# Efficient Estimation of Non-Linear Dynamic Panel Data Models with Application to Smooth Transition Models

Tue Gørgens* Christopher L. Skeels† Allan H. Würtz‡

February 16, 2008

Preliminary and incomplete.
Not to be quoted without the authors' permission.

Abstract: This paper explores estimation of a class of non-linear dynamic panel data models with additive unobserved individual specific effects. The models are specified by moment restrictions. The class includes the panel data AR(p) model and smooth transition autoregressive panel data (PSTAR) models. We derive an efficient set of moment restrictions for estimation and apply the results to estimation of panel smooth transition models with fixed effects, where the transition may be determined endogenously. The performance of the GMM estimator, both in terms of estimation precision and forecasting performance, is examined in a Monte Carlo experiment. We find that estimation of the parameters in the transition function can be problematic but that there may be significant benefits in terms of forecast performance.

Keywords: Dynamic panel data models, fixed effects, GMM estimation, smooth transition.
J.E.L. Classification Codes: C13, C23.

Acknowledgment: We wish to thank Timo Teräsvirta for helpful discussions. Tue Gørgens and Allan Würtz thank the Danish Social Science Research Council for financial support. Chris Skeels acknowledges the financial support of the Australian Research Council under grants DP0771445 and DP0770580.

*Economics Program, Research School of Social Sciences, The Australian National University, Canberra ACT 0200, Australia. Email: tue.gorgens@anu.edu.au

†Department of Economics, The University of Melbourne, Victoria 3010, Australia. Email: Chris.Skeels@unimelb.edu.au

‡School of Economics and Management, University of Aarhus, Building 1322, DK-8000 Aarhus C, Denmark. Email: awurtz@econ.au.dk

# 1 Introduction

The development of panel data models and, in particular, of dynamic panel data models continues to expand rapidly as the availability of panel data increases.[1] One of the challenges is how dynamic panel data models might incorporate more complex dynamic structures. The difficulty here arises because the nature of the non-linearity, and exactly how it interacts with stochastic assumptions, matters. For example, we see different considerations taking precedence when analysing random effects models on the one hand (e.g. Heckman, 1981, Bhargava and Sargan, 1983, Arellano and Carrasco, 2003, Wooldridge, 2005) and fixed effects models on the other (e.g. Wooldridge, 1997, Honoré and Kyriazidou, 2000, Hahn and Kuersteiner, 2002, Honoré, 2002, Hahn, Hausman, and Kuersteiner, 2007).

This paper explores efficient estimation in the class of non-linear dynamic panel data models with additive unobserved individual effects, where the models are specified by moment restrictions. Inasmuch as we are concerned with efficient estimation our contribution is in the tradition of papers such as Ahn and Schmidt (1995, 1997), Arellano and Bover (1995) and Hahn (1997). There are many different types of non-linearity considered in the literature. For example, non-linearities arise naturally when working with censored and limited dependent variables in a panel context (e.g. Honoré, 2002, and the references cited therein). In these models the unobserved individual specific effects are typically not additive separable and the specification of the models is often based on a likelihood approach. In contrast, the models considered

[1] For recent surveys, see Arellano and Honoré (2001), Hsiao (2001, 2003), Baltagi (2005) and Arellano and Hahn (2007).

here are specified by moment restrictions.

We apply our results to estimation of smooth transition models. Smooth transition models are common non-linear models in time series analysis. They have only been applied a few times with panel data (e.g. González, Teräsvirta, and van Dijk, 2005, Fok, Van Dijk, and Franses, 2005). We extend the literature by analysing smooth transitions in a dynamic model including the case where the transition function itself depends on the lagged endogenous variable. In a Monte Carlo experiment we find that estimation of the separate coefficients in the transition function may be difficult. The forecast performance, however, may be significantly improved despite the difficulty of estimating the separate coefficients.

The structure of the paper is as follows. In the next section we present a non-linear first-order autoregressive panel data model and develop a set of moment conditions upon which efficient GMM estimation might be based. We also extend our results to higher order autoregressive models. In Section 3 we focus attention on dynamic panel smooth transition models. We examine both exogenous and endogenous transitions and report a Monte Carlo experiment on estimation precision and forecasts. Concluding remarks appear in Section 4.

## 2 Efficient Estimation

In this section we introduce our model of interest and extend the arguments of Ahn and Schmidt (1995) to develop a set of moment conditions upon which an efficient GMM estimator can be based.

## 2.1 Dynamic panel data models

Our starting point is a first-order autoregressive panel data model of the form

$$y_{it} = \alpha_1 y_{i,t-1} + u_{it}, \tag{1}$$

$$u_{it} = \eta_i + v_{it}, \tag{2}$$

where $t = 1, \ldots, T$ and $i = 1, \ldots, N$. The observations are assumed to be independent across individuals ($i$), but not across time ($t$), and it is assumed that $y_0$ is known. In (1), $\alpha_1$ is an unknown parameter and $u_{it}$ is an unobserved "error" term. Equation (2) decomposes the error term into two unobserved individual-specific effects, one of which is constant over time, $\eta_i$, and the other, $v_{it}$, which varies with time. To fix the location, it is assumed that $\mathrm{E}(v_{it}) = 0$ for $t = 1, \ldots, T$. The model implies that $y_{i,t-1}$ is correlated with $\eta_i$.

