cemmap

centre for microdata methods and practice

Nonlinear panel data methods for dynamic heterogeneous agent models

Manuel Arellano
Stéphane Bonhomme

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP51/16

An ESRC Research Centre

Nonlinear Panel Data Methods for Dynamic Heterogeneous Agent Models*

Manuel Arellano† Stéphane Bonhomme‡

October 2016

Abstract

Recent developments in nonlinear panel data analysis allow identifying and estimating general dynamic systems. In this review we describe some results and techniques for nonparametric identification and flexible estimation in the presence of time-invariant and time-varying latent variables. This opens the possibility to estimate nonlinear reduced forms in a large class of structural dynamic models with heterogeneous agents. We show how such reduced forms may be used to document policy-relevant derivative effects, and to improve the understanding and facilitate the implementation of structural models.

JEL code: C23.
Keywords: dynamic models, structural economic models, panel data, unobserved heterogeneity.

*This work is prepared for the *Annual Review of Economics*. Arellano acknowledges research funding from the Ministerio de Economía y Competitividad, Grant ECO2016-79848-P.

†CEMFI, Madrid.
‡University of Chicago.

1 Introduction

Many economic settings are characterized by the presence of nonlinear relationships. Nonlinearity may arise due to the presence of risk aversion, substitution effects and complementarities in production, or nonlinear constraints such as borrowing restrictions or budget kinks, for example. Such nonlinearities are a pervasive feature of dynamic structural models.

On the other hand, dynamic econometric analysis has traditionally been based on linear methods. Vector autoregression (VAR) methods are commonly used to document the dynamic propagation of shocks in linear systems. As an example, in the analysis of earnings and consumption dynamics, covariance-based methods motivated by linearized representations are prominent in the literature (Hall and Mishkin, 1982, Blundell, Pistaferri and Preston, 2008).

In this paper we review recent work in panel data analysis which demonstrates the possibility to identify and estimate general nonlinear dynamic systems. Our main focus is on the development of flexible estimation methods, with the ability to allow for the presence of time-invariant or time-varying latent variables. Such latent variables are important to capture unobserved heterogeneity and unobserved state variables.

The motivation for those methods is twofold. A first aim is to relax linearity assumptions and reveal empirical nonlinearities in the data. Returning to the example of earnings and consumption, nonlinear persistence and propagation of shocks in earnings has been shown to be a feature of the PSID in the United States, and to be also present in Norwegian administrative data (Arellano, Blundell and Bonhomme, 2016). In addition, nonlinearities in the earnings process have implications for the nature of earnings risk, and thus for consumption and saving decisions (Arellano, 2014).

A second motivation for nonlinear panel data methods is to be able to model and estimate flexible reduced forms that are compatible with classes of structural dynamic models. We review several workhorse models in the literature, including life-cycle models of nondurable or durable consumption, intensive or extensive labor supply, saving behavior, and models of firm-level production functions. The nonlinear reduced forms of all these models may be analyzed and estimated using the methods reviewed in this paper.

The focus on the nonlinear reduced forms of dynamic models is useful for several reasons. We show that policy-relevant quantities such as average marginal propensities to consume or measures of insurability to income shocks may be recovered from the joint reduced-form distribution, without the need to fully specify a structural model. These quantities are obtained from the nonlinear policy rules of the dynamic problem, without resorting to linearized approximations.

Panel data-based nonlinear reduced-form methods offer a significant advantage compared to their cross-sectional or time-series counterparts in the way they allow for latent variables. Under suitable dynamic assumptions, panel data provides the opportunity to nonparametrically identify some latent variables that are key state variables in the decision problem, such as individual ability or preferences, firm productivity, or latent human capital profiles. This leads to a richer set of conditioning variables, identified from the panel dimension.

A well-known drawback of any reduced-form analysis compared to a structural approach is the inability to perform general counterfactual exercises. Fully structural approaches are commonly used in the settings that we take as examples in this paper (e.g., Gourinchas and Parker, 2002, Guvenen and Smith, 2014). However, structural estimation requires the researcher to specify all aspects of the model, including often tightly specified functional forms.

An important advantage of a *nonlinear* reduced form approach is to document moments and other features of the distribution of the observed data and the latent variables, which provide robust targets for a structural estimation exercise. In addition, both identification and estimation of a nonlinear reduced form may be a useful first step even when the final goal is to take a structural model to the data.

A central feature of the dynamic models we consider is the presence of dynamic restrictions on sequential conditioning sets. Those are typically the consequence of Markovian assumptions in the economic model. Such assumptions provide dynamic exclusion restrictions which are instrumental in establishing identification, as shown in Hu and Shum (2012) and the econometric literature on nonlinear models with latent variables recently reviewed in Hu (2015). Existing identification results allow for time-invariant heterogeneity (that is, for "fixed-effects"), but also for the presence of time-varying unobserved state variables under suitable conditions.

Nonlinear panel data approaches may help on the estimation side too. In linear models, panel data estimators often show unstable behavior; see for example Banerjee and Duflo (2003). One source of such instability is model misspecification, providing a motivation for relaxing linearity assumptions. The goal of a nonlinear panel data approach to estimation is to achieve flexibility in dynamic contexts, similarly to the aim of matching approaches in cross-sectional settings.

Flexible estimation of nonlinear dynamic reduced forms may be achieved by relying on sieve semi- or nonparametric approaches (Chen, 2007), where the dimension of the model grows with the sample size. In this review we focus in particular on the quantile-based estimator developed in Arellano and Bonhomme (2016) for continuous outcome variables.

The approach we advocate here has a wide scope, and this review only touches the tip of the iceberg of potential applications. In particular, we mostly focus on single agent models with *continuous* outcomes. In the last section we discuss discrete choice models, which have a similar structure and have been extensively studied. An important difference with the main focus of the paper concerns identification, which becomes more challenging in discrete outcomes models. We also briefly mention extending the approach beyond single agent settings, as has been recently done in Bonhomme, Lamadon and Manresa (2016a) in the context of models of workers and firms with two-sided heterogeneity.

The outline is as follows. In Section 2 we describe the reduced-form patterns of the dynamic models we study, and we provide specific economic examples in Section 3. In Section 4 we discuss what can be learned from reduced-form distributions in this context. Sections 5, 6, and 7 are devoted to identification, specification, and estimation of nonlinear dynamic models, respectively. Lastly, we discuss related approaches in Section 8, and we conclude in Section 9 with directions for future work.

2 Nonlinear dynamic econometric systems

In this section we first describe the general econometric pattern of the reduced form in a class of heterogeneous agent dynamic structural economic models. In the next section we will provide several examples of structural models which are special cases of the class we consider.

2.1 Models with time-invariant heterogeneity

The model has N agents (e.g., firms, households, or individuals) observed over T periods, and outcome variables Y_{it} and covariates X_{it}, observed for $i = 1, ..., N$ and $t = 1, ..., T$. The Y's typically include choice variables and payoff variables, while the X's may be state variables that are observed by the econometrician. Let $(Y_i, X_i) = (Y_{i1}, ..., Y_{iT}, X_{i1}, ..., X_{iT})$ be the vector of observations for agent i. In addition, there are determinants α_i that the researcher does not observe. Here we consider the case where α_i represents time-invariant unobserved heterogeneity. We will consider the case of time-varying state variables $\alpha_i = (\alpha_{i1}, ..., \alpha_{iT})$ in the next subsection.

Our aim is to characterize the nonlinear reduced form in a class of dynamic economic models. In the models we consider, the joint distribution of observables and unobservables has some key features, which we now discuss. Throughout we use f as a generic notation for a distribution function, and we denote $Z_i^t = (Z_{i1}, ..., Z_{it})$.

The first feature is one of *limited memory*. In models with time-invariant α's, we assume that

$$f(Y_{it} \,|\, Y_i^{t-1}, X_i^t, \alpha_i) = f\left(Y_{it} \,|\, Y_{i,t-1}, X_{it}, X_{i,t-1}, \alpha_i\right). \tag{1}$$

Under (1), the dependence of Y_{it} on past values of Y and X is limited to the last period. This could be generalized to allow for dependence on the last s periods, where $s \geq 1$. As we will see later in this paper, such Markovian conditional independence restrictions are natural in many economic settings where the relevant state variables are low-dimensional, and general identification results may be established under such conditions.[1] Note that, in the present setting, the Markov property holds conditionally on latent variables, and the models will generally not be unconditionally Markovian.

The second main feature of the models we consider is one of *sequential exogeneity*; that is, we do not rule out dynamic feedback conditionally on a fixed effect.[2] Moreover, the feedback process is also assumed to have limited memory given α_i. Specifically, we assume that

$$f(X_{it} \,|\, Y_i^{t-1}, X_i^{t-1}, \alpha_i) = f(X_{it} | Y_{i,t-1}, X_{i,t-1}, \alpha_i). \tag{2}$$

Such feedback effects are often key ingredients in dynamic structural models, since future state variables X (such as labor market experience) are typically affected by past choices Y (such as participation).

