Dynamic Panels with Threshold Effect and Endogeneity

Myung Hwan Seo
Department of Economics, Seoul National University, Kwan-Ak Ro 1, Kwan-Ak Gu, Seoul, Korea
Yongcheol Shin**
Department of Economics and Related Studies, University of York, York YO105DD, UK

Abstract

This paper addresses an important issue of modelling nonlinear asymmetric dynamics and unobserved individual heterogeneity in the threshold panel data framework, simultaneously. As a general approach, we develop the first-differenced GMM estimator, which allows both threshold variable and regressors to be endogenous. When the threshold variable becomes strictly exogenous, we propose a more efficient two-step least squares estimator. We provide asymptotic theory and develop the testing procedure for threshold effects and the threshold variable exogeneity. Monte Carlo studies provide a support for theoretical predictions. We present an empirical application investigating an asymmetric sensitivity of investment to cash flows.

JEL Classification: C13, C33, G31, G35
Key Words: Dynamic Panel Threshold Models, Endogenous Threshold Effects and Regressors, FD-GMM and FD-2SLS, Linearity and Exogeneity Tests, Investment.

** Corresponding author.

1 Introduction

The econometric literature on dynamic models has long been interested in the implications of the existence of nonlinear asymmetric dynamics. Examples include Markov-Switching, Smooth Transition and Threshold Autoregression Models. The popularity of these models lies in allowing us to draw inferences about the underlying data generating process or to yield reliable forecasts in a manner that is not possible using linear models. Until recently, however, most econometric analysis has stopped short of studying the issues of nonlinear asymmetric mechanisms explicitly within a dynamic panel data context. Hansen (1999) develops a static panel threshold model where regression coefficients can take on a small number of different values, depending on the value of exogenous stationary variable. González *et al.* (2005) generalize this approach and develop a panel smooth transition regression model which allows the coefficients to change gradually from one regime to another.[1] In a broad context these models are a specific example of the panel approach that allows coefficients to vary randomly over time and across cross-section units as surveyed by Hsiao (2003, Chapter 6).

These approaches are static, the validity of which has not yet been established in dynamic panels, though increasing availability of the large panel data sets has prompted more rigorous econometric analyses of dynamic heterogeneous panels. Surprisingly, there has been almost no rigorous study investigating an important issue of nonlinear asymmetric mechanism in dynamic panels, especially when time periods are short, though there is a huge literature on GMM estimation of linear dynamic panels with heterogeneous individual effects, *e.g.*, Holtz-Eakin *et al.* (1988), Arellano and Bond (1991), Ahn and Schmidt (1995), Arellano and Bover (1995), Blundell and Bond (1998), Alvarez and Arellano (2003), Bun and Windmeijer (2010), Hayakawa (2015) and Hsiao and Zhang (2015).

Another limitation is the maintained assumption of exogeneity of the regressors and/or the threshold variable. While the endogenous transition in the Markov-Switching model has been studied by Kim *et al.* (2008), much progress has not been made in the threshold regression literature. The standard least squares approach, such as Hansen (2000) and Seo and Linton (2007), requires exogeneity in all the covariates. Caner and Hansen (2004) relax this requirement by allowing for endogenous regressors, but they assume the threshold variable to be exogenous. See also Hansen (2011) for an extensive survey.

In the dynamic panel context, Dang *et al.* (2012) have proposed the generalized GMM estimator applicable for dynamic panel threshold models, which can provide consistent esti-

[1] See Fok *et al.* (2005) for a large T treatment of smooth transition regression, thus not requiring the fixed effect or first difference transformation to estimate the model.

mates of heterogeneous speeds of adjustment as well as a valid testing procedure for threshold effects in short dynamic panels with unobserved individual effects. Ramirez-Rondan (2013) has extended the Hansen's (1999) work to allow the threshold mechanism in dynamic panels, and proposed the maximum likelihood estimation techniques, following the approach by Hsiao *et al.* (2002). In order to allow endogenous regressors, Kremer *et al.* (2013) have considered a hybrid dynamic version by combining the forward orthogonal deviations transformation by Arellano and Bover (1995) and the instrumental variable estimation of the cross-section model by Caner and Hansen (2004). However, the crucial assumption in all of these studies is that either regressors or the transition variable or both are exogenous.[2]

We aim to fill this gap by explicitly addressing an important issue as how best to model nonlinear asymmetric dynamics and unobserved individual heterogeneity, simultaneously. To this end we extend the approaches by Hansen (1999, 2000) and Caner and Hansen (2004) to the dynamic panel data model with endogenous threshold variable and regressors. Specifically, following the main literature on the GMM, we consider the asymptotic experiment under large cross-section unit with a fixed time period.

We propose a general GMM approach based on the first-difference (FD) transformation. As we allow both threshold variable and regressors to be endogenous, the FD-GMM approach is expected to overcome the main limitation in the existing literature, namely, the assumption of exogeneity of regressors and/or the transition variable that may hamper the usefulness of threshold regression models in a general context. We develop the asymptotic theory through the diminishing threshold and the standard fixed threshold asymptotics (*e.g.* Hansen, 2000), and show that the FD-GMM estimator follows a normal distribution asymptotically. More importantly, the asymptotic normality holds true irrespective of whether the regression function is continuous or not. Hence, the standard inference on the threshold and other parameters based on the Wald statistic can be carried out. This is in contrast to the least squares approach in which the discontinuity of the regression function changes the asymptotic distribution in a dramatic way.

Next, we examine the special case where the threshold variable is strictly exogenous, and propose a more efficient two-step least squares (FD-2SLS) estimator. This generalizes Caner and Hansen's (2004) approach for the cross-section data to the dynamic panel data with a fixed effect. We establish that the FD-2SLS estimator satisfies the Oracle property because the threshold and the slope estimates are asymptotically independent. Furthermore, the FD-2SLS estimator of the threshold parameter is shown to be super-consistent. Though its infer-

[2] Recently, Yu and Phillips (2014) and Kourtellos *et al.* (2015) have also addressed an issue of endogenous threshold variable in the single equation context.

ence is non-standard, we show that a properly weighted LR statistic follows the same pivotal asymptotic distribution as in Hansen (2000).

We provide testing procedures for identifying the threshold effect, based on the supremum statistics, which follow non-standard asymptotic distributions due to the loss of identification under the null of no threshold effect. The critical values or the p-values of the tests can be easily evaluated by the bootstrap. Furthermore, we develop the Hausman type testing procedure for the validity of the null hypothesis that the threshold variable is exogenous.

Finite sample property of the FD-GMM estimator is examined through Monte Carlo studies. Specifically, we evaluate its bias and mean squared error, and the coverage probability of the confidence interval constructed by the asymptotic normal approximation. Overall results provide support for our theoretical predictions. Given that there are many different ways to compute the weight matrix in the first step, we propose an averaging of a class of the two-step FD-GMM estimators that are obtained by randomizing the weight matrix in the first step. This turns out to be successful in significantly reducing the sampling errors.

Using the UK company panel data, we demonstrate the usefulness of the proposed dynamic threshold panel data modelling by providing an empirical application investigating an asymmetric sensitivity of investment to cash flows. We consider three firm-specific variables as an endogenous threshold variable that potentially affects the investment dynamics. By employing a panel dataset of 560 UK firms over the period 1973-1987, we find that the cash flow sensitivity of investment is significantly stronger for cash-constrained, high-growth and high-leveraged firms, a consistent finding with an original hypothesis by Farazzi *et al.* (1988) that the sensitivity of investment to cash flows is an indicator of the degree of financial constraints.

The plan of the paper is as follows: Section 2 describes the model. Section 3 presents the detailed estimation steps for FD-GMM. Section 4 develops an asymptotic theory, including consistent and efficient estimation of the threshold parameter. Section 5 provides the inference for threshold effects and endogeneity of the transition variable. Finite sample performance of the FD-GMM estimator is examined in Section 6. Empirical application is presented in Section 7. Section 8 concludes. We provide two Appendices. Appendix A presents the estimation theory for FD-2SLS, which is shown to be more efficient in the special case where the threshold variable is exogenous. All the mathematical proofs are collected in Appendix B.

2 The Model

Consider the following dynamic panel threshold regression model:

$$y_{it} = \left(1, x_{it}'\right) \phi_1 1\left\{q_{it} \leq \gamma\right\} + \left(1, x_{it}'\right) \phi_2 1\left\{q_{it} > \gamma\right\} + \varepsilon_{it},\ i = 1, ..., n;\ t = 1, ..., T, \tag{1}$$

where y_{it} is a scalar stochastic variable of interest, x_{it} the $k_1 \times 1$ vector of time-varying regressors, that may include the lagged dependent variable, $1\{\cdot\}$ an indicator function, and q_{it} the transition variable. γ is the threshold parameter, and ϕ_1 and ϕ_2 are the slope parameters associated with different regimes. The error, ε_{it} consists of the error components:

$$\varepsilon_{it} = \alpha_i + v_{it}, \tag{2}$$

where α_i is an unobserved individual fixed effect and v_{it} is a zero mean idiosyncratic random disturbance. In particular, v_{it} is assumed to be a martingale difference sequence,

$$\mathrm{E}\left(v_{it}|\mathcal{F}_{t-1}\right) = 0,$$

where $\mathcal{F}_t$ is a natural filtration at time t. It is worthwhile to mention that we do not assume x_{it} or q_{it} to be measurable with respect to $\mathcal{F}_{t-1}$, say $\mathrm{E}\left(v_{it}x_{it}\right) \neq 0$ or $\mathrm{E}\left(v_{it}q_{it}\right) \neq 0$, thus allowing endogeneity in both the regressor, x_{it} and the threshold variable, q_{it}.

The estimation of dynamic panel data with a large number of individuals but with a fixed number of time periods has been commonplace, *e.g.* Holts-Eakin *et al.* (1988), Arellano and Bond (1991), Arellano and Bover (1995), Ahn and Schmidt (1995), Blundell and Bond (1998) and Alvarez and Arellano (2003). Following this tradition, we aim to extend the static panel threshold modelling advanced by Hansen (1999), and generalize the Arellano and Bond (1991) FD-GMM estimator to a new estimation approach applicable for dynamic panel threshold models. Specifically, we consider the asymptotic experiment under large n with a fixed T,[3] in which case the martingale difference assumption is just for expositional simplicity. The sample is generated from random sampling across i.

A leading example of interest is the self-exciting threshold autoregressive (SETAR) model popularized by Tong (1990), in which case x_{it} consists of the lagged y_{it}'s and $q_{it} = y_{i,t-d}$ for any $d \geq 1$.[4]

It is well-established in the linear dynamic panel data literature that the fixed effects estimator of the autoregressive parameters is biased downward (*e.g.* Nickell, 1981). To deal with the correlation of regressors with individual effects in (1) and (2), we consider the first-difference transformation of (1) as follows (*e.g.* Arellano and Bond, 1991):[5]

$$\Delta y_{it} = \beta' \Delta x_{it} + \delta' X_{it}' \mathbf{1}_{it}(\gamma) + \Delta\varepsilon_{it}, \tag{3}$$

[3] On the other hand, if $\frac{T}{N} \to c$ as $N \to \infty$, we conjecture (*e.g.* Alvarez and Arellano, 2003; Hsiao and Zhang, 2015) that our proposed FD-GMM estimator is asymptotically biased of order $\sqrt{c}$.

[4] We note in passing that all the results go through when $q_{it} = y_{it-d}$ for any $d \geq 1$, which covers the delayed SETAR mechanism. It is sufficient to check if the moment conditions hold with the particular choice of q_{it}.

[5] In (2), $\Delta\varepsilon_{it} = \Delta v_{it}$. For convenience we use $\Delta\varepsilon_{it}$ instead of Δv_{it} throughout the paper. Here, we decompose the parameters, $\phi_1 = (\phi_{11}, \phi_{12}')'$, $\phi_2 = (\phi_{21}, \phi_{22}')'$and $\delta = (\delta_1, \delta_2')'$, conformable with $(1, x_{it}')'$.

where Δ is the first difference operator, $\underset{k_1\times 1}{\beta} = \left(\phi_{12}, ..., \phi_{1,k_1+1}\right)'$, $\underset{(k_1+1)\times 1}{\delta} = \phi_2 - \phi_1$, and

$$\underset{2\times(1+k_1)}{X_{it}} = \begin{pmatrix} (1, x'_{it}) \\ \left(1, x'_{i,t-1}\right) \end{pmatrix} \text{ and } \underset{2\times 1}{\mathbf{1}_{it}(\gamma)} = \begin{pmatrix} 1\{q_{it} > \gamma\} \\ -1\{q_{it-1} > \gamma\} \end{pmatrix}.$$

Let $\theta = \left(\beta', \delta', \gamma\right)'$ and assume that θ belongs to a compact set, $\Theta = \Phi \times \Gamma \subset \mathbb{R}^k$, with $k = 2k_1 + 2$. Following convention, we let $\Gamma = \left[\underline{\gamma}, \overline{\gamma}\right]$, where $\underline{\gamma}$ and $\overline{\gamma}$ are two percentiles of the threshold variable. Typically, they are lower and upper tenth or fifteenth percentiles.

We allow for both "fixed threshold effect" and "diminishing or small threshold effect" for statistical inference for the threshold parameter, γ by defining (*e.g.* Hansen, 2000):

$$\delta = \delta_n = \delta_0 n^{-\alpha} \text{ for } 0 \leq \alpha < 1/2. \tag{4}$$

The OLS estimator obtained from (3) is biased since the transformed regressors are correlated with $\Delta\varepsilon_{it}$. To fix this problem we need to find an $l \times 1$ vector of instrument variables, $\left(z'_{it_0}, ..., z'_{iT}\right)'$ for $2 < t_0 \leq T$ with $l \geq k$ such that either

$$\mathrm{E}\left(z'_{it_0}\Delta\varepsilon_{it_0}, ..., z'_{iT}\Delta\varepsilon_{iT}\right)' = 0, \tag{5}$$

or

$$\mathrm{E}\left(\Delta\varepsilon_{it}|z_{it}\right) = 0, \text{ for each } t = t_0, ..., T. \tag{6}$$

Notice that z_{it} may include lagged values of (x_{it}, q_{it}) and lagged dependent variables and that the number of instruments may be different for each time t.[6]

3 FD-GMM Estimation

We allow for the threshold variable q_{it} to be endogenous; $E\left(q_{it}\Delta\varepsilon_{it}\right) \neq 0$ such that q_{it} does not belong to the set of instrumental variables, $\{z_{it}\}_{t=t_0}^{T}$. We consider the following l-dimensional column vector of the sample moment conditions:

$$\bar{g}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} g_i(\theta),$$

where

$$\underset{l\times 1}{g_i(\theta)} = \begin{pmatrix} z_{it_0}\left(\Delta y_{it_0} - \beta'\Delta x_{it_0} - \delta' X'_{it_0}\mathbf{1}_{it_0}(\gamma)\right) \\ \vdots \\ z_{iT}\left(\Delta y_{iT} - \beta'\Delta x_{iT} - \delta' X'_{iT}\mathbf{1}_{iT}(\gamma)\right) \end{pmatrix}. \tag{7}$$

[6] In practice, the choice of instruments is important. In Section 4.1 we present the order and the rank conditions (see Assumption 3 below) for the practitioners to check with their own choice of instruments.

Assume that $\mathrm{E}g_i(\theta) = 0$ if and only if $\theta = \theta_0$ and let $g_i = g_i(\theta_0) = \left(z'_{it_0}\Delta\varepsilon_{it_0}, ..., z'_{iT}\Delta\varepsilon_{iT}\right)'$, and $\Omega = \mathrm{E}(g_i g_i')$, where Ω is assumed to be positive definite. For a positive definite matrix, W_n such that $W_n \xrightarrow{p} \Omega^{-1}$, let

$$\bar{J}_n(\theta) = \bar{g}_n(\theta)' W_n \bar{g}_n(\theta). \tag{8}$$

Then, the GMM estimator of θ is given by

$$\widehat{\theta} = \arg\min_{\theta\in\Theta} \ \bar{J}_n(\theta). \tag{9}$$

Strictly speaking, $\widehat{\gamma}$ is given by an interval but we let $\widehat{\gamma}$ be the minimum of the interval.

Since the model is linear in ϕ for each $\gamma \in \Gamma$, and the objective function $\bar{J}_n(\theta)$ is not continuous in γ with $\theta = (\phi', \gamma)'$, the grid search algorithm is practical: for a fixed γ, let

$$\bar{g}_{1n} = \frac{1}{n}\sum_{i=1}^{n} g_{1i}, \quad \text{and} \quad \bar{g}_{2n}(\gamma) = \frac{1}{n}\sum_{i=1}^{n} g_{2i}(\gamma),$$

where

$$\underset{l\times 1}{g_{1i}} = \begin{pmatrix} z_{it_0}\Delta y_{it_0} \\ \vdots \\ z_{iT}\Delta y_{iT} \end{pmatrix}, \quad \underset{l\times(k-1)}{g_{2i}(\gamma)} = \begin{pmatrix} z_{it_0}\left(\Delta x_{it_0}, \mathbf{1}_{it_0}(\gamma)' X_{it_0}\right) \\ \vdots \\ z_{iT}\left(\Delta x_{iT}, \mathbf{1}_{iT}(\gamma)' X_{iT}\right) \end{pmatrix}.$$

Then, the GMM estimator of β and δ, for a given γ, is given by

$$\left(\widehat{\beta}(\gamma)', \widehat{\delta}(\gamma)'\right)' = \left(\bar{g}_{2n}(\gamma)' W_n \ \bar{g}_{2n}(\gamma)\right)^{-1} \bar{g}_{2n}(\gamma)' W_n \ \bar{g}_{1n}.$$

Denoting the objective function evaluated at $\widehat{\beta}(\gamma)$ and $\widehat{\delta}(\gamma)$ by $\widehat{J}_n(\gamma)$, we obtain the GMM estimator of θ by

$$\widehat{\gamma} = \underset{\gamma\in\Gamma}{\operatorname{argmin}} \widehat{J}_n(\gamma), \quad \text{and} \quad \left(\widehat{\beta}', \widehat{\delta}'\right)' = \left(\widehat{\beta}(\widehat{\gamma})', \widehat{\delta}(\widehat{\gamma})'\right)'.$$

The two-step optimal GMM estimator is obtained as follows:

1. Estimate the model by minimizing $\bar{J}_n(\theta)$ with either $W_n = I_l$ or

$$W_n = \begin{pmatrix} \frac{2}{n}\sum_{i=1}^{n} z_{it_0}z'_{it_0} & \frac{-1}{n}\sum_{i=1}^{n} z_{it_0}z'_{it_0+1} & 0 & \cdots \\ \frac{-1}{n}\sum_{i=1}^{n} z_{it_0+1}z'_{it_0} & \frac{2}{n}\sum_{i=1}^{n} z_{it_0+1}z'_{it_0+1} & \ddots & \ddots \\ 0 & \ddots & \ddots & \frac{-1}{n}\sum_{i=1}^{n} z_{iT-1}z'_{iT} \\ \vdots & \ddots & \frac{-1}{n}\sum_{i=1}^{n} z_{iT}z'_{iT-1} & \frac{2}{n}\sum_{i=1}^{n} z_{iT}z'_{iT} \end{pmatrix}^{-1} \tag{10}$$

and collect residuals, $\widehat{\Delta\varepsilon}_{it}$.

