Dynamic Panels with Threshold Effect and Endogeneity*

Myung Hwan Seo
London School of Economics

Yongcheol Shin
University of York

Revised, March 2014

Abstract

This paper addresses a challenging issue as how best to model nonlinear asymmetric dynamics and cross-sectional heterogeneity, simultaneously, in the dynamic threshold panel data framework, under which both threshold variable and regressors are allowed to be endogenous. Depending on whether the threshold variable is strictly exogenous or not, we propose two different estimation methods: the first-differenced two-step least squares and the first-differenced GMM. The former exploits the fact that the threshold variable is strictly exogenous to achieve the super-consistency of the threshold estimator. We provide asymptotic distributions of the two estimators with and without exogeneity of the threshold variable, respectively. The bootstrap-based testing procedure for the presence of threshold effect is also developed. Monte Carlo studies provide a support for our theoretical predictions. Finally, using the UK and the US company panel data, we provide two empirical applications investigating an asymmetric sensitivity of investment to cash flows and an asymmetric dividend smoothing.

JEL Classification: C13, C33, G31, G35

Key Words: Dynamic Panel Threshold Models, Endogenous Threshold Effects and Regressors, FD-GMM and FD-2SLS Estimation, Investment and Dividend Smoothing.

*We are grateful to Mehmet Caner, Yoosoon Chang, Jinseo Cho, Viet Anh Dang, Jinwook Jeong, Taehwan Kim, Matthew Greenwood-Nimmo, Myungjae Lee, Joon Park, Kevin Reilly, Laura Serlanga, seminar participants at Universities of Korea, Leeds, Sung Kyun Kwan and Yonsei, and conference delegates at Symposium on Econometric Theory and Applications, Kyoto University, August 2009, Computational and Financial Econometrics Annual Conference at University of London, December 2010, and Asia Meeting of Econometric Society, Korea University, Seoul, August 2011 for their helpful comments. We would also like to thank Minjoo Kim for excellent research assistance and helpful discussions. The second author acknowledges partial financial support from the ESRC (Grant No. RES-000-22-3161). The usual disclaimer applies.

1 Introduction

In the time series literature there have been many studies that examine the implications of the existence of a particular kind of nonlinear asymmetric dynamics. Examples are 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) generalise 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 data approach that allows coefficients to vary randomly over time and across cross-sectional 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 also prompted more rigorous econometric analyses of dynamic heterogeneous panels. Surprisingly, however, 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., Arellano and Bond (1991), Ahn and Schmidt (1995), Arellano and Bover (1995), Blundell and Bond (1998), Alvarez and Arellano (2003), and Hayakawa (2012).

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), not much progress has 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 still 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 recently proposed the generalised GMM estimator applicable for dynamic panel threshold models, which can provide consistent estimates of heterogeneous speeds of adjustment as well as a valid testing procedure for thresh-

[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.

old 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. We note in passing that the endogeneity in the threshold variable can occur in various examples, e.g., Kourtellos et al. (2009) and Yu (2013). Furthermore, in dynamic panels, it may naturally arise as a result of the first-difference (FD) transformation applied to remove unobserved heterogeneity even if the original threshold variable were exogenous.

We aim to fill this gap by explicitly addressing an important issue as how best to model nonlinear asymmetric dynamics and cross-sectional 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 regressors and threshold variable, and develop the general estimation and inference theory. We propose the two estimation methods based on the FD transformation, and evaluate their properties by the diminishing threshold effect asymptotics of Hansen (2000). Our approach is expected to avoid any sample selection bias problem (Hansen, 2000) associated with any arbitrary sample-splitting approach. More importantly, it will overcome the main limitedness in the existing literature, namely, the maintained assumption of exogeneity of regressors and\or the transition variable that may hamper the usefulness of threshold regression models in a general context.

As a general approach, we first develop the FD-GMM method in which the (time-varying) threshold variable can be endogenous or weakly exogenous. Next, considering that the least squares estimator is Oracle efficient in the standard regression, we also propose a more efficient two-step least squares (FD-2SLS) estimator in the special case where the threshold variable is strictly exogenous, but regressors are still allowed to be endogenous. The FD-2SLS approach generalizes the Caner and Hansen's the cross-section estimation to the dynamic panel data modelling. Furthermore, we identify cases where our FD-2SLS is able to estimate unknown parameters more efficiently than their method.

We develop the asymptotic theory for both the FD-GMM and the FD-2SLS estimators. First of all, the FD-GMM estimator is shown to be asymptotically normal. Thus, the standard inference based on the t- or Wald statistic is feasible, though the convergence rate is slower than $\sqrt{n}$, depending on an unknown quantity under the diminishing threshold framework.

Importantly, here, the asymptotic normality holds true irrespective of whether the regression function is continuous or not. This is in contrast to the least squares approach, where the discontinuity of the regression function changes the asymptotic distribution in a dramatic way. Hence, inference on the threshold parameter can be carried out in the standard manner. Next, we establish that the FD-2SLS estimator satisfies the oracle property where the threshold estimate and the slope estimate are asymptotically independent under the assumption of strict exogeneity of the threshold variable. We allow for general continuous or discontinuous non-linear regression models for the reduced form, and provide the corrected asymptotic variance formula for the estimator of the slope parameters. Although the FD-2SLS estimator of the threshold parameter is shown to be super-consistent, its inference is non-standard but can be easily conducted by inverting a properly weighted LR statistic, which follows a known pivotal asymptotic distribution (Hansen, 2000).

We also provide formal testing procedures for identifying the threshold effect. They are based on the supremum type statistics and they have non-standard asymptotic distributions due to the loss of identification under the null hypothesis of no threshold effect. But, the critical values or the p-values of the tests can be easily evaluated by the bootstrap.

Monte Carlo studies show that the overall simulation results, focusing on the bias, standard error, and mean square error of the two-step FD-GMM estimator, provide support for our theoretical predictions. Given that there are many different ways to compute the weight matrix in the first step, we also propose to consider an averaging of a class of the two-step FD-GMM estimators that are obtained by randomising the weight matrix in the first step. The averaging turns out to be quite successful in reducing the sampling errors, so that we recommend the use of the averaging method in practice, even in the other types of non-linear models applying the GMM techniques.

Using the UK and the US company panel data, we demonstrate the usefulness of the proposed dynamic threshold panel data modelling by providing two empirical applications investigating an asymmetric sensitivity of investment to cash flows and an asymmetric dividend smoothing. In the first application we employ a balanced panel dataset of 560 UK firms over the period 1973-1987, and 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. In the second application with the balanced panel dataset of 246 US firms over the period 1990 - 2001, we find that dividend smoothing is relatively stronger for firms that tend to pay the higher (target) dividend payout especially in the long-term perspective, a finding generally consistent with the survey evidence in Brav *et* *al.* (2005) and the micro empirical evidence in Leary and Michaely (2011).

The plan of the paper is as follows: Section 2 discusses the model set-up. Section 3 describes the detailed estimation steps for FD-GMM and FD-2SLS. Section 4 develops an asymptotic theory for both estimators, including consistent and efficient estimation of threshold parameter. Section 5 provides the bootstrap-based inference for threshold effects. Finite sample performance of the FD-GMM estimator is examined in Section 6. Two empirical applications are presented in Section 7. Section 8 concludes. Mathematical proofs are collected in an Appendix.

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} is the $k_1 \times 1$ vector of time-varying regressors, that may include the lagged dependent variable, $1(\cdot)$ is an indicator function, and q_{it} is the transition variable. γ is the threshold parameter, and ϕ_1 and ϕ_2 the slope parameters associated with different regimes. The regression 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 $\mathcal{F}_{t-1}$, thus allowing endogeneity in both the regressor, x_{it} and the threshold variable, q_{it}. But, as will be shown, efficient estimation depends on whether q_{it} is exogenous or not. As we will consider the asymptotic experiment under large n with a fixed T, 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 we have

$$x_{it} = q_{it} = y_{i,t-1}.$$

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):

$$\delta = \delta_n = \delta_0 n^{-\alpha} \text{ for } 0 \leq \alpha < 1/2. \tag{3}$$

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

$$\Delta y_{it} = \beta' \Delta x_{it} + \delta' X_{it}' \mathbf{1}_{it}(\gamma) + \Delta\varepsilon_{it}, \tag{4}$$

where $\underset{k_1\times 1}{\beta} = \left(\phi_{12}, ..., \phi_{1,k_1+1}\right)'$, $\underset{(k_1+1)\times 1}{\delta} = \phi_2 - \phi_1$, 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 $k = 2k_1+2$. It is worthwhile to note that the transformed model, (4) consists of 4 regimes, which are generated by two threshold variables, q_{it} and q_{it-1}. This change in the model characteristic is relevant in inference with the least squares estimation as discussed in Section 4.2.

The OLS estimator obtained from (4) will be biased since the transformed regressors are now 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 $2 < t_0 \leq T$, such that

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

or, for each $t = t_0, ..., T$,

$$\mathrm{E}\left(\Delta\varepsilon_{it} | z_{it}\right) = 0. \tag{6}$$

Notice that z_{it} may include lagged values of (x_{it}, q_{it}) and lagged dependent variables if not included in x_{it} or q_{it} already. The number of instruments may be different for each time t.

3 Estimation

Depending upon whether q_{it} is endogenous or not and whether the conditional moment restriction (6) holds or not, we will develop different estimation methods.

