmatrix} \kappa''_\eta(s_1+s_2)+\kappa''_{\varepsilon_1}(s_1) & \kappa''_\eta(s_1+s_2) \\ \kappa''_\eta(s_1+s_2) & \kappa''_\eta(s_1+s_2)+\kappa''_{\varepsilon_2}(s_2) \end{pmatrix}.$$

The identification argument for second derivatives now mimics the previous one for variances. Since $\kappa_{\varepsilon_j}(0) = 0$ and $\kappa'_{\varepsilon_j}(0) = 0$ (due to zero mean), identification of second derivatives ensures identification of the log *cf*, which can be obtained by successive integration. Given the *cf*, densities follow from the inversion formula. For example,

$$f_{\varepsilon_1}(\varepsilon) = \frac{1}{2\pi}\int e^{-\mathrm{i}s\varepsilon+\kappa_{\varepsilon_1}(s)}ds.$$

3.2. Linear Markov Model

Now consider a case where the permanent component is a nonstationary linear autoregression:

$$Y_{it} = \eta_{it} + \varepsilon_{it}$$

$$\eta_{it} = \rho_t \eta_{it-1} + v_{it},$$

where transitory and permanent shocks are mutually independent at all times.[15]

The covariance structure when $T = 4$ is

$$\mathrm{Var}\begin{pmatrix} Y_{i1} \\ Y_{i2} \\ Y_{i3} \\ Y_{i4} \end{pmatrix} = \{\omega_{ts}\} = \begin{pmatrix} \sigma^2_{\eta_1}+\sigma^2_{\varepsilon_1} & \rho_2\sigma^2_{\eta_1} & \rho_3\rho_2\sigma^2_{\eta_1} & \rho_4\rho_3\rho_2\sigma^2_{\eta_1} \\ & \sigma^2_{\eta_2}+\sigma^2_{\varepsilon_2} & \rho_3\sigma^2_{\eta_2} & \rho_4\rho_3\sigma^2_{\eta_2} \\ & & \sigma^2_{\eta_3}+\sigma^2_{\varepsilon_3} & \rho_4\sigma^2_{\eta_3} \\ & & & \sigma^2_{\eta_4}+\sigma^2_{\varepsilon_4} \end{pmatrix}.$$

The transitory shock variances $\sigma^2_{\varepsilon_2}, \sigma^2_{\varepsilon_3}$ and the parameters $(\sigma^2_{\eta_2}, \sigma^2_{\eta_3}, \rho_3)$ of the covariance matrix of (η_{i2}, η_{i3}) are identified as long as ρ_3 and ρ_4 are not zero. The last-period autoregressive parameter ρ_4 is also identified, but initial and terminal variances are not identified and neither is ρ_2, despite the model implying an overidentifying restriction.[16]

15. That is, η_{it} is independent of ε_{is} for all s, t.

16. Note that $\rho_3 = \omega_{13}/\omega_{12}$ and $\rho_4 = \omega_{24}/\omega_{23}$, $\sigma^2_{\eta_2} = \omega_{23}/\rho_3$ and $\sigma^2_{\eta_3} = \omega_{34}/\rho_4$, $\sigma^2_{\varepsilon_2} = \omega_{22} - \sigma^2_{\eta_2}$ and $\sigma^2_{\varepsilon_3} = \omega_{33} - \sigma^2_{\eta_3}$.

3.3. Nonlinear Markov Model

Finally, consider the nonlinear Markovian model

$$Y_{it} = \eta_{it} + \varepsilon_{it}$$

$$\eta_{it} = Q_t(\eta_{i,t-1}, V_{it}).$$

When $T = 4$ the densities of ε_{i2}, ε_{i3} and the joint density of (η_{i2}, η_{i3}) are nonparametrically identified subject to suitable distributional dependence.

The argument proceeds by showing that the characteristic function of ε_{i2} is identified in the first three waves. Similarly, the characteristic function of ε_{i3} is identified in waves two to four. Then, given these two characteristic functions, the joint characteristic function of (η_{i2}, η_{i3}) can be identified from the characteristic function of (Y_{i2}, Y_{i3}) using the relationship

$$\psi_{Y_2,Y_3}(s_2, s_3) = \psi_{\eta_2,\eta_3}(s_2, s_3)\psi_{\varepsilon_2}(s_2)\psi_{\varepsilon_3}(s_3).$$

Some intuition and a sketch of the argument in ABB follows. This arguments builds on and extends a result in Wilhelm (2012).

Some Intuition into how Identification Works. In the Gaussian linear case, identification of the Markov process for η_{it} amounts to identification of ρ_t and $\sigma^2_{v_t}$. From an instrumental-variable perspective, we may consider the following autoregressive model with measurement error:

$$Y_{it} = \rho_t Y_{i,t-1} + \left(v_{it} + \varepsilon_{it} - \rho_t \varepsilon_{i,t-1}\right). \tag{6}$$

Using $Y_{i,t-2}$ as instrument identifies ρ_t but not $\sigma^2_{v_t}$. Only the variance of the composite error $v_{it} + \varepsilon_{it} - \rho_t \varepsilon_{i,t-1}$ is identified in the IV regression (6). Note that the validity of $Y_{i,t-2}$ as instrument depends on the Markov assumption.

To identify $\sigma^2_{\varepsilon_{t-1}}$ (given $\rho_t \neq 0$) we consider another IV regression conditioned on $Y_{i,t-2}$:

$$Y_{i,t-1}Y_{it} = -\rho_t \sigma^2_{\varepsilon_{t-1}} + \rho_t Y^2_{i,t-1} + \zeta_{it} \tag{7}$$

where the error satisfies $E(\zeta_{it} \mid Y_{i,t-2}) = 0$ under the Markovian assumption.[17] A similar regression one period ahead identifies $\sigma^2_{\varepsilon_t}$ (given $\rho_{t+1} \neq 0$). Having identified $\sigma^2_{\varepsilon_t}$ and $\sigma^2_{\varepsilon_{t-1}}$, the variance of v_{it} can be obtained from the residual variance in the first IV equation.

17. The error is given by $\zeta_{it} = Y_{i,t-1}(v_{it} + \varepsilon_{it}) - \rho_t(Y_{i,t-1}\varepsilon_{i,t-1} - \sigma^2_{\varepsilon_{t-1}})$.

In the nonlinear case, we consider two nonparametric IV equations conditioned on $Y_{i,t-2}$ using Y_{it} or some bounded function $g(Y_{it})$:

$$g\left(Y_{it}\right) = s_t\left(Y_{i,t-1}\right) + e_{it} \qquad E\left(e_{it} \mid Y_{i,t-2}\right) = 0$$
$$Y_{i,t-1} g\left(Y_{it}\right) = \tilde{s}_t\left(Y_{i,t-1}\right) + \tilde{e}_{it} \qquad E\left(\tilde{e}_{it} \mid Y_{i,t-2}\right) = 0.$$

Here $s_t(\cdot)$ is the functional parameter in the nonparametric IV regression of $g(Y_{it})$ on $Y_{i,t-1}$ using $Y_{i,t-2}$ as a conditioning instrument, whereas $\tilde{s}_t(.)$ is a similar object when the outcome is $Y_{i,t-1}g(Y_{it})$.