[1] An interesting extension of (1) is to assume that period-t outcomes are conditionally independent of the past, given a sufficient statistic which is a low-dimensional function $Z_{it} = g(Y_i^{t-1}, X_i^t)$ of past Y's and X's, as in

$$f(Y_{it} \,|\, Y_i^{t-1}, X_i^t, \alpha_i) = f(Y_{it} \,|\, Z_{it}, \alpha_i).$$

[2] Specifically, the Chamberlain-Sims strict exogeneity condition does not hold (Chamberlain, 1982):

$$f(Y_{it} \,|\, Y_i^{t-1}, X_i^t, \alpha_i) \neq f(Y_{it} \,|\, Y_i^{t-1}, X_i^T, \alpha_i),$$

and the conditional distribution $f(Y_{it} \,|\, Y_i^{t-1}, X_i^t, \alpha_i)$ is regarded as the object of interest.

Equipped with (1) and (2), and continuing with the case where the unobserved α's are time-invariant, the conditional likelihood function for agent i given α_i and initial conditions takes the form

$$
\begin{aligned}
& f(Y_{i2}, X_{i2}, ..., Y_{iT}, X_{iT} \,|\, Y_{i1}, X_{i1}, \alpha_i) \\
& = \prod_{t=2}^{T} f\left(Y_{it} \,|\, Y_{i,t-1}, X_{it}, X_{i,t-1}, \alpha_i\right) f(X_{it}|Y_{i,t-1}, X_{i,t-1}, \alpha_i).
\end{aligned} \tag{3}
$$

Note that, in this first-order Markovian setup, initial conditions consist of the vector (Y_{i1}, X_{i1}).

The presence of the *latent* α's is a third key feature of the model. Examples of unobserved state variables abound in economics, such as individual ability or preferences, or firm productivity. In fixed-effects approaches, the time-invariant α_i's are conditioned upon and estimated. In random-effects approaches, which we will focus on in this paper, the conditional likelihood function in (3) is augmented with a specification for the conditional distribution of α_i given Y_{i1} and X_{i1},

$$
f(\alpha_i \,|\, Y_{i1}, X_{i1}). \tag{4}
$$

We will describe such approaches in some detail in Section 6.

2.2 Time-varying unobserved state variables

An important extension of the model is to allow for time-varying unobserved state variables α_{it}. In that case, we replace (1) and (2) with

$$
f(Y_{it} \,|\, Y_i^{t-1}, X_i^t, \alpha_i^t) = f\left(Y_{it} \,|\, Y_{i,t-1}, X_{it}, X_{i,t-1}, \alpha_{it}, \alpha_{i,t-1}\right), \tag{5}
$$

and

$$
f(X_{it}|Y_i^{t-1}, X_i^{t-1}, \alpha_i^t) = f(X_{it}|Y_{i,t-1}, X_{i,t-1}, \alpha_{it}, \alpha_{i,t-1}), \tag{6}
$$

respectively, and add the following assumption on the feedback process for α's:

$$
f(\alpha_{it}|Y_i^{t-1}, X_i^{t-1}, \alpha_i^{t-1}) = f(\alpha_{it}|Y_{i,t-1}, X_{i,t-1}, \alpha_{i,t-1}). \tag{7}
$$

In this case the *complete data* likelihood function, which is the joint likelihood of observed and latent variables, takes the following form:

$$
\begin{aligned}
& f(Y_{i2}, X_{i2}, \alpha_{i2}, ..., Y_{iT}, X_{iT}, \alpha_{iT} \,|\, Y_{i1}, X_{i1}, \alpha_{i1}) \\
& = \prod_{t=2}^{T} f\left(Y_{it} \,|\, Y_{i,t-1}, X_{it}, X_{i,t-1}, \alpha_{it}, \alpha_{i,t-1}\right) \\
& \qquad \times f(X_{it}|Y_{i,t-1}, X_{i,t-1}, \alpha_{it}, \alpha_{i,t-1}) f(\alpha_{it}|Y_{i,t-1}, X_{i,t-1}, \alpha_{i,t-1}).
\end{aligned}
$$

Fixed-effects approaches do not have the ability to handle time-varying unobservables, so they are not feasible in this case. In a random-effects approach the model is completed by specifying the conditional distribution of α_{i1} given Y_{i1} and X_{i1}.

We next illustrate through various examples the ability of this framework to capture dynamic relationships in a number of structural economic models.

3 Structural economic examples

In this section we consider several examples of dynamic structural models with heterogeneous agents, and we emphasize some features of their nonlinear reduced forms.

3.1 Models of earnings risk

We start by describing models for the dynamics of household earnings. The properties of earnings processes, such as the presence of nonlinearities and the form of heterogeneity, have important implications for the evolution of consumption and saving in life-cycle models (Arellano, 2014). We will later consider models of consumption and other household decisions.

Let us focus on two different processes for the log-earnings $\ln W_{it}$ of household i in period t, with distinct dynamic properties: a first-order Markov process, and a process with transitory innovations. In the first specification, log-earnings follow the process

$$\ln W_{it} = Q_W(\ln W_{i,t-1}, U_{it}), \tag{8}$$

where U_{it} are independent of W_i^{t-1} and independent over time. The nonlinear function Q_W is not restricted *a priori*. A simple, popular example of (8) is an autoregressive process of the form $\ln W_{it} = \mu + \rho \ln W_{i,t-1} + U_{it}$, with $|\rho| < 1$.

In the second specification, log-earnings are the sum of a persistent η component and a transitory ε component,

$$\ln W_{it} = \eta_{it} + \varepsilon_{it}, \quad \eta_{it} = Q_\eta(\eta_{i,t-1}, V_{it}), \tag{9}$$

where V_{it} are independent of η_i^{t-1} and all ε_{is}, the V's and ε's are independent over time, and Q_η is a nonlinear function. Persistent/transitory dynamic specifications such as (9) are very common in the literature; see for example Meghir and Pistaferri (2011). A prominent special case of (9) is the linear permanent/transitory model with $\eta_{it} = \eta_{i,t-1} + V_{it}$. Arellano *et al.* (2016) consider a more general, quantile-based version of (9) which allows for nonlinear transmission of earnings shocks. They provide evidence of nonlinear effects based on PSID data and Norwegian administrative data.

Those two earnings processes may easily be extended to allow for age or time variation, where the collinearity between age and time may be dealt with by pooling different cohorts. In addition, both (8) and (9) may be extended to allow for unobserved heterogeneity. As an example, the Markov transitions Q_η in (9) may allow for an unobserved time-invariant factor ζ_i, as in $\eta_{it} = Q_\eta(\eta_{i,t-1}, \zeta_i, V_{it})$.

Such nonlinear specifications may also be used to model the dynamic evolution of other state variables which are relevant to measuring the risk faced by households or individuals. Beyond income risk, similar statistical models may be suitable in the context of health risk or financial/wealth risk, for example.

3.2 Earnings and consumption dynamics

Our second example is a standard incomplete market model of consumption and saving over the life-cycle (e.g., Huggett, 1993, Kaplan Violante, 2010). An agent in the model is a household who earns W_{it} every period until retirement at age T. The household consumes C_{it}, and has access to a risk-free bond, the quantity A_{it} of which evolves according to the following budget constraint

$$A_{it} = (1+r)A_{i,t-1} + W_{i,t-1} - C_{i,t-1}, \tag{10}$$

with the possible addition of borrowing constraints.

Preferences are separable over time, with period-specific preferences $u(C_{it}, \nu_{it})$ over consumption. The time-varying taste shifters ν_{it} are not observed by the econometrician. In a specification without time-invariant heterogeneity in preferences, the ν_{it} are assumed to be independent over time. In addition, the utility function may also depend on a time-invariant household unobserved factor ξ_i, reflecting permanent heterogeneity in preferences.

Time is discounted at rate β. Household i maximizes the expected intertemporal discounted sum of utilities

$$\mathbb{E}_1 \left(\sum_{t=1}^{T} \beta^{t-1} u(C_{it}, \nu_{it}) \right),$$

subject to (10). Household log-earnings $\ln W_{it}$ evolve stochastically, following either the Markov process (8) or the persistent/transitory process (9).

To derive the form of the consumption policy rule in the model, let us focus on the case where agents' information sets and beliefs are standard, in that W_{it} (respectively, η_{it} and ε_{it}) are known to the agent at time t, while only the distribution of future W's (resp., η's and ε's) is known to them, not their specific realizations. Under standard conditions on the utility function, the consumption rule then takes the following form in the case of process (9):

$$C_{it} = g_t\left(A_{it}, \eta_{it}, \varepsilon_{it}, \nu_{it}\right), \tag{11}$$

where g_t is an age-specific function and the relevant state variables are period-t assets, the two earnings components, and the taste shifters, in addition to age. In the case of process (8) there is one state variable less, and consumption takes the form (for a different function g_t):

$$C_{it} = g_t\left(A_{it}, W_{it}, \nu_{it}\right). \tag{12}$$

When the taste shifter ν_{it} is scalar and the marginal utility of consumption is increasing with respect to it, then both consumption functions (11) and (12) are increasing in their last argument. In this particular case, the consumption function g_t may be shown to be identified under suitable conditions, up to normalizing the distribution of ν_{it}. In the absence of restrictions on the independent taste shifters ν_{it}, average derivative effects such as partial insurance coefficients will still be identified even though g_t is not. We will discuss identification in Section 5.