3.1 FD-GMM

We allow for the threshold variable q_{it} to be endogenous, and develop a two-step GMM estimation. To this end we consider the $l \times 1$ vector of the sample moment conditions:

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

where

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

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

$$\bar{J}_n(\theta) = \bar{g}_n(\theta)' W_n \bar{g}_n(\theta). \tag{2}$$

Then, the GMM estimator of θ is given by

$$\hat{\theta} = \arg\min_{\theta\in\Theta} \; \bar{J}_n(\theta). \tag{3}$$

Since the model is linear in β and δ for each γ and the objective function $\bar{J}_n(\theta)$ is not continuous in γ, the grid search algorithm is more practical. 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

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

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

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

The asymptotic property of the GMM estimator, $\hat{\gamma}$, which will be presented in Section 4, is different from the conventional least squares estimator, e.g., Chan (1993) and Hansen (2000).

The two-step optimal GMM estimator is obtained as follows: first, estimate the model by minimising $\bar{J}_n(\theta)$ with $W_n = I_l$ or

$$W_n = \begin{pmatrix} \frac{2}{n}\sum_{i=1}^n z_{it_0}z_{it_0}' & \frac{-1}{n}\sum_{i=1}^n z_{it_0}z_{it_0+1}' & 0 & \cdots \\ \frac{-1}{n}\sum_{i=1}^n z_{it_0+1}z_{it_0}' & \frac{2}{n}\sum_{i=1}^n z_{it_0+1}z_{it_0+1}' & \ddots & \ddots \\ 0 & \ddots & \ddots & \frac{-1}{n}\sum_{i=1}^n z_{iT-1}z_{iT}' \\ \vdots & \ddots & \frac{-1}{n}\sum_{i=1}^n z_{iT}z_{iT-1}' & \frac{2}{n}\sum_{i=1}^n z_{iT}z_{iT}' \end{pmatrix}^{-1} \tag{4}$$

and collect residuals, $\widehat{\Delta\varepsilon}_{it}$; second, re-estimate the parameter θ by minimising $\bar{J}_n(\theta)$ with

$$W_n = \left(\frac{1}{n}\sum_{i=1}^n \hat{g}_i\hat{g}_i' - \frac{1}{n^2}\sum_{i=1}^n \hat{g}_i\sum_{i=1}^n \hat{g}_i'\right)^{-1}, \tag{5}$$

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

3.2 FD-2SLS

This subsection considers the case where the threshold variables, q_{it} and $q_{i,t-1}$ in (4), are exogenous and the conditional moment restriction, (6) holds. That is, z_{it} includes q_{it} and $q_{i,t-1}$. In this case, we can improve upon the GMM estimator presented above. In particular, the threshold estimate, $\hat{\gamma}$ can achieve the efficient rate of convergence, as obtained in the classical regression model (e.g., Hansen, 2000), and the slope estimate, $\hat{\beta}$ and $\hat{\delta}$ can achieve the semi-parametric efficiency bound (Chamberlain, 1987) under homoskedasticity as if the true threshold value, γ_0, is known. This strong result can be obtained since the two sets of estimators are asymptotically independent. Our approach, while motivated by Caner and Hansen (2004), is different from their approach, as described below leaving aside the modification necessary due to the first-difference transformation.

The first-differenced model, (4) with the conditional moment condition (6) and the exogeneity of q, implies the following regression of Δy_{it} on z_{it}:

$$\mathrm{E}(\Delta y_{it}|z_{it}) = \beta'\mathrm{E}(\Delta x_{it}|z_{it}) + \delta'\mathrm{E}\left(X_{it}'|z_{it}\right)\mathbf{1}_{it}(\gamma). \tag{6}$$

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

$$\mathrm{E}\begin{pmatrix} 1, x_{it}' \\ 1, x_{it-1}' \end{pmatrix}|z_{it}\Bigg) = \begin{pmatrix} 1, F_{1t}'(z_{it}; b_{1t}) \\ 1, F_{2t}'(z_{it}; b_{2t}) \end{pmatrix} = \underset{2\times(1+k_1)}{F_t}(z_{it}; b_t), \tag{7}$$

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

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

For instance, Caner and Hansen (2004) considered the linear regression and the threshold regression for F_t.

Note that there are two regressions for x_{it} due to the first difference transformation and the possibility that z_{it} varies over time. Furthermore, it is not sufficient to consider the regression $\mathrm{E}(\Delta x_{it}|z_{it})$ only, due to the threshold effect in the structural form, (6).

The above representation in (6) and (7) motivates a two-step estimation procedure:

1. For each t, estimate the reduced form, (7) by the least squares, and obtain the parameter estimates, $\hat{b}_t,\ t = t_0, ..., T$, and the fitted values, $\hat{F}_{it} = F_t\left(z_{it}; \hat{b}_t\right)$.

2. Estimate θ by
$$\min_{\theta \in \Theta} \hat{\mathbb{M}}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}\sum_{t=t_0}^{T} e_{it}\left(\theta, \hat{b}_t\right)^2, \tag{8}$$
where
$$e_{it}(\theta, b_t) = \Delta y_{it} - \beta' H_t(z_{it}; b_t) - \delta' F_t(z_{it}; b_t)' \mathbf{1}_{it}(\gamma).$$
This step can be done by the grid search as the model is linear in β and δ for a fixed γ. Thus, $\hat{\beta}(\gamma)$ and $\hat{\delta}(\gamma)$ can be obtained from the pooled OLS of Δy_{it} on $\hat{H}_{it}$ and $\hat{F}_{it}'\mathbf{1}_{it}(\gamma)$, and $\hat{\gamma}$ is defined as the minimizer of the profiled sum of squares errors, $\hat{\mathbb{M}}_n(\gamma)$.

This procedure 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. The above two-step estimation yields the optimal estimate for β and δ provided that 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) under more general setup, it is reasonable to employ our two-step estimator and robustify the standard errors for the heteroskedasticity. We will provide a heteroskedasticity-robust standard error estimate for $\hat{\beta}$ and $\hat{\delta}$. Further, the standard error estimate is corrected for the estimation error in the first step estimation of b.

Remark 1 *We compare our procedure with Caner and Hansen (2004). As their work is not for the panel data, our comparison is based on the cross sectional regression framework and thus there is no first difference transformation. Their procedure consists of three steps. The first* *two steps are identical to ours. Then, they split the sample, $(y_i, x_i, z_i)_{i=1}^n$ into two according to whether q_i is greater than $\hat{\gamma}$ (obtained in step 2) or not, and they estimate ϕ_1 and ϕ_2 using each of two subsamples based on the standard GMM procedure for the linear regression. Their approach can be optimal under the following unconditional moment conditions:*

$$\mathrm{E}\begin{pmatrix} \varepsilon_i z_i 1\{q_i \leq \gamma\} \\ \varepsilon_i z_i 1\{q_i > \gamma\} \end{pmatrix} = 0. \tag{9}$$

However, this condition (9) *does not fully exploit the implications of conditional moment restriction* (6) *even under homoskedasticity.*

Remark 2 *As another way to appreciate the unconditional moment condition* (9), *observe that its reduced form is the threshold regression.*[2] *What is achieved through the three step estimation procedure is the special case, where* $\mathrm{E}(x_i|z_i)$ *is a threshold regression whose threshold variable and change point are the same as in the structural equation. However, two issues arise. First, we can estimate* γ *more efficiently. Consider the model for* $i = 1, ..., n$*:*

$$\begin{aligned} y_i &= \phi_1' x_i 1(q_i \leq \gamma) + \phi_2' x_i 1(q_i > \gamma) + \varepsilon_i, \\ x_i &= \Gamma_1' z_i 1(q_i \leq \gamma) + \Gamma_2' z_i 1(q_i > \gamma) + \eta_i, \end{aligned} \tag{10}$$

where $\mathrm{E}(\varepsilon_i, \eta_i'|z_i) = 0$*. Rewriting this system by substitution yields:*

$$y_i = \lambda_1' z_i 1(q_i \leq \gamma) + \lambda_2' z_i 1(q_i > \gamma) + e_i, \tag{11}$$

where $\mathrm{E}(e_i|z_i) = 0$*,* $\lambda_j = \Gamma_j \phi_j$*,* $j = 1, 2$*, and* $e_i = \varepsilon_i + \eta_i'(\phi_1 1(q_i \leq \gamma) + \phi_2 1(q_i > \gamma))$*. The transformed model has the same threshold value as in the original model, and we thus advocate the estimation of* γ *from the one-step regression,* (11) *rather than from the originally proposed two-step estimation. The regression,* (11) *is a standard exogenous threshold regression and the asymptotic distribution of the threshold estimate is provided by Hansen* (2000)*. Second, if the (true) reduced form follows* (10)*, the asymptotic distribution of* $\hat{\gamma}$ *provided in Caner and Hansen* (2004) *is not generally correct as the estimation error from the first step estimation of* γ *affects the second step estimation of* γ*.*[3] *Intuitively, the estimation error in the first step affects the second step estimation of* γ *since the true thresholds are restricted to be the same in both equations. The exact asymptotic characterisation is not of practical importance as we do not recommend the two-step estimation of* γ *in this case.*

[2] Note that the regression of $(x_i' 1\{q_i \leq \gamma\}, x_i' 1\{q_i > \gamma\})'$ on $(z_i' 1\{q_i \leq \gamma\}, z_i' 1\{q_i > \gamma\})'$ is equivalent to the regression of x_i on $(z_i' 1\{q_i \leq \gamma\}, z_i' 1\{q_i > \gamma\})'$.

[3] Lemma 1 in Caner and Hansen (2004) would be true with more restrictions. More specifically, their (A.7) is true only when the threshold estimate is n-consistent, which is not the case in the maintained diminishing threshold parameter setup. Accordingly, the high-level assumption (17) in their Assumption 2 is no longer satisfied.

