

Article

# Threshold Regression with Endogeneity for Short Panels

Tue Gørgens <sup>1,\*</sup> and Allan H. Würtz <sup>2,\*</sup><sup>1</sup> Research School of Economics, The Australian National University, Acton, ACT 2601, Australia<sup>2</sup> CREATES and Department of Economics and Business Economics, Aarhus University, Fuglesangs Allé 4, DK-8210 Aarhus V, Denmark

\* Correspondence: tue.gorgens@startmail.com (T.G.); awurtz@econ.au.dk (A.H.W.)

Received: 23 March 2019; Accepted: 16 May 2019; Published: 22 May 2019



**Abstract:** This paper considers the estimation of dynamic threshold regression models with fixed effects using short panel data. We examine a two-step method, where the threshold parameter is estimated nonparametrically at the  $N$ -rate and the remaining parameters are estimated by GMM at the  $\sqrt{N}$ -rate. We provide simulation results that illustrate advantages of the new method in comparison with pure GMM estimation. The simulations also highlight the importance of the choice of instruments in GMM estimation.

**Keywords:** threshold regression; dynamic models; endogeneity; panel data; GMM estimation; integrated difference kernel IDK estimator; superconsistency

**JEL Classification:** C23; C24; C26

## 1. Introduction

Threshold regression models allow for shifts in economic relationships when the threshold variable crosses the threshold parameter. This paper combines two recent econometric advances in estimating threshold regression models with endogeneity using short panel data sets.

[Seo and Shin \(2016\)](#) extended GMM estimation techniques for linear dynamic panel data models to threshold panel data models where both the regressors and the threshold variable may be endogenous. Their setup includes certain nonlinear dynamic panel data models such as the self-exciting threshold autoregressive (SETAR) model. We refer to this estimator as the pure GMM estimator. It has the usual properties, including  $\sqrt{N}$ -consistency and asymptotic normality, where  $N$  denotes the sample size.

[Yu and Phillips \(2018\)](#) considered the estimation of threshold regression models with endogenous regressors and threshold variable using i.i.d. data. They developed a (nonparametric) integrated difference kernel (IDK) estimator of the threshold parameter. They showed that the IDK estimator is  $N$ -consistent. Other parameters in the model can be estimated at the usual  $\sqrt{N}$ -rate by GMM, taking the estimated threshold parameter as given. The distribution of the IDK estimator is nonstandard.

In this paper, we explain how the ideas of [Yu and Phillips \(2018\)](#) can be adapted to the panel data context with fixed effects to obtain an  $N$ -consistent estimator of the threshold parameter. Following [Yu and Phillips](#), we estimate the threshold parameter using the IDK techniques and then the remaining parameters using standard GMM techniques, taking the estimated threshold parameter as given. The improvement in asymptotic efficiency of the threshold estimator spills over to the GMM estimators of the remaining parameters, since there is effectively one less parameter to estimate. The panel data context is different from the single structural equation with a single threshold variable considered by [Yu and Phillips \(2018\)](#). First, to avoid making assumptions about the fixed effects, we begin by eliminating them. This results in  $T - 2$  first-differenced structural

equations, and each equation involves two threshold variables, where  $T$  denotes the number of time periods. Second, to combine all the information available, we construct two estimators for each equation and then compute their overall average. The final step is to compute GMM estimates for the remaining parameters. Asymptotic theory for the IDK+GMM combination was provided by [Yu and Phillips \(2018\)](#) and no additional theoretical results are needed here.

We report results from a simulation study to illustrate advantages of the IDK+GMM combination over pure GMM estimation. The simulations confirm that the IDK+GMM estimator tend to have much smaller root mean square errors (RMSE) than the pure GMM estimator. For example, when  $N$  is equal to 800 the RMSE is 320% to 4630% higher for the pure GMM estimator of the threshold parameter. This reflects the fact that the IDK estimator is  $N$ -consistent while the pure GMM estimator is only  $\sqrt{N}$ -consistent.

We also investigated the importance of the choice of instruments. Even for estimating linear dynamic panel data models, the question of which moments to match remains largely unresolved (e.g., [Ahn and Schmidt 1995](#); [Arellano 2016](#)). [Seo and Shin \(2016\)](#) and [Yu and Phillips \(2018\)](#) offered different ad hoc suggestions for threshold models. Our simulations show that large reductions in RMSE are available by adding nonlinear transformations of lagged outcomes to the standard set of instruments. For example, the RSME in the baseline case is 100% to 730% higher than the RSME for an estimator that adds a constant and two percentile indicators of lagged outcomes as instruments.