The function $s_t(\cdot)$ is the solution to $E[g(Y_{it}) - s_t(Y_{i,t-1}) \mid Y_{i,t-2}] = 0$. The solution exists and is unique if both the conditional distributions $Y_{i,t} \mid Y_{i,t-1}$ and $Y_{i,t-1} \mid Y_{it}$ are complete. Completeness is in this context an instrument relevance assumption. It is the nonparametric counterpart to the nonzero covariance assumption $\text{Cov}(Y_{it}, Y_{i,t-1}) \neq 0$ that is required for identification of the covariance structure.[18]

It turns out that from $s_t(\cdot)$ and $\tilde{s}_t(.)$ we can identify the density of $\varepsilon_{i,t-1}$. Similarly, from $s_{t+1}(\cdot)$ and $\tilde{s}_{t+1}(\cdot)$ we can identify the density of $\varepsilon_{i,t}$. Finally, given the densities of $\varepsilon_{i,t-1}$ and $\varepsilon_{i,t}$, and the data distribution of $(Y_{i,t-1}, Y_{it})$ one can identify the joint density of $\eta_{i,t-1}$ and η_{it} by deconvolution.

A key step in the argument is that $s_t(\cdot)$ and $\tilde{s}_t(\cdot)$ are also the IV functional parameters in the same equations but using $\eta_{i,t-1}$ as the conditioning instrument. This is so due to conditional independence between $(Y_{i,t-1}, Y_{i,t})$ and $Y_{i,t-2}$ given $\eta_{i,t-1}$, provided the conditional distribution of $\eta_{i,t-1}$ given $Y_{i,t-2}$ is complete. This situation implies

$$E\left(g\left(Y_{it}\right) \mid \eta_{i,t-1}\right) = E\left(s_t\left(Y_{i,t-1}\right) \mid \eta_{i,t-1}\right)$$
$$\eta_{i,t-1} E\left(g\left(Y_{it}\right) \mid \eta_{i,t-1}\right) = E\left(\tilde{s}_t\left(Y_{i,t-1}\right) \mid \eta_{i,t-1}\right),$$

so that for every fixed η the following equation holds:

$$E_{\varepsilon_{i,t-1}}\left[\eta s_t\left(\eta + \varepsilon_{i,t-1}\right)\right] = E_{\varepsilon_{i,t-1}}\left[\tilde{s}_t\left(\eta + \varepsilon_{i,t-1}\right)\right]. \tag{8}$$

Using this equation, by a deconvolution-type argument the density of $\varepsilon_{i,t-1}$ is identified.

Sketch of the Deconvolution Argument. Let the Fourier transforms of $E_{\varepsilon_{i,t-1}}[s_t(\eta + \varepsilon_{i,t-1})]$ and $E_{\varepsilon_{i,t-1}}[\tilde{s}_t(\eta + \varepsilon_{i,t-1})]$ be $h_t(u)$ and $\tilde{h}_t(u)$, respectively.[19] They can be

18. The distribution of $Y_{i2} \mid Y_{i1}$ is complete if $E\left[\psi\left(Y_{i2}\right) \mid Y_{i1}\right] = 0$ implies that $\psi\left(Y_{i2}\right) = 0$ for all ψ in some space of functions (Newey and Powell 2003). In a linear conditional mean model with normal regression errors, completeness reduces to the usual rank condition for linear regressions.

19. That is, $h_t(u) = \int\int s_t(\eta + \varepsilon) f_{\varepsilon_{t-1}}(\varepsilon) e^{iu\eta} d\varepsilon d\eta$ with a similar expression for $\tilde{h}_t(u)$.

written as the following products of the Fourier transform of the IV functions and the characteristic function of $\varepsilon_{i,t-1}$:

$$h_t(u) = \varphi_{s_t}(u)\,\psi_{\varepsilon_{t-1}}(-u)$$
$$\tilde{h}_t(u) = \varphi_{\tilde{s}_t}(u)\,\psi_{\varepsilon_{t-1}}(-u)$$

where $\varphi_{s_t}(u) = \int s_t(y)e^{\mathrm{i}uy}dy$ and $\varphi_{\tilde{s}_t}(u) = \int \tilde{s}_t(y)e^{\mathrm{i}uy}dy$, both of which are data objects. Moreover, because of (8) it turns out that the first derivative of $h_t(u)$ satisfies

$$h_t'(u) = \mathrm{i}\tilde{h}_t(u),$$

so that we obtain the following first-order differential equation:

$$\varphi_{s_t}'(u)\,\psi_{\varepsilon_{t-1}}(-u) - \varphi_{s_t}(u)\,\psi_{\varepsilon_{t-1}}'(-u) = \mathrm{i}\varphi_{\tilde{s}_t}(u)\,\psi_{\varepsilon_{t-1}}(-u). \tag{9}$$

Using that $\psi_{\varepsilon_{t-1}}(0) = 1$, equation (9) can be solved in closed form for $\psi_{\varepsilon_{t-1}}(u)$ provided $\varphi_{s_t}(u)$ is nonvanishing.

When $T = 4$ the previous arguments identifies $\psi_{\varepsilon_2}(u)$ in waves one to three and $\psi_{\varepsilon_3}(u)$ in waves two to four. The distributions of $Y_{i4} \mid \eta_{i3}$ and $\eta_{i2} \mid Y_{i1}$ are also identified. However, the distributions of the initial and terminal components ε_{i1}, η_{i1}, ε_{i4} and η_{i4} are not identified.

A Normal Example. As an illustration consider jointly normal data with $T = 4$ and $g(Y_{i3}) = Y_{i3}$. In this case,

$$s(y_2) = \rho_3 y_2$$
$$\tilde{s}(y_2) = \rho_3\left(y_2^2 - \sigma_{\varepsilon_2}^2\right).$$

So we have

$$\varphi_s(u) = \rho_3\kappa_1(u)$$
$$\varphi_{\tilde{s}}(u) = -\rho_3\sigma_{\varepsilon_2}^2\kappa_0(u) + \rho_3\kappa_2(u),$$

where $\kappa_j(u) = \int x^j e^{\mathrm{i}ux}dx$. Using $\kappa_1'(u) = \mathrm{i}\kappa_2(u)$ and $\kappa_0(u) = -\mathrm{i}u\kappa_1(u)$, we obtain the following special case of the differential equation (9):

$$\frac{\psi_{\varepsilon_2}'(u)}{\psi_{\varepsilon_2}(u)} = \frac{\varphi_s'(-u) - \mathrm{i}\varphi_{\tilde{s}}(-u)}{\varphi_s(-u)} = \frac{-\rho_3\sigma_{\varepsilon_2}^2 u\kappa_1(-u)}{\rho_3\kappa_1(-u)} = -u\sigma_{\varepsilon_2}^2.$$

The result corresponds to the $\mathcal{N}\left(0, \sigma_{\varepsilon_2}^2\right)$ characteristic function $\psi_{\varepsilon_2}(u) = e^{-(1/2)\sigma_{\varepsilon_2}^2 u^2}$.