2. Estimate the parameter θ by minimizing $\bar{J}_n(\theta)$ with

$$W_n = \left(\frac{1}{n} \sum_{i=1}^{n} \widehat{g}_i \widehat{g}_i' - \frac{1}{n^2} \sum_{i=1}^{n} \widehat{g}_i \sum_{i=1}^{n} \widehat{g}_i' \right)^{-1}, \tag{11}$$

where $\widehat{g}_i = \left(\widehat{\Delta\varepsilon}_{it_0} z_{it_0}', ..., \widehat{\Delta\varepsilon}_{iT} z_{iT}' \right)'$.

Remark 1 *In the linear dynamic panel data literature the number of initial conditions on y_{i0} have been proposed to improve the efficiency of the FD-GMM estimator, e.g. Ahn and Schmidt (1995), Arellano and Bover (1995) and Blundell and Bond (1998).[7] In this paper we consider the dynamic panels with the length of time period not too small relative to the number of individuals as in our empirical application of firm's investment decision. In this regard, we adopt a more robust specification in which the distribution of y_{i0} given α_i is left unrestricted.*

4 Asymptotic Theory

This section develops an asymptotic theory for the FD-GMM estimator. There are two frameworks in the literature. One is the fixed threshold assumption (Chan, 1993) and the other the diminishing threshold assumption (Hansen, 2000). We also discuss the estimation of unknown quantities in the asymptotic distributions such as the asymptotic variances and the normalizing factors when an estimator is not asymptotically normal.

Partition $\theta = (\theta_1', \gamma)'$, where $\theta_1 = (\beta', \delta')'$. As the true value of δ is δ_n, the true values of θ and θ_1 are denoted by θ_n and θ_{1n}, respectively. Define

$$\underset{l \times k_1}{G_\beta} = \begin{bmatrix} -\mathrm{E}\left(z_{it_0} \Delta x_{it_0}'\right) \\ \vdots \\ -\mathrm{E}\left(z_{iT} \Delta x_{iT}'\right) \end{bmatrix}, \quad \underset{l \times (k_1+1)}{G_\delta}(\gamma) = \begin{bmatrix} -\mathrm{E}\left(z_{it_0} \mathbf{1}_{it_0}(\gamma)' X_{it_0}\right) \\ \vdots \\ -\mathrm{E}\left(z_{iT} \mathbf{1}_{iT}(\gamma)' X_{iT}\right) \end{bmatrix},$$

and

$$\underset{l \times 1}{G_\gamma}(\gamma) = \begin{bmatrix} \left\{ \mathrm{E}_{t_0-1}\left[z_{it_0} \left(1, x_{it_0-1}'\right) | \gamma \right] p_{t_0-1}(\gamma) - \mathrm{E}_{t_0}\left[z_{it_0} \left(1, x_{it_0}'\right) | \gamma \right] p_{t_0}(\gamma) \right\} \delta_0 \\ \vdots \\ \left\{ \mathrm{E}_{T-1}\left[z_{iT} \left(1, x_{iT-1}'\right) | \gamma \right] p_{T-1}(\gamma) - \mathrm{E}_{T}\left[z_{iT} \left(1, x_{iT}'\right) | \gamma \right] p_{T}(\gamma) \right\} \delta_0 \end{bmatrix},$$

where $\mathrm{E}_t[\cdot|\gamma]$ denotes the conditional expectation given $q_{it} = \gamma$ and $p_t(\cdot)$ the density of q_{it}.

[7] Bun and Windmeijer (2010) show for the covariance stationary AR(1) panel model that the system GMM estimator has a smaller bias and root mean square error than the FD-GMM when the series are persistent, but that this bias increases with increasing $\sigma_\alpha^2/\sigma_v^2$ and can become substantial.

Assumption 1 *The true value of* β *is fixed at* β_0 *while that of* δ *depends on* n *such that* $\delta_n = \delta_0 n^{-\alpha}$ *for some* $0 \leq \alpha < 1/2$ *and* $\delta_0 \neq 0$. θ_n *are interior points of* Θ. Ω *is finite and positive definite.*

This assumption allows for both the standard setup, $\delta_n = \delta_0 \neq 0$ for all n, and the diminishing setup, $\delta_n \to 0$ as $n \to \infty$. The latter has been widely used in the threshold model (without an endogenous regressor) to obtain a tractable asymptotic distribution for the least squares estimator of γ, see Hansen (2000). As shown below, however, the GMM estimate $\widehat{\gamma}$ is asymptotically normal whether or not $\delta_n \to 0$, implying that the inferential procedure is the same for any $0 \leq \alpha < 1/2$. Therefore, we do not need to consider the diminishing setup, though we keep it for an internal consistency of the expositions.

Assumption 2 (*i*) *The threshold variable,* q_{it} *has a continuous and bounded density,* $p_t(\cdot)$, *such that* $p_t(\gamma_0) > 0$ *for all* $t = 1, ..., T$; (*ii*) $\mathrm{E}_t\left(z_{it}\left(x_{it}', x_{i,t-1}'\right)|\gamma\right)$ *is continuous at* γ_0, *where* $\mathrm{E}_t(\cdot|\gamma) = \mathrm{E}(\cdot|q_{it} = \gamma)$.

The smoothness assumption on the distribution of the threshold variable and conditional moments are standard. Notice however that the distribution of GMM estimator of unknown threshold is invariant to the continuity of the regression function at the change point because our model does not require its discontinuity at the change point. This is a novel feature of the GMM. As a consequence, we do not need a prior knowledge on the continuity of the model to make inference for the threshold model.[8]

Assumption 3 *Let* $G = (G_\beta, G_\delta(\gamma_0), G_\gamma(\gamma_0))$, *then* G *is of full column rank.*

This is a standard rank condition in GMM for identification. Typically, the lagged variables are employed as instruments. For instance, if the model is the SETAR, the lagged dependent variables, y_{it-d}'s for $d > 1$ are valid instruments and thus the dimension of the moment conditions grows quickly as T increases to satisfy the required number of moments for identification.[9]

[8] The GMM criterion function can be viewed as an extreme form of smoothing in the sense of Seo and Linton (2007). The smoothed least squares implies moment conditions that include one of the type $\mathrm{E}\left(e_t(\theta)\, p\left(\frac{q_t - \gamma}{h_n}\right)\right) = 0$, where $e_t(\theta)$ is the error for a given θ, $p(\cdot)$ is a density function, and h_n is a smoothing parameter that goes to zero. The diminishing rate of h_n determines the degree of smoothing and the convergence rate of the threshold estimate, which is $\left(nh_n^{-1}\right)^{-1/2}$. The slower the rate is, the more smoothing it imples. The GMM criterion corresponds to the case where h_n is fixed, yielding the convergence rate of $n^{-1/2}$.

[9] Our moment conditions in (7) utilize the moments related to $\Delta\varepsilon_{it}$ only, but not those related to the level ε_{it} as in Blundell and Bond (1998).

Theorem 1 *Under Assumptions 1-3, as $n \to \infty$,*

$$\begin{pmatrix} \sqrt{n}\begin{pmatrix} \widehat{\beta}-\beta_0 \\ \widehat{\delta}-\delta_n \end{pmatrix} \\ n^{1/2-\alpha}\left(\widehat{\gamma}-\gamma_0\right) \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0, \left(G'\Omega^{-1}G\right)^{-1}\right).$$

Remark 2 *Theorem 1 establishes that the FD-GMM always follows the normal distribution asymptotically, irrespective of whether $\alpha = 0$ or not. It can be argued that such a normality result can be simply derived through applying the standard GMM asymptotics. However, for our models with non-smooth criterion functions, we still need to verify certain stochastic differentiability conditions, which is nontrivial and shown to be achieved by applying the empirical process theory, e.g. van der Vaart and Wellner (1996). We can also allow for $\alpha = 0$ unlike in the least squares of threshold regression (e.g. Hansen, 2000). Furthermore, our result does not require us to know a priori whether the regression function is continuous or not, the validity of which is confirmed by the Monte-Carlo studies below.*

The asymptotic variance matrix contains δ_0, and the convergence rate of $\widehat{\gamma}$ hinges on the unknown quantity, α. These two quantities cannot be consistently estimated in separation, but they cancel out in the construction of t-statistic. Thus, confidence intervals for θ can be constructed in the standard manner. Let

$$\widehat{\Omega} = \frac{1}{n}\sum_{i=1}^{n}\widehat{g}_i\widehat{g}_i' - \left(\frac{1}{n}\sum_{i=1}^{n}\widehat{g}_i\right)\left(\frac{1}{n}\sum_{i=1}^{n}\widehat{g}_i'\right),$$

be the sample variance of $\widehat{g}_i$, where $\widehat{g}_i = g_i\left(\widehat{\theta}\right)$, and

$$\widehat{G}_\beta = \begin{bmatrix} -\frac{1}{n}\sum_{i=1}^{n} z_{it_0}\Delta x_{it_0}' \\ \vdots \\ -\frac{1}{n}\sum_{i=1}^{n} z_{iT}\Delta x_{iT}' \end{bmatrix}, \quad \widehat{G}_\delta = \begin{bmatrix} -\frac{1}{n}\sum_{i=1}^{n} z_{it_0}\mathbf{1}_{it_0}\left(\widehat{\gamma}\right)' X_{it_0} \\ \vdots \\ -\frac{1}{n}\sum_{i=1}^{n} z_{iT}\mathbf{1}_{iT}\left(\widehat{\gamma}\right)' X_{iT} \end{bmatrix}.$$

G_γ may be estimated by the standard Nadaraya-Watson kernel estimator: for some kernel K and bandwidth h (*e.g.* the Gaussian kernel and Silverman's rule of thumb)[10], let

$$\widehat{G}_\gamma = \begin{bmatrix} \frac{1}{nh}\sum_{i=1}^{n} z_{it_0}\left[\left(1, x_{it_0-1}'\right)' K\left(\frac{\widehat{\gamma}-q_{it_0-1}}{h}\right) - \left(1, x_{it_0}'\right)' K\left(\frac{\widehat{\gamma}-q_{it_0}}{h}\right)\right]\widehat{\delta} \\ \vdots \\ \frac{1}{nh}\sum_{i=1}^{n} z_{iT}\left[\left(1, x_{iT-1}'\right)' K\left(\frac{\widehat{\gamma}-q_{iT-1}}{h}\right) - \left(1, x_{iT}'\right)' K\left(\frac{\widehat{\gamma}-q_{iT}}{h}\right)\right]\widehat{\delta} \end{bmatrix}. \qquad (12)$$

[10] For simplicity, we apply the same bandwidth to all the terms in $\widehat{G}_\gamma$, which is fine with the Silverman's rule of thumb under stationarity of q_{it}. In principle, a different bandwidth can be derived for each q_{it}.

Remark 3 *As $n \to \infty$, the consistency of $\widehat{\Omega}$ and $\widehat{G}$ follows from the standard uniform law of large numbers (ULLN) for iid data across i, and the consistency of the Nadaraya-Watson and the kernel density estimators. The existence of the absolute moment is sufficient to get ULLN. The convergence rate for $\widehat{G}$ follows the standard nonparameteric rate for the Nadaraya-Waston and the kernel density estimators. See Härdle and Linton (1994) for more details on the choice of kernel and bandwidth.*

Let $\widehat{V}_s = \widehat{\Omega}^{-1/2}\left(\widehat{G}_\beta, \widehat{G}_\delta\right)$ and $\widehat{V}_\gamma = \widehat{\Omega}^{-1/2}\widehat{G}_\gamma$. Then, the asymptotic variance matrix for the regression coefficient, $\theta_1 = \left(\beta', \delta'\right)'$ can be consistently estimated by $\left(\widehat{V}_s'\widehat{V}_s - \widehat{V}_s'\widehat{V}_\gamma\left(\widehat{V}_\gamma'\widehat{V}_\gamma\right)^{-1}\widehat{V}_\gamma'\widehat{V}_s\right)^{-1}$, while the t-statistic for $\gamma = \gamma_0$ defined by

$$t = \frac{\sqrt{n}\left(\widehat{\gamma} - \gamma_0\right)}{\widehat{V}_\gamma'\widehat{V}_\gamma - \widehat{V}_\gamma'\widehat{V}_s\left(\widehat{V}_s'\widehat{V}_s\right)^{-1}\widehat{V}_s'\widehat{V}_\gamma}, \tag{13}$$

converges to the standard normal distribution. Hence, the confidence intervals can be constructed in the standard manner. Alternatively, the nonparametric bootstrap can be employed to construct the confidence intervals, see Section 5.1 for details.

5 Testing

5.1 Testing for Linearity

The asymptotic results provide ways to make inference for unknown parameters and their functions. However, it is well-established that the test for linearity or threshold effects requires us to develop the different asymptotic theory due to the presence of unidentified parameters under the null (*e.g.* Davies, 1977). Specifically, recall the model specification (3) and consider the null hypothesis:

$$\mathcal{H}_0 : \delta = 0, \quad \text{for any } \gamma \in \Gamma, \tag{14}$$

against the alternative hypothesis

$$\mathcal{H}_1 : \delta \neq 0, \quad \text{for some } \gamma \in \Gamma.$$

Then, a natural test statistic for the null hypothesis, $\mathcal{H}_0$ is:

$$\text{supW} = \sup_{\gamma \in \Gamma} W_n\left(\gamma\right), \tag{15}$$

where $W_n\left(\gamma\right)$ is the standard Wald statistic for each fixed γ, that is,

$$W_n\left(\gamma\right) = n\widehat{\delta}\left(\gamma\right)'\widehat{\Sigma}_\delta\left(\gamma\right)^{-1}\widehat{\delta}\left(\gamma\right),$$

where $\widehat{\delta}(\gamma)$ is the FD-GMM estimate of δ, given γ, and $\widehat{\Sigma}_{\delta}(\gamma)$ is the consistent asymptotic variance estimator for $\widehat{\delta}(\gamma)$, given by $\widehat{\Sigma}_{\delta}(\gamma) = R\left(\widehat{V}_s(\gamma)\widehat{V}_s(\gamma)\right)^{-1} R'$, where $\widehat{V}_s(\gamma)$ is computed as in Section 4 with $\widehat{\gamma} = \gamma$ and $R = \left(\mathbf{0}_{(k_1+1)\times k_1}, I_{k_1+1}\right)$. The supremum statistic is an application of the union-intersection principle commonly used in the literature, *e.g.* Hansen (1996) and Lee *et al.* (2011).

We present the limiting distribution of the supW statistic below.

Theorem 2 *Let $G(\gamma) = (G_\beta, G_\delta(\gamma))$ and $D(\gamma) = G(\gamma)'\Omega^{-1}G(\gamma)$. Suppose that $\inf_{\gamma\in\Gamma}\det(D(\gamma)) > 0$ and Assumption 2 (i) holds. Then, under the null (14), we have:*

$$\text{supW} \xrightarrow{d} \sup_{\gamma\in\Gamma} Z'G(\gamma)'D(\gamma)^{-1}R'\left[RD(\gamma)^{-1}R'\right]^{-1}RD(\gamma)^{-1}G(\gamma)Z,$$

where $Z \sim \mathcal{N}\left(0, \Omega^{-1}\right)$.

Although the limiting distribution of supW is derived as the supremum of the square of a Gaussian process with a simpler covariance kernel, it is not straightforward to pivotalize the statistic and tabulate the critical values. Hence, we follow Hansen (1996) and bootstrap or simulate the asymptotic critical values or p-values as follows:

Let $\widehat{\theta}$ be the FD-GMM estimator and construct:

$$\widehat{\Delta\varepsilon_{it}} = \Delta y_{it} - \Delta x_{it}'\widehat{\beta} - \widehat{\delta}' X_{it}'\mathbf{1}_{it}(\widehat{\gamma}),$$

for $i = 1, ..., n$, and $t = t_0, ..., T$. Then,

1. Let i^* be a random draw from $\{1, ..., n\}$, and $X_{it}^* = X_{i^*t}$, $q_{it}^* = q_{i^*t}$, $z_{it}^* = z_{i^*t}$ and $\Delta\varepsilon_{it}^* = \widehat{\Delta\varepsilon_{i^*t}}$. Then, generate

$$\Delta y_{it}^* = \Delta x_{it}^{*\prime}\widehat{\beta} + \Delta\varepsilon_{it}^* \quad \text{for} \quad t = t_0, ..., T.$$

2. Repeat step 1 n times, and collect $\{(\Delta y_{it}^*, X_{it}^*, q_{it}^*, z_{it}^*) : i = 1, ..., n; t = t_0, ..., T\}$.