## 2. The SETAR Panel Data Model

For conciseness, we focus on the self-exciting threshold autoregressive (SETAR) model which is widely used in the time series literature (e.g., [Tong and Lim 1980](#); [Teräsvirta et al. 2011](#)). In the panel data terminology, the right-hand side variables in the SETAR model are predetermined rather than endogenous. Our results are easily extended to the case of endogenous regressors and an endogenous threshold variable, as we briefly discuss in the concluding remarks. For  $i = 1, \dots, N$  individuals and  $t = 1, \dots, T$  times, let  $y_{it}$  be a scalar observed random variable. The observations are assumed to be independent across individuals, but not across time. The basic SETAR panel data model is

$$\begin{aligned} y_{it} &= y_{it-1}\alpha_1^* + 1(y_{it-1} > \gamma^*)y_{it-1}\alpha_2^* + 1(y_{it-1} > \gamma^*)\alpha_3^* + u_{it}, & t = 2, \dots, T, \\ u_{it} &= c_i + v_{it}, \end{aligned} \quad (1)$$

where  $c_i$  is a time-invariant individual-specific unobserved random variable, and  $v_{it}$  is a time- and individual-specific unobserved random variable. The overall constant term is subsumed into  $c_i$  as usual. The lowercase Greek letters denote unknown parameters, and superscripts \* indicate “true” values. The threshold parameter is  $\gamma^*$ . For simplicity, define  $\xi = (\gamma, \alpha_1, \alpha_2, \alpha_3)$ . The parameter space consists of all  $\xi \in \mathbb{R}^4$ . Assume that all random variables have finite means and variances and that

$$E(v_{it}|c_i, y_{i1}, \dots, y_{it-1}) = 0, \quad t = 2, \dots, T. \quad (2)$$

An additional smoothness assumption will be introduced in Section 4. Some authors assume  $\alpha_3^* = 0$  from the outset (e.g., [Hansen 1999](#); [González et al. 2017](#)). In the time series and cross-section literatures  $\alpha_3^*$  is estimated (e.g., [Tong and Lim 1980](#); [Tong 2011](#); [Seo and Shin 2016](#); [Yu and Phillips 2018](#)).

## 3. GMM Estimator

We begin with the pure GMM estimator. Assumption (2) implies that for any function  $f : \mathbb{R} \times \mathbb{R}^4 \rightarrow \mathbb{R}$  we have

$$E(f(y_{is}, \xi)v_{it}) = 0, \quad \forall \xi \in \mathbb{R}^4, \quad s = 1, \dots, t-1, \quad t = 2, \dots, T. \quad (3)$$

Assumption (2) therefore implies an abundance of moment restrictions that can be used to estimate the unknown parameters.

Suppose a finite set has been selected and stacked in a  $M$ -vector, say  $p_{is}(\xi)$ . For example,  $p_{is}(\xi) = y_{is}$ ,  $p_{is}(\xi) = (y_{is}, y_{is}1(y_{is} > \gamma))'$ , or  $p_{is}(\xi) = (y_{is}, y_{is}^2, y_{is}^3)'$ . Holtz-Eakin et al. (1988) and Arellano and Bond (1991) proposed a set of linear moment restrictions on the second moments of the data for the linear dynamic panel data model ( $\alpha_2^* = 0$ ,  $\alpha_3^* = 0$ , and  $p_{is}(\xi) = y_{is}$ ). Generalising their set to the present context gives

$$E(p_{is}(\xi)\Delta u_{it}) = 0, \quad \forall \xi \in \mathbb{R}^4, \quad s = 1, \dots, t-2, \quad t = 3, \dots, T. \tag{4}$$

Crepon et al. (1997), Andrews and Lu (2001), Han and Kim (2014) and Gørgens et al. (2016) pointed out that there are also useful restrictions on the first moments of the data; namely

$$E(\Delta u_{it}) = 0, \quad t = 3, \dots, T. \tag{5}$$

In addition, Ahn and Schmidt (1995) analysed the quadratic restrictions on the second moments of the data

$$E(u_{iT}\Delta u_{it}) = 0, \quad t = 3, \dots, T-1. \tag{6}$$

Note  $\Delta u_{it}$  and  $u_{iT}$  are defined using the true parameter values and expectations are taken using the true parameter values.