Following the treatment of Ahn and Schmidt (1995), our assumptions for the first-order autoregressive panel data model are:

$$\mathrm{E}(y_{i0} v_{it}) = 0, \quad t = 1, \ldots, T; \tag{A1}$$

$$\mathrm{E}(\eta_i v_{it}) = 0, \quad t = 1, \ldots, T; \tag{A2}$$

$$\mathrm{E}(v_{is} v_{it}) = 0, \quad s = 1, \ldots, t-1; \quad t = 2, \ldots, T. \tag{A3}$$

This is a relatively weak set of assumptions. For instance, stationarity is not assumed nor is the relationship between $y_{i0}$ and $\eta_i$ restricted. Bond (2002, p6) and Hsiao (2003), inter alia, provide more complete discussions of initial conditions; see also the discussion of Ahn and Schmidt (1995, p.7). Equations

(1) and (2) together with assumptions A1–A3, which we shall hereafter refer to as Model 1, essentially comprise the model considered by Blundell and Bond (1998) and others.

In this paper we consider models which are not linear in the lagged dependent variable. The framework within which we shall work is

$$y_{it} = \alpha_1 y_{i,t-1} + \delta h(y_{i,t-1}, w_{it}, \theta) + u_{it}, \quad t = 2, \ldots, T, \tag{3}$$

where we assume that $h(y_{i,t-1}, w_{it}, \theta)$ is a scalar-valued function, differentiable with respect to the $p$-vector $\theta$, that $\delta$ is an unknown parameter and that $w_{it}$ is a $j$-vector of pre-determined variables.[2] We shall define Model 2 to be the set of equations (3) and (2), Assumptions A1–A3, together with Assumption A4 which is discussed below.

This structure extends the panel smooth transition regression (PSTR) model of González et al. (2005) to the dynamic case and also includes a variety of possible extensions of non-linear time series models to the panel data context. As a concrete example, consider the logistic transition function where $\theta = (\theta_1, \theta_2)'$ and

$$h(y_{i,t-1}, w_{it}, \theta) = \frac{1}{1 + \exp\{-\theta_1(w_{it} - \theta_2)\}} y_{i,t-1}.$$

Note that a threshold model obtains as a special case wherein $\theta_1 \to \infty$ and

[2]The parameterization of equation (3) is attractive in that a null hypothesis of linearity against the alternative of non-linearity is simply characterized as $H_0 : \alpha_2 = 0$ against $H_1 : \alpha_2 \neq 0$. The downside of this parameterization is that there is an inherent identification problem whereby $\theta$ is unidentified if $\alpha_2 = 0$. This means that inference on $\theta$ involves non-standard distribution theory along the lines discussed in Teräsvirta (1994, Section 3); see also Andrews and Ploberger (1994, 1995) for a very general treatment of this problem.

the threshold occurs at $\theta_2$.

Another class of models follows on choosing $h(y_{i,t-1}, w_{it}, \theta)$ to be a polynomial in $y_{i,t-1}$. For instance, setting $h(y_{i,t-1}, w_{it}, \theta) = y_{i,t-1}^2$, yields a quadratic model of the form

$$y_{it} = \alpha_1 y_{i,t-1} + \delta y_{i,t-1}^2 + u_{it}.$$

The quadratic model can be thought of as a second-order Taylor approximation to an arbitrary twice-differentiable non-linear regression function. It can also be seen as a class of dynamic models which includes the logistic map as a special case. This map, which was analysed in the seminal paper of May (1976), is known to exhibit complex dynamics, including chaotic behaviour for certain parameter values.

In addition to A1–A3, we make the following assumption for the non-linear dynamic panel data model:

$$\mathrm{E}\left[\begin{pmatrix} h(y_{i,s-1}, w_{is}, \theta) \\ \dfrac{\partial h(y_{i,s-1}, w_{is}, \theta)}{\partial \theta} \end{pmatrix} v_{it}\right] = 0, \quad s = 1, \ldots, t; \quad t = 1, \ldots, T. \qquad \text{A4}$$

It is possible to avoid involving the parameters by instead imposing the stronger conditional moment restriction $\mathrm{E}[v_{it}|y_{i,t-1}, w_{it}] = 0$ which implies that, for any function $g$, $\mathrm{E}[g(y_{i,t-1}, w_{it})v_{it}] = 0$. Rather than making this much stronger assumption, however, A4 only requires this implication to hold for the particular $g = [h, \partial h/\partial\theta']'$. Finally, we note in passing that A4 is analogous to assumptions that typically accompany non-linear regression models.

At this point we defer discussions of assumptions relating to $w_{it}$. This is primarily because, at this stage, $w_{it}$ is best thought of as a place-holder which provides our model with considerable flexibility, as illustrated above. Assumption A4 provides a mild assumption on $w_{it}$ not to be endogenous due to $v_{it}$. Until a particular $w_{it}$ is specified, however, it remains to be determined what further assumptions are appropriate and so they will need to be addressed on a case-by-case basis.