3. Construct the supW statistic, say supW*, from the bootstrap sample using the same estimation method for $\widehat{\theta}$.

4. Repeat steps 1-3 B times, and evaluate the bootstrap p-value by the frequency of supW* that exceeds the sample statistic, supW.

Note that when simulating the bootstrap samples, the null model is imposed in step 1.

5.2 Testing for Exogeneity

In this section we describe how to test for the exogeneity of the threshold variable. Recently, Kapetanios (2010) develops the exogeneity test of the regressors in threshold models, following the general principle of the Hausman (1978) test. Similarly, we can develop the Hausman type testing procedure for the validity of the null hypothesis that the threshold variable, q_{it} is exogenous. Indeed, this is a straightforward by-product obtained by combining FD-GMM and FD-2SLS estimators and their asymptotic results.

Specifically, we propose the following t-statistic for the null hypothesis that the FD-GMM estimate $\widehat{\gamma}$ of the unknown threshold is equal to the FD-2SLS estimate, $\widehat{\gamma}_{FD-2SLS}$ (see Appendix A for details of the FD-2SLS estimator):

$$t_H = \frac{\sqrt{n}\left(\widehat{\gamma} - \widehat{\gamma}_{FD-2SLS}\right)}{\widehat{V}_\gamma'\widehat{V}_\gamma - \widehat{V}_\gamma'\widehat{V}_s\left(\widehat{V}_s'\widehat{V}_s\right)^{-1}\widehat{V}_s'\widehat{V}_\gamma},$$

where the denominator is derived as in Section 4. Note that this t-statistic is identical to the t-statistic in (13) except that γ_0 is replaced by $\widehat{\gamma}_{FD-2SLS}$. However,

$$\widehat{\gamma}_{FD-2SLS} = \gamma_0 + o_p\left(n^{-1/2}\left(\widehat{V}_\gamma'\widehat{V}_\gamma - \widehat{V}_\gamma'\widehat{V}_s\left(\widehat{V}_s'\widehat{V}_s\right)^{-1}\widehat{V}_s'\widehat{V}_\gamma\right)\right)$$

due to its super-consistency. Then, it is easily seen that the asymptotic distribution of the t-statistic is the standard normal under the null hypothesis of strict exogeneity of q_{it}.

6 Monte Carlo Experiments

This section explores finite sample performance of the FD-GMM estimator. The finite sample property of the least squares estimators and the testing for the presence of threshold effect have been examined extensively in the literature (*e.g.* Hansen, 2000; Caner and Hansen, 2004), albeit in the single equation regression. Up to our knowledge, however, no existing studies have examined how the GMM estimator performs in this general context.

6.1 Bias and MSE

We consider the following two models:

$$y_{it} = (0.7 - 0.5y_{it-1})\,1\{y_{it-1} \le 0\} + (-1.8 + 0.7y_{it-1})\,1\{y_{it-1} > 0\} + \sigma_1 u_{it}, \qquad (16)$$

$$y_{it} = (0.52 + 0.6y_{it-1})\,1\{y_{it-1} \le 0.8\} + (1.48 - 0.6y_{it-1})\,1\{y_{it-1} > 0.8\} + \sigma_2 u_{it}, \qquad (17)$$

for $t = 1, ..., 10$, and $i = 1, ..., n$, where u_{it} are $iidN(0,1)$. The first model from Tong (1990) allows a jump in the regression function at the threshold point. The second is the continuous model considered by Chan and Tsay (1998). In both models the threshold is located around the center of the distribution of the threshold variable. In terms of the previous notations in (3), the unknown true parameter values are $\beta = -0.5$ and $(\delta_1, \delta_2) = (-2.5, 1.2)'$ in the first model and $\beta = 0.6$ and $(\delta_1, \delta_2) = (0.96, -1.2)'$ in the second. All the past levels of y_{it} are used as the instrumental variables.[11]

In addition we consider an averaging of a class of FD-GMM estimators, which is expected to be particularly relevant in finite sample. There are many different ways to compute the weight matrix, W_n in the first step, though there is no way to tell which is optimal. Provided that the first step estimators are consistent, all the second step estimators are asymptotically equivalent, suggesting that the averaging does not change the first order asymptotic distribution.[12] In this regard, we propose to randomize the weight matrix, W_n in the first step as follows: We compute W_n in (11) with $\widehat{g}_i = \left(\Delta\widetilde{\varepsilon}_{it_0} z'_{it_0}, ..., \Delta\widetilde{\varepsilon}_{iT} z'_{iT}\right)'$, where $\widetilde{\varepsilon}_{it}$s are randomly generated from $N(0,1)$. In our experiments we do this 100 times and take the average of the second step estimators. Our proposal follows the similar idea by Chamberlain and Imbens (2004) and Sun (2014), who demonstrate that randomizing initial draws are able to improve coverage rates leading to more accurate inference. Consistent with these expectations, the subsequent simulation results demonstrate that the variance of the averaging estimator is greatly reduced in small samples.

We examine the bias, standard error and mean square error (MSE) of the FD-GMM estimator with 1,000 iterations. For $n = 50$, 100 and 200, we set $\sigma_1 = 1$ and $\sigma_2 = 0.5$. The simulation results are reported in Tables 1 - 3. First, looking at the MSEs in Table 1, those of the FD-GMM for each parameter generally decreases as the sample size rises, but some parameters, particularly δ_1 and δ_2, are estimated with much larger MSEs. The continuous design yields higher MSEs for the regression coefficients, because it has the smaller change than the discontinuous design. When we compare MSEs of the FD-GMM with those of the averaging estimator, we find that the averaging significantly reduces MSEs. In some cases the gains are so large that MSEs of the FD-GMM estimator are as twice as those of the averaging estimator. As a rule of thumb, the reduction in MSEs by averaging becomes larger when the original MSEs are relatively large, though this gain becomes smaller as the sample size increases. Turning

[11] We have one IV for $t = 3$, two IVs for $t = 4$, and thus a total of 36 IVs for $T = 10$.

[12] Alternatively, we may consider the continuous updating GMM estimator (CUE) proposed by Hansen *et al.* (1996), which is supposed to be invariant to the initial weighting matrix. However, its evaluation goes beyond the scope of the current paper mainly due to the computational complexity and time. Furthermore, Hasuman *et al.* (2011) show that the CUE does not always perform well due to its no-moment problem that leads to wide dispersion of the estimates.

to biases and standard errors as reported in Tables 2 and 3, we observe that the averaging always reduces stand errors, but it has a mixed effect on biases. In particular, when the bias of the FD-GMM is large (those of δ_1 and δ_2), then the averaging reduces it and *vice versa*. As a result, the average bias of the FD-GMM is almost the same as that of the averaging whilst the standard deviation of the former is always larger than that of the latter. This implies that the averaging has positive MSE reduction effects on the FD-GMM estimator.

Tables 1-3 about here

We have also performed the same experiment by fixing the intercepts across the regimes:

$$y_{it} = 0.7 - 0.5y_{it-1}1\{y_{it-1} \leq 1.5\} + 0.7y_{it-1}1\{y_{it-1} > 1.5\} + \sigma_1 u_{it},$$

$$y_{it} = 0.52 + 0.6y_{it-1}1\{y_{it-1} \leq 0.4\} - 0.6y_{it-1}1\{y_{it-1} > 0.4\} + \sigma_2 u_{it},$$

where the threshold values are reset to stay in the middle of distribution. From Tables 4 - 6, we find that the averaging reduces MSEs and standard errors more substantially. Furthermore, biases are greatly reduced by the averaging for more than 70% of the cases.

Tables 4-6 about here

6.2 Coverage Probability

This section explores the coverage probability of the confidence intervals by inverting the t-statistic. We focus on the first two data generating processes (16) and (17). Table 7 reports empirical coverage probabilities of the 95% confidence intervals for the FD-GMM estimator and its averaging. In the averaging, both the estimator and the asymptotic variance estimator are averaged. We select the bandwidth for the asymptotic variance by the Silverman's rule of thumb multiplied by h, and report the results for $h = (0.5, 1, 1.5)$. Not surprisingly, as h rises, the coverage frequency inflates. The bandwidth with $h = 0.5$ yields too low coverage for the continuous design and that with $h = 1.5$ yields excessive coverage especially for the threshold parameter, γ. Thus, we follow the Silverman's rule.

Table 7 about here

The results for $h = 1$ appear to be more promising than the existing studies that document rather poor empirical coverage probabilities for γ and δ_2, *e.g.* Hansen (2000) and Caner and Hansen (2004). Importantly, the averaging results in much improved coverage, especially when n is small, in which case the FD-GMM tends to exhibit very poor coverage. Thus, subsequent discussions are focussed on the averaging results. For γ, the coverage improves steadily to the nominal 95% level as the sample size rises for both jump and continuous designs, from 99% at $n = 50$ to 98% or 95% at $n = 200$. For δ_2, we observe somewhat lower coverages in the continuous design, which improve as the sample size increases and look reasonable for $n = 200$. Finally, the results for β and δ_1 are better than those for δ_2.[13]

7 Empirical Application: A dynamic threshold panel data model of investment

An important research question in the investment literature is whether capital market imperfection affects the firm's investment behavior. Farazzi *et al.* (1988) find that investment spending by firms with low dividend payments is strongly affected by the availability of cash flows, rather than just by the availability of positive net present value projects.

One of the main methodological problems facing the conventional investment literature is that the distinction between constrained and unconstrained firms is routinely based on an arbitrary threshold of the measure used to split the sample. Furthermore, firms are not allowed to change groups over time since the split-sample is fixed for the complete sample period. We apply a threshold model of investment in dynamic panels to address this important issue. Most popular investment model takes the form of a Tobin's Q model in which the expectation of future profitability is captured by the forward-looking stock market valuation:

$$I_{it} = \beta_1 CF_{it} + \beta_2 Q_{it} + \varepsilon_{it}, \tag{18}$$

where I_{it} is investment, CF_{it} cash flows, Q_{it} Tobin's Q, and ε_{it} consists of the one-way error components, $\varepsilon_{it} = \alpha_i + v_{it}$.[14] The coefficient, β_1 represents the cash flow sensitivity of investment. If firms are not financially constrained, external finance can be raised to fund future investments without the use of internal finance. In this case, cash flows are least relevant to investment spending and β_1 is expected to be close to zero. In contrast, if firms were to face certain financial constraints, β_1 would be expected to be significantly positive. Extensions of this Tobin's Q model involve additional financing variables such as leverage to control for the effect of capital structure on investment (Lang *et al.*, 1996) as well as lagged investment to capture the accelerator effect of investment in which past investments have a positive effect

[13] More excessive coverage probabilities for γ are reported in Hansen (2000) and Caner and Hansen (2004), showing more than 98% coverage even for 90% nominal level. They also reported the lower coverages for δ_2.

[14] We have also estimated the model with the two-way error components by including the time dummies. The results, available upon request, are qualitatively similar.

on future investments (Aivazian *et al.*, 2005). Therefore, we consider the following augmented dynamic investment model:

$$I_{it} = \phi I_{it-1} + \theta_1 CF_{it} + \theta_2 Q_{it} + \theta_3 L_{it} + \varepsilon_{it}, \tag{19}$$

where L_{it} represents leverage.

We employ the same data set as used in Hansen (1999) and González *et al.* (2005). This dataset is a balanced panel of 565 US firms over the period 1973-1987. Hence, this study allows for comparisons with the existing literature. Following González *et al.* (2005), we exclude five companies with extreme data, and consider a final sample of 560 companies with 7840 company-year observations. To avoid the use of potentially persistent series, we normalize variables by the book value of assets. Namely, I_{it} is measured by investment to the book value of assets, CF_{it} by cash flow to the book value of assets, Q_{it} by the market value to the book value of assets, and L_{it} by the long-term debt to the book value of assets.

We then extend (19) into the dynamic panel data framework with threshold effects:

$$\begin{aligned} I_{it} = {} & (\phi_1 I_{it-1} + \theta_{11} CF_{it} + \theta_{21} Q_{it} + \theta_{31} L_{it}) \, 1\{q_{it} \leq \gamma\} \\ & (\phi_2 I_{it-1} + \theta_{12} CF_{it} + \theta_{22} Q_{it} + \theta_{32} L_{it}) \, 1\{q_{it} > \gamma\} + \alpha_i + v_{it}, \end{aligned} \tag{20}$$

where $1\{\cdot\}$ is an indicator function, q_{it} the transition variable and γ the threshold parameter. We estimate (20) by the proposed FD-GMM, which allows for both (contemporaneous) regressors and the transition variable to be endogenous. On the other hand, existing studies (*e.g.* Hansen, 1999; González *et al.*, 2005) employs the lagged values of CF, Q and L to avoid the potential problem of endogenous regressors and transition variable, which is a common practice in empirical corporate finance, *e.g.* Dang *et al.* (2012).

Table 8 summarizes the estimation results for the dynamic threshold model of investment, (20), with cash flow, leverage and Tobin's Q used as the transition variable, which are expected to proxy the certain degree of financial constraints. This choice of the transition variable is broader than Hansen (1999) who considers only leverage, and González *et al.* (2005) who employ leverage and Tobin's Q. The FD-GMM estimation results are reported respectively in the low and the high regimes.

When cash flow is used as the transition variable, the results for (20) show that the threshold estimate is 0.36 such that about 80% of observations fall into the lower cash-flow regime. The coefficient on lagged investment is significantly higher for firms with low cash flows, suggesting that the accelerator effect of investment is stronger for cash-constrained firms. The coefficient on Tobin's Q reveals an expected finding that firms respond to growth opportunities more quickly when they are cash-unconstrained than when they are constrained. Next, we find the more negative impacts of the leverage when firms are cash-constrained. This is consistent with our expectations that the leverage should have a stronger negative impact on investment for the constrained firms, which is in line with the overinvestment hypothesis about the role of leverage as a disciplining device that prevents firms from over-investing in negative net present value projects (*e.g.* Jensen, 1986). Finally and importantly, the sensitivity of investment to cash flow is significantly higher for cash-constrained firms than for cash-rich firms. Firms with limited cash resources are likely to face some forms of financial constraints (Kaplan and Zingales, 1997). Hence, this finding supports evidence for the role of financial constraints in the investment-cash flow sensitivity.

When the leverage is used as the transition variable, the threshold parameter is estimated at 0.10, lower than the mean leverage (0.24), with more than 73% of observations falling into the high-leverage regime. We find that past investment has a much higher positive impact on current investment for highly-levered firms, suggesting that firms with high leverage attempt to respond to growth options quickly, hence a higher accelerator effect. The effect of Tobin's Q on investment is higher for lowly-levered firms, which provides a support for the argument that by lowering the risky "debt overhang" to control underinvestment incentives *ex ante*, firms are able to take more growth opportunities and make more investments *ex post*, though these impacts are rather small. We also find the more negative impacts of the leverage when firms are highly levered. The coefficient on cash flow is significantly higher for firms in the high-leverage regime, a finding consistent with the prediction that cash flow should be more relevant and have a stronger effect on the level of investment for financially constrained firms.[15]

When using Tobin's Q as the transition variable, the threshold is estimated at 0.56 with 59% of observations falling into the higher growth regime. We find that past investment has a slightly stronger positive effect on current investment for firms with low Tobin's Q, but the differential impacts are statistically insignificant. The coefficient on Tobin's Q in the low regime is significantly higher, indicating that firms with low growth options respond more strongly to changes in their investment opportunities. Surprisingly, we find a negative relationship between leverage and investment only in the lower growth regime. The sensitivity of investment to cash flow is also relatively higher for high-growth firms than low-growth firms. This, therefore, supports the hypothesis that cash flow should be more relevant for firms with potentially high financial constraints.[16]

[15] Notice, however, that the non-dynamic threshold model of investment developed by Hansen (1999) fails to find conclusive evidence in favor of this prediction.

[16] When comparing our results with those reported in González *et al.* (2005), who apply the static panel smooth transition regression model, we find that their results are qualitatively similar to ours regarding the impacts on investment of both Tobin's Q and leverage. However, they document an opposite evidence that the

In order to check the validity of the final specifications employed above, we also report the test results for the null of no threshold effects and the validity of the overidentifying moment conditions in Table 8. First, we find that the bootstrap p-values of the supW test are all close to zero, providing strong evidence in favour of threshold effects. Next, the J-test results indicate that the null of valid instruments is not rejected for the cases with the leverage and the Tobin's Q used as the transition variable, though it is rejected at the 1% significance level for the case with the cash flow used as the transition variable. Given that the number of instruments rises quadratically with T, this evidence is relatively satisfactory.[17]

Table 8 about here

In sum, when examining a dynamic threshold panel data estimation of Tobin's Q model of investment by using the Tobin's Q, leverage and cash flow as a possible transition variable, we find that the results on the relationships between investment and past investment, as well as cash flow, Tobin's Q and leverage are generally consistent with theoretical predictions. More importantly, the cash flow sensitivity of investment is significantly stronger for cash-constrained, high-growth and high-leveraged firms, a consistent finding with an original hypothesis by Farazzi *et al.* (1988) that the sensitivity of investment to cash flows is an indicator of the degree of financial constraints facing the firms. Methodologically, our results clearly demonstrate the usefulness of the proposed dynamic panel data estimation with threshold effects despite the fact that the transition variables used in the current study may have caveats since these variables are imperfect measures of financial constraints.[18]

8 Conclusion

The investigation of nonlinear asymmetric dynamic modelling has recently assumed a prominent role. Increasing availability of the large and complex panel data sets has also prompted more rigorous econometric analyses of dynamic heterogeneous panels, especially when the time

coefficient on the (lagged) cash flow is positive but considerably smaller for the higher regime.

[17] To avoid the potential issue related to weak instrument or overfitting, we set the maximum lag order of y and x to be used as instruments to 4 (*e.g.* Roodman, 2009).