Define  $y_i = (y_{i1}, \dots, y_{iT})'$  and let  $g(y_i, \xi)$  be a vector of random variables such that the stacked moment restrictions can be written as  $E[g(y_i, \xi^*)] = 0$ . A necessary condition for the chosen moment restrictions to identify  $\xi^*$  is that  $E[g(y_i, \xi)] = 0$  if and only if  $\xi = \xi^*$ . A GMM estimator of  $\xi^*$  is defined as the global minimiser,  $\hat{\xi}$ , of the GMM objective function,

$$\hat{Q}(\xi) = \left[ N^{-1} \sum_{i=1}^N g(x_i, \xi) \right]' \hat{W} \left[ N^{-1} \sum_{i=1}^N g(x_i, \xi) \right], \tag{7}$$

where  $\hat{W}$  is a given weight matrix. The objective function attains its minimum on an interval of  $\gamma$  values. The ambiguity can be resolved by defining  $\hat{\gamma}$  as the midpoint (e.g., Yu 2015). Note that in general, the weight matrix  $\hat{W}$  may also be a function of the unknown parameters  $\xi$  (e.g., Hansen et al. 1996).

Despite nondifferentiability of the objective function with respect to  $\gamma$ , the asymptotic distribution of the GMM estimator is typically normal. Define the matrices  $G = D_\xi E[g(x_i, \xi^*)]$  and  $\Omega = E(g(y_i, \xi^*)g(y_i, \xi^*)')$ , where  $D_\xi$  denotes the partial derivative. Seo and Shin (2016) proved that if  $\hat{W} \rightarrow^p \Omega^{-1}$ ,  $G'\Omega^{-1}G$  is nonsingular, and other technical regularity conditions are satisfied, then

$$\sqrt{N}(\hat{\xi} - \xi^*) \rightarrow^d N(0, (G'\Omega^{-1}G)^{-1}). \tag{8}$$

In particular, the GMM estimator is  $\sqrt{N}$ -consistent.

#### 4. IDK Estimator

In this section we explain how the ideas of Yu and Phillips (2018) can be adapted to the panel data context with fixed effects to obtain an  $N$ -consistent estimator of the threshold parameter. We begin with eliminating the fixed effects by first-differencing the structural equation. Then we construct two estimators of the threshold parameter for each of the resulting  $T - 2$  equations. Finally, we obtain an overall estimator by taking the simple average of the basic estimators.

After first-differencing the structural Equation (1) and taking the conditional expectation, we get

$$E(\Delta y_{it} | y_{it-2}, y_{it-1}) = \Delta y_{it-1} \alpha_1^* + 1(y_{it-1} > \gamma^*)(y_{it-1} \alpha_2^* + \alpha_3^*) - 1(y_{it-2} > \gamma^*)(y_{it-2} \alpha_2^* + \alpha_3^*) + E(\Delta v_{it} | y_{it-2}, y_{it-1}), \quad t = 3, \dots, T. \tag{9}$$

Because the indicator functions are discontinuous, the conditional expectation is discontinuous when  $y_{it-1}$  or  $y_{it-2}$  equals  $\gamma^*$ . If the conditional expectation is smooth everywhere else, then these

discontinuities identify  $\gamma^*$ . The idea of the IDK estimator is to exploit the discontinuities for estimating  $\gamma^*$ . To rule out discontinuities occurring elsewhere, in addition to (2) assume that

$$E(\Delta v_{it} | y_{it-2} = a, y_{it-1} = b) \text{ is continuous in } (a, b), \quad t = 3, \dots, T. \tag{10}$$

To show that the discontinuities identify  $\gamma^*$ , let  $\gamma^-$  and  $\gamma^+$  indicate limits from the left and from the right, and define the functions  $A_t$  and  $B_t$  as the difference between the left and right limits of the conditional expectation function when  $y_{it-1}$  and  $y_{it-2}$  is near  $\gamma^*$ ; that is,

$$A_t(y, \gamma) = E(\Delta y_{it} | y_{it-2} = y, y_{it-1} = \gamma^+) - E(\Delta y_{it} | y_{it-2} = y, y_{it-1} = \gamma^-), \quad t = 3, \dots, T, \tag{11}$$

and

$$B_t(\gamma, y) = E(\Delta y_{it} | y_{it-2} = \gamma^-, y_{it-1} = y) - E(\Delta y_{it} | y_{it-2} = \gamma^+, y_{it-1} = y), \quad t = 3, \dots, T. \tag{12}$$