## 2.2 Efficient Moment Conditions for Estimation

In this section, we discuss moment conditions that can be used for estimation of Models 1 and 2. For Model 1, assumptions A1–A3 imply the following $T(T-1)/2$ linear moment conditions

$$\mathrm{E}(y_{is}\Delta u_{it}) = 0, \quad s = 0, \ldots, t-2, \quad t = 2, \ldots, T. \tag{4}$$

In addition, there are $T-2$ quadratic moment conditions

$$\mathrm{E}(u_{iT}\Delta u_{it}) = 0, \quad t = 2, \ldots, T-1. \tag{5}$$

Ahn and Schmidt (1995) prove that these are the only moment conditions implied by A1–A3.[3] Hence the set of moment conditions (4) and (5) provide a basis for efficient estimation.

For the non-linear dynamic panel data model, Model 2, Assumptions A1-A3 imply the same moment conditions (4) and (5) as in the linear dynamic

[3] Ahn and Schmidt (1995, p.9) also remark that these conditions are implied if A1–A3 are replaced by assumptions that the moments are constant over time instead of zero. That is, the values of these moments are not identified.

panel data model. Assumption A4 implies the $(p+1)T(T-1)/2$ moment conditions:

$$\mathrm{E}\left[\begin{pmatrix} h(y_{i,s-1}, w_{is}, \theta) \\ \dfrac{\partial h(y_{i,s-1}, w_{is}, \theta)}{\partial \theta} \end{pmatrix} \Delta u_{it}\right] = 0, \quad s = 1, \ldots, t-1; \quad t = 2, \ldots, T. \quad (6)$$

These results are summarized in the following theorem.

**Theorem 1.** *In Model 2, specified by equations* (3) *and* (2)*, assumptions A1–A4 imply that a complete set of moment conditions for efficient estimation are equations* (4)*,* (5) *and* (6)*. In total, there are* $T^2 - 2 + pT(T-1)/2$ *such moment conditions.*

**Proof.** See Appendix.

Efficient estimation must fully utilize the information in the set of moment conditions given in Theorem 1. One possibility is the GMM estimator based on an efficient weight matrix. Other possibilities include the empirical likelihood estimator (Owen, 1988) and the exponential tilting estimator (Kitamura and Stutzer, 1997, Imbens, Spady, and Johnson, 1998) which, in turn, are both special cases of the generalized empirical likelihood estimator (Smith, 1997). All of these estimators are efficient. In the next section, we investigate the GMM estimator with the panel smooth transition model.

There may be more moment conditions than listed in the Theorem 1 if additional assumptions about $w_{it}$ are appropriate. Some of the variables included in $w_{it}$ may be assumed to be strictly exogenous. For example, time trends and time dummies would add further moment conditions.

As mentioned earlier, assumption A4 could be replaced by the stronger conditional moment restriction $\mathrm{E}[v_{it}|y_{i,t-1}, w_{it}] = 0$. This restriction implies for any choice of $g$, $\mathrm{E}[g(y_{i,t-1}, w_{it})v_{it}] = 0$ and, thus, infinitely many unconditional moment restrictions. Under standard regularity conditions, it is possible to represent this conditional moment restriction by a finite number of unconditional moment restrictions by appropriate choice of $g$ functions, see e.g. Newey (1993). These $g$ functions, however, typically need to be estimated non-parametrically. The resulting unconditional moment restrictions could replace the unconditional moment restrictions in A4 and be used for efficient estimation as outlined in the Theorem. In this paper, we only impose the weaker unconditional moment restriction A4.

At this stage, there is no concern regarding stationarity of the dynamic models. This is a concern for estimation. In typical panel applications, where $T$ is small and $N$ is large, such concerns are often less important and the modelling richness afforded by the non-linearity may result in a superior approximation to the data. We will return to practical implications of the specifications when we discuss the case of the smooth transition model in the following sections.

## 2.3 Extension to higher-order dynamic model

In this subsection, we briefly describe how to extend the results to non-linear dynamic panel data models, which includes higher-order lags of the dependent variable.

Consider the model

$$y_{it} = \alpha_1 y_{i,t-1} + .. + \alpha_{q_l} y_{i,t-q_l} + \delta h(y_{i,t-1}, .., y_{i,t-q_h}, w_{it}, \theta) + u_{it}, \qquad (7)$$

for $t = q, .., T$, where $q = \max(q_l, q_h)$ is the largest lag included. We impose similar moment conditions as in Model 2. The new assumptions A1*–A4* are:

$$\mathrm{E}(y_{is}v_{it}) = 0, \qquad s = 0, \ldots, q-1; \quad t = q, \ldots, T. \tag{A1*}$$

$$\mathrm{E}(\eta_i v_{it}) = 0, \qquad t = q, \ldots, T; \tag{A2*}$$

$$\mathrm{E}(v_{is}v_{it}) = 0, \qquad s = q, \ldots, t-1; \quad t = q+1, \ldots, T. \tag{A3*}$$

$$\mathrm{E}\left[\begin{pmatrix} h(y_{i,s-1}, .., y_{i,s-q_h}, w_{is}, \theta) \\ \dfrac{\partial h(y_{i,s-1}, .., y_{i,s-q_h}, w_{is}, \theta)}{\partial \theta} \end{pmatrix} v_{it}\right] = 0, \qquad \begin{array}{c} s = q_h, \ldots, t; \\ t = q, \ldots, T. \end{array} \tag{A4*}$$