3.2.1 Threshold Regression in Reduced Form

Motivated by the preceding remarks, we propose another estimation procedure for a special case of (7) as follows:

$$\begin{aligned}
\begin{pmatrix} x_{it} \\ x_{it-1} \end{pmatrix} &= \begin{pmatrix} \Gamma_{1t}z_{it}1\{q_{it} \leq \gamma\} + \Gamma_{2t}z_{it}1\{q_{it} > \gamma\} + \eta_{1,it} \\ \Gamma_{3t}z_{it}1\{q_{it-1} \leq \gamma\} + \Gamma_{4t}z_{it}1\{q_{it-1} > \gamma\} + \eta_{2,it} \end{pmatrix} \\
\mathrm{E}\begin{pmatrix} \eta_{1,it} & |z_{it} \\ \eta_{2,it} & \end{pmatrix} &= 0. \qquad (12)
\end{aligned}$$

By substitution this yields a 4-regime threshold regression model:

$$\begin{aligned}
\Delta y_{it} &= \Lambda_{1t}'z_{it}1\{q_{it} \leq \gamma\} + \Lambda_{2t}'z_{it}1\{q_{it} > \gamma\} \\
&\quad + \Lambda_{3t}'z_{it}1\{q_{it-1} \leq \gamma\} + \Lambda_{4t}'z_{it}1\{q_{it-1} > \gamma\} + e_{it}, \qquad (13)
\end{aligned}$$

where $\mathrm{E}(e_{it}|z_{it}) = 0$, $\Lambda_{1t} = \Gamma_{1t}'\beta$, $\Lambda_{2t} = (\Gamma_{2t}'\beta + (1, \Gamma_{2t}')\delta)$, $\Lambda_{3t} = -\Gamma_{3t}'\beta$, $\Lambda_{4t} = -(\Gamma_{4t}'\beta + (1, \Gamma_{4t}')\delta)$, and

$$e_{it} = \Delta\varepsilon_{it} + \left(0, \mathbf{1}_{it}(\gamma)' \eta_{it}'\right)\phi_1 + \left(0, (\iota_2 - \mathbf{1}_{it}(\gamma))' \eta_{it}'\right)\phi_2,$$

with $\eta_{it} = \left(\eta_{1,it}, \eta_{2,it}\right)$ and $\iota_2 = (1, -1)'$. Thus, our proposal is:

1. Estimate γ by the pooled least square of (13), which can be done by the grid search,[4] and denote the estimate by $\tilde{\gamma}$.

2. Fix γ at $\tilde{\gamma}$ and estimate Γ_{jt}, $j = 1, ..., 4$, in (12) by OLS, for each t.

3. Estimate β and δ in (6) by OLS with γ and the reduced form parameters fixed at the estimates obtained from the preceding steps. Denote these estimates by $\tilde{\beta}$ and $\tilde{\delta}$ for later reference.

4 Asymptotic Distributions

This section develops an asymptotic theory for the FD-GMM and FD-2SLS estimators. There are two frameworks in the literature. One is Hansen's (2000) diminishing threshold assumption and the other is fixed threshold assumption as in Chan (1993). For the GMM estimator we present an asymptotics that accommodates both setups and for the 2SLS we develop the asymptotic distribution only under the diminishing threshold framework. 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.

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

4.1 FD-GMM

Partition $\theta = \left(\theta_1', \gamma\right)'$ with $\theta_1 = \left(\beta', \delta'\right)'$. 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:

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

and

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

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

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

This is a standard assumption for the threshold regression model as in Hansen (2000).

Assumption 2 (i) *The threshold variable, q_{it} has a continuous and bounded density, p_t, such that $p_t\left(\gamma_0\right) > 0$, 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 $\mathrm{E}_t\left(\cdot|\gamma\right) = \mathrm{E}\left(\cdot|q_{it} = \gamma\right)$ and $\mathrm{E}_t\left(z_{it}\left(x_{it}', x_{i,t-1}'\right)|\gamma\right)\delta_0 \neq 0$ for some t.*

The smoothness assumption on the distribution of the threshold variable and some conditional moments are standard. However, we do not require the discontinuity of the regression function 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.

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

This is a standard rank condition in GMM. Then, we have:

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

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

The asymptotic variance matrix contains δ_0, and the convergence rate of $\hat{\gamma}$ 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

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

where $\hat{g}_i = g_i \left(\hat{\theta} \right)$ and

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

Then, G_γ may be estimated by the standard Nadaraya-Watson kernel estimator: that is, for some kernel, K and bandwidth h, such as the Gaussian kernel and Silverman's rule of thumb (e.g., Hardle and Linton (1994) for more discussions on the choice of kernel and the bandwidth), let:

$$\hat{G}_\gamma = \left[\begin{array}{c} \frac{1}{nh} \sum_{i=1}^{n} z_{it_0} \left[\left(1, x_{it_0-1}\right)' K\left(\frac{\hat{\gamma} - q_{it_0-1}}{h} \right) - \left(1, x_{it_0}\right)' K\left(\frac{\hat{\gamma} - q_{it_0}}{h} \right) \right] \hat{\delta} \\ \vdots \\ \frac{1}{nh} \sum_{i=1}^{n} z_{iT} \left[\left(1, x_{iT-1}\right)' K\left(\frac{\hat{\gamma} - q_{iT-1}}{h} \right) - \left(1, x_{iT}\right)' K\left(\frac{\hat{\gamma} - q_{iT}}{h} \right) \right] \hat{\delta} \end{array} \right]. \tag{14}$$

Furthermore, let $\hat{V}_s = \hat{\Omega}^{-1/2} \left(\hat{G}_\beta, \hat{G}_\delta \right)$ and $\hat{V}_\gamma = \hat{\Omega}^{-1/2} \hat{G}_\gamma$. Then, the asymptotic variance-covariance matrix for the regression coefficient, θ_1 can be consistently estimated by

$$\left(\hat{V}_s' \hat{V}_s - \hat{V}_s' \hat{V}_\gamma \left(\hat{V}_\gamma' \hat{V}_\gamma \right)^{-1} \hat{V}_\gamma' \hat{V}_s \right)^{-1},$$

while the t-statistic for $\gamma = \gamma_0$ defined by

$$\left(\hat{V}_\gamma' \hat{V}_\gamma - \hat{V}_\gamma' \hat{V}_s \left(\hat{V}_s' \hat{V}_s \right)^{-1} \hat{V}_s' \hat{V}_\gamma \right) \sqrt{n} \left(\hat{\gamma} - \gamma_0 \right),$$

converges to the standard normal distribution. Therefore, the confidence intervals can be constructed as in the standard GMM case.

Alternatively, the standard nonparametric bootstrap, which resamples across i with replacement, can be employed to construct the confidence intervals, see Section 5 for details.

4.2 FD-2SLS

This section presents the asymptotic theory for the 2SLS estimator of θ. Here a few technical issues arise such as the multiple threshold variables as a result of the first difference transformation. We begin with the case where the reduced form is the regular nonlinear regression and the reduced form parameter estimates are asymptotically normal. Next, we consider the case where the reduced form is the threshold regression.

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. Thus, 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, $t = t_0, ..., T$. Accordingly, denote the collection of the corresponding elements of x_{it}'s by $\dot{x}_i$. Then, the reduced form regression can be written as:

$$\dot{x}_i = f\left(z_i, b\right) + \eta_i \text{ with } \mathrm{E}\left(\eta_i|z_i\right) = 0. \tag{15}$$

Let $\hat{b}$ denote the least squares estimate. We now follow the convention that $f_i\left(b\right) = f\left(z_i, b\right)$, $f_i = f\left(z_i, b_0\right)$, $\hat{f}_i = f\left(z_i, \hat{b}\right)$, *etc.*, where b_0 indicates the true value of b, when there is no confusion. We consider the two cases. The first case is where $\hat{b}$ is asymptotically normal and the second is the threshold regression.

4.2.1 Linearisable Reduced Form

This section considers the reduced forms, which allow for stochastic linearisation and thus the asymptotic normality of reduced form parameter estimates. We assume the asymptotic normality of $\hat{b}$, and the existence of a matrix-valued influence function, $\dot{f}$ below. More primitive conditions to yield this asymptotic normality of $\hat{b}$ are provided **in the Appendix. {yc: where??}** Notice that $|A|$ denotes the Euclidean norm if A is a vector, and the vector induced norm if A is a matrix.

Assumption 4 *There exists a matix-valued function, $\dot{f}\left(z_i, b\right)$ such that $\mathrm{E}\left|\dot{f}_i\right|^2 < \infty$ for some $a > 0$ and*

$$\sqrt{n}\left(\hat{b} - b_0\right) = \left(\mathrm{E}\dot{f}_i\dot{f}_i'\right)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\dot{f}_i\eta_i + o_p\left(1\right).$$

We begin with this high-level assumption because our main goal is to illustrate how the estimation error in the first step affects the asymptotic distribution of the estimator of the regression parameters, β and δ and of the threshold parameter, γ in the second step. We introduce some more notations. Let:

$$\underset{(2k_1+1)\times(T-t_0+1)}{\Xi_i(\gamma,b)} = \left(\Xi_{it_0}(\gamma,b_{t_0}),...,\Xi_{iT}(\gamma,b_T)\right),$$

where

$$\underset{(2k_1+1)\times 1}{\Xi_{it}(\gamma,b_t)} = \begin{bmatrix} H_{it}(b_t) \\ F_{it}(b_t)'\mathbf{1}_{it}(\gamma) \end{bmatrix}.$$

Also, 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

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

For $\hat{\gamma}$, introduce:

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

Following the notational convention, 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, for which we write $\delta_n = \delta_0 n^{-\alpha}$ for some $0<\alpha<1/2$ and $\delta_0 \neq 0$.*

If $\alpha = 0$, the asymptotic distribution for $\hat{\gamma}$ is different from the one obtained here. However, the convergence rate result in the proof of the theorem is still valid.