Other Identification Results for Nonlinear Latent Variable Models in the Literature. Hu and Schennach (2008) and Carroll, Chen, and Hu (2010) provide general nonparametric identification results for measurement error models, and Hu and Shum (2012) for Markovian dynamic models. They rely on completeness assumptions as ABB do. The results in Hu and Schennach cover the fixed-effect model (and much more) but not the nonlinear Markov model.

Wilhelm (2012) provides an identification result for nonparametric panel data models with measurement error in a continuous explanatory variable. He restricts either the structural or the measurement error to be independent over time in order to allow past covariates or past outcomes to be used as instruments.

Hu and Shum (2012) allow for a Markov latent process, while requiring that $T \geq 5$ ($T \geq 4$ under stationarity). In addition, their identification result depends on a monotonicity assumption, such as $E\left(Y_{it+1} \mid \eta_{it} = \eta\right)$ being increasing in η. In contrast, the identification proof in ABB exploits the additivity $Y_{it} = \eta_{it} + \varepsilon_{it}$ and only requires $T \geq 4$.

4. Estimation

One way of approaching estimation is by following the identification arguments described in the previous section. However, we shall follow an alternative simulation-based approach to directly estimate quantiles of the distributions of interest. A simulation-based approach is well suited to estimate models with many latent variables. As we saw in Section 2.2, natural measures of uncertainty and persistence can then be constructed from quantiles of permanent and transitory shocks.

4.1. Fixed Effect Model: Estimation

The error-components model with a fixed effect (4)-(5) is a well-known example for which alternative estimators exist, but one that serves as a simple illustration of the Arellano–Bonhomme (2013) estimation approach. We seek to estimate the quantiles of the fixed effect η_i and those of the transitory errors given data on $Y_i = \left(Y_{i1}, \ldots, Y_{iT}\right)$. Let $\delta(\tau)$ and $\mu_t(\tau)$ be the quantile functions of η_i and ε_{it} for $\tau \in (0, 1)$.

While the mean minimizes expected squared error, the median minimizes expected absolute error and other quantiles minimize expected asymmetric absolute loss, so that for example

$$\mu_t(\tau) = \arg\min_{\mu} \; E_{Y_i, \eta_i}\left[\rho_\tau\left(Y_{it} - \eta_i - \mu\right)\right],$$

where $\rho_\tau(u) = [\tau - 1(u < 0)]u$ is the so-called "check" function (see Koenker 2005).

If we had data on both Y_i and η_i we would calculate sample quantiles of $Y_{it} - \eta_i$ and η_i. However, since η_i is unobservable, we apply the law of iterated expectations and base estimation on the following expressions in which unobservables are integrated out:

$$\mu_t(\tau) = \arg\min_{\mu} \; E_{Y_i}\left[\int \rho_\tau\left(Y_{it} - \eta - \mu\right) f_i d\eta\right] \tag{10}$$

$$\delta(\tau) = \arg\min_{\delta} \; E_{Y_i}\left[\int \rho_\tau(\eta - \delta) f_i d\eta\right], \tag{11}$$

where $f_i = f(\eta \mid Y_i)$ is the posterior density of η_i given Y_i. Since f_i depends on the model's parameters $\mu(\cdot)$ and $\delta_t(\cdot)$, we proceed in an iterative fashion. Given some initial parameter values we compute f_i. Next we compute $\mu_t(\tau)$ and $\delta(\tau)$ from sample counterparts of (10) and (11) and use those to re-evaluate f_i, etc.

There are two difficulties with this approach: one is the need to compute integrals with respect to the η; the other is the fact that the model contains a continuum of parameters.

To address the first difficulty, given a value for f_i we draw M imputations per individual from f_i say $\eta_i^{(m)}$, $m = 1, \ldots, M$. Then we compute $\mu_t(\tau)$ as

$$\arg\min_{\mu} \sum_{i=1}^{N}\sum_{m=1}^{M} \rho_\tau\left(Y_{it} - \eta_i^{(m)} - \mu\right),$$

which is a standard quantile calculation. We compute $\delta(\tau)$ in a similar way. To draw values from f_i we use the random-walk Metropolis–Hastings algorithm.[20]

To address the second difficulty, we follow Wei and Carroll (2009) and approximate the posterior density of the latent variables using spline approximations of $\mu_t(\tau)$ and $\delta(\tau)$ with knots $0 < \tau_1 < \tau_2 < \cdots < \tau_L < 1$, under the assumption that $\mu_t(\tau)$ and $\delta(\tau)$ are smooth functions. When using piecewise-linear splines, f_i is available in closed form (up to a multiplicative constant).

The end intervals $(0, \tau_1)$ and $(\tau_L, 1)$ correspond to the tails of the distributions. We specify the tails of $\delta(\tau)$ as quantiles of exponential distributions:

$$\delta(\tau) = \begin{cases} \frac{1}{\underline{\lambda}} \ln\left(\frac{\tau}{\tau_1}\right) & \text{on } \left(0, \tau_1\right), \\ \frac{1}{\bar{\lambda}} \ln\left(\frac{1-\tau_L}{1-\tau}\right) & \text{on } \left(\tau_L, 1\right). \end{cases}$$

We proceed similarly for $\mu_1(\tau)$ and $\mu_2(\tau)$ with different parameters $\underline{\lambda}$ and $\bar{\lambda}$.

This estimation method is a special case of the EM algorithm for moment conditions[21] with a Monte Carlo E-step proposed in Arellano and Bonhomme (2013)

20. The Metropolis–Hastings algorithm is a simulation technique to sample a (posterior) distribution that proceeds by generating candidates that are either accepted or rejected according to some probability, which is driven by a ratio of posterior evaluations (see for example Chib and Greenberg 1995).

21. The moment equations are the first-order conditions of integrated check function minimization.

for panel quantile regression with unobserved individual effects. Sequential method-of-moments extensions of the EM algorithm are considered in Arcidiacono and Jones (2003) and Wei and Carroll (2009) amongst others. Arellano and Bonhomme (2013) give conditions under which their (integral-based) estimator is consistent as $N, L \to \infty$ and $L/N \to 0$. We conjecture that the simulation-based estimator with M fixed is also consistent under similar conditions.

The large-N fixed-L asymptotic distribution of the quantile estimator is easily obtained using results on GMM estimation with nonsmooth moments. This analysis will provide asymptotically valid inference provided the true parameters $\mu_t(\tau)$ and $\delta(\tau)$ are piecewise-linear functions with knots $\tau_1, \ldots, \tau_L$. This is less general than the consistency result where the number of knots was allowed to increase with N in order to approximate flexible population functions. The nonparametric error-components model has been widely studied (an early paper is Horowitz and Markatou, 1996). Hall and Lahiri (2008) show that in general quantiles of η_i and ε_{it} are not root-N estimable.