[18] Kaplan and Zingales (1997) find that the relationship between cash flows and investment is not monotonic with financial constraints. Consequently, a large body of the literature seeks to address the question of what measures can be used to classify firms as 'financially constrained' and 'unconstrained'. Several criteria have been suggested, including size, age, leverage, financial slack, dividend payout and bond rating (*e.g.* Hovikimian and Titman, 2006). An alternative approach would be to use indices computed to control for financial constraints, *e.g.* Whited and Wu (2006).

period is short. In this paper we have explicitly addressed this challenging issue by developing the dynamic threshold panel data model, which allows both regressors and threshold effect to be endogenous. We have proposed the FD-GMM estimation on the basis of FD transformation for removing unobserved individual effects, and derived its asymptotic properties through employing the diminishing threshold effect asymptotics and the empirical process theory. In the special case where the threshold variable is strictly exogenous, we have also proposed more efficient FD-2SLS estimation.

We note several avenues for further researches. First, it is uncertain if the FD-GMM is most efficient in the presence of an endogenous threshold variable, especially with respect to alternative initial conditions and potentially many weak instruments. Simultaneously, an extension to the large n, large T case would make an interesting future research topic. Next, given that estimation can be significantly affected by the presence of cross-sectionally correlated errors (*e.g.*, Pesaran, 2006; Bai, 2009), it would be desirable to explicitly control for the cross-section dependence in the dynamic threshold panels. Furthermore, researches to develop similar estimation algorithms for models with multivariate covariates, with multiple threshold variables and regimes, and with alternative nonlinear mechanisms will be under way.

A Appendix: the FD-2SLS Estimator

This Appendix considers the special case where the threshold variable, q_{it} in (3) , are exogenous and the conditional moment restriction (6) holds. That is, z_{it} includes q_{it} and $q_{i,t-1}$. In this case, the threshold estimate, $\widehat{\gamma}$ can achieve the efficient rate of convergence, as obtained in the classical regression model (*e.g.* Hansen, 2000), and the slope estimate, $\widehat{\phi}$ can achieve the semi-parametric efficiency bound (Chamberlain, 1987) under conditional homoskedasticity as if the true threshold value, γ_0, is known. This strong result can be obtained since the two sets of estimators are shown to be asymptotically independent.

A.1 Estimation

We consider two cases for the reduced form regression – the regression of endogenous regressors on the instrumental variables. The first type is a general non-linear regression where unknown parameters can be estimated by the standard $\sqrt{n}$ rate, and the second type is the threshold regression with a common threshold.

The second case was also considered by Caner and Hansen (2004), albeit in the cross-sectional regression. Their approach consists of three steps; the first two steps yield an estimate of the threshold value and the third step performs the standard GMM for the linear regression within each subsample divided by the estimated threshold. However, this split-sample GMM approach does not work with the panel data with a time varying threshold variable, q_{it}, because it generates multiple regimes with cross-regime restrictions. Importantly, we demonstrate below that the first step estimation error affects the asymptotic distribution of the threshold estimate in the second step. In this context, we will develop new consistent estimation algorithm for the threshold estimate.

A.1.1 Nonlinear Regression in Reduced Form

We consider general non-linear regressions for the reduced form and provide the asymptotic variance formula that corrects the estimation error stemming from the reduced form regression. This is practically relevant since the linear projection in the reduced form invalidates the consistency of $\widehat{\theta}$ when the structural form is the threshold regression, *e.g.* Yu (2013).

Under the conditional moment condition in (6) and the exogeneity of q, the first-differenced model in (3) implies the following regression of Δy_{it} on z_{it}:

$$\mathrm{E}\left(\Delta y_{it}|z_{it}\right) = \beta'\mathrm{E}\left(\Delta x_{it}|z_{it}\right) + \delta'\mathrm{E}\left(X_{it}'|z_{it}\right)\mathbf{1}_{it}\left(\gamma\right). \qquad (21)$$

Assume for each t that the reduced form regressions are given by

$$\mathrm{E}\left(\begin{array}{c} 1, x_{it}' \\ 1, x_{it-1}' \end{array} |z_{it}\right) = \left(\begin{array}{c} 1, F_{1t}'\left(z_{it}; b_{1t}\right) \\ 1, F_{2t}'\left(z_{it}; b_{2t}\right) \end{array}\right) = \underset{2\times(1+k_1)}{F_t}\left(z_{it}; b_t\right), \tag{22}$$

where $b_t = (b_{1t}', b_{2t}')'$ is an unknown parameter vector and F_t is a known function. Also let

$$H_t\left(z_{it}; b_t\right) = \mathrm{E}\left(\Delta x_{it}|z_{it}\right) = F_{1t}\left(z_{it}; b_t\right) - F_{2t}\left(z_{it}; b_t\right).$$

A few remarks are in order; (i) since all the elements of x_{it} or x_{it-1} are not endogenous, some elements of F_t would be fully known; (ii) we need to run two regressions for x_{it}, $\mathrm{E}\left(x_{it}|z_{it}\right)$ and $\mathrm{E}\left(x_{it}|z_{it+1}\right)$, as the instruments z_{it} are different for each t. This is due to the FD transformation and the fact that z_{it} varies over time; and (iii) it is not sufficient to consider the regression $\mathrm{E}\left(\Delta x_{it}|z_{it}\right)$ only, due to the last term in the structural form (21).

The representation in (21) and (22) motivates the following two-step estimation procedure:

1. For each t, estimate the reduced form, (22) by the least squares, and obtain the parameter estimates, $\widehat{b}_t,\ t = t_0, ..., T$, and the fitted values, $\widehat{F}_{it} = F_t\left(z_{it}; \widehat{b}_t\right)$ and $\widehat{H}_{it} = H_t\left(z_{it}; \widehat{b}_t\right)$.

2. Estimate θ by

$$\min_{\theta\in\Theta} \widehat{\mathbb{M}}_n\left(\theta\right) = \frac{1}{n}\sum_{i=1}^{n}\sum_{t=t_0}^{T} e_{it}\left(\theta, \widehat{b}_t\right)^2, \tag{23}$$

   where

$$e_{it}\left(\theta, b_t\right) = \Delta y_{it} - \beta' H_t\left(z_{it}; b_t\right) - \delta' F_t\left(z_{it}; b_t\right)' \mathbf{1}_{it}\left(\gamma\right).$$

   This step can be done by the grid search as the model is linear in β and δ for a fixed γ. Thus, $\widehat{\beta}\left(\gamma\right)$ and $\widehat{\delta}\left(\gamma\right)$ can be obtained from the pooled OLS of Δy_{it} on $\widehat{H}_{it}$ and $\widehat{F}_{it}' 1_{it}\left(\gamma\right)$, which are constructed in step 1. Finally, $\widehat{\gamma}$ is defined as the minimum of the minimizers of the profiled sum of squared errors, $\widehat{\mathbb{M}}_n\left(\gamma\right)$.

This produces a rate-optimal estimator for γ, implying that β and δ can be estimated as if γ_0 were known. In the special case with $T = t_0$, we end up estimating a linear regression model with a conditional moment restriction. This two-step estimation yields the optimal estimate for β and δ if the model is conditionally homoskedastistic, *i.e.*, $\mathrm{E}\left(\Delta\varepsilon_{it}^2|z_{it}\right) = \sigma^2$, see Chamberlain (1987). While it requires to estimate the conditional heteroskedasticity to fully exploit the implications of the conditional moment restriction, (6), in practice, it is reasonable to employ the two-step estimator and robustify the standard errors for heteroskedasticity. We will provide a heteroskedasticity-robust standard errors for $\widehat{\beta}$ and $\widehat{\delta}$. Note that these standard errors are also corrected for the estimation error stemming from the first step estimation of b.

A.1.2 Threshold Regression in Reduced Form

Consider the following (reduced) threshold regression:

$$\begin{aligned} x_{it} &= \Gamma_{1t} z_{it} 1\{q_{it} \le \gamma\} + \Gamma_{2t} z_{it} 1\{q_{it} > \gamma\} + \eta_{it}, \\ \mathrm{E}(\eta_{it} | z_{it}) &= 0, \end{aligned} \tag{24}$$

where $z_{it} = \left(1, x'_{it-1}\right)'$, and Γ_{1t} and Γ_{2t} are unknown parameters. This results in the following structural threshold regression:

$$\Delta y_{it} = \lambda'_{1t} z_{it} 1\{q_{it} \le \gamma\} + \lambda'_{2t} z_{it} 1\{q_{it} > \gamma\} - \lambda'_{3t} z_{it} 1\{q_{it-1} \le \gamma\} - \lambda'_{4t} z_{it} 1\{q_{it-1} > \gamma\} + e_{it}, \tag{25}$$

$\mathrm{E}(e_{it} | z_{it}) = 0,$

where $\lambda'_{1t} = \left(0, \beta' \Gamma_{1t}\right)$, $\lambda'_{2t} = \left(\delta_1, \phi'_{22} \Gamma_{2t}\right)$, $\lambda'_{3t} z_{it} = \beta' x_{it-1}$, $\lambda'_{4t} z_{it} = -\delta_1 + \phi'_{22} x_{it-1}$ and $e_{it} = \Delta \varepsilon_{it} + \eta'_{it} \left(\beta + 1\{q_{it} > \gamma\} \delta_2\right)$.[19] Since the estimates of λ and γ are asymptotically independent of each other, we do not need to impose any restrictions on λ to estimate γ.

Thus, we estimate the model as follows:

1. Estimate γ by the pooled least square of (25), which can be done by the grid search,[20] and denote the estimate by $\widetilde{\gamma}$.

2. For each t, fix γ at $\widetilde{\gamma}$ and estimate Γ_{jt}, $j = 1, 2$, in (24) by OLS.

3. Estimate β and δ in (21) by OLS with γ and the reduced form parameters fixed at the estimates obtained from the preceding steps. Denote these estimates by $\widetilde{\beta}$ and $\widetilde{\delta}$.

Remark 4 *Our approach is crucially different from that of Caner and Hansen (2004), who estimate the threshold parameter separately in the reduced and the structural form. Such an approach introduces dependency between separate threshold estimates, which violates the validity of their asymptotic results.*[21] *Intuitively, the estimation error in the first step will affect the second step estimation of γ since the true threshold is restricted to be the same in both reduced and structural forms. On the other hand, our FD-2SLS estimator is designed to remove asymptotic correlation between the threshold estimator and the first step estimator.*

[19] See footnote 5 for the definition of parameters.

[20] That is, fix γ and obtain $\widetilde{e}_{it}(\gamma)$ and $\widetilde{\lambda}_{jt}(\gamma)$, $j = 1, ..., 4$ by the OLS for each t. Then, $\widetilde{\gamma}$ is the minimizer of the profiled sum of squared errors, $\sum_{i,t} \widetilde{e}_{it}^2(\gamma)$ and $\widetilde{\lambda}_{jt} = \widetilde{\lambda}_{jt}(\widetilde{\gamma})$, $j = 1, ..., 4$.

[21] Lemma 1 in Caner and Hansen (2004) requires more restrictions. Specifically, their (A.7) is true only when the threshold estimate is n-consistent, which cannot be obtained under the maintained diminishing threshold parameter setup. Accordingly, the high-level assumption (17) in their Assumption 2 is no longer satisfied.

Remark 5 *We consider the common threshold case mainly because we highlight an important misspecification issue in Caner and Hansen (2004) that the first estimation of the threshold affect the second step estimation, which was not recognized properly in the literature. But, it would be more general to allow different thresholds in the structural and reduced-form equations. In principle, we may consider the multiple scenarios: the structural regression follows the threshold regression and the reduced form regression is symmetric and both structural and reduced regressions follow the threshold regression with the same threshold parameter and with different threshold parameters. Such an extension will be able to develop the framework of multiple thresholds with multiple threshold variables. Recently, in the single regression context, Chen et al. (2012) develops a threshold autoregressive model which contains two threshold variables. However, due to the more complicated specification issues associated with dynamic heterogeneous panel structure, we leave this important issue for future studies, see also Chong and Yan (2015) for the number of related technical issues.*

A.2 Asymptotic Distribution

This section presents the asymptotic theory for the FD-2SLS only under the diminishing threshold framework (Hansen, 2000). It is worthwhile to note that the transformed model, (3) consists of 4 regimes, which are generated by two threshold variables, q_{it} and q_{it-1}, while the threshold parameter is restricted to be the same. This change in the model characteristic from the original 2-regime threshold model complicates the estimation and statistical inference.

Since some elements of x_{it} may belong to z_{it}, in which case the reduced form is identity, and some elements of $E\left(x_{it}|z_{it}\right)$ may be identical to $E\left(x_{it}|z_{it+1}\right)$ for some t, we collect all distinct reduced form regression functions, F_t, $t = t_0, ..., T$, that are not identities, and denote it as $F\left(z_i, b\right)$, where z_i and b are the collections of all distinct elements of z_{it} and b_t, $t = t_0, ..., T$. We denote the collection of the corresponding elements of x_{it}'s by $\S_i$, and write the reduced form as the multivariate cross section regression as follows:

$$\S_i = F\left(z_i, b\right) + \eta_i \text{ with } \mathrm{E}\left(\eta_i|z_i\right) = 0. \tag{26}$$

Let $\widehat{b}$ denote the least squares estimate, and define $F_i\left(b\right) = F\left(z_i, b\right)$, $F_i = F\left(z_i, b_0\right)$ and $\widehat{F}_i = F\left(z_i, \widehat{b}\right)$, where b_0 indicates the true value of b.

A.2.1 Nonlinear Regression in Reduced Form

We first consider the case in which the reduced form is the regular nonlinear regression and the reduced form parameter estimate, $\widehat{b}$ is asymptotically normal.

Assumption 4 *The estimator $\widehat{b}$ is consistent. F is twice continuously differentiable in b in a neighborhood of b_0 almost surely and its first derivative matrix at b_0, a $k_b \times 2k_1(T-t_0+1)$ matrix-valued function, is denoted as $\mathbb{F}_i = \mathbb{F}(z_i)$. $\mathrm{E}|\mathbb{F}_i|^4$ and $\mathrm{E}|\eta_i|^4$ are finite, where $|A|$ denotes the Euclidean norm if A is a vector and the vector-induced norm if A is a matrix.*

Assumption 4 (which excludes the threshold regression) implies that

$$\sqrt{n}\left(\widehat{b}-b_0\right) = \left(\mathrm{E}\mathbb{F}_i\mathbb{F}_i'\right)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbb{F}_i\eta_i + o_p(1),$$

where η_i is given in (26). Here we illustrate how the estimation error in the first step affects the asymptotic distribution of the estimator of β, δ and γ in the second step. Recall the functions introduced in Section A.1.1 and let

$$\underset{(2k_1+1)\times 1}{\Xi_{it}(\gamma, b_t)} = \left[\begin{array}{c} H_{it}(b_t) \\ F_{it}(b_t)'\mathbf{1}_{it}(\gamma)\end{array}\right] \text{ for each } t; \quad \underset{(2k_1+1)\times(T-t_0+1)}{\Xi_i(\gamma, b)} = \left(\Xi_{it_0}(\gamma, b_{t_0}), ..., \Xi_{iT}(\gamma, b_T)\right). \tag{27}$$

Let e_i be the vector stacking $\left\{\Delta\varepsilon_{it} + \beta_0'\left(\Delta x_{it} - \mathrm{E}\left(\Delta x_{it}|z_{it}\right)\right)\right\}_{t=t_0}^{T}$. Then, define

$$\underset{(2k_1+1)\times(2k_1+1)}{M_1(\gamma)} = \mathrm{E}\left[\Xi_i(\gamma)\Xi_i(\gamma)'\right], \quad \text{and} \quad \underset{(2k_1+1)\times(2k_1+1)}{V_1(\gamma)} = A(\gamma)\Omega(\gamma,\gamma)A(\gamma)',$$

where

$$\begin{aligned}\underset{((2k_1+1)+k_b)\times((2k_1+1)+k_b)}{\Omega(\gamma_1,\gamma_2)} &= \mathrm{E}\left[\left(\begin{array}{c}\Xi_i(\gamma_1)e_i, \\ \mathbb{F}_i\eta_i\end{array}\right)\left(e_i'\Xi_i'(\gamma_2), \eta_i'\mathbb{F}_i'\right)\right], \\ \underset{(2k_1+1)\times((2k_1+1)+k_b)}{A(\gamma)} &= \left(I_{(2k_1+1)}, \ -\mathrm{E}\left[\frac{\partial}{\partial b'}\sum_{t=t_0}^{T}\left(H_{it}'\beta_0\right)\Xi_{it}(\gamma)\right]\left(\mathrm{E}\mathbb{F}_i\mathbb{F}_i'\right)^{-1}\right).\end{aligned}$$

For the asymptotic distribution of $\widehat{\gamma}$, we introduce:

$$\begin{aligned}M_2(\gamma) &= \sum_{t=t_0}^{T}\left[\mathrm{E}_t\left[\left(\left(1, F_{1,it}'\right)\delta_0\right)^2|\gamma\right]p_t(\gamma) + \mathrm{E}_{t-1}\left[\left(\left(1, F_{2,it}'\right)\delta_0\right)^2|\gamma\right]p_{t-1}(\gamma)\right], \\ V_2(\gamma) &= \sum_{t=t_0}^{T}\left(\mathrm{E}_t\left[\left(e_{it}\left(1, F_{1,it}'\right)\delta_0\right)^2|\gamma\right]p_t(\gamma) + \mathrm{E}_{t-1}\left[\left(e_{it}\left(1, F_{2,it}'\right)\delta_0\right)^2|\gamma\right]p_{t-1}(\gamma)\right) \\ &\quad + 2\sum_{t=t_0}^{T-1}\mathrm{E}_t\left[e_{it}e_{it+1}\left(1, F_{1,it}'\right)\delta_0\left(1, F_{2,it+1}'\right)\delta_0|\gamma\right]p_t(\gamma).\end{aligned}$$

As before, we write $V_j = V_j(\gamma_0)$ and $M_j = M_j(\gamma_0)$ for $j = 1, 2$.