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 $\left(\left(1, F'_{2,it}\right)\delta_0\right)^2$.*

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

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 for which we define the LR statistic as

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

Then, we present the main asymptotic results for the 2SLS estimator and the LR statistic in the following Theorem:

Theorem 2 *Let Assumptions 5-8 hold. Then,*

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

and

$$n^{1-2\alpha}\frac{M_2^2}{V_2}(\hat{\gamma} - \gamma_0) \xrightarrow{d} \underset{r\in\mathbb{R}}{\operatorname{argmin}}\left(\frac{|r|}{2} - W(r)\right), \tag{16}$$

where $W(r)$ is a two-sided standard Brownian motion and it is independent of the normal variate in the first limit. Furthermore,

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

where $\sigma_e^2 = \mathrm{E}\left(e_{it}^2\right)$.

The first step estimation error contributes the asymptotic variance of $\hat{\beta}$ and $\hat{\delta}$ through Ω, while it has no effect on the asymptotic distribution of $\hat{\gamma}$. Estimation of the asymptotic variances of $\hat{\beta}$ and $\hat{\delta}$ is standard, i.e., the same as in the linear regression due to the asymptotic independence. The asymptotic distribution for $\hat{\gamma}$ in (16) is symmetric around zero and has a known distribution function,

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

for $x \geq 0$, where Φ is the standard normal distribution function. see Bhattacharya and Brockwell (1976). The unknown norming factor, $n^{2\alpha}V_2^{-1}M_2^2$ can be estimated by $\hat{V}_2^{-1}\hat{M}_2^2$, where

$$\begin{aligned}
\hat{M}_2 &= \sum_{t=t_0}^{T}\frac{1}{nh}\sum_{i=1}^{n}\left[\left(\left(1,\hat{F}_{1,it}'\right)\hat{\delta}\right)^2 k\left(\frac{q_{it}-\hat{\gamma}}{h}\right)+\left(\left(1,\hat{F}_{2,it}'\right)\hat{\delta}\right)^2 k\left(\frac{q_{it-1}-\hat{\gamma}}{h}\right)\right],\\
\hat{V}_2 &= \sum_{t=t_0}^{T}\frac{1}{nh}\sum_{i=1}^{n}\left(\left(\hat{e}_{it}\left(1,\hat{F}_{1,it}'\right)\hat{\delta}\right)^2 k\left(\frac{q_{it}-\hat{\gamma}}{h}\right)+\left(\hat{e}_{it}\left(1,\hat{F}_{2,it}'\right)\hat{\delta}\right)^2 k\left(\frac{q_{it-1}-\hat{\gamma}}{h}\right)\right)\\
&\quad +2\sum_{t=t_0}^{T-1}\frac{1}{nh}\sum_{i=1}^{n}\hat{e}_{it}\hat{e}_{it+1}\left(1,\hat{F}_{1,it}'\right)\hat{\delta}\left(1,\hat{F}_{2,it+1}'\right)\hat{\delta}k\left(\frac{q_{it}-\hat{\gamma}}{h}\right).
\end{aligned}$$

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

4.2.2 Threshold Regression in Reduced Form

Now, consider the case where the reduced form is a threshold regression as described in (12). The estimator, $\hat{\theta}$ is obtained from the three-step procedure following (12). Despite the difference in the estimation procedure, the asymptotic distributions of $\hat{\theta}$ can be presented by a slight modification of Theorem 2.

Corollary 3 *Let Assumptions 1, 2, and 10-8 hold. Furthermore, assume that* $\Lambda_1-\Lambda_2=n^{-\alpha}\delta_1$ *and* $\Lambda_3-\Lambda_4=n^{-\alpha}\delta_2$. *Define* M_1, M_2, V_1 *and* V_2 *as in Theorem 2, where* $\dot{f}_{it}$ *in* V_2 *is given by*

$$\begin{aligned}
\dot{f}_{1,it} &= \left(z_{it}1\left\{q_{it}\leq\gamma_0\right\},z_{it}1\left\{q_{it}>\gamma_0\right\}\right)\otimes I_{k_1}\\
\dot{f}_{2,it} &= \left(z_{it}1\left\{q_{it-1}\leq\gamma_0\right\},z_{it}1\left\{q_{it-1}>\gamma_0\right\}\right)\otimes I_{k_1},
\end{aligned}$$

and $\delta_0'F_{1,it}$ *and* $\delta_0'F_{2,it}$ *in* M_2 *and* V_2 *are replaced with* $\delta_1'z_{it}$ *and* $\delta_2'z_{it}$, *respectively. Then, the asymptotic distribution of* $\tilde{\theta}$ *is the same as in Theorem 2.*

5 Testing for Linearity

The preceding asymptotic results provide ways to make inference for unknown parameters and their functions. However, it is well-established that the test for threshold effects requires us to develop the different asymptotic theory due to the presence of unidentified parameters under the null hypothesis. Specifically, we consider the null hypothesis of interest as

$$\mathcal{H}_0 : \delta_0 = 0, \quad \text{for any } \gamma \in \Gamma, \tag{17}$$

against the alternative

$$\mathcal{H}_1 : \delta_0 \neq 0, \quad \text{for some } \gamma \in \Gamma.$$

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

$$\text{supW} = \sup_{\gamma \in \Gamma} W_n(\gamma),$$

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

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

where $\hat{\delta}(\gamma)$ is the estimate of δ, given γ by either FD-GMM or FD-2SLS, and $\hat{\Sigma}_\delta(\gamma)$ is the consistent asymptotic variance estimator for $\hat{\delta}(\gamma)$. In the FD-GMM case, we may employ $\hat{\Sigma}_\delta(\gamma) = R\left(\hat{V}_s(\gamma)\hat{V}_s(\gamma)\right)^{-1} R'$, where $\hat{V}_s(\gamma)$ is computed as in Section 4 with $\hat{\gamma} = \gamma$ and $R = \left(\mathbf{0}_{(k_1+1)\times k_1}, I_{k_1+1}\right)$. In the FD-2SLS case, we can simply use the same formula for the estimation of the asymptotic variance of $\hat{\delta}(\gamma)$ since the estimation error in γ does not affect the estimation of δ. The supremum type statistic is an application of the union-intersection principle commonly used in the literature, e.g., Davies (1977), Hansen (1996), and Lee et al. (2011).

The limiting distribution of supW depends on the associated estimation methods. If δ were estimated by FD-2SLS, the limit is the supremum of the square of a Gaussian process with some unknown covariance kernel, yielding a non-pivotal asymptotic distribution. For FD-GMM, the Gaussian process is given by a simpler covariance kernel, though it seems not easy to pivotalise the statistic.

Theorem 4 *(i) Consider the FD-GMM estimation. Let* $G(\gamma) = (G_\beta, G_\delta(\gamma))$ *and* $D(\gamma) = G(\gamma)' \Omega^{-1} G(\gamma)$. *Suppose that* $\inf_\gamma \det(D(\gamma)) > 0$ *and Assumption 2 (i) holds. Then, under the null* (17), *we have*

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

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

(ii) Consider the 2SLS estimation. Suppose that Assumptions 6(i), 7, 8, 10 and 11 hold. Then, under the null (17),

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

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

When the reduced form is a threshold regression, our test can be performed using the model (13). A null hypothesis in this case is that both the reduced form and the structural equations are linear for all t; that is,

$$\mathcal{H}_0' : \Lambda_{1t} - \Lambda_{2t} = \Lambda_{3t} - \Lambda_{4t} = 0, \quad \text{for all } \gamma \in \Gamma \text{ and } t = t_0, ..., T. \tag{18}$$

As the model (13) is estimated by the pooled OLS for each γ, the construction of supW statistic is standard (e.g., Hansen, 1996).

Notice that these limiting distributions are not asymptotically pivotal and critical values cannot be tabulated. Hence, we bootstrap or simulate the asymptotic critical values or the p-values of the tests following Hansen (1996). Here we describe the bootstrap procedure in details.

Let $\hat{\theta}$ be either the FD-GMM or the FD-2SLS estimator, and construct:

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

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

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

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 $\hat{\theta}$.

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

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

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, albeit in the regression setup. However, up to our knowledge, the GMM estimator is first to be examined in this general context. In this section, we thus focus on the GMM estimator.

We consider the following two models:

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

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

for $t = 1, ..., 10$, and $i = 1, ..., n$, where u_{it} are iid $N(0, 1)$. 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 notation in (4), the unknown true parameter values are $\beta = -0.5$ and $\delta = (-2.5, 1.2)'$ in the first model and $\beta = 0.6$ and $\delta = (0.96, -1.2)'$ in the second . All the past levels of y_{it} are used as the instrumental variables.

In general, there are many different ways to compute the weight matrix, W_n in the first step. There is no way to tell which is optimal; provided that the first step estimators are all consistent, all the second step estimators are asymptotically equivalent. In this regard, we also consider an averaging of a class of FD-GMM estimators as defined in Section 3.1. The averaging does not change the first order asymptotic distribution but it is expected to be particularly relevant when the sample size is small. We propose to randomize the weight matrix, W_n in the first step as follows: We compute W_n in (5) with

$$\hat{g}_i = \left(\Delta\tilde{\varepsilon}_{it_0} z'_{it_0}, ..., \Delta\tilde{\varepsilon}_{iT} z'_{iT}\right)',$$

where $\tilde{\varepsilon}_{it}$'s are randomly generated from $\mathcal{N}(0, 1)$. In our experiment below, we do this 100 times and take the average of the second step estimators.

We examine the bias, standard error (s.e.), and mean square error (MSE) of the FD-GMM estimator with 1,000 iteration. 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 presented in Table 1, those of the FD-GMM estimator 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 estimation of the regression coefficients which is consistent with our theoretical finding. When we compare the MSEs of the original FD-GMM estimator with those of the averaging estimator, we find that the averaging significantly reduces the MSEs. In some cases the gains are so large that the MSEs of the original estimator are as twice as those of the averaging estimator.[5] As a rule of thumb, we find that the reduction in MSEs by averaging becomes larger when the original MSEs are rather big, though this gain becomes smaller as the sample size increases. Turning to biases and standard errors as reported in Tables 2 and 3, we observe that the averaging always reduces the stand errors, but it has a mixed effect on the biases. In particular, when the bias of the original FD-GMM estimator is large (those of δ_1 and δ_2), then the averaging reduces it and *vice versa*. As a result, the average biases of the FD-GMM estimator 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 bias reduction effects on the FD-GMM estimator.