These moment conditions imply the following moment conditions, which can be used for estimation:

$$\mathrm{E}(y_{is}\Delta u_{it}) = 0, \qquad s = 0, \ldots, t-2; \quad t = q+1, \ldots, T. \tag{8}$$

$$\mathrm{E}(u_{iT}\Delta u_{it}) = 0, \qquad t = q+1, \ldots, T-q. \tag{9}$$

$$\mathrm{E}\left[\begin{pmatrix} h(y_{i,s-1}, .., y_{i,s-q_h}, w_{is}, \theta) \\ \dfrac{\partial h(y_{i,s-1}, .., y_{i,s-q_h}, w_{is}, \theta)}{\partial \theta} \end{pmatrix} \Delta u_{it}\right] = 0, \qquad \begin{array}{c} s = q_h, \ldots, t-1; \\ t = q+1, \ldots, T. \end{array} \tag{10}$$

There are $((T-1)T-(q-1)q)/2$ moment restrictions in (8), $T-2q$ moment restrictions in (9) and $(p+1)((T-q_h)(T-q_h+1)-(q-q_h)(q-q_h+1))/2$ moment restrictions in (10). Hence, it is necessary that $T \geq q+1$ for the system to be identified, that is, at least $T+1+q$ observations on the individuals. In case $T = q+1$ and $q = q_h$, then the system is exactly identified. In any other case for $T \geq q+1$, including the case $T = q+1$ and $q > q_h$, the system is over-identified.

# 3 Dynamic Panel Smooth Transition Models

In the remaining part of the paper we analyse dynamic panel data models with a smooth transition on the lagged dependent variable. Though smooth transition models are familiar in time series analysis, they have so far only been applied to a limited extent with panel data.

To our knowledge, there are no paper on transition models with panel data that consider the case where the variables in the transition function are lagged endogenous variables. Fok et al. (2005) consider a dynamic model with smooth transition on the lagged endogenous variables. The transitions are determined by an exogenous variable and the analysis is done on large $N$ large $T$ asymptotics. Hence, they estimate the fixed effect and thereby avoid estimation of a model similar to Model 2. González et al. (2005) considers a non-dynamic transition model where all explanatory variables are assumed exogenous. He and Sandberg (2005) tests for a unit root against a first-order panel smooth transition autoregressive model. The transition function in their paper is determined by a time trend. Finally, the model considered by Hansen (1999) is a non-dynamic model and the explanatory variables included are exogenous.

We consider the non-linear dynamic panel data model, Model 2, with the function $h$ specified as a logistic smooth transition function in the next two subsections. The subsections are about the two cases where the variable in the transition function is exogenous or endogenous, respectively.

## 3.1 Exogenous Transitions

Consider the following logistic smooth transition model, where the transitions are determined by $w_{it}$:

$$y_{it} = \alpha_1 y_{i,t-1} + \delta h_w(w_{it}) y_{i,t-1} + \eta_i + v_{it}, \quad t = 2, \ldots, T, \tag{11a}$$

$$h_w(w_{it}) = \frac{1}{1 + \exp\{-\theta_1(w_{it} - \theta_2)\}}, \tag{11b}$$

As in many other non-linear models, identification of the parameters is non-trivial. There are several identification problems. First, the model with $(\alpha, \delta, \theta_1, \theta_2)$ is observationally equivalent to the model with $(\alpha+\delta, -\delta, -\theta_1, \theta_2)$. Secondly, $\alpha$, $\delta$ and $\theta_2$ are not identified when $\theta_1 = 0$. These two identification problems are resolved by the normalization $\theta_1 > 0$. This normalization is also convenient for the grid search used in the estimation program. Thirdly, $\theta_1$ and $\theta_2$ are not identified when $\delta = 0$. This can be resolved by setting $\theta_1$ and $\theta_2$ to some arbitrarily chosen values. Finally, identification breaks down in the limit as $\theta_1 \to \infty$ if $|\theta_2|$ is outside the support of $w_{it}$. This is resolved by assuming that the parameter space is bounded.

The Monte Carlo design is as follows. In all the experiments, $\eta_i \sim N(0, 1)$, $v_{it} \sim N(0, 1)$, $y_{i0} = \eta_i + v_{i0}$ with $v_{i0} \sim N(0, 1)$, and $w_{it} = 2^{-1/2}\eta_i + 2^{-1/2} r_{it}$ with $r_{it} \sim N(0, 1)$. To eliminate the effect of the initial observation, data for the first 100 time periods for each subject are discarded. With the parameter values we choose, this amounts to drawing the initial observations from the stationary distribution of $y$. Note, in the simulations $\eta_i$ and $w_{it}$ are independent of $v_{it}$. This could lead to further moment restrictions than the ones derived from A1–A4. We will, however, estimate the model using only assumptions A1–A4 to investigate the usefulness of these assumptions.

The values of $\alpha$, $\delta$, $\theta_1$ and $\theta_2$ vary across experiments as indicated in the tables of results. The parameters are estimated by performing a grid search over $\theta_1$ and $\theta_2$ and, for each value of $\theta_1$ and $\theta_2$, computing either IV or two-step GMM estimates of $\alpha_1$ and $\alpha_2$. The grid search is restricted to $0.2 \leq \theta_1 \leq 8.0$ and $-2.0 \leq \theta_2 \leq 2.0$. Initially, $11^2$ equally spaced points in the ranges $[\theta_1 - 0.5, \theta_1 + 0.5]$ and $(\theta_2 - 1.0, \theta_2 + 1.0)$ are evaluated. If the optimal point is on the boundary of a range, the search area is widened in the relevant direction (until the maximum range is reached). If the optimal point is in the interior, $5^2$ points are evaluated in the area between the optimal point and its neighbouring points. The proportion of samples where the best estimator is on the boundary of the maximum search range is indicated in the tables under the heading "%Fail". These boundary estimates are included in the calculation of the root mean square errors etc.