We further assume:

Assumption 5 *The true value of* β *is fixed at* β_0 *while that of* δ *depends on* n *such that* $\delta_n = \delta_0 n^{-\alpha}$ *for some* $0 < \alpha < 1/2$ *and* $\delta_0 \neq 0$.

This small δ assumption is to get a tractable asymptotic distribution. If $\alpha = 0$, the asymptotic distribution for $\widehat{\gamma}$ is different from the one obtained here, though the convergence rate result in the proof of Theorem 3 remains valid even if $\alpha = 0$. On the other hand, if $\alpha \geq 1/2$, the change is too small to identify the unknown threshold, γ_0.

Assumption 6 (i) *The threshold variable,* q_{it} *has a continuous and bounded density,* p_t*, such that* $p_t(\gamma_0) > 0$ *for all* $t = 1, ..., T$*;* (ii) $\mathrm{E}_t(w_{it}|\gamma)$ *is continuous at* γ_0 *for all* t*, and non-zero for some* t*, where* w_{it} *is either* $\left(e_{it}\left(1, F_{1,it}'\right)\delta_0 + e_{it+1}\left(1, F_{2,it+1}'\right)\delta_0\right)^2$, $\left(\left(1, F_{1,it}'\right)\delta_0\right)^2$*, or* $\left(\left(1, F_{2,it}'\right)\delta_0\right)^2$*;* (iii) $\mathrm{E}\,\mathrm{vec}\left(\Xi_i(\gamma, b)\right)\mathrm{vec}\left(\Xi_i(\gamma, b)\right)'$ *is continuously differentiable in* b *for all* γ *in a neighborhood of* γ_0.

Assumption 7 *For some* $\epsilon > 0$ *and* $\zeta > 0$, $\mathrm{E}\left(\sup_{t\leq T, |b-b_0|<\epsilon} |e_{it}F_t(z_{it}, b_t)|^{2+\zeta}\right) < \infty$*. For all* $\epsilon > 0$, $\mathrm{E}\left(\sup_{t\leq T, |b-b_0|<\epsilon} |e_{it}(F_t(z_{it}, b_t) - F_t(z_{it}))|^{2+\zeta}\right) = O\left(\epsilon^{2+\zeta}\right)$.

Assumption 8 *The minimum eigenvalue of the matrix,* $\mathrm{E}\Xi_{it}(\gamma)\Xi_{it}'(\gamma)$ *is bounded below by a positive value for all* $\gamma \in \Gamma$ *and* $t = 1, ..., T$.

The asymptotic confidence intervals can be constructed by inverting a test statistic. In particular, Hansen (2000) advocates the LR inversion for the construction of confidence intervals for the threshold value, γ_0, for which we define the LR statistic as

$$LR_n(\gamma) = n\frac{\widehat{\mathbb{M}}_n(\gamma) - \widehat{\mathbb{M}}_n(\widehat{\gamma})}{\widehat{\mathbb{M}}_n(\widehat{\gamma})}.$$

We present the main asymptotic results for the 2SLS estimator and the LR statistic below.

Theorem 3 *Let Assumptions 4-8 hold. Then,*

$$\sqrt{n}\begin{pmatrix} \widehat{\beta} - \beta_0 \\ \widehat{\delta} - \delta_n \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0, M_1^{-1}V_1M_1^{-1}\right), \tag{28}$$

and

$$n^{1-2\alpha}\frac{M_2^2}{V_2}(\widehat{\gamma} - \gamma_0) \xrightarrow{d} \underset{r\in\mathbb{R}}{\operatorname{argmin}}\left(\frac{|r|}{2} - W(r)\right), \tag{29}$$

where $W(r)$ *denotes the standard two-sided Brownian motion independent of the normal variate in* (28)*. Furthermore, for* $\sigma_e^2 = \mathrm{E}\left(e_{it}^2\right)$,

$$\frac{M_2\sigma_e^2}{V_2}\;LR(\gamma_0) \xrightarrow{d} \inf_{r\in\mathbb{R}}\left(|r| - 2W(r)\right).$$

Theorem 3 yields the asymptotic independence between $\widehat{\gamma}$ and the other estimates. The first estimation error does not affect the asymptotic distribution of $\widehat{\gamma}$, though it affects the asymptotic variance of $\widehat{\beta}$ and $\widehat{\delta}$ through V_1. However, estimation of the asymptotic variances of $\widehat{\beta}$ and $\widehat{\delta}$ is standard, i.e. the same as in the linear regression due to the aforementioned asymptotic independence.

Recall that $W(r) = W_1(-r) 1\{r \leq 0\} + W_2(r) 1\{r \geq 0\}$, where W_1 and W_2 are two independent Wiener processes. The asymptotic distribution for $\widehat{\gamma}$ in (29) is symmetric around zero with distribution function

$$1 + \sqrt{x/2\pi} \exp(-x/8) + (3/2) \exp(x) \Phi\left(-3\sqrt{x}/2\right) - ((x+5)/2) \Phi\left(\sqrt{x}/2\right) \text{ for } x \geq 0,$$

where Φ is the standard normal distribution function, see Bhattacharya and Brockwell (1976). The unknown normalizing factor, $n^{2\alpha} V_2^{-1} M_2^2$ can be consistently estimated by $\widehat{V}_2^{-1} \widehat{M}_2^2$, where

$$\widehat{M}_2 = \sum_{t=t_0}^{T} \frac{1}{nh} \sum_{i=1}^{n} \left[\left(\left(1, \widehat{F}_{1,it}'\right) \widehat{\delta} \right)^2 K\left(\frac{q_{it} - \widehat{\gamma}}{h} \right) + \left(\left(1, \widehat{F}_{2,it}'\right) \widehat{\delta} \right)^2 K\left(\frac{q_{it-1} - \widehat{\gamma}}{h} \right) \right],$$

$$\widehat{V}_2 = \sum_{t=t_0}^{T} \frac{1}{nh} \sum_{i=1}^{n} \left(\left(\widehat{e}_{it} \left(1, \widehat{F}_{1,it}'\right) \widehat{\delta} \right)^2 K\left(\frac{q_{it} - \widehat{\gamma}}{h} \right) + \left(\widehat{e}_{it} \left(1, \widehat{F}_{2,it}'\right) \widehat{\delta} \right)^2 K\left(\frac{q_{it-1} - \widehat{\gamma}}{h} \right) \right)$$
$$+ 2 \sum_{t=t_0}^{T-1} \frac{1}{nh} \sum_{i=1}^{n} \widehat{e}_{it} \widehat{e}_{it+1} \left(1, \widehat{F}_{1,it}'\right) \widehat{\delta} \left(1, \widehat{F}_{2,it+1}'\right) \widehat{\delta} K\left(\frac{q_{it} - \widehat{\gamma}}{h} \right).$$

The normalization factor, $V_2^{-1} M_2 \sigma_e^2$ for the LR statistic can be estimated by $\widehat{V}_2^{-1} \widehat{M}_2 \widehat{\sigma}_e^2$, where $\widehat{\sigma}_e^2 = (n(T - t_0 + 1))^{-1} \sum_{i=1}^{n} \sum_{t=t_0}^{T} \widehat{e}_{it}^2$. Notice that it becomes 1 under the conditional homoskedasticity and the martingale difference sequence assumption for e_{it}. Hansen (2000) provides the asymptotic distribution function of the LR_n statistic, which is $\left(1 - e^{-x/2}\right)^2$.

A.2.2 Threshold Regression in Reduced Form

Now, consider the case where the reduced form is the threshold regression, (24), which can be estimated via the three-step procedure described in Section A.1.2. It turns out that the asymptotic distributions of $\widehat{\theta}$ can be presented by a slight modification of Theorem 3. Thus, we state its asymptotic distribution as Corollary. Interestingly, the way how the covariance kernels are characterized in this case is illuminating. If we estimated the common threshold separately by the two-step approach as in Theorem 3, then the estimation error in the first step would affect the asymptotic distribution of the threshold estimate in the second step.

Corollary 4 *Let $\lambda_j = \left(\lambda_{jt_0}', ..., \lambda_{jT}'\right)'$, $j = 1, ..., 4$, and assume that $\lambda_1 - \lambda_2 = n^{-\alpha} \delta_1$ for some non-zero vector δ_1. Let Assumptions, 5, 6 and 8 hold with $F_{1,it} = \Gamma_{1t} z_{it} 1\{q_{it} \leq \gamma\} +$*

$\Gamma_{2t} z_{it} 1\{q_{it} > \gamma\}$, $F_{2,it} = x_{it-1}, \mathrm{E}|z_{it}|^4 < \infty$ *and* $\mathrm{E}e_{it}^4 < \infty$*. Then, the asymptotic distribution of* $\widehat{\theta}$ *estimated from (24) is the same as in Theorem 3.*

Notice that it would be desirable to relax certain conditions in Corollary 4 such as the common threshold across the reduced form and the structural form (see also Remark 5) or the same α to control the magnitude of the threshold effect.

A.3 Testing for Linearity

We present the asymptotic distribution of the supW statistic defined in (15), which tests the validity of the null hypothesis of no threshold effect (see (14)). If δ were estimated by the FD-2SLS, as is well-known in the literature, the limit is the supremum of the square of a Gaussian process with unknown covariance kernel, yielding non-pivotal asymptotic distribution.

Theorem 5 *Suppose that Assumptions, 6(i), 7, 8, and 4 hold. Then, under the null* (14),

$$\mathrm{supW} \xrightarrow{d} \sup_{\gamma\in\Gamma} B(\gamma)' M_1(\gamma)^{-1} R' \left[RM_1(\gamma)^{-1} V_1(\gamma) M_1(\gamma)^{-1} R'\right]^{-1} RM_1(\gamma)^{-1} B(\gamma),$$

where $B(\gamma)$ *is a mean-zero Gaussian process with the covariance kernel,* $A(\gamma_1)\,\Omega(\gamma_1,\gamma_2)\,A(\gamma_2)'$.

The p-values can be simulated following the same bootstrap steps as in Section 5.1. When the reduced form is a threshold regression, our test can be performed more efficiently based on the model, (25). In this case both reduced form and structural equations are linear under the null:

$$\mathcal{H}_0' : \lambda_{1t} - \lambda_{2t} = \lambda_{3t} - \lambda_{4t} = 0, \quad \text{for all } \gamma \in \Gamma \text{ and } t = t_0, ..., T. \tag{30}$$

As discussed earlier, the model, (25) can be estimated by the pooled OLS for each γ, and therefore, the construction of supW statistic is standard (*e.g.* Hansen, 1996).

A.4 Additional Simulation on Efficiency Comparison

To make an efficiency comparison of $\widehat{\gamma}$ estimated by FD-GMM and FD-2SLS, we have conducted an additional simulation. Too this end we modify the DGP in (16) as (DGP 1):

$$y_{it} = (0.7 - 0.5y_{it-1})\,1\{q_{it} \le 0\} + (-1.8 + 0.7y_{it-1})\,1\{q_{it} > 0\} + \sigma_1 u_{it},$$

where the transition variable, q_{it} is now randomly drawn from Uniform[-1,1], and independent of all u_{it}, $t = 1, 2, ..., T$. We also consider its restricted version with the common intercept (DGP 2):

$$y_{it} = 0.7 - 0.5y_{it-1}1\{q_{it} \le 0\} + 0.7y_{it-1}1\{q_{it} > 0\} + \sigma_1 u_{it}.$$

The FD-2SLS estimator is estimated by the 3-step procedure described in Section A.1.2, employing the following 4-regime threshold regression model:

$$\Delta y_{it} = \lambda_1' z_{it} 1\{q_{it} \leq \gamma\} + \lambda_2' z_{it} 1\{q_{it} > \gamma\} - \lambda_3' z_{it} 1\{q_{it-1} \leq \gamma\} - \lambda_4' z_{it} 1\{q_{it-1} > \gamma\} + \Delta u_{it},$$

where $z_{it} = (1, y_{it-2})$.

We examine bias and (logged) root mean square error (RMSE) of both estimators with 1,000 iterations. For $n = (50, 100, 200)$ and $T = 10$, we set $\sigma_1 = 1$. The simulation results reported in Table 9 demonstrates that the FD-2SLS displays clear dominance over the FD-GMM in terms of both magnitude and speed of decrease in RMSE as the sample size increases. This provides support for our theoretical prediction that the threshold estimate, $\widehat{\gamma}$, obtained by the FD-2SLS is super-consistent, as compared to the less efficient FD-GMM estimator.

Table 9 about here.

B Appendix: Proof of Theorems

B.1 GMM

Let

$$\xi_i = \begin{pmatrix} \Delta x_{it_0} z_{it_0}' \\ \vdots \\ \Delta x_{iT} z_{iT}' \end{pmatrix} \quad \text{and} \quad \zeta_i(\gamma) = \begin{pmatrix} X_{it_0}' \mathbf{1}_{it_0}(\gamma) z_{it_0}' \\ \vdots \\ X_{iT}' \mathbf{1}_{iT}(\gamma) z_{iT}' \end{pmatrix}.$$

Then, we can rewrite the moment indicator $g_i(\theta)$ given in (7) as

$$g_i(\theta) = g_i - \xi_i' (\beta - \beta_0) - \zeta_i' (\delta - \delta_n) - (\zeta_i(\gamma) - \zeta_i)' \delta, \tag{31}$$

where $\zeta_i = \zeta_i(\gamma_0)$ following the convention in this paper. Also recall that $g_i = g_i(\theta_n) = \left(z_{it_0}' \Delta\varepsilon_{it_0}, ..., z_{iT}' \Delta\varepsilon_{iT}\right)'$ and $\mathrm{E} g_i = 0$.

Proof of Theorem 1. We begin with the consistency of the estimator. First, we show that $\mathrm{E} g_i(\theta_n) = 0$ if and only if $\theta = \theta_n$. Suppose that $\beta = \beta_0$ and $\delta = \delta_n$ but $\gamma \neq \gamma_0$. Then,

$$\mathrm{E}(g_i(\theta)) = \delta_n' \left(\mathrm{E}\left(\mathbf{1}_{it}(\gamma)' X_{it} z_{it}'\right)' - \mathrm{E}\left(\mathbf{1}_{it}(\gamma_0)' X_{it} z_{it}'\right)' \right)_{t=t_0,\dots,T}' \neq 0$$

due to the rank condition in Assumption 3. Similarly, if either $\beta \neq \beta_0$ or $\delta \neq \delta_n$, but $\gamma = \gamma_0$,

$$\mathrm{E}(g_i(\theta)) = \left(-\mathrm{E}\left(\Delta x_{it} z_{it}'\right)' (\beta - \beta_0), -\mathrm{E}\left(\mathbf{1}_{it}(\gamma_0)' X_{it} z_{it}'\right)' (\delta - \delta_n) \right)_{t=t_0,\dots,T} \neq 0.$$

If $\phi \neq \phi_0$ and $\gamma \neq \gamma_0$, the rank condition is sufficient since $\left((\beta - \beta_0)', (\delta - \delta_n)', \delta'\right) \neq 0$.

Next, given the linearity in the slope parameters for a fixed γ, we can write

$$\begin{pmatrix} \widehat{\beta}(\gamma)-\beta_0 \\ \widehat{\delta}(\gamma)-\delta_n \end{pmatrix} = \left(A_n(\gamma)' W_n A_n(\gamma)\right)^{-1} A_n(\gamma)' W_n \left(\bar{g}_n + \frac{1}{n}\sum_{i=1}^{n}\left(\zeta_i - \zeta_i(\gamma)\right)\delta_n\right), \qquad (32)$$

where $A_n(\gamma) = \frac{1}{n}\sum_{i=1}^{n}\left[\xi_i', \zeta_i(\gamma)'\right]$ and $\bar{g}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} g_i(\theta)$. As convention, $\bar{g}_n = \bar{g}_n(\theta_n)$. Since $W_n \xrightarrow{p} \Omega^{-1}$ and $A_n(\gamma) \xrightarrow{p} A(\gamma) = \mathrm{E}\left[\xi_i', \zeta_i(\gamma)'\right]$ uniformly, which follows from the standard uniform law of large numbers (ULLN),

$$n^{\alpha}\begin{pmatrix} \widehat{\beta}(\gamma)-\beta_0 \\ \widehat{\delta}(\gamma)-\delta_n \end{pmatrix} \xrightarrow{p} \left(A(\gamma)'\Omega^{-1}A(\gamma)\right)^{-1}\left(A(\gamma)'\Omega^{-1}\left(\mathrm{E}\zeta_i - \mathrm{E}\zeta_i(\gamma)\right)\delta_0\right),$$

as $\bar{g}_n = O_p\left(n^{-1/2}\right)$ due to the CLT. Since $\bar{g}_n(\theta)$ is continuous in β and δ for any given γ, the continuous mapping theorem and standard algebra yield that

$$n^{\alpha}\bar{g}_n\left(\widehat{\beta}(\gamma), \widehat{\delta}(\gamma), \gamma\right) \xrightarrow{p} \left(I + A(\gamma)\left(A(\gamma)'\Omega^{-1}A(\gamma)\right)^{-1}A(\gamma)'\Omega^{-1}\right)\left(\mathrm{E}\zeta_i - \mathrm{E}\zeta_i(\gamma)\right)\delta_0.$$

The term in the first brackets in the right hand side is positive definite and $\mathrm{E}\zeta_i(\gamma) = \mathrm{E}\zeta_i$ if and only if $\gamma = \gamma_0$. Therefore, $p\lim_{n\to\infty} n^{2\alpha}\bar{J}_n\left(\widehat{\beta}(\gamma), \widehat{\delta}(\gamma), \gamma\right)$ is continuous and uniquely minimized at $\gamma = \gamma_0$ and the convergence is uniform, which implies the consistency of γ.