4.2. Markovian Model: Estimation

Estimation of the model in which the permanent component is a Markov process instead of a fixed effect proceeds along similar lines. Instead of following the identification argument in estimation, ABB use the simulation-based sequential quantile approach, which turns out to be a simple way of dealing with the fact that permanent shocks are unobservable latent variables. We now wish to estimate the model:

$$\begin{aligned} Y_{it} &= \eta_{it} + \varepsilon_{it} \\ \eta_{it} &= Q_t\left(\eta_{i,t-1}, V_{it}\right) \quad t = 2, \ldots, T \\ \eta_{i1} &= \delta\left(V_{i1}\right) \\ \varepsilon_{it} &= \mu_t\left(U_{it}\right) \quad t = 1, \ldots, T, \end{aligned}$$

where $V_{i1}, \ldots, V_{iT}, U_{i1}, \ldots, U_{iT}$ are mutually independent, marginally uniform on $(0, 1)$.

We seek to estimate the quantile functions that describe households' uncertainty about their future income. Namely, (i) conditional quantiles of permanent shocks given the past $Q_t(\eta, \tau)$, (ii) unconditional quantiles of the initial condition: $\delta(\tau)$, and (iii) unconditional quantiles of transitory shocks: $\mu_t(\tau)$. Given $Q_t(\eta, \tau)$, measures of persistence can be readily calculated as explained before.

Anonymous Models. To estimate $Q_t(\eta, \tau)$ we wish to use a flexible model that is capable of approximating any quantile function arbitrarily well as a series approximation with a sufficient number of terms. The model is

$$Q_t(\eta, \tau) = \gamma_t(\tau)' \varphi_t(\eta),$$

where the K components of $\varphi_t(\cdot)$ belong to a dictionary of functions, and

$$\sup_{(\eta,\tau)} \left| Q_t(\eta,\tau) - \sum_{k=1}^{K} \gamma_{tk}(\tau)\,\varphi_{tk}(\eta) \right| \to 0 \text{ as } K \to \infty.$$

This approach can be described as based on "anonymous" models (borrowing the term from Ellers and Marx 1996) in the sense that the model parameters have no scientific interpretation. The interpretable objects will be summary measures of derivative effects constructed from the estimated $Q_t(\eta,\tau)$ function. See Chen (2007) for a review of the literature on nonparametric estimation using the method of sieves.

Quantile Regression. Once again we approach estimation as an incomplete data problem. If we had data on Y_{it} and η_{it}, we would calculate sample quantiles of $Y_{it} - \eta_{it}$ and η_{i1} to estimate $\mu_t(\tau)$ and $\delta(\tau)$, and quantile regressions of η_{it} on $\eta_{i,t-1}$ to estimate $Q_t(\eta,\tau)$. The difference with the fixed-effect model is that now we not only compute unconditional quantiles but also quantile regressions to estimate the conditional quantile function of the Markov process.

Quantile regression does not enforce monotonicity in estimation of conditional quantile functions (the quantile crossing problem). We address this potential complication by rearranging the estimated nonmonotonic curve into a monotonic curve (Chernozhukov, Fernández-Val, and Galichon 2010).

The Simulation-Based Sequential Quantile Approach in Action. Since $\eta_i = (\eta_{i1},\ldots,\eta_{iT})$ is unobserved, we construct instead M imputed values $\eta_i^{(m)}$, $m = 1,\ldots,M$, for each individual in the panel. Then we minimize sample check functions averaged over imputed values:

$$\min_{\mu} \sum_{i=1}^{N}\sum_{m=1}^{M} \rho_\tau\left(Y_{it} - \eta_{it}^{(m)} - \mu\right) \qquad t = 1,\ldots,T.$$

$$\min_{\delta} \sum_{i=1}^{N}\sum_{m=1}^{M} \rho_\tau\left(\eta_{i1}^{(m)} - \delta\right)$$

$$\min_{\gamma} \sum_{i=1}^{N}\sum_{m=1}^{M} \rho_\tau\left(\eta_{it}^{(m)} - \gamma'\varphi_t\left(\eta_{i,t-1}^{(m)}\right)\right) \qquad t = 2,\ldots,T$$

For the imputed values to be valid they have to be draws from the distribution of η_i conditioned on the data $f_i = f(\eta \mid Y_i)$. As before, to draw from f_i we use the random walk Metropolis–Hastings algorithm. Variances of proposals are calibrated to yield an acceptance rate of ≈ 0.4. Post-simulation, a thinned sequence of draws is retained in order to reduce autocorrelation.

The trouble is that f_i depends on the distributions of $\eta_{it} \mid \eta_{i,t-1}$, η_{i1}, and ε_{it}, which in turn depend on all the parameters we are hoping to estimate: $\mu_t(\cdot)$, $\gamma_t(\cdot)$, and $\delta(\cdot)$. In fact, by Bayes rule

$$f_i \propto \prod_{t=1}^{T} f_{\varepsilon_t}\left(Y_{it} - \eta_{it}\right) \prod_{t=2}^{T} f_{\eta_t|\eta_{t-1}}\left(\eta_{it} \mid \eta_{i,t-1}\right) f_{\eta_1}\left(\eta_{i1}\right).$$

We solve this problem using an EM-type iteration, which alternates between QR estimation and generation of imputed values based on an approximation to f_i. To approximate f_i we use the Wei–Carroll spline approximations of $\mu_t(\tau)$, $\gamma_t(\tau)$, and $\delta(\tau)$. When using piecewise-linear splines, f_i is available in closed form (up to a multiplicative constant) also in the regression case.

The end intervals $(0, \tau_1)$ and $(\tau_L, 1)$ correspond to the tails of the various distributions. The tails of $\mu_t(\tau)$ and $\delta(\tau)$ are specified as quantiles of exponential distributions. We do the same for the QR intercepts, while the other coefficients are kept constant on the end intervals $(0, \tau_1)$ and $(\tau_L, 1)$.

To sum up, the general idea is to pack together all estimates for each τ and t into an updated estimate of the posterior (E step). Given this, do τ by τ estimation for each t (M step) and so on. Using the check function allows us to decompose the M step into L different subproblems. Moreover, using the check function we get a globally convex objective function in each M step. It should be noted that this is not a calculation of maximum likelihood estimates but a method-of-moments, and that the numerical convergence of the EM algorithm for moment equations is not guaranteed in general.