To summarize the description of the model, the evolution of log-earnings, consumption, and assets is given by either (8)-(12) or (9)-(11) depending on the earnings process considered, and the budget constraint (10). Given the stochastic assumptions that we have made, the reduced form of the model is thus a particular example of the setup introduced in Section 2, with $Y_{it} = C_{it}$ and $X_{it} = (W_{it}, A_{it})$. In the case of the permanent/transitory earnings process (9), the time-varying unobserved state variables are the persistent earnings components $\alpha_{it} = \eta_{it}$. In the case of process (8) with unobserved time-invariant earnings heterogeneity $\alpha_i = \zeta_i$, the latter is the latent state variable.

As shown in Arellano *et al.* (2016), this simple framework may be generalized in a number of ways, through a simple modification of the arguments of the consumption function. As already discussed, an empirically relevant extension is to allow for time-invariant unobserved heterogeneity in households' preferences (or discount factors) ξ_i. The resulting consumption function then takes the following form in the case of process (9):

$$C_{it} = g_t\left(A_{it}, \eta_{it}, \varepsilon_{it}, \xi_i, \nu_{it}\right). \tag{13}$$

As a further extension, one may allow for the presence of consumption habits, leading to the following form for the consumption rule:

$$C_{it} = g_t\left(C_{i,t-1}, A_{it}, \eta_{it}, \varepsilon_{it}, \xi_i, \nu_{it}\right). \tag{14}$$

Other extensions include allowing for advance information on future earnings shocks, or for different types of assets with different returns, possibly differing in their degree of liquidity as in Kaplan and Violante (2014). In the last case, the composite consumption rule becomes an age-specific function of all types of assets, in addition to the earnings components and unobserved tastes.

3.3 Consumption and labor supply

Consider now an extension of the life-cycle model of the previous subsection which allows for individual labor supply decisions. Household i comprises two individuals, who work H_{1it} and H_{2it} hours in period t at hourly wages w_{1it} and w_{2it}, respectively. As in Blundell, Pistaferri and Saporta-Eksten (2016), a unitary household maximizes the expected discounted sum

$$\mathbb{E}_1\left(\sum_{t=1}^{T}\beta^{t-1}u(C_{it}, H_{1it}, H_{2it}, \nu_{it})\right),$$

subject to

$$A_{it} = (1+r)A_{i,t-1} + w_{1i,t-1}H_{1i,t-1} + w_{2i,t-1}H_{2i,t-1} - C_{i,t-1},$$

with additional constraints on assets and hours worked.

Log-wages $\ln w_{1it}$ and $\ln w_{2it}$ follow dynamic processes of the form (8) or (9), with different parameters for the two household members. In addition, wage shocks such as (U_{1it}, U_{2it}) in (8), and (V_{1it}, V_{2it}) and $(\varepsilon_{1it}, \varepsilon_{2it})$ in (9), are allowed to be dependent within households.

In this model the household consumption rule and labor supply rules take the following form in the case where individual log-wages follow permanent/transitory processes (9):

$$C_{it} = g_t\left(A_{it}, \eta_{1it}, \eta_{2it}, \varepsilon_{1it}, \varepsilon_{2it}, \nu_{it}\right), \tag{15}$$

$$H_{1it} = h_{1t}\left(A_{it}, \eta_{1it}, \eta_{2it}, \varepsilon_{1it}, \varepsilon_{2it}, \nu_{it}\right), \tag{16}$$

$$H_{2it} = h_{2t}\left(A_{it}, \eta_{1it}, \eta_{2it}, \varepsilon_{1it}, \varepsilon_{2it}, \nu_{it}\right), \tag{17}$$

for some age-specific functions g_t, h_{1t} and h_{2t}, with a similar expression when log-wages follow process (8). In the absence of restrictions on ν_{it} and its dimensionality (beyond the fact that the ν_{it} are independent over time and independent of the state variables), one may identify general average derivative effects in (15)-(16)-(17), as we shall see in the next section, although it is generally not possible to fully identify the functions g_t, h_{1t}, and h_{2t}.

This model may be generalized by allowing for an extensive participation margin. In the absence of state-dependent costs of participation, the resulting policy rules are very similar to (15)-(16)-(17), except that some of the labor supply outcome variables are binary participation indicators. In the presence of costs of participation, the two lagged participation indicators enter as additional state variables, hence as additional arguments in the consumption and labor supply rules.

3.4 Durable consumption

Following Berger and Vavra (2015), consider a standard incomplete market model with a durable consumption margin, subject to fixed costs of adjustment. Let D_{it} denote the durable stock of household i in period t. The household maximizes

$$\mathbb{E}_1\left(\sum_{t=1}^{T}\beta^{t-1}u(C_{it},D_{it},\nu_{it})\right),$$

subject to

$$A_{it}=(1+r)A_{i,t-1}+W_{i,t-1}-C_{i,t-1}+(1-\delta)D_{i,t-2}-D_{i,t-1}-F(D_{i,t-1},D_{i,t-2}), \qquad (18)$$

and subject to lower bounds on D_{it} and A_{it}. In (18), δ and F denote the depreciation rate on durables and the fixed cost to adjust the durable stock, respectively.

The nondurable and durable consumption rules then take the following form, in the case of the persistent/transitory earnings process (9):

$$C_{it} = g_t\left(A_{it},D_{i,t-1},\eta_{it},\varepsilon_{it},\nu_{it}\right), \qquad (19)$$

$$D_{it} = h_t\left(A_{it},D_{i,t-1},\eta_{it},\varepsilon_{it},\nu_{it}\right), \qquad (20)$$

with again a similar expression under process (8).[3]

[3] Note that, in this model, one may also be interested in the *potential* consumption decisions $C_{it}(1)$ and $C_{it}(0)$ associated with the household adjusting or not adjusting the durable stock. In general, identifying the joint reduced-form distribution of consumption of durables and nondurables, and possibly earnings components and assets, will not be sufficient to identify the distributions of those potential consumption choices. It would be of interest to provide sufficient (albeit model-specific) conditions for the identification of such objects based on the reduced form.

In addition to a partial equilibrium model, Berger and Vavra (2015) consider a general equilibrium version of the model, where wages and interest rates are endogenous and subject to aggregate uncertainty. Their focus is on the effect of business cycle fluctuations on durable consumption expenditure patterns. As the consumption policy rules in (19) and (20) are t-dependent, their general forms are unchanged in this case, although the specific forms of the functions g_t and h_t differ between the partial and general equilibrium versions of the model. In that case the reduced-form distributions vary with calendar time in addition to age. However, in the approach pursued in this paper the effect of aggregate shocks, while allowed for, is left un-modeled. Disentangling the effects of micro-level and macro-level shocks in dynamic systems is an interesting question for future work.

3.5 Production function and unobserved productivity

In our last example, let now Y_{it} denote the output of a firm i at time t. Let K_{it} denote capital input, and ω_{it} denote latent productivity, where we abstract from labor inputs for simplicity. Production is given by

$$Y_{it} = Q_Y(K_{it}, \omega_{it}, \varepsilon_{it}), \tag{21}$$

where ε_{it} are independent of all inputs, and independent over time. An example is a multiplicative specification of the form $Y_{it} = \omega_{it} Q_Y(K_{it}, \varepsilon_{it})$. The laws of motion of capital and productivity are given by

$$\begin{aligned} K_{it} &= (1-\delta)K_{i,t-1} + I_{i,t-1}, \\ \omega_{it} &= Q_\omega(\omega_{i,t-1}, V_{it}), \end{aligned} \tag{22}$$

where δ is the depreciation rate, $I_{i,t-1}$ is firm's investment chosen at time $t-1$ which becomes productive at t, and V_{it} are independent shocks. According to (22), latent productivity follows a nonlinear first-order Markov process.

As in Olley and Pakes (1996), firms choose investment in each period in order to maximize expected profits net of investment costs. The future ε and ω values are not observed to the firm. The state variables at t are K_{it}, ω_{it}, t itself (which reflects the economic environment faced by the firm), and some stochastic determinants of costs ν_{it} which we assume to be independent of other state variables and independent over time. The investment rule then takes the form

$$I_{it} = g_t(K_{it}, \omega_{it}, \nu_{it}), \tag{23}$$

for a nonlinear function g_t. In the absence of ν's, and under monotonicity of g_t in (23) with respect to ω_{it}, Olley and Pakes (1996) propose to invert that relationship so as to proxy for unobserved productivity using observed quantities in a linear version of (21).[4]

The model of output, capital, and investment outlined here is a special case of the setup discussed in Section 2. Estimating such a model makes it possible to take nonlinear production functions with unobserved inputs to firm-level panel data. In addition, the present setting may be generalized to allow for an R&D decision R_{it} influencing the evolution of the latent productivity process, as in

$$\omega_{it} = Q_{\omega}(\omega_{i,t-1}, R_{i,t-1}, V_{it}), \tag{24}$$

along the lines of Doraszelski and Jaumandreu (2013).