Convergence rate and asymptotic normality: Recall the definition of $\bar{J}_n(\theta)$ in (8) and let $J_n(\theta) = \mathrm{E}\left(g_i(\theta)\right)' W_n \mathrm{E}\left(g_i(\theta)\right)$. Also recall Assumption 3 and the definition of G in it and note that $G'\Omega^{-1}G$ is nonsingular and finite and that

$$\underset{l\times k}{G} = \left(G_\beta, G_\delta, G_\gamma\right) = \left(-\mathrm{E}\xi_i', -\mathrm{E}\zeta_i', -\frac{\partial}{\partial\gamma}\mathrm{E}\zeta_i(\gamma_0)'\delta_n\right).$$

And let $D_n = 2\kappa_n^{-1}G'W_n\bar{g}_n$, where κ_n is a $2k_1+2$ dimensional diagonal matrix whose first $2k_1+1$ diagonals are ones and the other element is n^{α}. We first claim that for any $h_n \to 0$

$$\sup_{|\theta-\theta_n|\le h_n} \frac{\sqrt{n}R_n(\theta)}{1+\sqrt{n}\left|\theta-\theta_n\right|} = o_p(1), \qquad (33)$$

where

$$R_n(\theta) = \bar{J}_n(\theta) - \bar{J}_n(\theta_n) - J_n(\theta) - D_n'(\theta-\theta_n).$$

Note that $\kappa_n D_n = O_p\left(n^{-1/2}\right)$ from CLT and $J_n(\theta) = 2(\theta-\theta_n)'\kappa_n^{-1}G'W_nG\kappa_n^{-1}(\theta-\theta_n) + o\left(\left|\theta-\theta_n\right|^2\right)$. Then, using $\kappa_n^{-1}\left(\widehat{\theta}-\theta_n\right)$ instead of $\widehat{\theta}-\theta_0$, the same line of argument as in the proof of Theorem 7.1 in Newey and McFadden (1994) yields that $\kappa_n^{-1}\left(\widehat{\theta}-\theta_n\right) = O_p\left(n^{-1/2}\right)$. Let $\widetilde{\theta}-\theta_n = \left(G'W_nG\right)^{-1}G'W_n\bar{g}_n$, then it follows that $\widetilde{\theta}-\theta_n-\kappa_n^{-1}\left(\widehat{\theta}-\theta_n\right) = o_p\left(n^{-1/2}\right)$. Therefore, we obtain the limit distribution as that of $\sqrt{n}\left(\widetilde{\theta}-\theta_n\right)$, that is, $\mathcal{N}\left(0, \left(G'\Omega^{-1}G\right)^{-1}\right)$.

Proof of (33) Define a centered empirical process

$$\varepsilon_n(\theta) = \sqrt{n}\left(\bar{g}_n(\theta) - \mathrm{E}g_i(\theta) - \bar{g}_n\right)$$

and decompose R_n to obtain a bound (see the proof of Theorem 7.2 of Newey and McFadden for details) such that

$$\frac{\sqrt{n}R_n(\theta)}{1+\sqrt{n}\left|\theta-\theta_n\right|} \leq \sum_{j=1}^{5} r_{jn}(\theta),$$

where

$$\begin{aligned}
r_{1n}(\theta) &= \left(2 + \left|\theta-\theta_n\right|/\sqrt{n}\right)\left|\varepsilon_n(\theta)' W_n \varepsilon_n(\theta)\right| / \left(1+\sqrt{n}\left|\theta-\theta_n\right|\right) \\
r_{2n}(\theta) &= \left|\left(\mathrm{E}g_i(\theta) - G\kappa_n^{-1}(\theta-\theta_n)\right)' W_n \sqrt{n}\bar{g}_n\right| / \left[\left|\theta-\theta_n\right|\left(1+\sqrt{n}\left|\theta-\theta_n\right|\right)\right] \\
r_{3n}(\theta) &= \left|\sqrt{n}\left(\mathrm{E}g_i(\theta) + \bar{g}_n\right)' W_n \varepsilon_n(\theta)\right| / \left(1+\sqrt{n}\left|\theta-\theta_n\right|\right) \\
r_{4n}(\theta) &= \left|\mathrm{E}g_i(\theta)' W_n \varepsilon_n(\theta)\right| \acute{/} \left|\theta-\theta_n\right| \\
r_{5n}(\theta) &= \sqrt{n}\left|\mathrm{E}g_i(\theta)'\left(W_n - W\right)\mathrm{E}g_i(\theta)\right| / \left[\left|\theta-\theta_n\right|\left(1+\sqrt{n}\left|\theta-\theta_n\right|\right)\right].
\end{aligned}$$

Let $h_n \to 0$ be any arbitrary sequence. First, note that $\sup_{|\theta-\theta_n|\leq h_n}\left|\varepsilon_n(\theta)\right| = o_p(1)$ if the empirical process $\sqrt{n}\left(\bar{g}_n(\theta) - \mathrm{E}g_i(\theta)\right)$ is stochastically equicontinuous. However, $g_i(\theta)$ is a sum of four terms and the first is free of θ and the next two are linear in β and δ, leaving only the last term to check for the stochastic equicontinuity. Since δ is bounded and each element in $\zeta_i(\gamma)$ is of the type, $\zeta_{it}1\{q_{it} > \gamma\}$, we need to show that the empirical process indexed by the type is stochastically equicontinuous. However, the indicator functions of half intervals constitute a Vapnik-Chervonenkis (VC) class and Theorem 2.14.1 of van der Vaart and Wellner (1996) yields the desired result by choosing an envelope function, $\left|\zeta_{it}\right| 1\{\left|q_{it}-\gamma_0\right| \leq h_n\}$.

Next, note that

$$\sup_{|\theta-\theta_n|\leq h_n} \sqrt{n}\mathrm{E}g_i(\theta) / \left(1+\sqrt{n}\left|\theta-\theta_n\right|\right) \leq \sup_{|\theta-\theta_n|\leq h_n} \left|\mathrm{E}g_i(\theta)\right| / \left|\theta-\theta_n\right| = O(1),$$

due to the differentiability of $\mathrm{E}g_i(\theta)$. For the same reason, $\sup_{|\theta-\theta_n|\leq h_n}\left|\mathrm{E}g_i(\theta) - G\kappa_n^{-1}(\theta-\theta_n)\right| / \left|\theta-\theta_n\right| = o(1)$. Therefore, these and the Cauchy-Schwarz inequality yield that $\sup_{|\theta-\theta_n|\leq h_n}\left|r_{jn}(\theta)\right| = o_p(1)$ for all j. ■

B.2 2SLS

In this section, many variables and processes are indexed by two different types of parameters, the reduced form parameter b and the structural form parameter θ, for instance, $e_{it}(\theta, b)$, $H_{it}(\theta, b)$, $\mathbb{M}_n(\theta, b)$, $M_n(\theta, b)$, and so on. As in previous sections, we make the following notational convention, where we write for instance $e_{it} = e_{it}(\theta_0, b_0)$, $e_{it}(\theta) = e_{it}(\theta, b_0)$, and $\widehat{e}_{it}(\theta) = e_{it}\left(\theta, \widehat{b}\right)$ and the same for the other terms.

Now, we turn to the proof of main theorem.

Proof of Theorem 3. We follows the standard three-step approach of establishing consistency, convergence rate, and asymptotic distribution in sequel.

Consistency We show that $\widehat{\theta} = \theta_n + o_p(1)$. Recall that

$$e_{it}(\theta) = e_{it} - (\beta - \beta_0)' H_{it} - (\delta - \delta_n)' \left(F_{it}' \mathbf{1}_{it}\right) - \left[\mathbf{1}_{it}(\gamma) - \mathbf{1}_{it}\right]' F_{it}\delta, \tag{34}$$

and let

$$M_n(\theta) = \sum_{t=t_0}^{T} \mathrm{E}\left(e_{it}^2(\theta)\right).$$

Then, it is sufficient to show (i) $\sup_{\theta\in\Theta}\left|\widehat{\mathbb{M}}_n(\theta) - M_n(\theta)\right| \xrightarrow{p} 0$ and (ii) $M_n(\theta)$ is continuous and has a unique minimum at θ_n. For (ii), note that $M_n(\theta)$ is continuous everywhere, twice differentiable everywhere but $\gamma = \gamma_0$, and the second derivative with respect to β and δ is positive definite uniformly in γ by Assumption 8. Furthermore, direct calculation reveals that $\partial M_n(\theta)/\partial\gamma$ is positive if $\gamma > \gamma_0$ and negative if $\gamma < \gamma_0$ in a neighborhood of θ_n. Since the conditional mean is the minimizer of the mean squared errors, θ_n becomes the unique minimizer of $M_n(\theta)$ in the compact set, Θ. For (i), note that

$$\sup_{\theta\in\Theta}\left|\widehat{\mathbb{M}}_n(\theta) - M_n(\theta)\right| \le \sup_{\theta\in\Theta}\left|\widehat{\mathbb{M}}_n(\theta) - \mathbb{M}_n(\theta)\right| + \sup_{\theta\in\Theta}\left|\mathbb{M}_n(\theta) - M_n(\theta)\right| \xrightarrow{p} 0$$

Convergence of the first term following the inequality is delegated to the proof on convergence rate below, while the convergence of the second is a standard ULLN, *e.g.* Newey and McFadden's (1994, Lemma 2.4). Thus, the consistency proof is complete.

Convergence rate We verify the conditions of Theorem 3.4.1 in van der Vaart and Wellner (1996) with the distance function defined by

$$d_n(\theta, \theta_n) = |\beta - \beta_0| + |\delta - \delta_n| + |\gamma - \gamma_0|^{1/(2-4\alpha)}.$$

In particular, in terms of maximization, we need to show that (using their notation), for $\delta_n < \delta < \eta$

$$\sup_{\delta/2 < d_n(\theta,\theta_n) \le \delta} \widetilde{M}_n(\theta) - \widetilde{M}_n(\theta_n) \le -\delta^2, \tag{35}$$

and

$$\mathrm{E}\sup_{\delta/2 < d_n(\theta,\theta_n) \le \delta} \sqrt{n}\left[\left(\widetilde{\mathbb{M}}_n - \widetilde{M}_n\right)(\theta) - \left(\widetilde{\mathbb{M}}_n - \widetilde{M}_n\right)(\theta_n)\right] \le C\phi_n(\delta)$$

for functions ϕ_n such that $\delta \to \phi_n(\delta)/\delta^{\alpha}$ is decreasing on (δ_n, η).Then, for $r_n \leq C\delta_n^{-1}$ and $r_n^2\phi_n(r_n^{-1}) \leq \sqrt{n}$, and for any $\widehat{\theta}$ such that

$$\widetilde{\mathbb{M}}_n\left(\widehat{\theta}\right) \geq \widetilde{\mathbb{M}}_n(\theta_n) + O_p\left(r_n^{-2}\right),$$

$d\left(\widehat{\theta}, \theta_n\right) = O_p\left(r_n^{-1}\right)$. For our case, we set $r_n = \sqrt{n}$, $\delta_n = n^{-1/2}$, and $\phi_n(\delta) = \delta$. Because any estimator $\widehat{\theta}$ satisfying $\widetilde{\mathbb{M}}_n\left(\widehat{\theta}\right) \geq \widetilde{\mathbb{M}}_n(\theta_n) + O_p\left(r_n^{-2}\right)$ has the convergence rate of r_n^{-1} and $r_n = \sqrt{n}$, the maximizer of $\widehat{\mathbb{M}}_n(\theta)$ such that $\left|\widetilde{\mathbb{M}}_n(\theta) - \widehat{\mathbb{M}}_n(\theta)\right| = O_p\left(n^{-1}\right)$ has the same convergence rate of r_n^{-1} in terms of the distance d_n.

Define

$$\begin{aligned}
r_{it}(\theta, b) &:= e_{it}(\theta, b) - e_{it}(\theta) \\
&= (H_{it}(b) - H_{it})'\beta_0 - \mathbf{1}_{it}'(F_{it}(b) - F_{it})\delta_n \\
&- (H_{it}(b) - H_{it})'(\beta - \beta_0) - \mathbf{1}_{it}'(F_{it}(b) - F_{it})(\delta - \delta_n) \\
&- (\mathbf{1}_{it}(\gamma) - \mathbf{1}_{it})'(F_{it}(b) - F_{it})\delta,
\end{aligned}$$

then,

$$\widehat{\mathbb{M}}_n(\theta) - \mathbb{M}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}\sum_{t=t_0}^{T}\left(r_{it}^2\left(\theta, \widehat{b}\right) + 2e_{it}(\theta)\, r_{it}\left(\theta, \widehat{b}\right)\right).$$

However, the first term $\frac{1}{n}\sum_{i=1}^{n}\sum_{t=t_0}^{T} r_{it}^2\left(\theta, \widehat{b}\right) = O_p\left(n^{-1}\right)$ uniformly in θ in a neighborhood of θ_0 by applying the ULLN, the $\sqrt{n}$-consistency of $\widehat{b}$ and the differentiability of F in Assumption 4. For the second term, note that, proceeding similarly by expansion of F and H and applying the CLT and ULLN, $\frac{1}{n}\sum_{i=1}^{n}\sum_{t=t_0}^{T} e_{it} r_{it}\left(\theta, \widehat{b}\right) = O_p\left(n^{-1}\right)$ uniformly in θ in a neighborhood of θ_0, where e_{it} is the first term in the expansion of $e_{it}(\theta)$ in (34). Then,

$$\widehat{\mathbb{M}}_n(\theta) = \mathbb{M}_n(\theta) - \mathbb{R}_n\left(\theta, \widehat{b}\right) + O_p\left(n^{-1}\right), \tag{36}$$

where $\mathbb{R}_n(\theta, b) = \frac{2}{n}\sum_{i=1}^{n}\sum_{t=t_0}^{T} r_{it}(\theta, b)\left((\beta - \beta_0)' H_{it} + (\delta - \delta_n)'(F_{it}'\mathbf{1}_{it}) + [\mathbf{1}_{it}(\gamma) - \mathbf{1}_{it}]' F_{it}\delta\right)$.

Since $\widehat{b}$ is square root n consistent, we may consider the process over the expanded parameter space $\psi \in \Theta_n \times B_n$, where $\Theta_n = \{\theta : d_n(\theta, \theta_n) \leq \epsilon\}$ for some $\epsilon > n^{-1/2}$ and $B_n = \{b : |b - b_0| \leq K/\sqrt{n}\}$ for some $K < \infty$. Note that $\psi_n = \left(\theta_n', b_0'\right)'$ should correspond to θ_n in van der Vaart and Wellner's Theorem 3.4.1. Accordingly, from (36) we define

$$\widetilde{\mathbb{M}}_n(\psi) = -\mathbb{M}_n(\theta) + \mathbb{R}_n(\theta, b), \tag{37}$$

for which we multiplied -1 to make it a maximization problem. Then, it is sufficient to verify the above conditions of Theorem 3.4.1 for $\widetilde{\mathbb{M}}_n(\psi)$. Accordingly, let

$$\widetilde{M_n}(\psi) = \mathrm{E}\widetilde{\mathbb{M}}_n(\psi).$$

and check the first condition (35). Note that $\widetilde{M}_n(\psi_n) = -M_n(\theta_n)$, and

$$\widetilde{M}_n(\psi) = -M_n(\theta) + 2\mathrm{E}\sum_{t=t_0}^{T} r_{it}(\theta, b)\left((\beta-\beta_0)' H_{it} + (\delta-\delta_n)'\left(F_{it}'\mathbf{1}_{it}\right) + \left[\mathbf{1}_{it}(\gamma) - \mathbf{1}_{it}\right]' F_{it}\delta\right),$$

whose last term is $O\left(n^{-1/2}\right)$ due to Assumption 4 and the fact that $|b-b_0| \le K/\sqrt{n}$. Thus, it is enough to consider $M_n(\theta)$. However, as shown in the consistency proof, $M_n(\theta)$ is quadratic around θ_n in terms of the distance d_n and it satisfies the condition (35).

The maximal inequality for the empirical process $\sqrt{n}\left(\left(\widetilde{\mathbb{M}}_n - \widetilde{M}_n\right)(\psi) - \left(\widetilde{\mathbb{M}}_n - \widetilde{M}_n\right)(\psi_n)\right)$ is the second condition to check. Consider $\mathbb{M}_n(\theta)$, the first term of $\widetilde{\mathbb{M}}$ in (37). Then, we need to check the maximal inequality for the centered empirical process:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{t=t_0}^{T}\left[e_{it}^2(\theta) - e_{it}^2 - \mathrm{E}e_{it}^2(\theta) + \mathrm{E}e_{it}^2\right].$$

The function $e_{it}^2(\theta) - e_{it}^2$ is the sum of linear and quadratic functions of β and δ multiplied by $\left[\mathbf{1}_{it}(\gamma) - \mathbf{1}_{it}\right]$. This is a VC class of functions. In this case, a maximal inequality bound is given by the L^2 norm of an envelope. We choose the following envelope:

$$2|e_{it}||F_{it}|\epsilon + |F_{it}|^2\epsilon^2 + 2|e_{it}||\mathbf{1}_{it}(\gamma) - \mathbf{1}_{it}||F_{it}|(|\delta_n| + \epsilon) + |\mathbf{1}_{it}(\gamma) - \mathbf{1}_{it}||F_{it}|^2(|\delta_n| + \epsilon)^2,$$

for some $C < \infty$. The first two terms are clearly $O(\epsilon)$ in L^2 norm. As the last two terms can be treated in a similar way, we only need to show that

$$\mathrm{E}^{1/2}\left\{|e_{it}|^2|F_{it}|^2\left(1\left(|q_{it}-\gamma_0| \le \epsilon^{2-4\alpha}\right) + 1\left(|q_{it-1}-\gamma_0| \le \epsilon^{2-4\alpha}\right)\right)\right\}(|\delta_n| + \epsilon) = O(\epsilon).$$

But, the standard algebra using the change-of-variables yields that

$$\mathrm{E}^{1/2}|e_{it}|^2|F_{it}|^2 1\left(|q_{it}-\gamma_0| \le \epsilon^{2-4\alpha}\right)|\delta_n| = O\left(\epsilon^{1-2\alpha}|\delta_0| n^{-\alpha}\right) = O(\epsilon),$$

where the last equality follows since $\epsilon > n^{-1/2}$.