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

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

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

where the threshold values also change such that they stay at the middle of distribution. From the results reported in Tables 4-6, we find that the averaging reduces MSEs and standard errors even more substantially. Furthermore, the biases are greatly reduced by the averaging for more than 70% of the cases. Hence, we recommend the practitioner to apply the averaging method to reduce the sampling errors associated with the two-step FD-GMM estimators.

7 Empirical Applications

7.1 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 behaviour. 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. Their empirical findings support the hypothesis that cash flow has a significantly positive effect on investment for financially constrained firms, suggesting that the sensitivity of investment to cash flows is an indicator of the degree of financial constraints.

[5] Only in the case of the jump DGP with $n = 50$, the averaging is slightly worse than the original estimator, but the difference is negligibly small.

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 level 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. Hence, we apply a threshold model of investment in dynamic panels to address this problem. 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:

$$I_{it} = \beta_1 CF_{it} + \beta_2 Q_{it} + \varepsilon_{it} \tag{19}$$

where I_{it} is investment, CF_{it} cash flows, Q_{it} Tobin's Q, and ε_{it} consists of the one-way error components, $\varepsilon_{it} = \alpha_i + v_{it}$.[6] 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 on future investments (Aivazian et al., 2005). Therefore, we consider the following augmented dynamic investment model:

$$I_{it} = \phi I_{it-1} + \theta_1 CF_{it} + \theta_2 Q_{it} + \theta_3 L_{it} + \varepsilon_{it}, \tag{20}$$

where L_{it} represents leverage. We then extend (20) into the dynamic panel data framework with threshold effects as:

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

where $1_{\{q_{it} \le c\}}$ and $1_{\{q_{it} > c\}}$ are an indicator function, q_{it} is the transition variable and γ the threshold parameter.

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, which is extracted from an original data set constructed by Hall and Hall (1993). Hence, this study allows for comparisons with the existing literature. Following González et al. (2005), we exclude five

[6] 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.

companies with extreme data values, and consider a final sample of 560 companies with 7840 company-year observations.[7]

Table 7 summarises the estimation results for the dynamic threshold model of investment, (21), 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 is broader than Hansen (1999) who considers only leverage and González et al. (2005) who employ both leverage and Tobin's Q. In each case we only report the FD-GMM estimation results which allow for both (contemporaneous) regressors and the transition variable to be (possibly) endogenous.[8] The estimation results are reported respectively in the low and high regimes.

When cash flow is used as the transition variable, the results for (21) show that the threshold estimate is 0.36 such that about 80% of observations fall into the lower cash-constrained 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 negative impact on investment and a stronger impact 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 the evidence for the role of financial constraints in the investment-cash flow sensitivity.

When the leverage is used as the transition variable, we find that the threshold estimate is 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*,

[7] An exact definition of the variables is as follow: Investment is measured by investment to the book value of assets, Tobin's Q the market value to the book value of assets, leverage long-term debt to the book value of assets, cash flow is cash flow to the book value of assets.

[8] Notice that the previous empirical studies (e.g. Hansen, 1999; González et al., 2005) use the lagged values of Q and CF to avoid the potential endogenous regressors.

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. Notice, however, that Hansen (1999), who considers leverage as the transition variable, fails to find conclusive evidence in favor of this prediction by estimating a non-dynamic threshold model of investment.[9]

When using Tobin's Q as the transition variable, the threshold is estimated at 0.56 with about 59% of observations falling into the higher growth regime. We now 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.[10]

Table 7 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

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

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

since these variables are imperfect measures of financial constraints.[11]

7.2 A dynamic threshold panel data model of dividend smoothing

In a seminal study on dividend policy, Lintner (1956) suggests that firms gradually adjust dividends in response to changes in earnings, implying that firm managers make dividend adjustment in response to unanticipated (permanent) changes in firms' earnings towards a long-run target payout ratio. The number of empirical studies generally find evidence in favour of such dividend smoothing at both firm and aggregate levels, e.g. Fama and Babiak (1968), Marsh and Merton (1987), Skinner (2008) and Cho et al. (2013).

However, the adjustment of dividends may be asymmetric as managers react differently to earnings shocks across different market conditions. In particular, Brav *et al.* (2005) provide recent survey evidence that firms are more likely to increase their dividend than to cut it whilst the magnitude of the average cut is more severe than the magnitude of the average dividend increase. Applying the two-stage approach by Fama and Babiak (1968) to the data at the firm level in the US, Leary and Michaely (2011) find that a firm is less likely to smooth dividends and move towards the target when its dividend is below the target whilst it is more likely to smooth dividends and leave them unchanged when its dividend is above target. Alternatively, at the aggregate level employing the SP500 data over 1871Q1 - 2004Q2, Kim and Seo (2010) estimate the threshold VECM for the (log) dividend-price relationship (assuming that real stock prices are proxy for permanent earnings) and find that the upward stickiness (smoothing) in the lower regime (when its dividend is below the target) is a far more prominent than the downward stickiness in the upper regime. Notice that there is a conflict between the results of smoothing asymmetry at the disaggregate and the aggregate level, though the micro-evidence in Leary and Michaely (2011) is more consistent with the survey evidence reported in Brav *et al.* (2005).

Hence, we examine the issue of asymmetric dividend smoothing by extending the Lintner's (1956) partial adjustment model into the following dynamic panel data threshold model:

$$\Delta d_{it} = \left(\phi_1 d_{i,t-1} + \theta_1 e_{it}\right) 1_{\{q_{it} \leq \gamma\}} + \left(\phi_2 d_{i,t-1} + \theta_2 e_{it}\right) 1_{\{q_{it} > \gamma\}} + \alpha_i + v_{it}. \quad (22)$$

We follow Skinner (2008) and construct the annual firm data on dividend per share real price

[11] Kaplan and Zingales (1997) find that the relationship between cash flows and investment is not monotonic with respect to 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, market to book value, 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). Nonetheless, all these issues are beyond the scope of the current paper.

(d), earnings per share (e) and return on asset (ROA) over the period 1990 - 2001 from CRSP/Compustat. By excluding companies with non-paying dividend observations and keeping the companies with the full period observations over 12 years, we obtain the final balanced panel dataset for 246 firms with 2,952 company-year observations. We also follow the literature and consider the two candidates for $q_{it} = \{ROA_{it}, e_{it}\}$. Both measures are expected to provide a reasonable proxy for the market conditions (sentiments). Hence, this study is expected to contribute to the existing literature on dividend policy by incorporating asymmetries in dividend adjustment at the disaggregate firm level.

Table 8 presents the estimation results for the dynamic threshold model of the asymmetric dividend smoothing, (22). When return on asset is used as the transition variable, the results for (22) show that the threshold estimate is 0.148 such that 61% of observations falling into the higher ROA regime. The coefficient on lagged dividend is significantly higher for firms with the higher ROA (0.905 vs 0.804), suggesting that the dividend smoothing is stronger for firms with the higher ROA. As expected, the impact reaction of dividend to earning is stronger for the higher ROA regime at 0.038 than for the lower regime at 0.005, but it is statistically significant only at the upper regime. Furthermore, we find that the long-run target payout coefficients, estimated by $\hat{\beta}_1 = \hat{\theta}_1 / \left(1 - \hat{\phi}_1\right)$ and $\hat{\beta}_2 = \hat{\theta}_2 / \left(1 - \hat{\phi}_2\right)$, are 0.007 and 0.43 respectively for firms with lower and higher ROA. Next, when earnings per share (EPS) is used as the transition variable, we find that the threshold is estimated at 0.605, lower than the median, with more than 64% of observations falling into the high-EPS regime. Here the results are qualitatively similar to those when ROA is used as the transition variable. In particular, the coefficient on lagged dividend is significantly higher for firms with higher EPS, suggesting that the dividend smoothing is stronger for firms with higher EPS.

Table 8 about here

These results, combined together, suggest that dividend smoothing is substantially stronger for firms that tend to pay the higher (target) dividend payout especially in the long-term perspective, a finding generally consistent with survey evidence in Brav *et al.* (2005).[12]

[12] McMillan (2007) applies the asymmetric ESTR model to the data for a number of countries, and provides similar empirical evidence that the log dividend yields are characterised by an inner random walk regime and the reverting outer regimes where the speed of reversion differs between positive and negative dividend-yield changes, such that price rises greater than the level supported by dividends exhibit a greater degree of persistence than price falls relative to dividends. However, this type of asymmetry persistence may arise from the interaction of noise and fundamental traders in terms of the positive feedback trading, e.g. Shleifer (2000)

8 Conclusion

The investigation of nonlinear asymmetric dynamic modelling has assumed a prominent role. It is clear that misclassifying a stable nonlinear process as linear can be misleading in time series and dynamic panel data analysis. Increasing availability of the large and complex panel data sets has also prompted rigorous econometric analyses of dynamic heterogeneous panels, especially when the time period is short. Recently, some progress has been made, e.g., Dang et al. (2012), Kremer et al. (2013) and Ramirez-Rondan (2013). However, all of these studies maintain the assumption that the regressors and/or the threshold variable are exogenous. This limitation may hamper the usefulness of threshold regression models in a general context. In this paper we have explicitly addressed this challenging issue by extending the approaches by Hansen (1999, 2000) and Caner and Hansen (2004) and developing the dynamic threshold panel data model, which allows both regressors and threshold effect to be endogenous.