The results are shown in table 1. Estimation of the individual parameters can be hard as also noted in Teräsvirta (1994) for the case of time series data. In the table the root mean square error (RMSE) for each parameter estimator is reported. RMSE varies considerable depending on the true value of the parameters. Estimation of $\alpha$ and $\delta$ can be considerably less precise than is estimation of their sum $\alpha + \delta$. In the extreme where $h_w(w_{it}) = 1$ ("regime 1"), the model is an AR(1) process with $\alpha + \delta$ being the parameter on the lagged endogenous variable. Similarly, in the other extreme with $h_w(w_{it}) = 0$ ("regime 0"), the model is an AR(1) process with $\alpha$ being the parameter on the lagged endogenous variable. Hence, it is often easier to estimate the

Table 1: Simulation results for exogenous smooth transition model (11)

| $\theta_1$ | $\theta_2$ | Pr($H$) | RMSE | | | | | A | A | B | %Fail |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $\theta_1$ | $\theta_2$ | $\alpha$ | $\delta$ | $\alpha+\delta$ | | LIN | | |
| | | | | | $T=2$, IV estimation | | | | | | |
| $\alpha=0.0$, $\delta=0.7$ | | | | | | | | | | | |
| 1.5 | 0.0 | 0.95 | 1.62 | 0.78 | 0.31 | 0.51 | 0.38 | 0.32 | 0.49 | 0.19 | 0.11 |
| 2.0 | 0.0 | 0.86 | 1.64 | 0.47 | 0.19 | 0.36 | 0.33 | 0.35 | 0.57 | 0.21 | 0.06 |
| 3.0 | 0.0 | 0.68 | 1.76 | 0.28 | 0.12 | 0.33 | 0.33 | 0.38 | 0.66 | 0.23 | 0.04 |
| 4.0 | 0.0 | 0.54 | 1.81 | 0.25 | 0.10 | 0.32 | 0.33 | 0.40 | 0.71 | 0.24 | 0.10 |
| $\alpha=0.7$, $\delta=-0.7$ | | | | | | | | | | | |
| 1.5 | 0.0 | 0.95 | 1.67 | 0.72 | 0.48 | 0.62 | 0.35 | 0.35 | 0.49 | 0.22 | 0.10 |
| 2.0 | 0.0 | 0.86 | 1.71 | 0.42 | 0.38 | 0.45 | 0.24 | 0.37 | 0.56 | 0.23 | 0.03 |
| 3.0 | 0.0 | 0.68 | 1.84 | 0.23 | 0.31 | 0.33 | 0.14 | 0.39 | 0.65 | 0.24 | 0.05 |
| 4.0 | 0.0 | 0.54 | 1.94 | 0.19 | 0.29 | 0.30 | 0.12 | 0.40 | 0.69 | 0.24 | 0.13 |
| | | | | | $T=4$, GMM estimation | | | | | | |
| $\alpha=0.0$, $\delta=0.7$ | | | | | | | | | | | |
| 1.5 | 0.0 | 0.95 | 1.34 | 0.53 | 0.33 | 0.59 | 0.30 | 0.11 | 0.45 | 0.07 | 0.10 |
| 2.0 | 0.0 | 0.86 | 1.04 | 0.28 | 0.13 | 0.20 | 0.12 | 0.11 | 0.52 | 0.07 | 0.02 |
| 3.0 | 0.0 | 0.68 | 1.12 | 0.12 | 0.06 | 0.13 | 0.09 | 0.11 | 0.61 | 0.07 | 0.02 |
| 4.0 | 0.0 | 0.54 | 1.34 | 0.06 | 0.05 | 0.09 | 0.07 | 0.11 | 0.65 | 0.07 | 0.05 |
| $\alpha=0.7$, $\delta=-0.7$ | | | | | | | | | | | |
| 1.5 | 0.0 | 0.95 | 1.37 | 0.37 | 0.26 | 0.47 | 0.24 | 0.11 | 0.45 | 0.07 | 0.04 |
| 2.0 | 0.0 | 0.86 | 1.44 | 0.27 | 0.21 | 0.32 | 0.13 | 0.11 | 0.52 | 0.07 | 0.04 |
| 3.0 | 0.0 | 0.68 | 1.29 | 0.09 | 0.07 | 0.09 | 0.05 | 0.11 | 0.61 | 0.07 | 0.02 |
| 4.0 | 0.0 | 0.54 | 1.39 | 0.06 | 0.06 | 0.07 | 0.04 | 0.11 | 0.66 | 0.07 | 0.04 |