Turning to $\mathbb{R}_n(\theta, b)$, we note that $\mathbb{R}_n(\theta_n, b_0) = 0$ and apply the Taylor series expansion to $r_{it}(\theta, b)$ with respect to b. Then, as $\widehat{b}$ is a $\sqrt{n}$-consistent estimator, the ULLN is sufficient to satisfy the maximal inequality in (35).

The last condition to be checked is:

$$\widetilde{\mathbb{M}}_n\left(\widehat{\theta}, \widehat{b}\right) \ge \widetilde{\mathbb{M}}_n(\theta_n, b_0) + O_p\left(n^{-1}\right).$$

But, for $\left|\widehat{b}-b_0\right| \leq K/\sqrt{n}$, we have

$$\begin{aligned}
\widetilde{\mathbb{M}}_n\left(\widehat{\theta},\widehat{b}\right) &= \mathbb{M}_n\left(\widehat{\theta},\widehat{b}\right)+O_p\left(n^{-1}\right)\\
&\geq \mathbb{M}_n\left(\theta_n,\widehat{b}\right)+O_p\left(n^{-1}\right)\\
&= \widetilde{\mathbb{M}}_n\left(\theta_n,\widehat{b}\right)+O_p\left(n^{-1}\right)\\
&= \widetilde{\mathbb{M}}_n\left(\theta_n,b_0\right)+O_p\left(n^{-1}\right),
\end{aligned}$$

where we have shown the first and third equality in (36), the second inequality by construction, and the last equality follows because $\mathbb{M}_n(\theta,b)$ does not depend on b for $\theta=\theta_n$. Thus,

$$\sqrt{n}d_n(\theta,\theta_0)=\sqrt{n}\left(|\theta_1-\theta_{10}|+|\gamma-\gamma_0|^{1/(2-4\alpha)}\right)=O_p(1).$$

Asymptotic distribution: Let h be a k-dimensional vector and r_n be the k-dimensional vector whose first $k-1$ elements are $\sqrt{n}$ and the last element is $n^{1-2\alpha}$. Accordingly, partition $h=(h_c',h_\gamma)'$. We derive the weak convergence of the centered and rescaled criterion function

$$n\left(\widehat{\mathbb{M}}_n\left(\theta_n+h./r_n,\widehat{b}\right)-\mathbb{M}_n\left(\theta_n,\widehat{b}\right)\right) \tag{38}$$

on $\{h:|h|\leq K\}$ for an arbitrary $K<\infty$, where ./ is the elementwise division. Then, the argmax continuous mapping theorem (*e.g.* van der Vaart and Wellner, 1996) will yield the desired result.

Let $e_i=(e_{it_0},...,e_{iT})'$, $h_n=h./r_n$, and $\Xi_{2i}\left(h_\gamma n^{2\alpha-1},b\right)$ denote the bottom k_1+1 rows of $\Xi_i(\gamma,b)$ evaluated at $\gamma=\gamma_0+h_\gamma n^{2\alpha-1}$, and define

$$\begin{aligned}
m_{ni}(h,b) &= \sqrt{n}\left[e_i(\theta_n+h_n,b)-e_i(b)\right]\\
&= \Xi_i(b)'h_c-\sqrt{n}\left(\Xi_{2i}\left(h_\gamma n^{2\alpha-1},b\right)-\Xi_{2i}(b)\right)'\left(\delta_n+h_\delta/\sqrt{n}\right).
\end{aligned}$$

Writing $\widehat{e}_i=e_i\left(\widehat{b}\right)$, $\widehat{m}_{ni}(h)=m_{ni}\left(h,\widehat{b}\right)$, and $e_i=e_i(b_0)$ as before, we have:

$$n\left(\mathbb{M}_n\left(\theta_n+h_n,\widehat{b}\right)-\mathbb{M}_n\left(\theta_n,\widehat{b}\right)\right)=\frac{1}{n}\sum_{i=1}^n|\widehat{m}_{ni}(h)|^2-\frac{2}{\sqrt{n}}\sum_{i=1}^n\widehat{e}_i'\widehat{m}_{ni}(h). \tag{39}$$

Consider the last term in (39). By Assumption 4 and (26), we apply the mean value theorem to get an expansion:

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{i=1}^n\widehat{m}_{ni}(h)'\widehat{e}_i &= \frac{1}{\sqrt{n}}\sum_{i=1}^n\widehat{m}_{ni}(h)'\Delta\varepsilon_i\\
&\quad+\frac{1}{n}\sum_{i=1}^n\widehat{m}_{ni}(h)'\frac{\partial\Xi_i\left(\widetilde{b}\right)'\theta_{10}}{\partial b'}\left(\frac{\mathrm{E}\left(\mathbb{F}_i\mathbb{F}_i'\right)^{-1}}{\sqrt{n}}\sum_{i=1}^n\mathbb{F}_i\eta_i+o_p(1)\right).
\end{aligned} \tag{40}$$

Next, expand its first term in (40):

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widehat{m}_{ni}\left(h\right)'\Delta\varepsilon_i\\
&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(h_c'\Xi_i\left(\widehat{b}\right)-n^{\frac{1}{2}-\alpha}\left(\delta_0+o\left(1\right)\right)'\left(\Xi_{2i}\left(h_\gamma n^{2\alpha-1},\widehat{b}\right)-\Xi_{2i}\left(\widehat{b}\right)\right)\right)\Delta\varepsilon_i\\
&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}h_c'\Xi_i\Delta\varepsilon_i-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}n^{\frac{1}{2}-\alpha}\delta_0'\left(\Xi_{2i}\left(h_\gamma n^{2\alpha-1}\right)-\Xi_{2i}\right)\Delta\varepsilon_i+o_p\left(1\right),
\end{aligned}
\tag{41}
$$

where the last equality is due to the asymptotic normality of $\widehat{b}$. The CLT applies for the first term in (41). For the weak convergence of the second term, we need to consider a sequence of classes of functions:

$$
\mathcal{G}_n=\left\{g_n\left(h_\gamma\right)=n^{\frac{1}{2}-\alpha}\delta_0'\left(\Xi_{2i}\left(h_\gamma n^{2\alpha-1}\right)-\Xi_{2i}\right)\Delta\varepsilon_i:\left|h_\gamma\right|<K\right\},
$$

with a sequence of envelope functions,

$$
G_n=n^{\frac{1}{2}-\alpha}\left|\delta_0\right|\left|\Delta\varepsilon_i\right|\left|F\left(z_i\right)\right|\left|\mathbf{1}_i\left(\gamma\right)-\mathbf{1}_i\left(\gamma_0\right)\right|,
$$

and apply Theorem 2.11.22 of van der Vaart and Wellner (1996). Recall that $\Xi_{2i}\left(h_\gamma n^{2\alpha-1}\right)$ is the collection of $F_{it}'\mathbf{1}_{it}\left(h_\gamma n^{2\alpha-1}\right)$ over all t. As the indicator functions (and those multiplied by a random variable) constitute a VC class of functions, they satisfy the uniform entropy condition of Theorem 2.11.22. Since $\Xi_{2i}\left(\gamma\right)$ has continuous first and second moments, it remains to verify the conditions on the envelope G_n. It is clear that $\mathrm{E}G_n^2=O\left(1\right)$ and the Lindeberg condition is satisfied since

$$
\begin{aligned}
&\mathrm{E}\left(G_n^2 1\left(\left|G_n\right|>\eta\sqrt{n}\right)\right)\\
&\leq \mathrm{E}2n^{1-2\alpha}\left|\delta_0\right|^2\sum_{t=t_0-1}^{T}1\left(\left|q_{it}-\gamma_0\right|\leq h_\gamma n^{-1+2\alpha}\right)\\
&\times\left(\left|\Delta\varepsilon_i\right|^2\left|F\left(z_i\right)\right|^2\right)1\left(\left|\Delta\varepsilon_i\right|\left|F\left(z_i\right)\right|>\frac{\eta n^{\alpha}}{2\left(T+1\right)\left|\delta_0\right|}\right)\\
&\leq O\left(n^{-\alpha\zeta}\right)=o\left(1\right).
\end{aligned}
$$

due to Assumption 7. We will specify the covariance kernel below after noting that the second term in (40) expands by the standard Taylor series expansion to yield

$$
\begin{aligned}
&\frac{2}{\sqrt{n}}\sum_{i=1}^{n}\widehat{e}_i'\widehat{m}_{ni}\left(h\right)\\
&=\left(I-\mathrm{E}\widetilde{m}_{ni}\left(h\right)\left[I_T\otimes\left(\iota\otimes\beta_0\right)\right]'\mathbb{F}_i'\mathrm{E}\left(\mathbb{F}_i\mathbb{F}_i'\right)^{-1}\right)\frac{2}{\sqrt{n}}\sum_{i=1}^{n}\left[\begin{array}{c}\widetilde{m}_{ni}\left(h\right)'\Delta\varepsilon_i\\ \mathbb{F}_i\eta_i\end{array}\right]+o_p\left(1\right),
\end{aligned}
$$

where $\widetilde{m}_{ni}(h) = h_c'\Xi_i - n^{\frac{1}{2}-\alpha}\delta_0'\left(\Xi_{2i}\left(h_\gamma n^{2\alpha-1}\right) - \Xi_{2i}\right)$. Turning back to the covariance kernel of the empirical process indexed by $\mathcal{G}_n$ above and the covariance between the process indexed by h_c and the process indexed by h_γ,we note that the latter vanishes due to the difference in the convergence rates. For this, it is enough to observe that each element in the matrix $\mathrm{E}\left(\Xi_{2i}\left(h_\gamma n^{2\alpha-1}\right) - \Xi_{2i}\right)$ is bounded by, up to a constant,

$$\mathrm{E}1\left\{|q_{it}-\gamma_0| \le h_\gamma n^{2\alpha-1}\right\} = \int 1\{|q|\le 1\}\, p\left(h_\gamma n^{2\alpha-1}q+\gamma_0\right) h_\gamma n^{2\alpha-1}dq = O\left(n^{2\alpha-1}\right),$$

due to Assumption 2, where the change-of-variable is applied for the first equality. By the same reasoning,

$$\mathrm{E}\widetilde{m}_{ni}(h)\frac{\partial\Xi_i\left(\widetilde{b}\right)'\theta_{10}}{\partial b'} = h_c'\mathrm{E}\Xi_i\frac{\partial\Xi_i'\theta_{10}}{\partial b'} + o(1),$$

and the limit of $\frac{1}{n}\sum_{i=1}^n |\widehat{m}_{ni}(h)|^2$ is the sum of a quadratic function of h_c and a function of h_γ without any interaction term. This implies the asymptotic independence between $\left(\widehat{\beta}',\widehat{\delta}'\right)$ and $\widehat{\gamma}$. For the former, note that $g_n(h_\gamma)\, g_n\left(\acute{h}_\gamma\right) = 0$ unless h_γ and $\acute{h}_\gamma$ have the same sign. For $h_\gamma > \acute{h}_\gamma \ge 0$,

$$\begin{aligned}
&n^{-1+2\alpha}\mathrm{E}\left(g_n(h_\gamma)\, g_n\left(\acute{h}_\gamma\right)\right)\\
&= \delta_0'\sum_{r,t=t_0}^{T}\mathrm{E}\left[\Delta\varepsilon_{it}\Delta\varepsilon_{ir}F_{it}'\left[\mathbf{1}_{it}\left(\gamma_0+h_\gamma n^{2\alpha-1}\right)-\mathbf{1}_{it}\right]\left[\mathbf{1}_{ir}\left(\gamma_0+\acute{h}_\gamma n^{2\alpha-1}\right)-\mathbf{1}_{ir}\right]'F_{ir}\right]\delta_0.
\end{aligned} \tag{42}$$

The evaluation of the expectation can be done in the same way as above. Thus, those expectations involving the products of indicators of q_{it} and $q_{it'}$ with $t \ne t'$ will vanish. After some algebra, we can show that the limit of (42) is $\delta_0' V_2(\gamma_0)\,\delta_0\left(h_\gamma - \acute{h}_\gamma\right)$, and more generally

$$\delta_0' V_2(\gamma_0)\,\delta_0\left|h_\gamma - \acute{h}_\gamma\right| 1\left\{\mathrm{sgn}(h_\gamma) = \mathrm{sgn}\left(\acute{h}_\gamma\right)\right\},$$

where $V_2(\gamma)$ is given in Section 4. This functional form of the covariance kernel implies that the limit Gauss process is a two-sided Brownian motion originating from zero.

Now, applying a standard ULLN to $\frac{1}{n}\sum_{i=1}^n\sum_{t=t_0}^T m_{it}(h,b)^2$, and using the consistency of $\widehat{b}$ and the same line of arguments as above, we may conclude that

$$\frac{1}{n}\sum_{i=1}^n |\widehat{m}_{ni}(h)|^2 \xrightarrow{p} h_c'\mathrm{E}\Xi_i\Xi_i'h_c + M_2(\gamma_0)\,|h_\gamma|.$$

Given the structure of the weak limit of (38), the minimizer $\widehat{h}_c$ is normally distributed and the argmin $\widehat{h}_\gamma$ is that of a two-sided Brownian motion added by a linear trend. The representation in main body of the theorem follows from Hansen (2000), in which it is shown for a two-sided standard Brownian motion W and for any positive constants c_1 and c_2 that

$$\underset{\gamma\in\mathbb{R}}{\operatorname{argmin}}\left[c_1\left|\gamma\right|-2\sqrt{c_2}W\left(\gamma\right)\right]=\frac{c_2}{c_1^2}\underset{\gamma\in\mathbb{R}}{\operatorname{argmin}}\left[\frac{\left|\gamma\right|}{2}-W\left(\gamma\right)\right].$$

Furthermore, the same line of proof as in Theorem 2 of Hansen (2000) applies to the convergence of $LR_n\left(\gamma_0\right)$ given the results obtained above about $\widehat{\theta}_1$ and $\widehat{\gamma}$. This completes the proof. ■

Proof of Corollary 4. The consistency proof is almost identical to Theorem 3, and thus omitted. For the convergence rate of the estimator, recall that we need to verify two conditions, one is the condition on the limit criterion function and the other is the condition on the maximal inequality of the empirical process part. The latter is identical to that in Theorem 3 since the sum of two VC classes of functions is VC. For the former note that the current case has another component in the regression function than in Theorem 3, which is $1\left\{q_{it-1}>\gamma\right\}$. This generates a kink in the limit criterion function at γ_0 as $1\left\{q_{it}>\gamma\right\}$ does. Therefore, the limit criterion function has the same feature as the one in Theorem 3. Thus, we get the same rate of convergence as in Theorem 3.

Finally, turning to the asymptotic distribution, we note that the argument for the stochastic equicontinuity of the rescaled criterion function is the same as in Theorem 3. To get the covariance kernel of the limit Gaussian process note that, as discussed in (42), the covariances between two terms involving two indicators of q_{it} and $q_{it'}$ with $t\neq t'$ vanish, yielding the covariance kernel as desired. Details are omitted to avoid repetition. ■

B.3 Testing

Proof of Theorem 2. Applying the standard ULLN and the continuous mapping theorem to (32), we have:

$$W_n\left(\gamma\right)\Rightarrow\left[\begin{array}{c}Z'\Omega^{-1/2}G\left(\gamma\right)'\left(G\left(\gamma\right)'\Omega^{-1}G\left(\gamma\right)\right)^{-1}R'\left[R\left(G\left(\gamma\right)'\Omega^{-1}G\left(\gamma\right)\right)^{-1}R'\right]^{-1}\\ \times R\left(G\left(\gamma\right)'\Omega^{-1}G\left(\gamma\right)\right)^{-1}G\left(\gamma\right)\Omega^{-1/2}Z,\end{array}\right]$$

where $G\left(\gamma\right)=\left(G_\beta,G_\delta\left(\gamma\right)\right)$ and Z is the standard normal variate of dimension l, which is the number of moment conditions. ■

Proof of Theorem 5. As the model is linear for each γ, the marginal convergence of $\sqrt{n}\widehat{\delta}\left(\gamma\right)$ is standard and the asymptotic distribution is given as in Theorem 3. The finite dimensional convergence for any finite collection of γ values is then also standard. Therefore, it remains to show the stochastic equicontinuity of the process. Recall the expression from (23) that

$$\begin{pmatrix} \widehat{\beta}(\gamma) \\ \widehat{\delta}(\gamma) \end{pmatrix} = \left(\frac{1}{n} \sum_{i=1}^{n} \sum_{t=t_0}^{T} \mathbb{X}_{it}\left(\widehat{b}_t, \gamma\right) \mathbb{X}_{it}\left(\widehat{b}_t, \gamma\right)' \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} \sum_{t=t_0}^{T} \mathbb{X}_{it}\left(\widehat{b}_t, \gamma\right) \Delta y_{it} \right),$$

where $\mathbb{X}_{it}(b_t, \gamma) = \left(H_t(z_{it}; b_t)' - \left(F_t(z_{it}; b_t)' \mathbf{1}_{it}(\gamma) \right)' \right)'$. The uniform convergence of the first sum can be derived as in the proof of Theorem 3 using the ULLN and the consistency of $\widehat{b}$ in Assumption 4. Thus, the stochastic equicontinuity of $\sqrt{n}\left(\left(\widehat{\beta}(\gamma) - \beta(\gamma) \right)', \widehat{\delta}(\gamma)' \right)$ implies that of $\sqrt{n}\widehat{\delta}(\gamma)$. Since the functions H_t and F_t are twice continuously differentiable in b_t,it ends up with verifying the stochastic equicontinuity of the empirical process of the types of functions $f(z_{it}) 1\{q_{it} > \gamma\}$, where f is some known transformation of z_{it}. However, this is a VC class of function, which implies the stochasting equicontinuity of the empiricl process of this class of functions, see e.g. van der Vaart and Wellner's (1996) Section 2.6. ∎

Acknowledgement

We are mostly grateful to the editor, Han Hong, the associate editor and three anonymous referees for their helpful comments. We are also grateful to Mini Ahn, Heather Anderson, Mehmet Caner, Jinseo Cho, In Choi, Viet Anh Dang, Robert Faff, Matthew Greenwood-Nimmo, Jinwook Jeong, Taehwan Kim, Jay Lee, Myungjae Lee, James Morley, Joon Park, Kevin Reilly, Laura Serlenga, seminar participants at Universities of Canterbury, Korea, Leeds, Melbourne, New South Wales, Queensland, Sogang, Sung Kyun Kwan and Yonsei, and conference delegates at the AMES at Korea University, Seoul, August 2011, the 20th Panel Data Conference at Hitotsubashi University, Tokyo, July 2014 and the ESEM, Toulouse, August 2014 for their helpful comments. We would like to thank Minjoo Kim for excellent research assistance. The first author acknowledges support by Promising-Pioneering Researcher Program by Seoul National University (SNU) in 2015 and partial support by Jewon research institute. The second author acknowledges partial financial support from the ESRC (Grant No. RES-000-22-3161). The usual disclaimer applies.