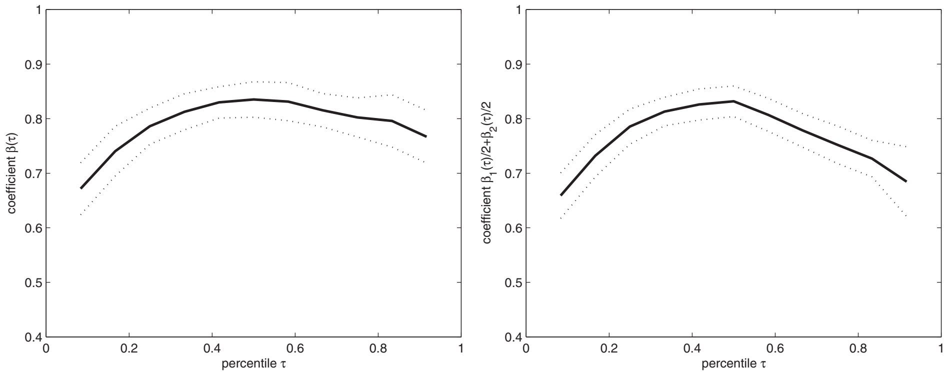
5. Empirical Results: Nonlinear Earnings Dynamics in the PSID

This section reports some estimates of a nonlinear permanent–transitory model of household labor income using data from the PSID. The emphasis is on exploring the consequences of unusual shocks (V_{it} that are close to 0 or 1) for the persistence of latent earnings (η_{it}). Unusual shocks are important for measuring risk. A flexible model may reveal empirical patterns of shock transmission that would pass unnoticed to location or location-scale models.

Data. The estimates presented here are based on PSID data for the years 1998–2008 (every other year, six waves). We select a balanced subsample of 892 households with nonzero income during the period. The PSID was collected each year until 1996 and biannually since 1997. PSID after 1999 has the advantage of including detailed information on consumption expenditures and asset holdings.[22]

We regress log total household labor income (after taxes and transfers) on dummies for year, year of birth, education, race, employment status (both members), number of

22. See Blundell, Pistaferri, and Saporta-Eksten (2012).



(a) $Q(Y_{i,t-1}, \tau) = \alpha(\tau) + \beta(\tau) Y_{i,t-1}$

(b) $Q(Y_{i,t-1}, \tau) = \alpha(\tau) + \beta_1(\tau) Y_{i,t-1}^{+} + \beta_2(\tau) Y_{i,t-1}^{-}$

FIGURE 1. Quantile autoregressions of log income: derivative effects. PSID, balanced subsample, 1998–2008, $N = 892$. Dashed lines: pointwise 95% confidence bands clustered by individual. Left graph is $\beta(\cdot)$, right graph is $(\beta_1(\cdot) + \beta_2(\cdot))/2$.

kids, number of adults, income recipient other than husband/wife, state, big city, and kids not in family unit. The variable we study, Y_{it}, is the residual in that regression, for household i in period t.[23]

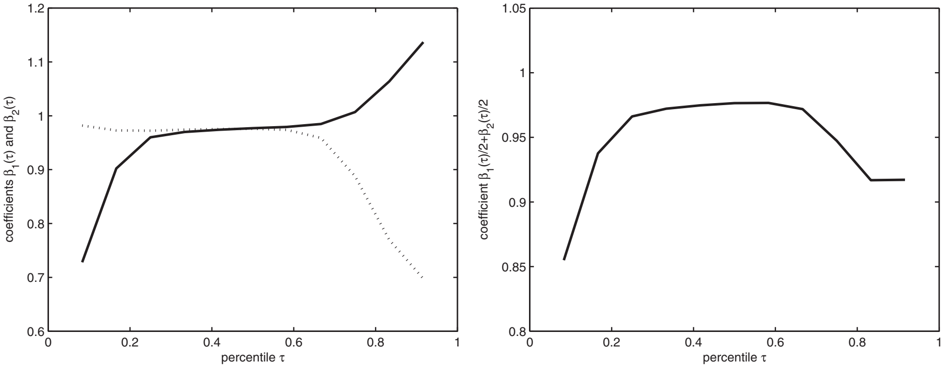
Quantile Autoregressions. Figure 1 shows plots of autoregressive effects at different percentiles for two quantile autoregressive models. The first one is linear in lagged income $Y_{i,t-1}$, while the second estimates separate effects for $Y_{i,t-1}^{+} = \max(Y_{i,t-1}, 0)$ and $Y_{i,t-1}^{-} = \min\ (Y_{i,t-1}, 0)$. The results show strong nonlinearities. A marked inverted U shape pattern is present, specially in the more flexible specification. Persistence peaks around the median at about 0.8, but declines significantly outside the central part of the distribution.

However, these results make no attempt to control for transitory shocks or measurement error. Neglected transitory shocks would be expected to attenuate nonlinearities and overall persistence.

Nonlinear Permanent-Transitory Estimates. We now take the following simple nonlinear unobserved-components model to the PSID sample:

$$Y_{it} = \eta_{it} + \varepsilon_{it},$$

23. See Blundell, Pistaferri, and Preston (2008) for a more detailed discussion of a similar specification of income residuals.



$$Q(\eta_{i,t-1}, \tau) = \alpha(\tau) + \beta_1(\tau)\eta^{+}_{i,t-1} + \beta_2(\tau)\eta^{-}_{i,t-1}$$

FIGURE 2. Permanent–transitory model: coefficient estimates and average persistence measure. Left graph shows $\beta_1(\cdot)$ (solid) and $\beta_2(\cdot)$ (dashed). Right graph shows $(\beta_1(\cdot) + \beta_2(\cdot))/2 \approx \rho(\cdot)$.

where

$$\begin{aligned} \varepsilon_{it} &= \mu(U_{it}), \\ \eta_{i1} &= \delta(V_{i1}), \end{aligned}$$

and

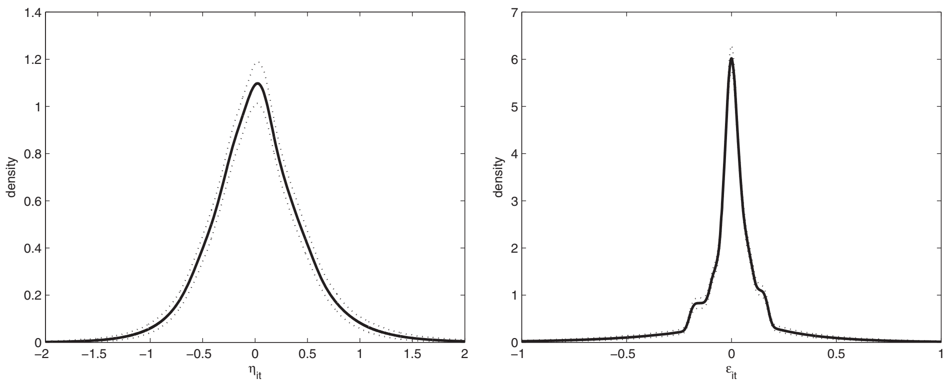
$$\eta_{it} = \alpha(V_{it}) + \beta_1(V_{it})\eta^{+}_{i,t-1} + \beta_2(V_{it})\eta^{-}_{i,t-1}, \quad t = 2, \ldots, T.$$

The functions $\alpha(\cdot)$, $\beta_1(\cdot)$, $\beta_2(\cdot)$, $\delta(\cdot)$, $\mu(\cdot)$, are piecewise-linear on $(0, 1)$ with eleven knots $\tau_1 = 1/12, \ldots, \tau_{11} = 11/12$.