*Legend:* Pr($H$): Pr($0.05 < h_w(w_{it}) < 0.95$); RMSE: root mean square error; A: mean root mean square regression function error; B: mean absolute regression function prediction error; %Fail: proportion of samples with estimates of $\theta_1$ or $\theta_2$ on the boundary of the grid search range or with negative definite GMM weight matrix (failed samples included in the calculation of RMSE and MARFPE). *Notes:* Cor($y_{it}, \eta_i$) $\simeq$ 0.82 and Cor($y_{it}, v_{it}$) $\simeq$ 0.45 in all cases. The standard deviation of $y_{it}$ increases from about 2.10 to 2.35 as $\theta_1$ increases from 1.5 to 4.0. Sample size 1000 and 100 samples. The efficient GMM estimator is implemented as the two-step estimator of $\alpha$ and $\delta$ taking $\theta_1$ and $\theta_2$ as fixed in the computation of moment conditions and weight matrix, and performing a grid search over $\theta_1$ and $\theta_2$.

extreme regimes than the regimes in between. Not surprisingly, the larger time dimension, the more precise estimators.

Table 1 also reports the forecast performance of the model. Since the models include a fixed effect, we compare forecasts of the regression function, that is, the systematic part of the model. The forecast is compared with the linear first-order autoregressive model. The exogenous smooth transition model is better at forecasting the regression function. This is especially the case for $T = 4$. Hence, even though it may be hard to precisely estimate the individual parameters in the exogenous smooth transition model, there is a considerable gain in forecast performance.

For a few of the designs there is a probability about 10 percent of obtaining estimates on the parameter boundary in the grid search for $\theta_1$ and $\theta_2$. As is seen in the forecast performance this does not mean that the model cannot forecast. It simply means that it is hard to pinpoint the individual parameters.

## 3.2 Endogenous Transitions

In this subsection, we discuss the logistic transition model, where the transitions are determined by the lag endogenous variable. This model is also known as a smooth transition first-order autoregressive (STAR(1)) model. It is given by

$$y_{it} = \alpha_1 y_{i,t-1} + \delta h_y(y_{i,t-1}) y_{i,t-1} + \eta_i + v_{it}, \quad t = 2, \ldots, T, \tag{12a}$$

$$h_y(y_{i,t-1}) = \frac{1}{1 + \exp\{-\theta_1(y_{i,t-1} - \theta_2)\}}, \tag{12b}$$

The identification of the parameters is equivalent to identification of the parameters in the model above with an exogenous variable in the transition function.

Table 2 shows the results. For our designs, it is harder to estimate the smooth transition model when the transition is determined by the lagged dependent variable. This is seen by the percentage of parameter estimates of $\theta_1$ and $\theta_2$ on the boundary of the grid search. The RMSE on some of the parameter estimators is also quite high. For example, it can be hard to determine whether the effect on the lagged dependent variable comes from $\alpha$ or $\delta$ but the sum of the two can be estimated much more precise. It is, however, not possible to give these conclusions uniformly over the various experiments.

The forecasts of the regression function show that the correct smooth transition specification may not be superior to the misspecified AR(1) model. This is the case for $T = 2$. For $T = 4$, however, the smooth transition model is considerable better at forecasting than the AR(1) model.

# 4 Conclusion

In this paper we have explored estimation of a class of non-linear dynamic panel data models with additive unobserved individual specific effects. The models are specified by moment restrictions. The class includes the panel data AR(p) model, polynomial dynamic models and smooth transition autoregressive panel data (PSTAR) models amongst others. By extending the

Table 2: Simulation results for endogenous smooth transition model (12)

| $\theta_1$ | $\theta_2$ | Pr($H$) | RMSE | | | | | A | A | B | %Fail |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $\theta_1$ | $\theta_2$ | $\alpha$ | $\delta$ | $\alpha+\delta$ | | LIN | | |
| | | | | | $T=2$, IV estimation | | | | | | |
| $\alpha=0.0$, $\delta=0.7$ | | | | | | | | | | | |
| 1.5 | 0.0 | 0.66 | 1.75 | 0.95 | 0.38 | 2.76 | 2.62 | 1.19 | 0.65 | 0.66 | 0.08 |
| 2.0 | 0.0 | 0.53 | 1.84 | 0.81 | 0.31 | 2.62 | 2.50 | 1.21 | 0.70 | 0.70 | 0.06 |
| 3.0 | 0.0 | 0.37 | 2.01 | 0.63 | 0.20 | 1.29 | 1.20 | 0.90 | 0.73 | 0.60 | 0.05 |
| 4.0 | 0.0 | 0.28 | 2.06 | 0.56 | 0.19 | 1.02 | 0.94 | 0.82 | 0.73 | 0.59 | 0.15 |
| $\alpha=0.7$, $\delta=-0.7$ | | | | | | | | | | | |
| 1.5 | 0.0 | 0.66 | 1.98 | 0.98 | 1.14 | 1.20 | 0.31 | 0.88 | 0.66 | 0.51 | 0.13 |
| 2.0 | 0.0 | 0.53 | 1.91 | 0.73 | 0.80 | 0.84 | 0.19 | 0.85 | 0.71 | 0.53 | 0.06 |
| 3.0 | 0.0 | 0.36 | 1.99 | 0.57 | 0.68 | 0.75 | 0.24 | 0.81 | 0.74 | 0.56 | 0.08 |
| 4.0 | 0.0 | 0.27 | 2.05 | 0.42 | 0.45 | 0.52 | 0.22 | 0.70 | 0.74 | 0.53 | 0.17 |
| | | | | | $T=4$, GMM estimation | | | | | | |
| $\alpha=0.0$, $\delta=0.7$ | | | | | | | | | | | |
| 1.5 | 0.0 | 0.66 | 2.09 | 0.65 | 0.30 | 0.43 | 0.19 | 0.15 | 0.55 | 0.09 | 0.12 |
| 2.0 | 0.0 | 0.53 | 1.84 | 0.49 | 0.29 | 0.42 | 0.16 | 0.15 | 0.59 | 0.10 | 0.08 |
| 3.0 | 0.0 | 0.37 | 2.41 | 0.29 | 0.14 | 0.17 | 0.07 | 0.12 | 0.61 | 0.10 | 0.14 |
| 4.0 | 0.0 | 0.28 | 2.55 | 0.18 | 0.06 | 0.09 | 0.06 | 0.12 | 0.60 | 0.10 | 0.26 |
| $\alpha=0.7$, $\delta=-0.7$ | | | | | | | | | | | |
| 1.5 | 0.0 | 0.66 | 2.65 | 0.64 | 0.16 | 0.41 | 0.29 | 0.13 | 0.55 | 0.09 | 0.15 |
| 2.0 | 0.0 | 0.53 | 2.57 | 0.44 | 0.14 | 0.34 | 0.23 | 0.13 | 0.59 | 0.10 | 0.15 |
| 3.0 | 0.0 | 0.36 | 2.72 | 0.27 | 0.06 | 0.13 | 0.11 | 0.12 | 0.61 | 0.10 | 0.18 |
| 4.0 | 0.0 | 0.27 | 2.76 | 0.29 | 0.06 | 0.16 | 0.12 | 0.11 | 0.60 | 0.10 | 0.33 |