3.6 Other examples

Dynamic economic models which have similar dynamic reduced-form implications as the ones reviewed above are very common in the literature. Consider as an example a model where unobserved state variables α_{it} are dynamically affected by past decisions $Y_{i,t-s}$. Such dynamic feedback effects are present in models of endogenous human capital accumulation, such as in the classic Ben Porath (1967) model. While the latent earnings component $\alpha_{it} = \eta_{it}$ is assumed to be strictly exogenous in (9), in the class of models considered in Section 2 the latent η_{it} may be affected by past choices such as labor supply or investment. The latent productivity process in (24) is also sequentially exogenous but not strictly exogenous, due to it being affected by past R&D expenditures. In the identification discussion in Section 5 we will show the possibility to identify such dynamic feedback effects based on Markovian assumptions.

In the analysis of firm decisions, models of investment, inventory and markups have a similar structure (e.g., Aguirregabiria, 1999). The large literature on dynamic discrete choice models also studies settings with related reduced-form implications; see, e.g., Rust (1994) or the survey by Aguirregabiria and Mira (2010), and see Section 8 below. Many other examples may be found in the literature on recursive macroeconomic models (e.g., Ljungqvist and Sargent, 2004).

[4]See Levinsohn and Petrin (2003), Ackerberg *et al.* (2015), and also Huang and Hu (2011) for related approaches to estimating production functions.

4 Learning from nonlinear reduced forms

In this section we turn to the question of how to interpret nonlinear dynamic reduced forms such as the ones we introduced in Section 2.

To fix ideas, let us start by focusing on the simple life-cycle consumption and saving model of Subsection 3.2, in the presence of the persistent-transitory earnings process (9) and permanent unobserved heterogeneity in preferences ξ_i. In this case the reduced form is a joint distribution of consumption, assets, earnings, and latent earnings components, over time, with the addition of the time-invariant unobserved heterogeneity.

A first observation is that average derivative effects on consumption, assets and earnings may be recovered from the joint reduced-form distribution. As an example, the following average marginal propensity to consume out of the persistent earnings component η_{it} is identified from the reduced-form, as

$$\begin{aligned}\phi_C(a,\eta,\varepsilon,\xi) &= \mathbb{E}\left[\frac{\partial g_t\left(a,\eta,\varepsilon,\xi,\nu_{it}\right)}{\partial\eta}\right] \\ &= \frac{\partial}{\partial\eta}\mathbb{E}\left[C_{it} \,|\, A_{it}=a,\eta_{it}=\eta,\varepsilon_{it}=\varepsilon,\xi_i=\xi\right]. \end{aligned} \tag{25}$$

Equation (25) is a consequence of the independence between the taste shifters ν_{it} and the state variables $(A_{it},\eta_{it},\varepsilon_{it},\xi_i)$; see Matzkin (2013), for example. In general, the conditional distribution of the individual derivative effect $\partial g_t\left(a,\eta,\varepsilon,\xi,\nu_{it}\right)/\partial\eta$ (over realizations of the taste shifters ν_{it}) is *not* identified, although identification holds in the special case where the consumption function is increasing with respect to a scalar ν_{it}.[5]

The average derivative effect in (25) is indicative of the degree of "partial insurance" in the sense of Blundell *et al.* (2008), the quantity $1-\phi_C(a,\eta,\varepsilon,\xi)$ being a measure of consumption insurability of shocks to the persistent earnings component η_{it}. It is worth noting that, in the fully nonlinear setup considered here, $1-\phi_C(a,\eta,\varepsilon,\xi)$ is a *nonlinear* measure of partial insurance, which is heterogeneous in various dimensions: with respect to assets, the two earnings components, and unobserved household heterogeneity ξ_i. Moreover, other average derivative effects may be similarly recovered, such as the average marginal propensity of consume out of assets $\mathbb{E}\left[\partial g_t\left(a,\eta,\varepsilon,\xi,\nu_{it}\right)/\partial a\right]$.

The ability to recover average derivative effects based on the nonlinear reduced form extends to models with multiple choices, such as the model of consumption and labor supply

[5]This condition is a rank preservation assumption, as made in Chernozhukov and Hansen (2005), for example.

outlined in Subsection 3.3. In that model, a particularly interesting effect is the following average derivative (for $k = 1, 2$ and $\ell = 1, 2$ denoting the two household members):

$$\begin{aligned}\phi_{H_{k\ell}}(a, \eta_1, \eta_2, \varepsilon_1, \varepsilon_2, \xi) &= \mathbb{E}\left[\frac{\partial h_{kt}\left(a, \eta_1, \eta_2, \varepsilon_1, \varepsilon_2, \xi, \nu_{it}\right)}{\partial \eta_\ell}\right] \\ &= \frac{\partial}{\partial \eta_\ell}\mathbb{E}\left[H_{kit} \,|\, A_{it} = a, \eta_{1it} = \eta_1, \eta_{2it} = \eta_2, \varepsilon_{1it} = \varepsilon_1, \varepsilon_{2it} = \varepsilon_2, \xi_i = \xi\right].\end{aligned} \tag{26}$$

The quantity $1 - \phi_{H_{k\ell}}(a, \eta_1, \eta_2, \varepsilon_1, \varepsilon_2, \xi)$ is a measure of an individual's labor supply insurability to persistent shocks to her own income (when $k = \ell$) or to her spouse's income (when $k \neq \ell$). When adding a labor force participation decision to the model, one may similarly define measures of insurability of *total* labor supply, including both positive and zero hours, to income shocks.

In addition to average derivatives of choice variables such as consumption or labor supply, one can also document nonlinear measures of *persistence* of the state variables. An important example is the persistent/transitory model (9) of earnings risk. Following Arellano *et al.* (2016) one may compute the following measure of persistence of the latent component η_{it},

$$\rho(\eta_{i,t-1}, \tau) = \mathbb{E}\left[\frac{\partial Q_\eta(\eta_{i,t-1}, \tau)}{\partial \eta}\right],$$

at different values of the state $\eta_{i,t-1}$ and the innovation τ. This quantity will be identified under the assumption that Q_η be strictly increasing in its second argument, and V_{it} be normalized to follow a standard uniform random variable. In that case τ belongs to the unit interval. Note that making those assumptions on Q_η and V_{it} amounts to expressing the shocks to the persistent earnings component in rank form. Similar measures could be computed for latent firm-level productivity in the production function example discussed in Subsection 3.5.

The availability of a nonlinear dynamic reduced-form specification also allows one to perform impulse response analysis. While it is very common to document impulse responses based on linear vector autoregressions (VARs), a disadvantage of such analyses is that VAR models are linear ones, and may thus not be appropriate to capture the rich empirical nonlinearities predicted by a structural model. With this motivation in mind, Berger and Vavra (2014) revisit the question of the cyclical behavior of durable consumption expenditures using a nonlinear VAR method proposed in Auerbach and Gorodnichenko (2012). This analysis provides a complement to the structural model in Berger and Vavra (2015), which we outlined in Subsection 3.4. The nonlinear panel data methodology described in this paper provides a flexible alternative to such approaches, which (when nonparametric) has the ability to fully nest the structural model of interest, while it also allows one to account for the presence of state variables that are not observed by the econometrician.

However, in contrast to fully structural approaches, identifying and estimating the nonlinear reduced form in a dynamic economic model is not sufficient to assess the counterfactual predictions of the model. As an example, in the life-cycle consumption and saving model of Subsection 3.2, consider a counterfactual exercise consisting in a modification of the process of earnings dynamics, such as an increase in its degree of persistence. In such a case the consumption rule will most likely be different in the new regime, due to the fact that consumption decisions are partly determined by the nature of earnings risk. Hence the nonlinear reduced-form distribution of consumption is not invariant to the counterfactual change. In other words, reduced-form exercises, even flexible ones, are subject to the Lucas (1976) critique.

Nevertheless there are important benefits to a nonlinear dynamic approach based on flexible reduced forms in order to identify, estimate, and understand structural economic models. Regarding identification, as we will describe in the next section, the recent literature on nonlinear models with latent variables provides conditions under which dynamic panel data models are nonparametrically identified based on Markovian restrictions. Existing results allow for the presence of time-invariant heterogeneity, but also time-varying unobserved state variables. Hence, starting from a structural model, these results may be useful in a first step to identify the joint distribution of observed variables and unobserved heterogeneity, before turning in a second step to the specific analysis of the identification of the structural parameters.

A general nonlinear specification of the dynamic relationships may be useful to estimate structural models too. Estimates of the nonlinear reduced form provide robust targets for the structural model, which could be used in simulated method of moments or indirect inference estimation, for example. They also provide tools to test specific implications of a structural model. Many economic models have predictions for the full distribution of the data, beyond first and second moments, for which the nonlinear approach described here is well-suited. In addition, the fact that the nonlinear reduced form contains observed *and* unobserved variables (to the econometrician) may also be useful for structural estimation.