Using assumption (10), we then have

$$A_t(y, \gamma) = B_t(\gamma, y) = \{1(\gamma^+ > \gamma^*) - 1(\gamma^- > \gamma^*)\}(\gamma^* \alpha_2^* + \alpha_3^*) = \begin{cases} 0 & \text{if } \gamma \neq \gamma^*, \\ \gamma^* \alpha_2^* + \alpha_3^* & \text{if } \gamma = \gamma^*, \end{cases} \quad t = 3, \dots, T. \tag{13}$$

It follows that  $\gamma^* = \arg \max_{\gamma} A_t(y, \gamma)^2$  and  $\gamma^* = \arg \max_{\gamma} B_t(y, \gamma)^2$  for all  $y \in \mathbb{R}$ . Furthermore,  $\gamma^* \alpha_2^* + \alpha_3^* \neq 0$  is a necessary condition for (13) to uniquely identify  $\gamma^*$ .

While it is possible to base estimation of  $\gamma^*$  on  $A_t(y, \cdot)$  or  $B_t(\cdot, y)$  with a fixed value of  $y$ , such an estimator will not have good properties. To achieve  $N$ -consistency, our estimators of  $\gamma^*$  are based on density-weighted averages of  $A_t$  and  $B_t$ . Let  $r_t$  denote the joint density of  $(y_{it-2}, y_{it-1})$  and let  $p_t$  denote the marginal density of  $y_{it}$ . Define the objective function  $R_t^A$  by

$$R_t^A(\gamma) = E[A_t(y_{it-2}, \gamma)^2 r_t(y_{it-2}, \gamma)^2] = \int_{-\infty}^{\infty} A_t(y, \gamma)^2 r_t(y, \gamma)^2 p_{t-2}(y) dy, \quad t = 3, \dots, T, \tag{14}$$

and the objective function  $R_t^B$  by

$$R_t^B(\gamma) = E[B_t(\gamma, y_{it-1})^2 r_t(\gamma, y_{it-1})^2] = \int_{-\infty}^{\infty} B_t(\gamma, y)^2 r_t(\gamma, y)^2 p_{t-1}(y) dy, \quad t = 3, \dots, T. \tag{15}$$

The discontinuity points of  $R_t^A$  and  $R_t^B$  are the same as those of  $A_t(y, \cdot)$  and  $B_t(\cdot, y)$  provided certain technical regularity conditions hold, including that  $r_t$  is continuous and bounded away from 0 in an open neighbourhood where  $y_{it-2} = \gamma^*$  or  $y_{it-1} = \gamma^*$ . That is, we generally have that  $\gamma^* = \arg \max_{\gamma} R_t^A(\gamma)$  and  $\gamma^* = \arg \max_{\gamma} R_t^B(\gamma)$ .

We define “basic” IDK estimators as the arg max of each of the sample analogues of  $R_t^A$  and  $R_t^B$  for  $t = 3, \dots, T$ . The estimators of  $R_t^A$  and  $R_t^B$  are implemented using generalised kernels. Let  $k$  be a univariate kernel function with support  $[-1, 1]$ , and let  $h$  denote the bandwidth. To keep the notation simple, we use the same bandwidth everywhere. Then estimator of  $R_t^A$  and  $R_t^B$  are

$$\hat{R}_t^A(\gamma) = \frac{1}{N} \sum_{i=1}^N \left( \frac{1}{N-1} \sum_{\substack{j=1 \\ j \neq i}}^N \Delta y_{jt} K_h(y_{jt-2} - y_{it-2}) k_h^+(y_{jt-1} - \gamma) - \frac{1}{N-1} \sum_{\substack{j=1 \\ j \neq i}}^N \Delta y_{jt} K_h(y_{jt-2} - y_{it-2}) k_h^-(y_{jt-1} - \gamma) \right)^2, \quad t = 3, \dots, T, \tag{16}$$