In particular, depending upon whether a threshold variable is endogenous or not, we have proposed the two alternative estimation procedures, respectively called FD-GMM and FD-2SLS, on the basis of the FD transformation for removing unobserved individual effects. Their asymptotic properties are derived through employing the diminishing threshold effect and the empirical process theory. The FD-GMM approach works well in the general case where both threshold variable and regressors are allowed to be endogenous. Furthermore, FD-2SLS is shown to be a more efficient estimation method in the special case when the threshold variable is strictly exogenous.

Our proposed approaches are expected to avoid any sample selection bias problem associated with an arbitrary sample-splitting or the dummy variable approach and greatly extend the scope of the applicability of the dynamic threshold panel data model in Economics and Finance, as demonstrated in our two empirical applications to assessing an asymmetric sensitivity of investment to cash flows and an asymmetric dividend smoothing.

Finally, we note several avenues for further researches following the current study. First, the FD-2SLS is more efficient than the FD-GMM if the exogeneity condition of the threshold variable is met, though it is still uncertain if the FD-GMM is most efficient in case of the endogenous threshold variable. This will be an interesting future research topic. Next, given that conventional estimation procedures 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 panel data framework. Furthermore, researches to develop similar estimation algorithms for models with multivariate stochastic covariates and for alternative nonlinear models will be under way.

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, 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] Brav, A., J. Graham, C. Harvey and R. Michaely (2005): "Payout Policy in the 21st Century," *Journal of Financial Economics* 77, 483-527.

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

[13] Cho, J.S., T.H. Kim and Y. Shin (2013): "Quantile Cointegration in the Autoregressive Distributed Lag Modelling Framework," *mimeo.*, University of York.

[14] 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.

[15] Davies, R.B. (1977): "Hypothesis testing when a nuisance parameter is present only under the alternative," *Biometrika* 64, 247-254.

[16] Fama, E.F. and H. Babiak (1968): "Dividend Policy: An Empirical Analysis," *Journal of the American Statistical Association* 63, 1132-1161.

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

[18] 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.

[19] 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.

[20] Hall, B.H. and R.E. Hall (1993): "The Value and Performance of U.S. Corporations," *Brookings Papers on Economic Activity* 6, 1-34.

[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] Hayakawa, K. (2012): "The Asymptotic Properties of the System GMM Estimator in Dynamic Panel Data Models when Both N and T are Large," *mimeo.*, Hiroshima University.

[26] 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.

[27] Hsiao, C. (2003): *Analysis of Panel Data*. Cambridge: Cambridge University Press.

[28] 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.

[29] Jensen, M. (1986): "Agency Costs of Free Cash Flow, Corporate Finance and Takeovers," *American Economic Review* 76, 323-339.

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

[31] 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.

[32] Kim, S. and B. Seo (2010): "Asymmetric Dividend Smoothing in the Aggregate Stock Market," *Quantitative Finance* 10, 349-355.

[33] Kourtellos, A., T. Stengos and C.M. Tan (2009): "Structural Threshold Regression," *mimeo.*, University of Cyprus.

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

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

[36] Leary, M.T. and R. Michaely (2011): "Determinants of Dividend Smoothing: Empirical Evidence," *Review of Financial Studies* 24, 3197-3249.

[37] 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.

[38] Lintner, J. (1956): "Distribution of Incomes of Corporations among Dividends, Retained Earnings, and Taxes," *American Economic Review* 46, 97-113.

[39] Marsh, T. and R. Merton (1987): "Dividend Behaviour for the Aggregate Stock Market," *Journal of Business* 60, 1-40.

[40] McMillan, D. (2007): "Bubbles in the Dividend-Price Ratio? Evidence from an Asymmetric Exponential Smooth-Transition Model," *Journal of Banking and Finance* 31, 787-804.

[41] Newey, W and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics* IV, 2111-2245, Elsevier.

[42] Nickell, S. (1981): "Biases in Dynamic Models with Fixed Effects," *Econometrica* 49, 1417-1426.

[43] Pesaran, M.H. (2006): "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure," *Econometrica* 74, 967-1012.

[44] Seo, M.H. and O. Linton (2007): "A Smoothed Least Squares Estimator for Threshold Regression Models," *Journal of Econometrics* 141, 704-735.

[45] Shleifer, A. (2000): *Inefficient Markets: An Introduction to Behavioural Finance. Clarendon Lectures in Economics.* Oxford: Oxford University Press.

[46] Skinner, D. (2008): "The Evolving Relation between Earnings , Dividends, and Stock Repurchases," *Journal of Financial Economics* 87, 582-609.

[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): "Threshold Regression with Endogeneity," *mimeo.*, University of Auckland.

[50] Whited, T.M and G. Wu (2006): "Financial Constraints Risk," *Review of Financial Studies* 19, 531-559.

[51] 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.

A Proof of Theorems

A.1 GMM

In this section we derive the asymptotic normality of the FD-GMM estimator, which allows for multiple threshold variables and multiple regimes. As described in Section 2, the FD transformation changes the model characteristic such that the number of threshold variables is more than one. Specifically, the moment indicator, $g_i(\theta)$ defined in (1) contains the indicator function, $1\{q_{it} > \gamma\}$ for $t = t_0 - 1, t_0, ..., T$, although the jumps arise at the same value, γ. Here, we allow for γ to vary over t, which may prove useful in some applications. With this generalization, and imposing (4) and Assumption 1, we consider a more general form of moment condition than that presented in (1), that is,

$$g_n(w_i;\theta) = g_i - \xi_i'(\beta - \beta_0) - \zeta_i(\boldsymbol{\gamma}_0)'(\delta - \delta_n) - (\zeta_i(\boldsymbol{\gamma}) - \zeta_i(\boldsymbol{\gamma}_0))'\delta, \tag{23}$$

where w_i stands for the data of an i-th individual, $\zeta_i(\boldsymbol{\gamma}) = \sum_{j=t_0-1}^{T} \zeta_{ij} 1(q_{ij} > \gamma_j)$, $\boldsymbol{\gamma}$ is the collection of all γ_t's, ξ_i and ζ_{ij}'s are the $k_1 \times l$ and the $(k_1+1) \times l$ matrix transformations of w_i, respectively. The function, $\zeta_i(\boldsymbol{\gamma})$ is introduced due to the FD transformation, and we index $g(\cdot,\cdot)$ by subscript n to make explicit the dependence of the true value δ_n on the sample size n, reflecting the shrinking threshold assumption.

Next, we assume that

Assumption 9 *(i) $\delta_n = \delta_0 n^{-\alpha}$ for some $0 \le \alpha < 1/2$ and $\delta_0 \neq 0$, and all θ_n are interior points of Θ, which is compact.*
(ii) For all $n = 1, 2, ...,$ $\mathrm{E}g_n(w_i;\theta) = 0$ if and only if $\theta = \theta_n$.
(iii) The threshold variable q_{it} has continuous and bounded density at γ_0 for all t, $\mathrm{E}\zeta_i(\boldsymbol{\gamma})$ is continuously differentiable at $\boldsymbol{\gamma}_0$ and $G'\Omega^{-1}G$ is nonsingular and finite, where

$$\underset{l\times k}{G} = \left(-\mathrm{E}\xi_i', -\mathrm{E}\zeta_i(\boldsymbol{\gamma}_0)', -\frac{\partial}{\partial \boldsymbol{\gamma}'}\mathrm{E}\zeta_i(\boldsymbol{\gamma}_0)'\delta_0\right),$$

where $k = (2k_1+1) + T - t_0 + 2$.
(iv) Ω is finite and positive definite.

Then, we have

Lemma 5 *Let Assumption 9 hold and denote by $\hat{\theta}$ the GMM estimator of θ, which is the minimizer of $\bar{J}_n(\theta)$ in* (2) *with g as defined in* (23). *Then,*

$$\begin{pmatrix} \sqrt{n}\begin{pmatrix} \hat{\beta}-\beta_0 \\ \hat{\delta}-\delta_n \end{pmatrix} \\ n^{1/2-\alpha}(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0) \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0,\left(G'\Omega^{-1}G\right)^{-1}\right).$$

Proof of Lemma 5. We fix $W=\Omega^{-1}$ hereafter. Write $g_{ni}(\theta)$ for $g_n(w_i;\theta)$ and thus $\bar{g}_n(\theta)=\frac{1}{n}\sum_{i=1}^{n}g_{ni}(\theta)$ and let $\bar{g}_n$ indicate $\bar{g}_n(\theta)$ evaluated at the true value θ_n, i.e. $\bar{g}_n(\theta_n)$. We proceed in two steps by first establishing consistency and then deriving convergence rate and asymptotic normality.