*Legend:* Pr($H$): Pr($0.05 < h_y(y_{i,t-1}) < 0.95$); RMSE: root mean square error; A: mean root mean square regression function error; B: mean absolute regression function prediction error; %Fail: proportion of samples with estimates of $\theta_1$ or $\theta_2$ on the boundary of the grid search range or with negative definite GMM weight matrix (failed samples included in the calculation of RMSE and MARFPE). *Notes:* For Cor($y_{it}, \eta_i$) $\simeq 0.85$ and Cor($y_{it}, v_{it}$) $\simeq 0.40$ in all cases. The standard deviation of $y_{it}$ increases from about 2.40 to 2.55 as $\theta_1$ increases from 1.5 to 4.0. The $\theta_1$ and $\theta_2$ are scaled by dividing $y_{i,t-1}$ in $h_y$ with the standard deviation of $y_{it}$ (found by simulation). Sample size 1000 and 100 samples.

analysis of Ahn and Schmidt (1995) we derive a set of moment restrictions which provide a basis for efficient estimation. We subsequently extend these results to allow for higher-order non-linear dynamics.

Having established results that apply to the entire class of models under consideration we then specialize our analysis to consider estimation of panel smooth transition models with fixed effects, where the transition may be determined endogenously. The performance of the GMM estimator, both in terms of estimation precision and forecasting performance, is examined in a Monte Carlo experiment. We find that estimation of the parameters in the transition function can be problematic but that there may be significant benefits in terms of forecast performance.

# References

Ahn, S. C. and P. Schmidt (1995). Efficient estimation of models for dynamic panel data. *Journal of Econometrics 68*, 5–27.

Ahn, S. C. and P. Schmidt (1997). Efficient estimation of dynamic panel data models: Alternative assumptions and simplified estimation. *Journal of Econometrics 76*(1–2), 309–321.

Andrews, D. W. K. and W. Ploberger (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica 62*, 1383–1414.

Andrews, D. W. K. and W. Ploberger (1995). Admissibility of the likelihood ratio test when a nuisance parameter is present only under the alternative. *Annals of Statistics 23*, 1609–1629.

Arellano, M. and O. Bover (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics 68*(1), 29–51.

Arellano, M. and R. Carrasco (2003). Binary choice panel data models with predetermined variables. *Journal of Econometrics 115*(1), 125–157.

Arellano, M. and J. Hahn (2007). Understanding bias in nonlinear panel models: Some recent developments. In R. Blundell, W. K. Newey, and T. Persson (Eds.), *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress*, Volume III, Cambridge. Cambridge University Press.

Arellano, M. and B. Honoré (2001). Panel data models: Some recent developments. In J. J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 5, Chapter 53, pp. 3229–3296. Amsterdam, The Netherlands: Elsevier Science B.V.

Baltagi, B. H. (2005). *Econometric Analysis of Panel Data* (Third ed.). Chichester: John Wiley & Sons.

Bhargava, A. and J. D. Sargan (1983). Estimating dynamic random effects models from panel data covering short time periods. *Econometrica 51*(6), 1635–1659.

Blundell, R. and S. Bond (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics 87*, 115–143.

Bond, S. (2002). Dynamic panel data models: A guide to micro data methods and practice. Cemmap working paper CWP09/02, The Institute for Fiscal Studies, Department of Economics, UCL, London.

Fok, D., D. Van Dijk, and P. H. Franses (2005). A multi-level panel star model for us manufacturing sectors. *Journal of Applied Econometrics 20*, 811–827.