and

$$\hat{R}_t^B(\gamma) = \frac{1}{N} \sum_{i=1}^N \left( \frac{1}{N-1} \sum_{\substack{j=1 \\ j \neq i}}^N \Delta y_{jt} K_h(y_{jt-1} - y_{it-1}) k_h^-(y_{jt-2} - \gamma) - \frac{1}{N-1} \sum_{\substack{j=1 \\ j \neq i}}^N \Delta y_{jt} K_h(y_{jt-1} - y_{it-1}) k_h^+(y_{jt-2} - \gamma) \right)^2, \quad t = 3, \dots, T, \quad (17)$$

where

$$K_h(a) = \frac{1}{h} k\left(\frac{a}{h}\right), \quad (18)$$

$$k_h^-(a) = \frac{1(-1 < \frac{a}{h} < 0) \frac{1}{h} k(\frac{a}{h})}{\int_{-1}^0 k(v) dv}, \quad (19)$$

$$k_h^+(a) = \frac{1(0 < \frac{a}{h} < 1) \frac{1}{h} k(\frac{a}{h})}{\int_0^1 k(v) dv}. \quad (20)$$

Define the estimators  $\hat{\gamma}_t^A = \arg \max_{\gamma} \hat{R}_t^A(\gamma)$  and  $\hat{\gamma}_t^B = \arg \max_{\gamma} \hat{R}_t^B(\gamma)$  for  $t = 3, \dots, T$ . Finally, we construct an overall estimator  $\hat{\gamma}$  by taking the average of all  $\hat{\gamma}_t^A$  and  $\hat{\gamma}_t^B$ .

Having estimated  $\gamma^*$ , the  $\alpha$ 's can be estimated in a second step at the  $\sqrt{N}$ -rate by GMM as described in Section 3 after redefining  $\xi = (\alpha_1, \alpha_2, \alpha_3)$ . Since  $\hat{\gamma}$  converges at the  $N$ -rate, the asymptotic distribution is the same as if  $\gamma^*$  is known.

The setup here differs somewhat from that of [Yu and Phillips \(2018\)](#), who considered a single structural equation with a single threshold variable. Here we have  $T - 2$  first-differenced structural equations, and each equation involves two threshold variables. The latter means that it is necessary to condition on both  $y_{it-2}$  and  $y_{it-1}$  in (9), and gives rise to the two distinct estimators based on  $A_t$  and  $B_t$ , respectively.

[Yu and Phillips \(2018\)](#) proved that the basic IDK estimator is  $N$ -consistent under certain technical regularity conditions. The asymptotic distribution is nonstandard. Their results apply directly to each of our basic estimators,  $\hat{\gamma}_t^A$  and  $\hat{\gamma}_t^B$  for  $t = 3, \dots, T$ . Taking the overall average does not affect the  $N$ -consistency and reduces the variance. [Yu and Phillips \(2018\)](#) did not provide standard errors in their empirical illustration. Arguably, we are interested in making inferences about the regression function in most empirical applications, not about individual parameters, and the former is dominated by the variance of  $\hat{\alpha}$ s, while the variance of  $\hat{\gamma}$  is negligible in comparison. Inference methods for the threshold parameter are developed by [Liao et al. \(2018\)](#).

## 5. Simulation Results

To illustrate the advantage of the IDK+GMM estimator over pure GMM and to investigate the importance of the choice of instruments, we conducted a small simulation study for one of the designs used by [Seo and Shin \(2016\)](#). The DGP is defined in the table note. For simplicity, all results for the GMM estimators presented here are one-step estimators using the optimal weight matrix.

Panel A of Table 1 shows our baseline results which use only the untransformed lagged outcome variables as instruments, as suggested by [Seo and Shin \(2016\)](#). The RMSE for the pure GMM estimator are monotonically decreasing at rates suggesting  $\sqrt{N}$ -consistency, as expected. The RMSE for the IDK+GMM estimator are much lower, especially for  $\gamma$ , and the convergence rates are compatible with  $N$ -consistency for  $\gamma$  and  $\sqrt{N}$ -consistency for the  $\alpha$ s.

Given the disparate convergence rates we expect the RMSE ratio of pure GMM to the IDK+GMM combination for  $\gamma$  to diverge, while the RMSE ratios for the  $\alpha$ s should converge to finite limit values corresponding to the ratio of the asymptotic variances of the respective GMM estimators. The numbers

shown in the right-most four columns in Table 1 are compatible with these expectations. In panel A, when  $N = 800$ , the efficiency gain for  $\gamma$  is huge, more than a factor of 27. The gains for the  $\alpha$ s are also large, with RMSE for pure GMM more than twice the RMSE for the IDK+GMM estimator.