Consistency: Given the linearity in the slope parameters for a fixed $\boldsymbol{\gamma}$, we can write

$$\begin{pmatrix} \hat{\beta}(\boldsymbol{\gamma})-\beta_0 \\ \hat{\delta}(\boldsymbol{\gamma})-\delta_n \end{pmatrix} = \left(\bar{\zeta}_n(\boldsymbol{\gamma})'W_n\bar{\zeta}_n(\boldsymbol{\gamma})\right)^{-1}\bar{\zeta}_n(\boldsymbol{\gamma})'W_n\left(\bar{g}_n+\frac{1}{n}\sum_{i=1}^{n}\left(\zeta_i(\boldsymbol{\gamma}_0)-\zeta_i(\boldsymbol{\gamma})\right)\delta_n\right), \quad (24)$$

where $\bar{\zeta}_n(\boldsymbol{\gamma})=\frac{1}{n}\sum_{i=1}^{n}\left[\xi_i',\zeta_i(\boldsymbol{\gamma})'\right]$. Let $\bar{\zeta}_n(\boldsymbol{\gamma})\xrightarrow{p}\zeta(\boldsymbol{\gamma})$ uniformly, which follows from the standard uniform law of large numbers (ULLN). Thus,

$$n^{\alpha}\begin{pmatrix} \hat{\beta}(\boldsymbol{\gamma})-\beta_0 \\ \hat{\delta}(\boldsymbol{\gamma})-\delta_n \end{pmatrix} \xrightarrow{p} \left(\zeta(\boldsymbol{\gamma})'W\,\zeta(\boldsymbol{\gamma})\right)^{-1}\left(\zeta(\boldsymbol{\gamma})'W\ \left(\zeta_2(\boldsymbol{\gamma}_0)-\zeta_2(\boldsymbol{\gamma})\right)\delta_0\right),$$

as $\bar{g}_n=O_p\left(n^{-1/2}\right)$ due to Assumption 9 (iv). Since $\bar{g}_n(\theta)$ is continuous in β and δ for any given γ, the continuous mapping theorem and standard algebra yield that

$$n^{\alpha}\bar{g}_n\left(\hat{\beta}(\boldsymbol{\gamma}),\hat{\delta}(\boldsymbol{\gamma}),\boldsymbol{\gamma}\right)\xrightarrow{p}\left(I+\zeta(\boldsymbol{\gamma})\left(\zeta(\boldsymbol{\gamma})'W\,\zeta(\boldsymbol{\gamma})\right)^{-1}\zeta(\boldsymbol{\gamma})'W\right)\left(\zeta_2(\boldsymbol{\gamma}_0)-\zeta_2(\boldsymbol{\gamma})\right)\delta_0.$$

The term in the first brackets in the right hand side is positive definite and $\zeta_2(\boldsymbol{\gamma})=\zeta_2(\boldsymbol{\gamma}_0)$ if and only if $\boldsymbol{\gamma}=\boldsymbol{\gamma}_0$. Therefore, $p\lim_{n\to\infty}\ n^{2\alpha}\bar{J}_n\left(\hat{\beta}(\boldsymbol{\gamma}),\hat{\delta}(\boldsymbol{\gamma}),\boldsymbol{\gamma}\right)$ is continuous and uniquely minimzed at $\boldsymbol{\gamma}=\boldsymbol{\gamma}_0$ and the convergence is uniform, which implies the consistency of $\hat{\boldsymbol{\gamma}}$.

Convergence rate and Asymptotic normality: Let $J_n(\theta)=\mathrm{E}\left(g_{ni}(\theta)\right)'W_n\mathrm{E}\left(g_{ni}(\theta)\right)$ and $D_n=2\kappa_n^{-1}G'W_n\bar{g}_n$, where κ_n is a diagonal matrix whose first $2k_1+1$ diagonals are ones and the other k_q elements are n^{α}'s. We first claim that for any $h_n\to 0$,

$$\sup_{|\theta-\theta_n|\le h_n}\frac{\sqrt{n}R_n(\theta)}{1+\sqrt{n}\left|\theta-\theta_n\right|}=o_p(1), \quad (25)$$

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 the CLT and $J_n(\theta) = 2(\theta - \theta_n)' \kappa_n^{-1} G' W_n G \kappa_n^{-1} (\theta - \theta_n) + o\left(|\theta - \theta_n|^2\right)$. Following the same line of arguments as in the proof of Theorem 7.1 in Newey and McFadden (1994), we can establish that $\kappa_n^{-1}\left(\hat{\theta} - \theta_n\right) = O_p\left(n^{-1/2}\right)$, where we use $\kappa_n^{-1}\left(\hat{\theta} - \theta_n\right)$ instead of $\hat{\theta} - \theta_0$. Let $\tilde{\theta} - \theta_n = (G'WG)^{-1} G' W_n \bar{g}_n$, then it again follows from the same proof that $\tilde{\theta} - \theta_n - \kappa_n^{-1}\left(\hat{\theta} - \theta_n\right) = o_p\left(n^{-1/2}\right)$. Therefore, we obtain the limit distribution as in this Lemma.

Proof of (25) Define a centered empirical process

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

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

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

where

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

First note that $\sup_{|\theta - \theta_n| \leq h_n} |\varepsilon_n(\theta)| = o_p(1)$ if the empirical process $\sqrt{n}\left(\bar{g}_n(\theta) - \mathrm{E}g_{ni}(\theta)\right)$ is stochastically equicontinuous. However, $g_n(w_i, \theta)$ is a sum of four terms, of which the first is free of θ and the next two are linear in θ_1. For the last term, note that δ is bounded and $\zeta_i(\boldsymbol{\gamma})$ is the sum of $\zeta_{ij} 1\left\{q_{ij} > \gamma_j\right\}$'s. Notice, however, that this function, as indexed by $\boldsymbol{\gamma} \in \left\{|\boldsymbol{\gamma} - \boldsymbol{\gamma}_0| \leq h_n\right\}$ and centered at $\zeta_{ij} 1\left\{q_{ij} > \gamma_{j0}\right\}$, constitutes a Vapnik-Chervonenkis (VC) class. Thus, Theorem 2.14.1 of van der Vaart and Wellner (1996) yields the desired result by choosing an envelope function as $\left|\zeta_{ij}\right| 1\left\{\left|q_{ij} - \gamma_{j0}\right| \leq h_n\right\}$. Next, note that

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

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

Proof of Theorem 1. We check the regularity conditions in Lemma 5. First we demonstrate that $g_n(\theta_n) = 0$ if and only if $\theta = \theta_n$. That is, suppose $\beta = \beta_0$ and $\delta = \delta_n$ but $\gamma \neq \gamma_0$, then

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

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

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

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

The other conditions in Assumption 9 are readily satisfied. ■

A.2 2SLS

Recall the notational convention that we write g for $g(\theta_0, b_0)$ or $g(\theta)$ for $g(\theta, b_0)$ for a given random function $g(\cdot,\cdot)$ when there is no confusion. This is repeatedly used in this section.

Before we prove Theorem 2, we discuss a set of more primitive sufficient conditions for the asymptotic normality in Assumption 4. One way to characterize the asymptotic property of the reduced form regression, F_t, is through the empirical process theory. Let $\|\cdot\|_{Q,2}$ indicate the L^2-norm with respect to a probability measure, Q, and denote the covering number and the bracketing number, respectively, by $N(\cdot,\cdot,\cdot)$ and by $N_{[]}(\cdot,\cdot,\cdot)$. The notation, P is reserved for the true probability measure. The entropy (with bracketing) is the log of the covering number (the bracketing number). Either of the following entropy integral conditions is imposed to achieve the asymptotic tightness. The first is the uniform-entropy condition:

$$\int_0^\infty \sup_{\xi<\xi_0} \sup_Q \sqrt{\log N\left(\varepsilon \left\|\bar{f}_\xi\right\|_{Q,2}, \mathcal{F}_\xi, L_2(Q)\right)} d\varepsilon < \infty, \tag{26}$$

where the supremum is taken over all the finitely discrete measure, Q on the sample space, $\mathcal{F}_\xi$ a class of functions with an envelope $\bar{f}_\xi$. The second is the bracketing entropy integral condition given by

$$\int_0^\infty \sup_{\xi<\xi_0} \sqrt{\log N_{[]}\left(\varepsilon \left\|\bar{f}_\xi\right\|_{P,2}, \mathcal{F}_\xi, L_2(\mathrm{P})\right)} d\varepsilon < \infty. \tag{27}$$

Then, we introduce the following assumptions, which are sufficient to obtain Assumption 4.

Assumption 10 *The estimator,* $\hat{b}$ *is consistent such that* $\hat{b} - b_0 = O_p\left(n^{-1/2}\right)$*. The class of functions,* $\mathcal{F}_\epsilon = \{f(\cdot, b) - f(\cdot, b_0) : |b - b_0| < \epsilon$*, for some* $\epsilon > 0\}$*, with an envelope function* $\bar{f}_\epsilon$*, satisfies either of the entropy integral conditions,* (26) *and* (27) *and* $\mathrm{E}\bar{f}_\epsilon^{4+a}(z_i) = o\left(\epsilon^{4+a}\right)$ *for every* $\eta > 0$ *and some* $a > 0$.

Assumption 11 *There exists a* $k_b \times 2k_1(T - t_0 + 1)$ *matrix-valued function* $\dot{f}$ *such that*

$$\mathrm{E}\left[f(z_i, b) - f(z_i, b_0) - \dot{f}(z_i)'(b - b_0)\right]^2 = o\left(|b - b_0|^2\right)$$

for any $f_b - f_{b_0} \in \mathcal{F}_\epsilon$.

Assumption 10 is very general and allows for non-regular regression, such as threshold model, as well as regular cases, where $\hat{b}$ is $\sqrt{n}$-consistent and asymptotically normal. Assumption 11 is a differentiability condition for the regression function, f in mean square, which excludes the threshold regression.

Now, we turn to the proof of main theorem.

Proof of Theorem 2. First, we establish the consistency of the estimators. Recall that

$$e_{it}(\theta) = e_{it} - (\beta - \beta_0)' H_{it} - (\delta - \delta_n)'\left(F_{it}' \mathbf{1}_{it}\right) - [\mathbf{1}_{it}(\gamma) - \mathbf{1}_{it}]' F_{it}\delta, \tag{28}$$

and let $M_n(\theta) = \sum_{t=t_0}^{T} \mathrm{E}\left(e_{it}^2(\theta)\right)$. Then, $M_n(\theta)$ is 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 $\gamma < \gamma_0$. Therefore, $M_n(\theta)$ is globally minimized and continuous at $\theta = \theta_n$. Furthermore, $\sup_{\theta\in\Theta}\left|\mathbb{M}_n\left(\theta, \hat{b}\right) - M_n(\theta)\right| \xrightarrow{p} 0$ as $\sup_{\theta\in\Theta}\left|\mathbb{M}_n\left(\theta, \hat{b}\right) - \mathbb{M}_n(\theta)\right| \xrightarrow{p} 0$, as shown in the following rate proof and the uniform convergence of $\mathbb{M}_n(\theta)$ is standard. Thus, the consistency proof is complete.