5 Identification based on short panels

As illustrated in Section 2, dynamic economic models typically feature conditional independence restrictions, which hold as a result of Markovian assumptions on state dependence. They also feature the presence of latent variables, either time-invariant or time-varying. The literature on nonlinear models with latent variables has established general conditions guaranteeing nonparametric identification in such settings, based on short panels. In linear models with independent errors, nonparametric identification follows from the Kotlarski (1967) lemma and related results based on nonparametric deconvolution. Hall and Zhou (2003), Hu (2008), Hu and Schennach (2008), and Hu and Shum (2012), among others, extend the identification argument to nonlinear models; see Hu (2015) for a recent survey of this literature.

The general setup analyzed in Hu and Shum (2012, HS hereafter), which builds on the static case studied in Hu and Schennach (2008), is useful for our purpose since it is based on Markovian dynamic restrictions. The setup applies to cases where latent variables α_{it} are time-varying. Models with time-invariant α_i, leaving the conditional distribution of α_i given initial conditions unrestricted, will also be covered as special cases of the general framework. The analysis in HS relies on the following two conditional independence assumptions, where we denote $Z_{it} = (Y_{it}, X_{it})$.

Assumption 1

(i) (Z_{it}, α_{it}) *is conditionally independent of* $(Z_i^{t-2}, \alpha_i^{t-2})$ *given* $(Z_{i,t-1}, \alpha_{i,t-1})$.

(ii) Z_{it} *is conditionally independent of* $\alpha_{i,t-1}$ *given* $(Z_{i,t-1}, \alpha_{it})$.

Hu and Shum refer to *(i)* and *(ii)* in Assumption 1 as "first-order Markov" and "limited feedback" assumptions, respectively. Part *(i)* is a consequence of the Markovian restrictions in (5), (6), and (7). In contrast, part *(ii)* is an additional assumption on the reduced-form system. Note that this assumption is satisfied in all the examples described in Subsections 3.1 to 3.5. Under the two parts in Assumption 1, HS establishes nonparametric identification of the Markovian law of motion $f(Y_{it}, X_{it}, \alpha_{it} \,|\, Y_{i,t-1}, X_{i,t-1}, \alpha_{i,t-1})$ in a fully nonstationary setting, based on $T = 5$ periods of data, under a number of additional assumptions which we now briefly discuss.

The identification argument in HS relies on invertibility, or "completeness" conditions which certain operators are assumed to satisfy. Such assumptions are commonly made in the literature on nonparametric instrumental variables (Newey and Powell, 2003). Completeness intuitively requires some nonparametric counterpart to a rank condition to hold. It is worth noting, however, that completeness conditions are high-level ones, and that they are not testable in general (Canay *et al.*, 2013). Primitive conditions for completeness have only been derived so far in a handful of cases (D'Haultfoeuille, 2011, Hu and Shiu, 2012). In addition, the conditions in HS require $f(Y_{it}, X_{it} \,|\, Y_{i,t-1}, X_{i,t-1}, \alpha_{it})$ to depend on (Y_{it}, X_{it}), $(Y_{i,t-1}, X_{i,t-1})$, and α_{it}, in such a way that a certain spectral decomposition is unique. Lastly, in HS a scaling assumption is imposed, which consists in assuming that, for example, for some (y, x) the conditional expectation $\mathbb{E}(Y_{it} \,|\, Y_{i,t-1} = y, X_{i,t-1} = x, \alpha_{i,t-1} = \alpha)$ is strictly increasing in α. In that case, identification is defined up to an arbitrary monotonone transformation of $\alpha_{i,t-1}$.

Earnings and consumption dynamics (continued). Consider as an example the life-cycle model of earnings, consumption and saving described in Subsection 3.2, in the presence of the persistent/transitory earnings process (9). In that case, (Y_{it}, X_{it}) contains log-earnings, consumption, and assets, α_{it} is the latent persistent earnings component, and Assumption 1 is satisfied due to the dynamic properties of the model. Hence the model falls into the general setup considered in HS, so identification follows provided their other assumptions are also satisfied; in particular, $T \geq 5$ periods of continuous or mixed discrete/continuous observations are needed for their result to hold.

Alternatively, in this setting, Wilhelm (2015) provides a constructive identification argument that may be applied to the earnings process alone. The consumption function may then be identified by solving recursive nonparametric instrumental variables problems using lags and leads of earnings as instruments; see Arellano *et al.* (2016). Those restrictions extend the instrumental variables restrictions used in Hall and Mishkin (1982) and the subsequent literature on consumption partial insurance to a fully nonlinear setting. These authors also provide conditions for identification in the presence of latent time-invariant preference heterogeneity ξ_i.

6 Specifying nonlinear dynamic systems

The reduced-form distributions in the dynamic models we consider consist of two main parts: the conditional distributions of outcomes (or covariates) given past values, and the distribution of unobserved heterogeneity. We now describe how each part may be specified in a flexible, yet tractable manner.

6.1 Flexible conditional distributions

Flexible models of distributions may be constructed based on sieve approaches, such as orthogonal polynomial or spline specifications; see Gallant and Nychka (1987), and Chen (2007) for a review of sieve methods in econometrics. As a starting case, consider modeling the distribution of a binary X_{it} given past values of X and Y, and some factors α. A series specification is as follows

$$\Pr\left(X_{it} = 1 \,|\, Y_{i,t-1}, X_{i,t-1}, \alpha_i\right) = \Lambda\left(\sum_{k=1}^{K} a_k \varphi_k(Y_{i,t-1}, X_{i,t-1}, \alpha_i)\right), \tag{27}$$

where Λ is a strictly increasing function mapping the real line to the unit interval. An example is $\Lambda(u) = (1+\exp(-u))^{-1}$, corresponding to a series logit specification. The functions φ_k belong to a pre-specified family such as ordinary polynomials, orthogonal polynomials, or splines, for example. Similar specifications may be used for discrete X's more generally.

In the presence of continuously distributed outcomes or covariates, a convenient alternative approach is to model conditional distributions via their inverses; that is, through conditional *quantile* functions. This approach was recently proposed by Arellano and Bonhomme (2016). Consider a linear quantile specification for the outcome variables

$$Y_{it} = \sum_{k=1}^{K} a_k(U_{it}) \varphi_k(Y_{i,t-1}, X_{it}, X_{i,t-1}, \alpha_i), \tag{28}$$

with a similar specification for the feedback process of sequentially exogenous covariates

$$X_{it} = \sum_{k=1}^{K} b_k(V_{it}) \varphi_k(Y_{i,t-1}, X_{i,t-1}, \alpha_i), \tag{29}$$

for different choices of K and functions φ_k for Y and X, where U's and V's are independent standard uniform random variables, independent over time.

It is assumed that the right-hand sides in (28) and (29) are strictly increasing in U_{it} and V_{it}, respectively, so these equations effectively model the conditional quantile functions of Y and X. This allows for a flexible modeling of the dependence of Y and X on their past values and unobserved heterogeneity, for example through polynomial interaction terms. In addition, in this specification conditioning variables may have a nonlinear effect on outcomes at various points of the distributions, through the U's and V's. All a and b parameters may be allowed to vary unrestrictedly with time, thus allowing for general nonstationarity. Arellano *et al.* (2016) illustrate the ability of such specifications to reveal nonlinear empirical relationships.

The functions $a_k(\tau)$ and $b_k(\tau)$ in (28) and (29) are defined on the unit interval, for $\tau \in (0,1)$. Wei and Carroll (2013) proposed to specify them as piecewise-linear splines. That is, the unit interval is divided into $L+1$ sub-intervals, with knots $\tau_1 < \tau_2 < ... < \tau_L$. On each $(\tau_\ell, \tau_{\ell+1})$, $a_k(\tau)$ and $b_k(\tau)$ are linear. These functions of τ are also continuous at τ_ℓ and $\tau_{\ell+1}$. The parameters of the model are then the $a_{k\ell} = a_k(\tau_\ell)$ and $b_{k\ell} = b_k(\tau_\ell)$, for $k = 1, ..., K$ and $\ell = 1, ..., L$. The functions may be extended to the tail subintervals $(0, \tau_1)$ and $(\tau_L, 1)$ using a parametric form for the intercept parameters in a's and b's, as detailed in Arellano and Bonhomme (2016).

In the presence of multiple covariates, this approach can be extended using triangular specifications such as

$$X_{mit} = \sum_{k=1}^{K} b_{mk}(V_{mit})\varphi_k(Y_{i,t-1}, X_{i,t-1}, X_{1it}, ..., X_{m-1,it}, \alpha_i),$$

based on a sequential ordering of variables. Note that the ordering would be irrelevant in a fully nonparametric specification.