The parameters $\beta_1(\cdot)$ and $\beta_2(\cdot)$ are constant on $(0, \tau_1)$ and $(\tau_{11}, 1)$, respectively, while $\alpha(\cdot)$, $\delta(\cdot)$, and $\mu(\cdot)$ are modeled as quantiles of exponential distributions, with shape parameters that are estimated together with the other parameters.

Figures 2 and 3 show the results. Figure 2 plots separate and average persistence measures. In the central range of the distribution, measured persistence for positive and negative values of $\eta_{i,t-1}$ are of similar magnitude and not far from unity, so that the unit root model would be an acceptable description for this part of the distribution. However, a very negative shock reduces the persistence of a "positive history" (a positive lagged level of η) but preserves the persistence of a negative history. At the other end, a very positive shock reduces the persistence of a negative history but preserves (or actually increases) the persistence of a good history. On average the model reproduces the inverted U shape that we saw in the quantile autoregressions, only that now the nonlinearity is more accentuated and the overall persistence is greater.

These results suggest a richer view of persistence, away from the conventional unit root versus mean reversion dichotomy.



Persistent component η_{it} Transitory shock ε_{it}

FIGURE 3. Estimated densities. Dashed lines are pointwise 95% confidence bands across 1,000 simulations. $\sigma(0.75) = 0.13$, $\sigma_\varepsilon(0.75) = 0.11$, $E[\mathrm{Var}(\eta_{it} \mid \eta_{i,t-1})] = 0.038$, $\mathrm{Var}(\varepsilon_{it}) = 0.055$.

Figure 3 shows estimated densities of η_{it} and ε_{it}. There is evidence of non-normality, especially in the density of transitory shocks, which is consistent with other deconvolution estimates in the literature (e.g., Bonhomme and Robin 2010).

Further work is needed to better understand nonlinear persistence in earnings. First, further robustness checks are needed (estimation by cohort, nonstationary specifications, additional covariates, fixed effects). Secondly, more flexible specifications of the Markov process need to be estimated. The current two-segment model of the permanent component is unnecessarily restrictive, and it has the undesirable feature of a kink at zero as a function of the lagged permanent component. Finally, so far transitory shocks and measurement error are bundled together. An external estimate of the measurement error distribution from validation data could be imported to separate one from the other by deconvolution—extending the Arellano and Bonhomme (2013) approach.

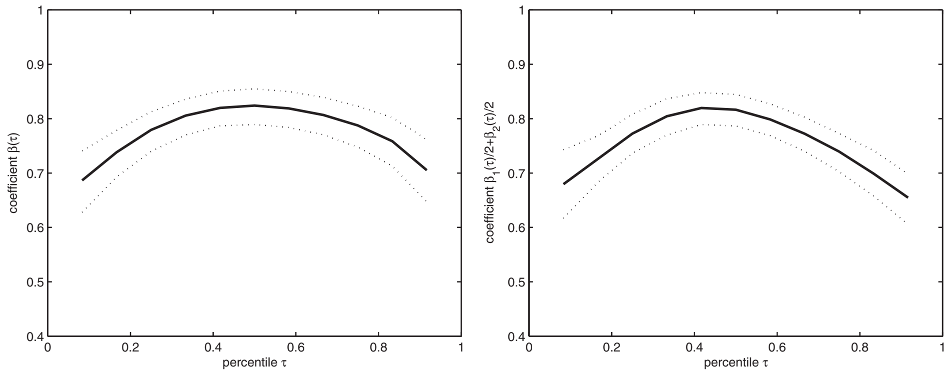
Assessing the Evidence from Simulated Data. We generated data from the estimated nonlinear permanent–transitory model and used these data to estimate quantile autoregressions. The results, reported in Figure 4, closely mimic those obtained on the original data.

Next, we estimated the canonical model on our PSID sample:

$$\begin{aligned} Y_{it} &= \eta_{it} + \varepsilon_{it} \\ \eta_{it} &= \eta_{i,t-1} + v_{it} \end{aligned}$$

Using a covariance-based minimum distance approach we find

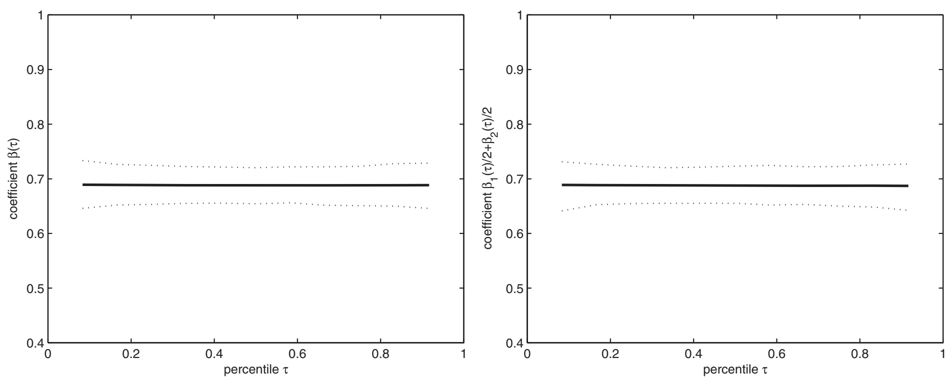
$$\mathrm{Var}\left(\eta_{i1}\right) = 0.147, \quad \mathrm{Var}\left(v_{it}\right) = 0.018, \quad \text{and} \quad \mathrm{Var}\left(\varepsilon_{it}\right) = 0.083.$$



(a) $Q(Y_{i,t-1}, \tau) = \alpha(\tau) + \beta(\tau) Y_{i,t-1}$

(b) $Q(Y_{i,t-1}, \tau) = \alpha(\tau) + \beta_1(\tau) Y^{+}_{i,t-1} + \beta_2(\tau) Y^{-}_{i,t-1}$

FIGURE 4. Dynamic QR estimates from simulated data (nonlinear permanent–transitory model). Dashed lines are pointwise 95% confidence bands across 1,000 simulations of the nonlinear model.



(a) $Q(Y_{i,t-1}, \tau) = \alpha(\tau) + \beta(\tau) Y_{i,t-1}$

(b) $Q(Y_{i,t-1}, \tau) = \alpha(\tau) + \beta_1(\tau) Y^{+}_{i,t-1} + \beta_2(\tau) Y^{-}_{i,t-1}$

FIGURE 5. Dynamic QR estimates from simulated data (canonical model). Dashed lines are pointwise 95% confidence bands across 1,000 simulations of the nonlinear model.

The variance of ε_{it} will include a measurement error component, so it is expected to be larger than the value in a standard calibration.