Table 1. Simulation results.

N	RMSE Pure GMM				RMSE IDK+GMM				Pure/IDK			
	$\gamma$	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\gamma$	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\gamma$	$\alpha_1$	$\alpha_2$	$\alpha_3$
(A) Instruments: $y_{is}$												
50	0.47	0.30	0.57	1.45	0.07	0.24	0.33	0.58	6.6	1.3	1.8	2.5
100	0.39	0.26	0.48	1.10	0.04	0.19	0.27	0.47	9.0	1.3	1.8	2.3
200	0.36	0.23	0.44	1.09	0.03	0.15	0.21	0.38	14.1	1.6	2.1	2.8
400	0.31	0.20	0.38	0.85	0.01	0.11	0.16	0.30	21.1	1.8	2.4	2.9
800	0.22	0.17	0.30	0.47	0.01	0.08	0.12	0.23	27.6	2.1	2.6	2.1
(B) Instruments: constant, $y_{is}$												
50	0.37	0.25	0.45	1.11	0.07	0.21	0.28	0.54	5.3	1.2	1.6	2.0
100	0.31	0.21	0.36	0.77	0.04	0.17	0.23	0.45	7.2	1.2	1.6	1.7
200	0.28	0.19	0.35	0.81	0.03	0.13	0.17	0.37	10.9	1.4	2.0	2.2
400	0.24	0.16	0.30	0.55	0.01	0.10	0.13	0.29	16.4	1.6	2.4	1.9
800	0.18	0.13	0.25	0.47	0.01	0.07	0.09	0.22	22.4	1.9	2.7	2.1
(C) Instruments: constant, $y_{is}, y_{is}1(y_{is} > \hat{\gamma})$												
50	0.56	0.37	0.70	1.95	0.07	0.15	0.17	0.46	7.9	2.5	4.2	4.2
100	0.52	0.34	0.61	1.72	0.04	0.12	0.12	0.38	12.3	2.8	5.2	4.6
200	0.50	0.29	0.56	1.63	0.03	0.10	0.08	0.31	19.2	3.0	7.1	5.2
400	0.42	0.26	0.47	1.24	0.01	0.08	0.05	0.25	28.8	3.3	9.0	4.9
800	0.38	0.23	0.44	1.10	0.01	0.06	0.03	0.20	47.3	4.0	12.5	5.6
(D) Instruments: constant, $y_{is}, y_{is}^2, y_{is}^3$												
50	0.15	0.14	0.21	0.40	0.07	0.13	0.17	0.33	2.2	1.1	1.2	1.2
100	0.09	0.10	0.14	0.24	0.04	0.09	0.12	0.23	2.2	1.1	1.2	1.0
200	0.06	0.07	0.10	0.16	0.03	0.06	0.08	0.16	2.4	1.1	1.3	1.0
400	0.05	0.05	0.07	0.12	0.01	0.04	0.05	0.11	3.1	1.2	1.4	1.0
800	0.04	0.04	0.06	0.08	0.01	0.03	0.04	0.08	4.6	1.3	1.6	1.0
(E) Instruments: constant, $y_{is}, y_{is}1(y_{is} > y_{0.33}), y_{is}1(y_{is} > y_{0.67})$												
50	0.06	0.13	0.16	0.26	0.07	0.12	0.16	0.28	0.9	1.0	1.0	0.9
100	0.05	0.09	0.12	0.19	0.04	0.08	0.11	0.19	1.1	1.1	1.1	1.0
200	0.04	0.07	0.10	0.14	0.03	0.06	0.07	0.14	1.7	1.2	1.3	1.0
400	0.04	0.06	0.07	0.11	0.01	0.04	0.05	0.10	2.5	1.3	1.5	1.0
800	0.03	0.05	0.06	0.08	0.01	0.03	0.03	0.07	4.2	1.5	1.9	1.1
(F) Ratio of panel A over panel E												
50	7.62	2.42	3.48	5.51		2.01	2.01	2.04				
100	8.01	2.82	3.88	5.74		2.29	2.42	2.41				
200	8.15	3.21	4.61	7.56		2.46	2.81	2.72				
400	8.31	3.54	5.05	7.89		2.58	3.09	2.89				
800	6.60	3.68	4.75	6.01		2.71	3.41	3.17				

RMSE: root mean square error;  $s = 1, \dots, t - 2$ ;  $y_b$ : percentile  $b$  of  $y_{it}$ . The DGP is  $\gamma^* = 0, \alpha_1^* = -0.5, \alpha_2^* = 1.2, \alpha_3^* = -2.5, c_i = 0.7, v_{it} \sim \text{Normal}(0, 1)$ , and  $y_{i,-30} \sim \text{Normal}(0, 1)$ . The estimations use data for  $1 \leq t \leq 10$ . All experiments have 5000 simulated samples. The bandwidths are 6.5 times the standard deviation of regressors.