González, A., T. Teräsvirta, and D. van Dijk (2005). Panel smooth transition models. SSE/EFI Working Paper Series in Economics and Finance, No. 604, Stockholm School of Economics.

Hahn, J. (1997). Efficient estimation of panel data models with sequential moment restrictions. *Journal of Econometrics 1*(79), 1–21.

Hahn, J., J. Hausman, and G. Kuersteiner (2007). Long difference instrumental variables estimation for dynamic panel models with fixed effects. *Journal of Econometrics 127*(2), 574–617.

Hahn, J. and G. Kuersteiner (2002). Asymptotically unbiased inference for a dynamic panel model with fixed effects when both "n" and "t" are large. *Econometrica 70*(4), 1639–1657.

Hansen, B. E. (1999). Threshold effects in non-dynamic panels: Estimation, testing, and inference. *Journal of Econometrics 93*, 345–368.

He, C. and R. Sandberg (2005). Inference for unit roots in a panel data smooth transition autoregressive model where the time dimension is fixed. SSE/EFI Working Paper Series in Economics and Finance, No. 581, Stockholm School of Economics.

Heckman, J. J. (1981). Statistical models for discrete panel data. In C. F. Manski and D. M. Fadden (Eds.), *Structural Analysis of Discrete Data with Econometric Applications*. MIT Press.

Honoré, B. (2002). Nonlinear models with panel data. *Portuguese Economic Journal 1*, 163–179.

Honoré, B. E. and E. Kyriazidou (2000). Panel data discrete choice models with lagged dependent variables. *Econometrica 68*, 839–874.

Hsiao, C. (2001). Panel data models. In B. H. Baltagi (Ed.), *A Companion to Theoretical Econometrics*, Chapter 16, pp. 349–365. Malden, Mass.: Blackwell Publishers.

Hsiao, C. (2003). *Analysis of Panel Data* (Second ed.). Econometric Society Monographs No. 34. Cambridge, New York and Melbourne: Cambridge University Press.

Imbens, G. W., R. H. Spady, and P. Johnson (1998). Information theoretic approaches to inference in moment condition models. *Econometrica 66*, 333–357.

Kitamura, Y. and M. Stutzer (1997). An information-theoretic alternative to generalized method of moments estimation. *Econometrica 65*, 861–874.

May, R. M. (1976). Simple mathematical models with very complicated dynamics. *Nature 261*(5560), 459–467.

Newey, W. K. (1993). Efficient estimation of models with conditional moment restrictions. In G. S. Maddala, C. R. Rao, and H. D. Vinod (Eds.), *Handbook of Statistics*, Volume 11. Amsterdam: North-Holland.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika 75*, 237–249.

Smith, R. J. (1997). Alternative semi-parametric likelihood approaches to generalized method of moments estimation. *Economic Journal 107*, 503–519.

Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association 89*, 208–218.

Wooldridge, J. M. (1997). Multiplicative panel data models without the strict exogeneity assumption. *Econometric Theory 13*, 667–678.

Wooldridge, J. M. (2005). Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of Applied Econometrics 20*(1), 39–54.

# Appendix: Proof of Theorem 1

Define $\tau = (p+1)T$ and let

$$z_{it} = \begin{bmatrix} h(y_{i,t-1}, w_{it}, \theta) \\ \partial h(y_{i,t-1}, w_{it}, \theta)/\partial\theta \end{bmatrix}$$

Following the arguments of Ahn and Schmidt (1995), we see that the model is comprised of $\tau + T + 1$ functions of data; namely $y_{i0},\ y_{i1},\ \ldots,\ y_{iT}$ and $z_{i1},\ z_{i2},\ \ldots,\ z_{i,T}$. The unrestricted variance matrix of these variables have $(\tau+T+1)(\tau+T+2)/2$ distinct components. By assumptions A1–A4, these components can be written in terms of the fundamental parameters:

(i) $\alpha_1$, $\delta$, $\mathrm{E}(\eta_i^2)$, $\mathrm{E}(y_{i0}\eta_i)$, $\mathrm{E}(y_{i0}^2)$; (5 parameters)

(ii) $\mathrm{E}(v_{it}^2), \quad t = 1, \ldots, T;$ ($T$ parameters)

(iii) $\mathrm{E}(y_{i0}z_{it}), \quad t = 1, \ldots, T;$ ($\tau$ parameters)

(iv) $\mathrm{E}(z_{is}z_{it}'), \quad s = 1, \ldots, t; \quad t = 1, \ldots, T;$ ($\tau(\tau+1)/2$ parameters)

(v) $\mathrm{E}(v_{is}z_{it}), \quad s = 1, \ldots, t-1; \quad t = 2, \ldots, T;$ ($\tau(T-1)/2$ parameters)

(vi) $\mathrm{E}(\eta_i z_{it}), \quad t = 1, \ldots, T.$ ($\tau$ parameters)

Hence the number of restrictions on the variance matrix is the difference between the number of its distinct components and the number of fundamental parameters, namely $(1+p/2)T^2 - (p/2)T - 4$. It follows that the number of moment conditions available for the estimation of $\alpha_1$ and $\delta$ is no more than $(1+p/2)T^2 - (p/2)T - 2$. The proof is complete in observing that there are $(1+p/2)T^2 - (p/2)T - 2$ moment conditions in (4), (5) and (6).