Assuming that the shocks and initial condition are normally distributed, we then simulate 1,000 datasets according to the canonical model, and re-estimate the log-income quantile regressions. The results are in Figure 5, which reassuringly show no sign of nonlinearities, thus suggesting that the evidence is inconsistent with the canonical model. Overall, these simulation results help us increase the confidence in our estimation methods, since they suggest that the nonlinearities that we find are not a statistical artifact of the estimation method.

6. Conclusions

Income processes are routinely used as a representation of uncertainty in macro and empirical micro analysis of consumption and savings. Typically the process is univariate and it is fit to realized income data alone. I expect progress on measuring risk to come from the use of more data but also from a more exhaustive use of income data. One important avenue of progress is to use multivariate models that combine income data with other sources of information about the uncertainty that households face. The other important avenue is to use subjective expectations data in conjunction with income histories. Using panel data on actual incomes will remain a central ingredient to both lines of development.

I have argued that going from an econometrics of autocovariances to an econometrics of flexible distributions is feasible and has the potential to reveal richer aspects of risk—for example, nonlinear persistence of unusual shocks. An econometrics of distributions will be also important in the way forward for finding practical ways of combining income data with discrete-event data and data on subjective expectations.

I have focused on a simple unobserved components model with a nonlinear Markov process. Relying on results in ABB, I have argued that the model is nonparametrically identified under standard conditions in the nonparametric literature. I have also explained how a simulation-based sequential quantile regression method developed in Arellano and Bonhomme (2013) can be used to estimate this model. Finally, I have presented some evidence of nonlinear persistence of unusual income shocks in the PSID.

The next step is to assess the implications of nonlinear income risk for consumption behavior. An important methodological step is to extend the nonlinear latent structure framework to joint examination of income and consumption. In particular, to the estimation of the degree of partial insurance to permanent and transitory income shocks. One would expect interesting interactions between nonlinear persistence of income shocks and consumption insurance opportunities. Another interesting question is to explore the extent to which consumers fully respond to nonlinearities, or whether consumers that have experienced extreme shocks respond differently to those having experienced only mid-range shocks. Further research on these issues will be hopefully reviewed in another lecture.

References

Abowd, John M. and David Card (1989). “On the Covariance Structure of Earnings and Hours Changes.” *Econometrica*, 57, 411–445.

Altonji, Joseph G., Anthony A. Smith, and Ivan Vidangos (2013). “Modelling Earnings Dynamics.” *Econometrica*, 81, 1395–1454.

Alvarez, Javier and Manuel Arellano (2004). “Robust Likelihood Estimation of Dynamic Panel Data Models.” CEMFI Working Paper No. 0421.

Arcidiacono, Peter, and John Bailey Jones (2003). “Finite Mixture Distributions, Sequential Likelihood and the EM Algorithm.” *Econometrica*, 71, 933–946.

Arellano, Manuel (2003). *Panel Data Econometrics*. Oxford University Press.

Arellano, Manuel and Stéphane Bonhomme (2013). “Random Effects Quantile Regression.” Unpublished.

Arellano, Manuel, Richard Blundell, and Stéphane Bonhomme (2014). “Household Earnings and Consumption: A Nonlinear Framework.” Unpublished.

Attanasio, Orazio (2009). “Expectations and Perceptions in Developing Countries: Their Measurement and Their Use.” *American Economic Review: Papers and Proceedings*, 99, 87–92.

Attanasio, Orazio and Britta Augsburg (2012). “ Subjective Expectations and Income Processes in Rural India.” Unpublished.

Blundell, Richard and Ian Preston (1998). “Consumption Inequality and Income Uncertainty.” *Quarterly Journal of Economics*, 113, 603–640.

Blundell, Richard, Luigi Pistaferri, and Ian Preston (2008). “Consumption Inequality and Partial Insurance.” *American Economic Review*, 98, 1887–1921.

Blundell, Richard, Luigi Pistaferri, and Itay Saporta-Eksten (2012). “Consumption Inequality and Family Labor Supply.” NBER Working Paper No. 18445.

Bonhomme, Stéphane, and Jean-Marc Robin (2010). “Generalized Nonparametric Deconvolution with an Application to Earnings Dynamics.” *Review of Economic Studies*, 77, 491–533.

Bound, John, Charles Brown, and Nancy Mathiowetz (2001). “Measurement Error in Survey Data.” In *Handbook of Econometrics*, Vol. 5, edited by James J. Heckman and Edward E. Leamer. Elsevier, pp. 3705–3843.

Browning, Martin, Mette Ejrnæs, and Javier Alvarez (2010). “Modelling Income Processes with Lots of Heterogeneity.” *Review of Economic Studies*, 77, 1353–1381.

Carroll, Raymond J., Xiaohong Chen, and Yingyao Hu (2010). “Identification and Estimation of Nonlinear Models using Two Samples with Nonclassical Measurement Errors.” *Journal of Nonparametric Statistics*, 22, 379–399.

Castañeda, Ana, Javier Díaz-Giménez, and José-Víctor Ríos-Rull (2003). “Accounting for the U.S. Earnings and Wealth Inequality.” *Journal of Political Economy*, 111, 818–857.

Chamberlain, Gary and Keisuke Hirano (1999). “Predictive Distributions Based on Longitudinal Earnings Data.” *Annales d’Economie et de Statistique*, 55-56, 211–42.

Chen, Xiaohong (2007). “Large Sample Sieve Estimation of Semi-nonparametric Models.” In *Handbook of Econometrics*, Vol. 6B, edited by James J. Heckman and Edward E. Leamer. Elsevier, pp. 5549–5632.

Chernozhukov, Victor, Iván Fernández-Val, and Alfred Galichon (2010). “Quantile and Probability Curves Without Crossing.” *Econometrica*, 78, 1093–1125.

Chib, Siddhartha and Edward Greenberg (1995). “Understanding the Metropolis–Hastings Algorithm.” *The American Statistician*, 49, 327–335.

Cunha, Flavio, James J. Heckman, and Salvador Navarro (2005). “Separating Uncertainty from Heterogeneity in Life Cycle Earnings.” *Oxford Economic Papers*, 57, 191–261.

Deaton, Angus and Christina Paxson (1994). “Intertemporal Choice and Inequality.” *Journal of Political Economy*, 102, 437–467.

Dominitz, Jeff and Charles F. Manski (1997). “Using Expectations Data To Study Subjective Income Expectations.” *Journal of the American Statistical Association*, 92, 855–867.

Ellers, Paul H. C. and Brian D. Marx (1996). “Flexible Smoothing with *B*-splines and Penalties.” *Statistical Science*, 11, 89–121.

Engle, Robert F. and Simone Manganelli (2004). “CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles.” *Journal of Business and Economic Statistics*, 22, 367–381.

Evans, Martin and Paul Wachtel (1993). "Inflation Regimes and the Sources of Inflation Uncertainty." *Journal of Money, Credit, and Banking*, 25, 475–511.