In the remainder of Table 1 we consider different sets of instruments. Panel B shows big reductions in RMSE for the pure GMM estimator when a constant term is also used as an instrument. Han and Kim (2014) and Gørgens et al. (2016) found similar improvements for the linear model. The improvements are relatively less for the IDK+GMM estimator.

Since the structural equation is nonlinear, one might expect that nonlinear transformations of lagged outcomes could be useful instruments. Based on the suggestion by Yu and Phillips (2018),

we added  $y_{is}1(y_{is} > \hat{\gamma})$  to the set of instruments. Panel C in Table 1 shows that this does not improve the RMSE for the pure GMM estimator. On the contrary, the estimation noise in the instruments adds significantly to the RMSE. The results are more promising for the IDK+GMM estimator, where substantial reductions in RMSE are observed.

In panel D, we have added quadratic and cubic transformations of the lagged dependent variable, and in panel E we have added threshold functions where the threshold depends on percentiles of the data rather than the structural parameter. As shown in panel F, when  $N = 800$  the RMSE for the pure GMM estimator drops by factors of 3.6–6.6, while the RMSE for the IDK+GMM estimator drops by factors of 2.7–3.4.

## 6. Concluding Remarks

This paper has shown how the ideas of [Yu and Phillips \(2018\)](#) can be adapted to the panel data context with fixed effects. Theoretically, the advantage of the IDK+GMM combination is that the estimator of the threshold parameter is  $N$ -consistent, while the pure GMM estimator converges only at the  $\sqrt{N}$ -rate. In simulation exercises, we confirmed that the IDK+GMM combination offers a huge practical advantage over pure GMM estimation, even when the former is implemented relatively simply. We also investigated the importance of the choice of instruments and showed that adding fixed nonlinear transformations of the lagged dependent variable can be highly effective when estimating nonlinear equations.

We have focused on the SETAR model in this paper. A more general threshold regression panel data model is

$$\begin{aligned} y_{it} &= x'_{it}\alpha_1^* + 1(q_{it} > \gamma^*)x'_{it}\alpha_2^* + 1(q_{it} > \gamma^*)\alpha_3^* + u_{it}, & t = 1, \dots, T, \\ u_{it} &= c_i + v_{it}, \end{aligned} \quad (21)$$

where  $x_{it}$  is a vector of possibly endogenous variables,  $q_{it}$  is a possibly endogenous scalar variable, and  $\alpha_1^*$ ,  $\alpha_2^*$  and  $\alpha_3^*$  are conformable parameter vectors. It is straightforward to construct an IDK+GMM estimator analog to the SETAR case, and similar efficiency gains are available.

The IDK estimator we have described utilises discontinuities in the conditional expectation function given in (9). It will fail if  $\gamma^*\alpha_2^* + \alpha_3^* = 0$ , because then (9) is continuous. However, in this case the partial derivatives of (9) may be discontinuous at  $y_{it-2} = \gamma^*$  or  $y_{it-1} = \gamma^*$ , so IDK estimation is still possible (e.g., [Yu and Phillips 2018](#); [Porter and Yu 2015](#)).

If it is known that  $E(c_i | y_{it-1} = y)$  is a smooth function of  $y$ , then we can construct an estimator of  $\gamma^*$  directly based on Equation (1), without first-differencing and without assumption (10). Since an extra time period is available for estimation and since we only need to smooth in one dimension ( $y_{it-1}$ ) instead of two ( $y_{it-1}, y_{it-2}$ ) when defining  $R_t$ , this estimator is expected to be more efficient.

For simplicity, we have constructed an overall estimator by taking a simple average of multiple estimators based on separate equations. It is a topic for future research to investigate how best to combine the information. One could consider weighted averages or, instead of averaging separate estimators, one could base an estimator on a (weighted) average over the objective functions. Which is better may depend on e.g., the time pattern of  $\text{Var}(v_{it})$ .