An alternative approach, which does not require postulating such an ordering and may thus be particularly well-suited to model the joint distribution of multiple choice variables, is based on a copula specification. To illustrate this approach, consider modelling the distribution of two outcome variables (for example, consumption and hours of work) as follows:

$$\begin{aligned} Y_{1it} &= \sum_{k=1}^{K} a_{1k}(U_{1it})\varphi_k(Y_{i,t-1}, X_{it}, X_{i,t-1}, \alpha_i), \\ Y_{2it} &= \sum_{k=1}^{K} a_{2k}(U_{2it})\varphi_k(Y_{i,t-1}, X_{it}, X_{i,t-1}, \alpha_i), \end{aligned}$$

where U_{1it} and U_{2it} both follow standard uniform marginal distributions, and are jointly independent of the state variables $(Y_{i,t-1}, X_{it}, X_{i,t-1}, \alpha_i)$. This quantile-based modeling of the two marginal distributions is completed by specifying a copula C for the bivariate random variable (U_{1it}, U_{2it}). In practice, a parametric specification for C may be based on a low-dimensional family (such as Frank, Gumbel, or Gaussian) or on a more flexible choice such as the Bernstein family (Sancetta and Satchell, 2004); see Nelsen (1999) and Joe (1997) for references on copulas.

6.2 Unobserved heterogeneity

Allowing for unobserved heterogeneity in estimation may be based on two general approaches: fixed-effects or random-effects. In a fixed-effects approach the α_i's are conditioned upon, and estimated together with the other parameters of the model (the $a_{k\ell}$ and $b_{k\ell}$ in the previous subsection). Evidently, a fixed-effects approach is not able to deal with time-varying unobserved heterogeneity such as α_{it}.

In a correlated random-effects approach the researcher specifies the conditional distribution of unobserved heterogeneity. Consider the case of time-invariant heterogeneity α_i. A possibility is to model the conditional distribution of α_i given covariates and initial conditions as a Gaussian distribution with linear mean and constant variance (Chamberlain, 1980). Other common specifications include letting α_i be discretely distributed, possibly with covariate-dependent type probabilities.

A different, quantile-based approach is introduced in Arellano and Bonhomme (2016). They specify α_i using a quantile-based model, as follows:

$$\alpha_i = \sum_{k=1}^{K} c_k(W_i)\varphi_k\left(Y_{i1}, X_{i1}\right), \tag{30}$$

where the W_i's are independent standard uniform random variables, independent of all other random variables. The $c_k(\tau)$ functions are specified in a similar way as in (28) and (29). The aim of this specification is to allow for flexible dependence between unobserved heterogeneity and initial conditions. Misspecifying the form of this dependence is a well-known source of bias in dynamic models (Heckman, 1981), so a flexible approach is appealing in this context. Distributional specifications for multiple unobservables may be obtained through a triangular approach or a copula modeling, as we outlined above. A triangular specification based on a sequential ordering may be appealing in this context.

Lastly, a similar approach can be used to deal with the presence of time-varying unobservables. As an example, one may specify the feedback process of unobserved state variables α_{it} as

$$\alpha_{it} = \sum_{k=1}^{K} c_k(W_{it})\varphi_k\left(Y_{i,t-1}, X_{i,t-1}, \alpha_{i,t-1}\right), \tag{31}$$

where the W_{it}'s are independent standard uniform random variables. An analogous specification may be used to model the conditional distributions of the initial α's.

7 Algorithms for flexible estimation

The main challenge to estimate the nonlinear dynamic systems considered here is due to the presence of latent variables. In this section we describe how recently developed econometric methods may provide tractable estimators in those settings. We focus on correlated random-effects methods, although fixed-effects methods could also be a possibility in models with time-invariant unobserved heterogeneity.

The likelihood functions in Section 2 are mixtures of likelihoods, with respect to an underlying time-invariant α_i or a time-varying sequence $(\alpha_{i1}, ..., \alpha_{iT})$. As a result, estimation methods for mixture models are well-suited. A natural approach is the Expectation-Maximization (EM) algorithm of Dempster, Laird and Rubin (1977). Related alternative methods may be based on Markov Chain Monte Carlo (MCMC) techniques; see for example Lancaster (2004) for a review of panel data applications of those techniques. EM and MCMC methods both alternate between updates of two types of parameters: the ones that enter the conditional distributions of outcome variables and covariates (such as the a's in (28)), and the ones that enter the distributions of latent variables (such as the c's in (30)). The E-step in EM requires computing integrals with respect to latent variables, a task which may be challenging in models with time-varying unobserved states α_{it}. MCMC methods do not require computing such integrals, as they only require drawing from the posterior distribution of latent variables, and may thus be easier to apply in those settings.

Recently, Arellano and Bonhomme (2016, AB hereafter) introduce an estimation method tailored to the quantile-based specifications which we described in Section 6. Their approach is based on a stochastic EM algorithm (Celeux and Diebolt, 1993), which shares a number of features with EM and MCMC. The algorithm proceeds by iteratively repeating the following two steps until convergence to a stationary regime, parameter estimates being computed as means of a large number of realizations of the resulting chain.

In the first step, the latent variables α_i are drawn from their posterior distribution, with M draws per individual. Given some values of the parameters, coming from the previous iteration of the algorithm, the joint *complete data* likelihood function implied by a model such as (28)-(29)-(30) is easy to compute, so one can readily draw from the associated posterior distribution using a Metropolis Hastings sampler.

In the second step, parameter updates are computed given the latent draws. That is, in the case of the quantile-based model (28)-(29)-(30), the a's, b's and c's are estimated using simple linear quantile regressions. As an example, the $a_{k\ell} = a_k(\tau_\ell)$ in (28) for $k = 1, ..., K$ are estimated through a quantile regression of outcome variables Y_{it} on functions of state variables $\varphi_k(Y_{i,t-1}, X_{it}, X_{i,t-1}, \alpha_i^{(m)})$, at each percentile τ_ℓ, where the $\alpha_i^{(m)}$ are the imputed values of the unobserved component drawn from the posterior distribution in the first step. The quantile regression objective is convex and efficient optimization routines are available (e.g., Koenker and Bassett, 1978, Koenker, 2005), making this second step computationally tractable too.

The second step easily accommodates the presence of specifications other than quantile-based ones. As an example, one could model the outcome distribution in (28) through a nonlinear conditional mean model instead. In such a case, the a_k's would be updated through a nonlinear regression of outcomes on functions of state variables and imputed values. Likewise, in models with binary or other discrete outcome variables, such as a durable consumption decision or a participation margin in labor supply, parameters in the discrete choice model may be updated through (series) logit or probit, for example. As another example, when using a copula modeling for multivariate outcome variables or covariates, the copula parameters may be updated via a maximum likelihood step.

AB provide details on the implementation of this stochastic EM algorithm. The method can readily be generalized to allow for time-varying α_{it}'s, as done in an application to earnings and consumption dynamics in Arellano *et al.* (2016). A difference with the time-invariant heterogeneity case is that one then needs to draw M *sequences* $(\alpha_{i1}^{(m)}, ..., \alpha_{iT}^{(m)})$ for each individual. Efficient simulation methods such as particle filtering (e.g., Herbst and Schorfheide, 2015) may be used for this purpose.

Statistical properties. In parametric settings, the asymptotic properties of estimators based on stochastic EM have been characterized in Nielsen (2000), who provides conditions for root-N consistency and asymptotic normality, and gives the expression of asymptotic variances. The algorithm outlined in this section differs from the standard stochastic EM algorithm since it is not based on likelihood functions but on quantile-based estimating equations; see Elashoff and Ryan (2004) for properties of the EM algorithm based on estimating equations. AB adapts the asymptotic derivations in Nielsen (2000) to this setting.

Using quantile-based steps as opposed to likelihood steps for parameter updates in AB's algorithm is motivated by computational considerations, since doing so allows one to split the parameter updates into τ_ℓ-specific updates, and since this exploits the convexity of the objective function of quantile regression. As in related settings based on partial likelihood functions (e.g., Arcidiacono and Jones, 2003), this sequential approach is in general less efficient than full maximum likelihood. In practice, inference may be based on empirical counterparts to the analytical variance-covariance matrix, or on re-sampling methods such as the bootstrap or subsampling.

Given the goal of the approach described in this paper, which aims at achieving flexible estimation of nonlinear reduced forms of economic models, it is conceptually appealing to see the parametric specification as an approximation to a nonparametric joint distribution which becomes more accurate in larger samples. This means that one should conduct the asymptotic analysis in a setting where K (the number of functions of state variables) and L (the number of knots in the spline model for quantile specifications) tend to infinity as the number of individuals N increases. Some progress has recently been made in this direction. AB provide conditions for consistency of their stochastic EM-based estimator in this joint asymptotic. Belloni *et al.* (2016) develop inference methods for the whole quantile process in series quantile regression models. Extending the latter results to provide joint inference on all reduced-form parameters in a fully nonparametric setting in the presence of latent variables is still an unsolved question.

8 Related approaches

In this last section we briefly review recent work in two directions which we have not yet considered in this paper: dynamic discrete choice models, and models with interactions between agents and multi-sided heterogeneity.

8.1 Discrete outcomes

The focus of this review is on models with continuous or mixed discrete/continuous outcomes, such as consumption and extensive labor supply, for example. When all outcomes of interest are discrete, related methods have been proposed in the literature on structural dynamic discrete choice models. Classic examples are Rust (1987) and Keane and Wolpin (1997). See

Aguirregabiria and Mira (2010) for a survey of those methods.