References

[1] Ahn, S.C. and P. Schmidt, 1995, Efficient Estimation of Models for Dynamic Panel Data. Journal of Econometrics 68, 5-27.

[2] Aivazian, V.A., Y. Ge and J. Qiu, 2005, The Impact of Leverage on Firm Investment: Canadian Evidence. Journal of Corporate Finance 11, 277-291.

[3] Alvarez, J. and M. Arellano, 2003, The Time Series and Cross-section Asymptotics of Dynamic Panel Data Estimators. Econometrica 71, 1121-1159.

[4] Andrews, D.W.K., 1994, Empirical Process Methods in Econometrics. in Handbook of Econometrics IV, eds by R.F. Engle and D.L. McFadden, 2247-2294, Elsevier.

[5] Arellano, M. and S. Bond, 1991, Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. Review of Economic Studies 58, 277-297.

[6] Arellano, M. and O. Bover, 1995, Another Look at the Instrumental Variable Estimation of Error Components Models. Journal of Econometrics 68, 29-51.

[7] Bai, J., 2009, Panel Data Models with Interactive Fixed Effects. Econometrica 77, 1229-1279.

[8] Blundell, R. and S. Bond, 1998, Initial Conditions and Moment Restrictions in Dynamic Panel Data Models. Journal of Econometrics 87, 115-143.

[9] Bun, M.J.G. and F. Windmeijer, 2010, The Weak Instrument Problem of the System GMM Estimator in Dynamic Panel Data Models. Econometrics Journal, 95–126.

[10] Caner, M. and B.E. Hansen, 2004, Instrumental Variable Estimation of a Threshold Model. Econometric Theory 20, 813-843.

[11] Chamberlain, G., 1987, Asymptotic Efficiency in Estimation with Conditional Moment Restrictions. Journal of Econometrics 34, 305–334.

[12] Chamberlain, G. and G. Imbens, 2004, Random Effects Estimators with Many Instrumental Variables. Econometrica 72, 295-306.

[13] Chan, K.S., 1993, Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model. Annals of Statistics 21, 520-33.

[14] Chen, H., T.T.L. Chong and J. Bai, 2012, Theory and Applications of TAR Model with Two Threshold Variables. Econometric Reviews 31, 142–170.

[15] Chong, T.T.L. and I.K.M. Yan, 2015, A New Threshold Regression Approach to Predict Currency Crises. mimeo., Chinese University of Hong Kong.

[16] Dang, V.A., M. Kim and Y. Shin, 2012, Asymmetric Capital Structure Adjustments: New Evidence from Dynamic Panel Threshold Models. Journal of Empirical Finance 19, 465-482.

[17] Davies, R.B., 1977, Hypothesis Testing when a Nuisance Parameter is Present only under the Alternative. Biometrika 64, 247-254.

[18] Fazzari, S.M., R.G. Hubbard and B.C. Petersen, 1988, Financing Constraints and Corporate Investment. Brookings Papers on Economic Activity 1, 141–195.

[19] 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.

[20] González, A., T. Teräsvirta and D. van Dijk, 2005, Panel Smooth Transition Model and an Application to Investment Under Credit Constraints. Working Paper, Stockholm School of Economics.

[21] Hansen, B.E., 1996, Inference when a Nuisance Parameter is not Identified under the Null Hypothesis. Econometrica 64, 414-30.

[22] Hansen, B.E., 1999, Threshold Effects in Non-dynamic Panels: Estimation, Testing and Inference. Journal of Econometrics 93, 345-368.

[23] Hansen, B.E., 2000, Sample Splitting and Threshold Estimation. Econometrica 68, 575-603.

[24] Hansen, B.E., 2011, Threshold Autoregression in Economics. Statistics and Its Interface 4, 123-127.

[25] Hansen, L., J. Heaton and A. Yaron, 1996, Finite-sample Properties of Some Alternative GMM Estimators. Journal of Business and Economic Statistics 14, 262–280.

[26] Hausman, J.A., 1978, Specification Tests in Econometrics. Econometrica 46, 1251–1271.

[27] Hausman, J., R. Lewis, K. Menzel and W. Newey, 2011, Properties of the CUE Estimator and a Modification with Moments. Journal of Econometrics 165, 45–57.

[28] Hayakawa, K., 2015, The Asymptotic Properties of the System GMM Estimator in Dynamic Panel Data Models when Both N and T are Large. Econometric Theory 31: 647–667.

[29] Hovakimian, G. and S. Titman, 2006, Corporate Investment with Financial Constraints: Sensitivity of Investment to Funds from Voluntary Asset Sales. Journal of Money, Credit, and Banking 38, 357-374.

[30] Hsiao, C., 2003, Analysis of Panel Data. Cambridge: Cambridge University Press.

[31] Hsiao, C., M.H. Pesaran and K. Tahmiscioglu, 2002, Maximum Likelihood Estimation of Fixed Effects Dynamic Panel Data Models Covering Short Time Periods. Journal of Econometrics 109, 107-150.

[32] Hsiao, C. and J. Zhang, 2015, IV, GMM or Likelihood Approach to Estimate Dynamic Panel Models when either N or T or both are Large. Journal of Econometrics 187, 312–322.

[33] Holtz-Eakin, D., W.K. Newey and H.S. Rosen, 1988, Estimating Vector Autoregressions with Panel Data. Econometrica 56, 1371–1395

[34] Jensen, M., 1986, Agency Costs of Free Cash Flow, Corporate Finance and Takeovers. American Economic Review 76, 323-339.

[35] Kapetanios, G., 2010, Testing for Exogeneity in Threshold Models. Econometric Theory 26, 231-259.

[36] Kaplan, S. and L. Zingales, 1997, Do Financing Constraints Explain Why Investment is Correlated with Cash Flow?" Quarterly Journal of Economics 112, 169-216.

[37] Kim, C.J. and J. Piger and R. Startz, 2008, Estimation of Markov Regime-switching Regression Models with Endogenous Switching. Journal of Econometrics 143, 263-273.

[38] Kourtellos, A., T. Stengos and C.M. Tan, 2015, Structural Threshold Regression. forthcoming in Econometric Theory.

[39] Kremer, S., A. Bick and D. Nautz, 2013, Inflation and Growth: New Evidence from a Dynamic Panel Threshold Analysis. Empirical Economics 44, 861-878.

[40] Lang, L., E. Ofek and R.M. Stulz, 1996, Leverage, Investment, and Firm Growth. Journal of Financial Economics 40, 3-29.

[41] Lee, S., M.H. Seo, and Y. Shin, 2011, Testing for Threshold Effects in Regression Models. Journal of the American Statistical Association 106, 220-231.

[42] Newey, W. and D.L. McFadden, 1994, Large Sample Estimation and Hypothesis Testing. in Handbook of Econometrics IV, eds by R.F. Engle and D.L. McFadden, 2111-2245, Elsevier.

[43] Nickell, S., 1981, Biases in Dynamic Models with Fixed Effects. Econometrica 49, 1417-1426.

[44] Pesaran, M.H., 2006, Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure. Econometrica 74, 967-1012.

[45] Seo, M.H. and O. Linton, 2007, A Smoothed Least Squares Estimator for Threshold Regression Models. Journal of Econometrics 141, 704-735.

[46] Sun, Y., 2014, Fixed-smoothing Asymptotics in a Two-step Generalized Method of Moments Framework. Econometrica 82: 2327–2370.

[47] Tong, H., 1990, Nonlinear Time Series: A Dynamical System Approach. Oxford: Oxford University Press.

[48] van der Vaart, A.W. and J.A. Wellner, 1996, Weak Convergence and Empirical Process. New York: Springer.

[49] Yu, P., 2013, Inconsistency of 2SLS Estimators in Threshold Regression with Endogeneity. Economics Letters 120, 532-536.

[50] Yu, P. and P.C.B. Phillips, 2014, Threshold Regression with Endogeneity. Cowels Foundation Discussion Paper No. 1966.

[51] Whited, T.M and G. Wu, 2006, Financial Constraints Risk. Review of Financial Studies 19, 531-559.

[52] Zilak, J., 1997, Efficient Estimation with Panel Data When Instruments Are Predetermined: An Empirical Comparison of Moment-Condition Estimators. Journal of Business and Economic Statistics 15, 419-431.

Table 1: MSE of FD-GMM estimators

		FD-GMM				Averaging			
DGP	n	γ	β	δ_1	δ_2	γ	β	δ_1	δ_2
Jump	50	0.063	0.077	0.179	0.498	0.115	0.096	0.185	0.566
	100	0.089	0.075	0.207	0.600	0.087	0.066	0.172	0.517
	200	0.066	0.068	0.174	0.536	0.067	0.056	0.144	0.474
Cont.	50	0.077	0.320	0.588	0.863	0.009	0.112	0.292	0.273
	100	0.079	0.383	0.677	1.002	0.041	0.203	0.439	0.591
	200	0.083	0.383	0.662	0.963	0.060	0.289	0.542	0.743

Table 2: Bias of FD-GMM estimators

		FD-GMM				Averaging			
DGP	n	γ	β	δ_1	δ_2	γ	β	δ_1	δ_2
Jump	50	−0.041	0.005	−0.044	0.100	−0.269	0.199	−0.151	−0.390
	100	−0.047	0.007	−0.044	0.095	−0.106	0.073	−0.070	−0.093
	200	−0.029	−0.011	−0.018	0.098	−0.060	0.016	−0.034	0.033
Cont.	50	0.057	0.180	-0.288	0.184	0.055	0.105	-0.198	0.163
	100	0.064	0.145	-0.271	0.199	0.057	0.099	-0.231	0.210
	200	0.074	0.190	-0.298	0.162	0.067	0.158	-0.270	0.170

Table 3: Standard Error of FD-GMM estimators

		FD-GMM				Averaging			
DGP	n	γ	β	δ_1	δ_2	γ	β	δ_1	δ_2
Jump	50	0.247	0.277	0.421	0.699	0.207	0.238	0.402	0.644
	100	0.294	0.273	0.452	0.769	0.275	0.246	0.409	0.713
	200	0.255	0.261	0.417	0.726	0.252	0.236	0.377	0.688
Cont.	50	0.272	0.537	0.711	0.911	0.080	0.317	0.503	0.497
	100	0.274	0.601	0.777	0.981	0.194	0.440	0.621	0.739
	200	0.279	0.589	0.757	0.968	0.236	0.514	0.685	0.845

Table 4: MSE of FD-GMM estimators (restricted)

		FD-GMM			Averaging		
DGP	n	γ	β	δ	γ	β	δ
Jump	50	0.105	0.102	0.124	0.050	0.095	0.132
	100	0.106	0.116	0.142	0.075	0.097	0.122
	200	0.095	0.080	0.102	0.076	0.070	0.088
Cont.	50	0.033	0.075	0.155	0.019	0.067	0.143
	100	0.039	0.094	0.192	0.030	0.085	0.177
	200	0.039	0.082	0.170	0.034	0.080	0.168

Table 5: Bias of FD-GMM estimators (restricted)

		FD-GMM			Averaging		
DGP	n	γ	β	δ	γ	β	δ
Jump	50	0.009	0.051	-0.008	-0.029	-0.082	0.143
	100	0.012	0.064	-0.047	0.021	0.031	-0.010
	200	0.028	0.052	-0.047	0.025	0.041	-0.035
Cont.	50	0.013	-0.049	0.103	0.092	-0.008	0.038
	100	0.021	-0.081	0.144	0.052	-0.053	0.098
	200	0.014	-0.064	0.116	0.028	-0.051	0.094

Table 6: Standard Error of FD-GMM estimators (restricted)

		FD-GMM			Averaging		
DGP	n	γ	β	δ	γ	β	δ
Jump	50	0.324	0.315	0.352	0.222	0.297	0.335
	100	0.325	0.334	0.374	0.273	0.310	0.350
	200	0.307	0.278	0.316	0.275	0.261	0.295
Cont.	50	0.182	0.270	0.380	0.102	0.259	0.376
	100	0.196	0.295	0.414	0.164	0.286	0.409
	200	0.197	0.279	0.396	0.183	0.278	0.399

Table 7: Coverage Frequency of FD-GMM estimators

			FD-GMM				Averaging			
DGP	h	n	γ	β	δ_1	δ_2	γ	β	δ_1	δ_2
Jump	1/2	50	0.876	0.731	0.736	0.647	0.878	0.641	0.753	0.705
		100	0.931	0.895	0.897	0.847	0.942	0.884	0.907	0.871
		200	0.937	0.917	0.950	0.897	0.939	0.930	0.956	0.914
	1	50	0.960	0.857	0.886	0.716	0.995	0.821	0.894	0.778
		100	0.978	0.962	0.971	0.899	0.991	0.946	0.973	0.928
		200	0.979	0.963	0.967	0.933	0.983	0.969	0.979	0.947
	3/2	50	0.986	0.882	0.928	0.814	1.000	0.805	0.910	0.867
		100	0.995	0.968	0.977	0.936	0.998	0.969	0.980	0.954
		200	1.000	0.971	0.982	0.971	1.000	0.973	0.985	0.974
Cont.	1/2	50	0.427	0.473	0.621	0.518	0.904	0.700	0.744	0.694
		100	0.525	0.716	0.804	0.698	0.772	0.819	0.857	0.798
		200	0.585	0.796	0.894	0.798	0.691	0.847	0.926	0.839
	1	50	0.811	0.592	0.745	0.624	0.990	0.780	0.871	0.799
		100	0.898	0.795	0.916	0.806	0.980	0.881	0.947	0.876
		200	0.900	0.862	0.947	0.868	0.947	0.905	0.965	0.894
	3/2	50	0.965	0.680	0.810	0.669	0.999	0.847	0.904	0.865
		100	0.997	0.892	0.944	0.843	1.000	0.937	0.970	0.916
		200	1.000	0.917	0.969	0.889	1.000	0.941	0.980	0.914

Note: These are empirical coverage frequencies of 95% nominal confidence intervals. The bandwidth for the asymptotic variance estimation in equation (12) is selected by h times Silverman's rule of thumb.

Table 8: A dynamic threshold panel data model of investment

$\mathbf{x}_{it} \setminus q_{it}$	Cash Flow	-Leverage	Tobin Q
	Lower Regime ($\boldsymbol{\phi}_1$)		
I_{-1}	0.580 (0.132)	0.590 (0.123)	0.382 (0.226)
CF	0.245 (0.121)	0.600 (0.118)	−0.044 (0.209)
Q	−0.017 (0.016)	−0.013 (0.014)	0.368 (0.173)
L	−0.128 (0.049)	−0.029 (0.087)	−0.386 (0.184)
	Upper Regime ($\boldsymbol{\phi}_2$)		
I_{-1}	−0.215 (0.480)	0.253 (0.158)	0.365 (0.142)
CF	0.012 (0.128)	−0.043 (0.146)	0.217 (0.084)
Q	0.028 (0.021)	0.021 (0.014)	−0.031 (0.010)
L	0.825 (0.195)	2.968 (0.725)	0.194 (0.095)
	Difference ($\boldsymbol{\delta}$)		
I_{-1}	−0.796 (0.561)	−0.336 (0.439)	−0.016 (0.325)
CF	−0.233 (0.154)	−0.643 (0.203)	0.261 (0.264)
Q	0.045 (0.035)	0.034 (0.024)	−0.401 (0.175)
L	0.953 (0.207)	2.998 (0.745)	0.581 (0.147)
Threshold	0.358 (0.039)	0.100 (0.033)	0.561 (0.244)
Upper Regime (%)	19.4	73.6	58.9
Linearity (p-value)	0.0	0.0	0.0
J-test (p-value)	60.1 (0.004)	33.3 (0.185)	45.4 (0.091)
No. of IVs	36	36	43

Table 9: Efficiency Comparison of FD-GMM and FD-2SLS Estimators

		FD-GMM		FD-2SLS	
	n	bias	$\ln(RMSE)$	bias	$\ln(RMSE)$
DGP 1	50	-0.002	-2.6	0.002	-4.7
	100	0.003	-2.6	0.001	-4.9
	200	0.002	-2.7	0.0	-5.0
DGP 2	50	0.007	-1.7	0.009	-3.2
	100	-0.001	-1.8	0.003	-4.0
	200	-0.006	-1.9	0.002	-4.6