Finally, to illustrate the advantage of the IDK+GMM estimator over the pure GMM estimator, our simulations focused on the design considered by [Seo and Shin \(2016\)](#). To further investigate the properties of the IDK+GMM estimator in future research, it would be interesting to consider simulation designs where endogeneity is more severe (e.g.,  $c_i$  is correlated with  $y_{i1}$ ) and where the number of time periods is smaller (i.e.,  $T$  is small). Also, in practice the optimal weight matrix is not known, and it would be useful to compare two-step estimation of the weight matrix and continuous updating (e.g., [Hansen et al. 1996](#)).

**Author Contributions:** Both authors contributed equally to the paper.

**Funding:** Tue Gørgens' research was supported in part by Australian Research Council Grant DP1096862.

**Conflicts of Interest:** The authors declare no conflict of interest. The funder had no role in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## References

- Ahn, Seung Chan, and Peter Schmidt. 1995. Efficient estimation of models for dynamic panel data. *Journal of Econometrics* 68: 5–27. [CrossRef]
- Andrews, Donald W. K., and Biao Lu. 2001. Consistent model and moment selection procedures for GMM estimation with application to dynamic panel data models. *Journal of Econometrics* 101: 123–64. [CrossRef]
- Arellano, Manuel. 2016. Modelling optimal instrumental variables for dynamic panel data models. *Research in Economics* 70: 238–61. [CrossRef]
- Arellano, Manuel, and Stephen Bond. 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies* 58: 277–97. [CrossRef]
- Crepon, Bruno, Francis Kramarz, and Alain Trognon. 1997. Parameters of interest, nuisance parameters and orthogonality conditions: An application to autoregressive error component models. *Journal of Econometrics* 82: 135–56. [CrossRef]
- González, Andrés, Timo Teršvirta, Dick Van Dijk, and Yukai Yang. 2017. *Panel Smooth Transition Regression Models*. CREATES Research Paper #2017-36. Aarhus: Aarhus University.
- Gørgens, Tue, Chirok Han, and Sen Xue. 2016. *Moment Restrictions and Identification in Linear Dynamic Panel Data Models*. ANU Working Papers in Economics and Econometrics #633. Canberra: Australian National University.
- Han, Chirok, and Hyoungjong Kim. 2014. The role of constant instruments in dynamic panel estimation. *Economics Letters* 124: 500–503. [CrossRef]
- Hansen, Bruce E. 1999. Threshold effects in non-dynamic panels: Estimation, testing, and inference. *Journal of Econometrics* 93: 345–68. [CrossRef]
- Hansen, Lars Peter, John Heaton, and Amir Yaron. 1996. Finite-sample properties of some alternative GMM estimators. *Journal of Business & Economic Statistics* 14: 262–80.
- Holtz-Eakin, Douglas, Whitney Newey, and Harvey S. Rosen. 1988. Estimating vector autoregressions with panel data. *Econometrica* 56: 1371–95. [CrossRef]
- Liao, Qin, Peter C. B. Phillips, and Ping Yu. 2018. Inferences and Specification Testing in Threshold Regression with Endogeneity. Available online: <https://sites.google.com/view/qinliao/research> (accessed on 3 May 2019).
- Porter, J., and P. Yu. 2015. Regression discontinuity with unknown discontinuity points: Testing and estimation. *Journal of Econometrics* 189: 132–47. [CrossRef]
- Seo, Myung Hwan, and Yongcheol Shin. 2016. Dynamic panels with threshold effect and endogeneity. *Journal of Econometrics* 195: 169–86. [CrossRef]
- Teräsvirta, Timo, Dag Tjøstheim, and Clive W. J. Granger. 2011. *Modelling Nonlinear Economic Time Series*. Advanced Texts in Econometrics. Oxford: Oxford University Press.
- Tong, Howell. 2011. Threshold models in time series analysis—30 years on. *Statistics and Its Interface* 4: 107–18. [CrossRef]
- Tong, H., and K. S. Lim. 1980. Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society, Series B* 42: 245–92. [CrossRef]
- Yu, Ping. 2015. Adaptive estimation of the threshold point in threshold regression. *Journal of Econometrics* 189: 83–100. [CrossRef]
- Yu, Ping, and Peter C. B. Phillips. 2018. Threshold regression with endogeneity. *Journal of Econometrics* 203: 50–68. [CrossRef]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).