Discrete outcomes models with continuous latent variables are generally not point identified, however. Kasahara and Shimotsu (2009) and Browning and Carro (2014) provide conditions for identification under the assumption that the latent variables are time-invariant and have a finite (known) number of points of support. Establishing identification of reduced-form distributions in the presence of latent variables may be useful as a step toward establishing identification of the structural model. Partial identification results have been obtained in simple discrete choice panel data models in Honoré and Tamer (2006). Recent work by Connault (2016) considers discrete choice models with time-varying unobserved state variables. Studying identification further in those discrete settings seems an important research avenue.

On the estimation side, alternatives to full-solution estimation of structural models have been proposed in the literature (e.g., Hotz and Miller, 1988, Aguirregabiria and Mira, 2002, Su and Judd, 2012). The approach advocated in this paper is closest to the first stage in the two-stage estimator proposed by Arcidiacono and Miller (2011), where the conditional choice probabilities and the probabilities of the unobserved discrete types are estimated jointly. Arcidiacono and Miller then propose estimating the structural parameters in a second stage, motivating this approach on computational grounds.

8.2 Beyond single agent models

The main focus of this review is on single agent models. Extending these models to allow for interactions between agents (such as husband and wife, village members, or workers and firms) is an active research area.

Bonhomme *et al.* (2016a) propose a model of wages and worker/firm sorting for matched employer-employee panel data. They consider a setup with two-sided unobserved heterogeneity. In a similar spirit to the general approach advocated in this paper, they model the joint distribution of wages and mobility decisions under certain dynamic assumptions which they show to hold in a number of theoretical models of sorting such as wage posting models or models with wage bargaining. The state variables of the economic model are the time-invariant worker and firm latent types, as well as the wages, thus allowing for a relaxation of network exogeneity assumptions commonly made in this literature. Similarly as in Section 2, the models they consider imply Markovian conditional independence restrictions which are used to establish nonparametric identification under suitable rank conditions. The estimated reduced-form distribution may then be used to perform variance decomposition exercises in the spirit of Abowd, Kramarz and Margolis (1999), or more generally distributional decomposition exercises quantifying the effects of worker heterogeneity, firm heterogeneity, and allocation patterns of workers to firms, on the wage distribution.

In a setting with two-sided latent heterogeneity, a correlated random-effects approach to estimation is challenging to implement, due to the complex structure of the likelihood function in this case. Bonhomme *et al.* (2016a) propose to treat firm heterogeneity as fixed-effects, while modeling worker heterogeneity using a correlated random-effects specification. The main insight is that, conditional on the firm effects, the structure of the model is analogous to a single agent model such as the ones we have focused on in the previous sections. In addition, in order to reduce dimensionality and make the approach tractable in short panels they rely on a discretization of firm-level heterogeneity. The statistical properties of this approach are studied in Bonhomme *et al.* (2016b), in a setting where population unobserved heterogeneity is continuously distributed and the discretization is seen as an approximation.

More generally, there are many important dynamic economic models for which the single agent focus of this paper is not appropriate. Examples can be found the literature on dynamic games (e.g., Aguirregabiria and Mira, 2007, Pesendorfer and Schmidt-Dengler, 2008). Other related examples may be found in the literature on dynamic models of economic networks (e.g., Jackson, 2009). Generalizing the approach presented in this paper to such settings seems a promising avenue.

9 Conclusion

Increased data availability provides opportunities to document novel nonlinear economic relationships. Examples where nonlinearities have been shown or suggested to matter empirically are the analysis of earnings, consumption and wealth (Arellano *et al.*, 2016, Guvenen *et al.*, 2016), dynamic public finance models (Golosov and Tsyvinski, 2015), or models of asset pricing (Constandinides and Gosh, 2016, Schmidt, 2015).

A large econometric literature has developed methods to achieve robustness to functional forms, including semi-parametric and nonparametric methods, and bounds approaches. However, these developments have so far mostly been limited to cross-sectional settings. In contrast, dynamic panel data models have typically been analyzed in tightly parametric settings, most often linear ones. In this perspective, the aim of the recent work reviewed in this paper is to develop such a robust approach for dynamic systems, in the presence of nonlinearities and unobserved heterogeneity.

The tools we have reviewed concern both identification and estimation. Regarding the former, economic assumptions on the relevant state variables and their evolution imply dynamic exclusion restrictions which may be used to establish identification, similarly as in linear models. Regarding the latter, flexible estimation methods based on quantile specifications or other sieves make it possible to take rich nonlinear models to panel data. Importantly, these methods allow for the presence of time-invariant heterogeneity or time-varying latent variables, which are often key state variables in the economic model. We have reviewed recent advances based on simulation methods. More work is needed on their computational and statistical properties.

Since the nonlinear methods do not rely on linear approximations, there is no mismatch between the joint distribution under study and the dynamic implications of a nested structural model. Hence the methods reviewed here may be used in combination with structural approaches, in particular in order to establish identification and improve estimation. In addition, as the examples mentioned in this review demonstrate, policy-relevant average derivative effects may be recovered without the need for additional functional form assumptions.

Among the many questions for future work, an important one concerns robustness. While consistent with large classes of economic models, dynamic conditional independence assumptions are instrumental to establish identification. It would be interesting to assess the impact of relaxing some of these assumptions, for example in the spirit of Chen *et al.* (2011). Lastly, extending the methods reviewed here to models of economic networks or risk sharing, and to identify the effects of macroeconomic risk, are also important tasks.

References

[1] Abowd, J., F. Kramarz, and D. Margolis (1999): "High Wage Workers and High Wage Firms", *Econometrica*, 67(2), 251–333.

[2] Ackerberg, D. A., K. Caves, and G. Frazer (2015): "Identification Properties of Recent Production Function Estimators," *Econometrica*, 83(6), 2411–2451.

[3] Aguirregabiria, V., and P. Mira (2002): "Swapping the Nested Fixed-Point Algorithm: A Class of Estimators for Discrete Markov Decision Models," *Econometrica*, 70(4), 1519–1543.

[4] Aguirregabiria, V., and P. Mira (2007): "Sequential Estimation of Dynamic Discrete Games," *Econometrica*, 75(1), 1–53.

[5] Aguirregabiria, V., and P. Mira (2010): "Dynamic discrete choice structural models: A survey," *Journal of Econometrics*, 156, 38–67.

[6] Arcidiacono, P., and J. B. Jones (2003): 'Finite Mixture Distributions, Sequential Likelihood and the EM Algorithm", *Econometrica*, 71(3), 933–946.

[7] Arcidiacono, P., and R. Miller (2011): 'Conditional Choice Probability Estimation of Dynamic Discrete Choice Models With Unobserved Heterogeneity", *Econometrica*, 79(6), 1823–1867.

[8] Arellano, M. (2014): "Uncertainty, Persistence, And Heterogeneity: A Panel Data Perspective," *Journal of the European Economic Association*, 12(5), 1127–1153.

[9] Arellano, M., R. Blundell, and S. Bonhomme (2016): "Earnings and Consumption Dynamics: A Nonlinear Panel data Framework," unpublished working paper.

[10] Arellano, M., and S. Bonhomme (2016): "Nonlinear Panel Data Estimation via Quantile Regressions," to appear in *Econometrics Journal.*

[11] Auerbach, A. J., and Y. Gorodnichenko (2012): "Measuring the Output Responses to Fiscal Policy," *American Economic Journal: Economic Policy*, 4(2), 1–27.

[12] Banerjee, A. V., and E. Duflo (2003): "Inequality and Growth: What Can the Data Say?" *Journal of Economic Growth*, 8(3), 267-299.

[13] Belloni, A., Chernozhukov, V., Chetverikov, D., and I. Fernández-Val (2016): "Conditional Quantile Processes based on Series or Many Regressors," unpublished manuscript.

[14] Ben-Porath, Y. (1967): "The Production of Human Capital and the Life Cycle of Earnings," *Journal of Political Economy*, 352–365.

[15] Berger, D., and J. Vavra (2014): "Measuring How Fiscal Shocks Affect Durable Spending in Recessions and Expansions," *American Economic Review* papers and proceedings, 104(5), 112–115.

[16] Berger, D., and J. Vavra (2015): "Consumption Dynamics During Recessions," *Econometrica*, 83(1), 101–154.

[17] Blundell, R., L. Pistaferri, and I. Preston (2008): "Consumption Inequality and Partial Insurance," *American Economic Review*, 98(5): 1887–1921.

[18] Blundell, R., L. Pistaferri, and I. Saporta-Eksten (2016): "Consumption Smoothing and Family Labor Supply," *American Economic Review*, 106(2), 387–435.

[19] Bonhomme, S., T. Lamadon, and E. Manresa (2016a): "A Distributional Framework for Matched Employer-Employee Data," unpublished manuscript.

[20] Bonhomme, S., T. Lamadon, and E. Manresa (2016b): "Discretizing Unobserved Heterogeneity: Approximate Clustering Methods for Dimension Reduction," unpublished manuscript.