Evdokimov, Kirill and Halbert White (2012). "Some Extensions of a Lemma of Kotlarski." *Econometric Theory*, 28, 925–932.

Geweke, John and Michael Keane (2000). "An Empirical Analysis of Earnings Dynamics among Men in the PSID: 1968-1989." *Journal of Econometrics*, 96, 293–356.

Guiso, Luigi, Tullio Jappelli, and Daniele Terlizzese (1992). "Earnings Uncertainty and Precautionary Saving." *Journal of Monetary Economics*, 30, 307–337.

Guvenen, Fatih (2007). "Learning Your Earning: Are Labor Income Shocks Really Very Persistent." *American Economic Review*, 97, 687–712.

Guvenen, Fatih and Anthony Smith (2010). "Inferring Labor Income Risk from Economic Choices: An Indirect Inference Approach." Unpublished.

Hall, Peter and Soumendra N. Lahiri (2008). "Estimation of Distributions, Moments and Quantiles in Deconvolution Problems." *Annals of Statistics*, 36, 2110–2134.

Hall, Robert E. and Frederic S. Mishkin (1982). "The Sensitivity of Consumption to Transitory Income: Estimates from Panel Data on Households." *Econometrica*, 50, 461–481.

Hause, John C. (1977). "The Covariance Structure of Earnings and the On-the-Job Training Hypothesis." *Annals of Economic and Social Merasurement*, 6, 335–365.

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L. Violante (2009). "Quantitative Macroeconomics with Heterogeneous Households." *Annual Review of Economics*, 1, 319–354.

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L. Violante (2012). " Consumption and Labor Supply with Partial Insurance: An Analytical Framework." Unpublished.

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L. Violante (2013). "From Wages to Welfare: Decomposing Gains and Losses from Rising Inequality." In *Advances in Economics and Econometrics, Tenth World Congress*, Vol. 2, edited by Daron Acemoglu, Manuel Arellano, and Eddie Dekel. Cambridge University Press, pp. 235–280.

Horowitz, Joel L. and Marianthi Markatou (1996). "Semiparametric Estimation of Regression Models for Panel Data." *Review of Economic Studies*, 63, 145–168.

Hospido, Laura (2012). "Modelling Heterogeneity and Dynamics in the Volatility of Individual Wages." *Journal of Applied Econometrics*, 27, 386–414.

Hu, Yingyao and Susanne M. Schennach (2008). "Instrumental Variable Treatment of Nonclassical Measurement Error Models." *Econometrica*, 76, 195–216.

Hu, Yingyao and Matthew Shum (2012). "Nonparametric Identification of Dynamic Models with Unobserved State Variables." *Journal of Econometrics*, 171, 32–44.

Johnson, David S., Jonathan A. Parker, and Nicholas S. Souleles (2006). "Household Expenditure and the Income Tax Rebates of 2001." *American Economic Review*, 96, 1589–1610.

Kaplan, Greg and Giovanni L. Violante (2010). "How Much Consumption Insurance Beyond Self-Insurance?" *American Economic Journal: Macroeconomics*, 2, 53–87.

Kapteyn, Arie and Jelmer Y. Ypma (2007). "Measurement Error and Misclassification: A Comparison of Survey and Administrative Data." *Journal of Labor Economics*, 25, 513–551.

Kaufmann, Katja and Luigi Pistaferri (2009). "Disentagling Insurance and Information in Intertemporal Consumption Choices." *American Economic Review: Papers and Proceedings*, 99, 387–392.

Keane, Michael P. and Kenneth I. Wolpin (1997). "The Career Decisions of Young Men." *Journal of Political Economy*, 105, 473–522.

Kim, Tae-Hwan and Halbert White (2004). "On More Robust Estimation of Skewness and Kurtosis." *Finance Research Letters*, 1, 56–73.

Koenker, Roger (2005). Quantile Regression, *Econometric Society Monographs*. Cambridge University Press.

Koenker, Roger and Zhijie Xiao (2006). "Quantile Autoregression." *Journal of the American Statistical Association*, 101, 980–990.

Kotlarski, Ignacy (1967). "On Characterizing the Gamma and the Normal Distribution." *Pacific Journal of Mathematics*, 20, 69–76.

Kuan, Chung-Ming, Yu-Lieh Huang, and Ruey S. Tsay (2005). "An Unobserved-Component Model with Switching Permanent and Transitory Innovations." *Journal of Business and Economic Statistics*, 23, 443–454.

Li, Tong and Quang Vuong (1998). "Nonparametric Estimation of the Measurement Error Model Using Multiple Indicators." *Journal of Multivariate Analysis*, 65, 139–165.

Lillard, Lee A. and Robert J. Willis (1978). "Dynamic Aspects of Earnings Mobility." *Econometrica*, 46, 985–1012.

Low, Hamish, Costas Meghir and Luigi Pistaferri (2010). "Wage Risk and Employment Risk over the Life Cycle." *American Economic Review*, 100, 1432–1467.

MaCurdy, Thomas E. (1982). "The Use of Time Series Processes to Model the Error Structure of Earnings in a Longitudinal Data Analysis." *Journal of Econometrics*, 18, 83–114.

Manski, Charles F. (2004). "Measuring Expectations." *Econometrica*, 72, 1329–1376.

Meghir, Costas and Luigi Pistaferri (2004). "Income Variance Dynamics and Heterogeneity." *Econometrica*, 72, 1–32.

Meghir, Costas and Luigi Pistaferri (2011). "Earnings, Consumption and Life Cycle choices." In *Handbook of Labor Economics*, Vol. 4B, edited by O. Ashenfelter and D. Card. Elsevier, pp 774–854.

Moffitt, Robert A. and Peter Gottschalk (1995). "Trends in the Covariance Structure of Earnings in the US: 1969–1987." Unpublished.

Newey, Whitney K. and James L. Powell (2003). "Instrumental Variable Estimation of Nonparametric Models." *Econometrica*, 71, 1565–1578.

Pischke, Jörn-Steffen (1995). "Measurement Error and Earnings Dynamics: Some Estimates from the PSID Validation Study." *Journal of Business and Economic Statistics*, 13, 305–314.

Quah, Danny (1990). "Permanent and Transitory Movements in Labor Income: An Explanation for 'Excess Smoothness' in Consumption." *Journal of Political Economy*, 98, 449–475.

Storesletten, Kjetil, Chris I. Telmer, and Amir Yaron (2004). "Consumption and Risk Sharing Over the Life Cycle." *Journal of Monetary Economics*, 51, 609–633.

Teräsvirta, Timo (1994). "Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models." *Journal of the American Statistical Association*, 89, 208–218.

Wei, Ying and Raymond J. Carroll (2009). "Quantile Regression with Measurement Error." *Journal of the American Statistical Association*, 104, 1129–1143.

Wilhelm, Daniel (2012). " Identification and Estimation of Nonparametric Panel Data Regressions with Measurement Error." Unpublished.