[21] Browning, M., and J. M. Carro (2014): "Dynamic Binary Outcome Models with Maximal Heterogeneity," *Journal of Econometrics*, 178(2), 805–823.

[22] Canay, I. A., A. Santos, and A. Shaikh (2013): "On the Testability of Identification in Some Nonparametric Models with Endogeneity," *Econometrica*, 81(6), 2535–2559.

[23] Celeux, G., and J. Diebolt (1993): "Asymptotic Properties of a Stochastic EM Algorithm for Estimating Mixing Proportions," *Comm. Statist. Stochastic Models*, 9, 599-613.

[24] Chamberlain, G. (1980): "Analysis of Covariance with Qualitative Data," *Review of Economic Studies*, 47, 225–238.

[25] Chamberlain, G. (1982): "The General Equivalence of Granger and Sims Causality," *Econometrica*, 50, 569–581.

[26] Chen, X. (2007): "Sieve Methods in Econometrics," *Handbook of Econometrics*.

[27] Chen, X., E. T. Tamer, and A. Torgovitsky (2011): "Sensitivity Analysis in Semiparametric Likelihood Models," unpublished working paper.

[28] Chernozhukov, V., and C. Hansen (2005): "An IV model of Quantile Treatment Effects," *Econometrica*, 73, 245–262.

[29] Connault, B. (2016): "Hidden Rust Models," unpublished working paper.

[30] Constandinides, G., and A. Gosh (2016): "Asset Pricing with Countercyclical Household Consumption Risk , " to appear in the *Journal of Finance*.

[31] Dempster, A. P., N. M. Laird, and D. B. Rubin (1977): "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society*, B, 39, 1–38.

[32] D'Haultfoeuille, X. (2011): "On the Completeness Condition for Nonparametric Instrumental Problems," *Econometric Theory*, 27, 460–471

[33] Doraszelski, U., and J. Jaumandreu (2013): "R&D and Productivity: Estimating Endogenous Productivity," *Review of Economic Studies*, 80(4), 1338–1383.

[34] Elashoff, M. and L. Ryan (2004): "An EM Algorithm for Estimating Equations," *Journal of Computational and Graphical Statistics*, 13(1), 48–65.

[35] Gallant, A. R., and D. W. Nychka (1987): "Semi-Nonparametric Maximum Likelihood Estimation," *Econometrica*, 363–390.

[36] Golosov, M., and A. Tsyvinski (2015): "Policy Implications of Dynamic Public Finance," *Annual Reviews of Economics*, 7, pp.147-171

[37] Gourinchas, P.O. and J. A. Parker (2002): "Consumption over the Life Cycle." *Econometrica*, 70, 47-91.

[38] Guvenen, F., and A. Smith (2014): "Inferring Labor Income Risk from Economic Choices: An Indirect Inference Approach," *Econometrica*, November, 82(6), 2085–2129.

[39] Guvenen, F., F. Karahan, S. Ozcan, and J. Song (2016): "What Do Data on Millions of U.S. Workers Reveal about Life-Cycle Earnings Risk?" unpublished working paper.

[40] Hall, R., and F. Mishkin (1982): "The sensitivity of Consumption to Transitory Income: Estimates from Panel Data of Households," *Econometrica*, 50(2): 261–81.

[41] Hall, P., and X. H. Zhou (2003): "Nonparametric Estimation of Component Distributions in a Multivariate Mixture," *Annals of Statistics*, 201–224.

[42] Heckman, J. J. (1981): "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process," in Manski, and McFadden (Eds.), *Structural Analysis of Discrete Data with Econometric Applications*. MIT Press.

[43] Herbst, E. P., and F. Schorfheide (2015): *Bayesian Estimation of DSGE Models*. Princeton University Press.

[44] Honoré, B. E., and E. Tamer (2006): "Bounds on Parameters in Panel Dynamic Discrete Choice Models," *Econometrica*, 74(3), 611–629.

[45] Hotz, J., and R. Miller (1988): "Conditional Choice Probabilities and the Estimation of Dynamic Models", *Review of Economic Studies*, 60(3), 497–529.

[46] Hu, Y. (2008): "Identification and Estimation of Nonlinear Models with Misclassification Error Using Instrumental Variables: A General Solution," *Journal of Econometrics*, 144(1), 27–61.

[47] Hu, Y. (2015): "Microeconomic Models with Latent Variables: Applications of Measurement Error Models in Empirical Industrial Organization and Labor Economics," Technical report, Cemmap Working Papers, CWP03/15.

[48] Hu, Y. and S. M. Schennach (2008): "Instrumental Variable Treatment of Nonclassical Measurement Error Models," *Econometrica*, 76, 195–216.

[49] Hu, Y. and J.-L. Shiu (2012): "Nonparametric Identification Using Instrumental Variables: Sufficient Conditions for Completeness," unpublished manuscript.

[50] Hu, Y. and M. Shum (2012): "Nonparametric Identification of Dynamic Models with Unobserved State Variables," *Journal of Econometrics*, 171, 32–44.

[51] Huang, G., and Y. Hu (2011): "Estimating Production Functions with Robustness Against Errors in the Proxy Variables," Cemmap working paper CWP35/11.

[52] Huggett, M. (1993), "The Risk-Free Rate in Heterogeneous-Agent Incomplete-Insurance Economies". *Journal of Economic Dynamics and Control*, 17, 953–969.

[53] Jackson, M. O. (2009): "Networks and Economic Behavior," *Annual Review of Economics*, 1(1), 489–511.

[54] Joe, H. (1997): *Multivariate Models and Dependence Concepts*. London: Chapman & Hall.

[55] Kaplan, G., and G. Violante (2010): "How Much Consumption Insurance Beyond Self-Insurance" *American Economic Journal*, 2(4), 53–87.

[56] Kaplan, G., and G. Violante (2014): "A Model of the Consumption Response to Fiscal Stimulus Payments," *Econometrica*, 82(4), 1199-1239.

[57] Kasahara, H., and K. Shimotsu (2009): "Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choices," *Econometrica*, 77(1), 135–175.

[58] Keane, M., and K. Wolpin (1997): "The Career Decisions of Young Men," *Journal of Political Economy*, 105(3), 473–522.

[59] Koenker, R. (2005): *Quantile Regression*, Econometric Society Monograph Series, Cambridge: Cambridge University Press.

[60] Koenker, R. and G. J. Bassett (1978): "Regression quantiles, " *Econometrica*, 46, 33–50.

[61] Kotlarski, I. (1967): "On Characterizing the Gamma and Normal Distribution," *Pacific Journal of Mathematics*, 20, 69–76.

[62] Lancaster, T. (2004): *An Introduction to Modern Bayesian Econometrics*, Blackwell.

[63] Levinsohn, J., and A. Petrin (2003): "Estimating Production Functions Using Inputs to Control for Unobservables," *Review of Economic Studies*, 70(2), 317–341.

[64] Ljungqvist, L., and T. Sargent (2004): *Recursive Macroeconomic Theory*. MIT Press.

[65] Lucas, R. E. (1976): "Econometric Policy Evaluation: A Critique," Carnegie-Rochester conference series on public policy, Vol. 1. North-Holland.

[66] Matzkin, R. L. (2013): "Nonparametric Identification in Structural Economic Models," Annual Review of Economics, 5(1), 457–486.

[67] Meghir, C., and L. Pistaferri, (2011): "Earnings, Consumption and Life Cycle Choices," *Handbook of Labor Economics*, Elsevier.

[68] Nelsen, R. B. (1999): *An Introduction to Copulas*. New-York: Springer Verlag.

[69] Newey, W. , and J. Powell (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*.

[70] Nielsen, S. F. (2000): "The Stochastic EM Algorithm: Estimation and Asymptotic Results," *Bernoulli*, 6(3): 457–489.

[71] Pesendorfer, M., and P. Schmidt-Dengler (2008): "Asymptotic Least Squares Estimators for Dynamic Games," *Review of Economic Studies*, 75(3), 901–928.

[72] Rust, J. (1987): "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 999–1033.

[73] Rust, J. (1994): "Structural Estimation of Markov Decision Processes," *Handbook of econometrics*, 4(4), 3081–3143.

[74] Sancetta, A., and S. Satchell (2004): "The Bernstein Copula and its Applications to Modeling and Approximations of Multivariate Distributions," *Econometric Theory*, 20, 535–562.

[75] Schmidt, L. (2015): "Climbing and Falling Off the Ladder: Asset Pricing Implications of Labor Market Event Risk," unpublished manuscript.

[76] Su, C. and K. Judd (2012): "Constrained Optimization Approaches to Estimation of Structural Models", *Econometrica*, 80(5), 2213–2230.

[77] Wei, Y. and R. J. Carroll (2009): "Quantile Regression with Measurement Error," *Journal of the American Statistical Association*, 104, 1129–1143.

[78] Wilhelm, D. (2012): "Identification and Estimation of Nonparametric Panel Data Regressions with Measurement Error," unpublished manuscript.