Convergence rate We verify the conditions of Theorem 3.4.1 in van der Vaart and Wellner (1996). We do so with $r_n = \sqrt{n}$, $\delta_n = n^{-1/2}$, and $\phi_n(\delta) = \delta$. Since $r_n = \sqrt{n}$, the terms in the expansion of $\mathbb{M}_n\left(\theta, \hat{b}\right)$ that are $O_p\left(n^{-1}\right)$ are irrelevant in the verification of the conditions in the theorem.

Define

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

and write

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

The first term can be shown to be $O_p\left(n^{-1}\right)$ uniformly in θ by applying a ULLN, the $\sqrt{n}$-consistency of $\hat{b}$ in Assumption 10, and the mean square differentiability of F in Assumption 11. Furthermore, for any $K < \infty$,

$$\sup_{\theta \in \Theta, |b-b_0| \le K/\sqrt{n}} \left| \frac{1}{n}\sum_{i=1}^{n}\sum_{t=t_0}^{T} e_{it} r_{it}(\theta, b) \right| = O_p\left(n^{-1}\right), \tag{29}$$

where e_{it} is the first term in the expansion of $e_{it}(\theta)$ in (28). This can be verified by applying Theorem 2.11.22 or 2.11.23 in van der Vaart and Wellner (1996). The uniform entropy integral conditions in these theorems are easily satisfied since the class, $\mathcal{I} = \{1(q > \gamma) : \gamma \in \Theta\}$ is a VC-class of functions, satisfying two entropy conditions (26) and (27), and the class $\mathcal{F} = \{(f(\cdot, b) - f(\cdot, b_0)) : |b - b_0| < \epsilon\}$ is assumed to satisfy either of them. We may recall that the entropy results are preserved under the product and summations (e.g., Andrews, 1994). Thus, it remains to verify the conditions in (2.11.21). The first requirement of the continuity in the second mean is obvious. The second one is the conditions on the envelope. Noting that all the terms in r_{it} are bounded by a constant multiple of $\left|F_{it}\left(b_0 + hn^{-1/2}\right) - F_{it}\right|$, we set $|e_{it}|\sqrt{n}\bar{f}_{K/\sqrt{n}}$ as an envelope function, which then satisfies the second condition in (2.11.21) due to Assumption 7.

Due to (29), it remains to verify the conditions of Theorem 3.4.1 for

$$\tilde{\mathbb{M}}_n(\psi) = -\mathbb{M}_n(\theta) + \mathbb{R}_n(\theta, b), \tag{30}$$

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

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

Accordingly, let

$$\tilde{M}_n(\psi) = -\mathrm{E}\tilde{\mathbb{M}}_n(\psi).$$

Assume $\psi \in \Theta_n \times B_n$, 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 $\psi_n = (\theta_n', b_0')'$ should correspond to θ_n in Theorem 3.4.1 in van der Vaart and Wellner (1996).

We now verify the conditions in the theorem with the preceding definitions. The first condition to check is:

$$\sup_{\epsilon/2 < d_n(\psi, \psi_n) < \epsilon} \tilde{M}_n(\psi) - \tilde{M}_n(\psi_n) \leq -\epsilon^2,$$

which follows because $\tilde{M}_n(\psi_n) = -M_n(\theta_n)$, and

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

whose last term is $O\left(n^{-1/2}\right)$ due to Assumption 11 and the fact that $|b - b_0| \leq K/\sqrt{n}$.

The maximal inequality for $\sqrt{n}\left(\left(\tilde{\mathbb{M}}_n - \tilde{M}_n\right)(\psi) - \left(\tilde{\mathbb{M}}_n - \tilde{M}_n\right)(\psi_n)\right)$ is the second condition to check. Begin with $\mathbb{M}_n(\theta)$, which is the first term of $\tilde{\mathbb{M}}$, given in (30). That is, 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 $[\mathbf{1}_{it}(\gamma) - \mathbf{1}_{it}]$. 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 may 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 show:

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

However, the standard algebra using the change-of-variables yields:

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

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

We proceed similarly for $\mathbb{R}_n(\theta, b)$. As $\mathbb{R}_n(\theta_n, b_0) = 0$, the centering is not necessary. We have already described that a class of functions of the type, $r_{it}(\theta, b)\left((\beta - \beta_0)' H_{it} +\right.$

$(\delta-\delta_n)'(F_{it}'\mathbf{1}_{it})+[\mathbf{1}_{it}(\gamma)-\mathbf{1}_{it}]'F_{it}\delta)$ satisfy either of the entropy integral conditions. An envelope for this class might be:

$$C\left|F_{it}(b)-F_{it}\right|\left|F_{it}\right|\left(\epsilon+\left|\mathbf{1}_{it}(\gamma)-\mathbf{1}_{it}\right|\left(|\delta_n|+\epsilon\right)\right),$$

for a finite constant C depending on $\Theta_n\times B_n$. Then, it is clear that this envelope has L^2 norm of order $O(\epsilon)$ by the same reasoning as in the preceding discussion.

The last condition to be checked is:

$$\tilde{\mathbb{M}}_n\left(\hat{\theta},\hat{b}\right)\geq\tilde{\mathbb{M}}_n(\theta_n,b_0)+O_p\left(n^{-1}\right).$$

But, we may assume that $\hat{b}\in B_n$ without loss of generality. Then,

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

where the first and third equalities are due to (29), the second inequality by construction,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 $n^{1-2\alpha}$. Also, let $h_n=h./r_n$, where ./ is the elementwise division. We first derive the weak convergence of

$$n\left(\mathbb{M}_n\left(\theta,\hat{b}\right)-\mathbb{M}_n\left(\theta_n,\hat{b}\right)\right) \tag{31}$$

on $\Theta_n=\{\theta:\theta=\theta_n+h_n \text{ for } |h|\leq K\}$ for an arbitrary $K<\infty$. Then, the argmax continuous mapping theorem yields the desired result. As we already proved that the classes of functions in $\mathbb{M}_n$ satisfy either the uniform-entropy condition or the bracketing entropy integral condition, it remains to verify the conditions on envelope functions and specify the covariance kernels of the limit process.

To begin with, let $e_i=(e_{it_0},...,e_{iT})'$, $h=(h_c',h_\gamma)'$, 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

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

Also keeping the notational convention, we write $\hat{e}_i = e_i\left(\hat{b}\right)$, $\hat{m}_{ni}(h) = m_{ni}\left(h, \hat{b}\right)$, and $e_i = e_i(b_0)$, etc. Then,

$$n\left(\mathbb{M}_n\left(\theta_n + h_n, \hat{b}\right) - \mathbb{M}_n\left(\theta_n, \hat{b}\right)\right) = \frac{1}{n}\sum_{i=1}^{n}|\hat{m}_{ni}(h)|^2 - \frac{2}{\sqrt{n}}\sum_{i=1}^{n}\hat{e}_i'\hat{m}_{ni}(h). \quad (32)$$

We begin with the last term. Due to Assumption 4 and (15), we may apply the mean value theorem to get

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{m}_{ni}(h)'\hat{e}_i &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{m}_{ni}(h)'\Delta\varepsilon_i \\ &\quad + \frac{1}{n}\sum_{i=1}^{n}\hat{m}_{ni}(h)'\frac{\partial\Xi_i\left(\tilde{b}\right)'\theta_{10}}{\partial b'}\left(\frac{\mathrm{E}\left(\dot{f}_i\dot{f}_i'\right)^{-1}}{\sqrt{n}}\sum_{i=1}^{n}\dot{f}_i\eta_i + o_p(1)\right), \quad (33)\end{aligned}$$

and

$$\begin{aligned}&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{m}_{ni}(h)'\Delta\varepsilon_i \\ = &\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(h_c'\Xi_i\left(\hat{b}\right) - n^{\frac{1}{2}-\alpha}(\delta_0 + o(1))'\left(\Xi_{2i}\left(h_\gamma n^{2\alpha-1}, \hat{b}\right) - \Xi_{2i}\left(\hat{b}\right)\right)\right)\Delta\varepsilon_i. \quad (34)\end{aligned}$$

The first part of this expansion, $h_c'\Xi_i(b)\Delta\varepsilon_i$, easily satisfies either of the entropy conditions (26) and (27) by Assumption 10 as a class of functions indexed by b in a neighborhood of b_0. Then, with a proper moment condition, the first part of the empirical process above is stochastically equicontinuous. For the second part, we need to consider a sequence of classes of functions

$$\mathcal{G}_n = \left\{g_n(b, h_\gamma) = n^{\frac{1}{2}-\alpha}\delta_0'\left(\Xi_{2i}\left(h_\gamma n^{2\alpha-1}, b\right) - \Xi_{2i}(b)\right)\Delta\varepsilon_i : |b - b_0| < \epsilon, |h_\gamma| < K\right\},$$

with an envelope function,

$$G_n = n^{\frac{1}{2}-\alpha}|\delta_0|\left(\sup_{|b-b_0|<\epsilon}|\Delta\varepsilon_i||f(z_i, b)|\right)|\mathbf{1}_i(\gamma) - \mathbf{1}_i(\gamma_0)|.$$

Due to the permanence of the entropy conditions with respect to the product, as discussed when deriving the rate, one of the two entropy conditions is satisfied for this sequence of classes, which allows us to apply Theorem 2.11.22 or 2.11.23 in van der Vaart and Wellner (1996). It is sufficient to verify the conditions on the envelope G_n. The Lindeberg condition is satisfied since

$$
\begin{aligned}
& \mathrm{E}\left(G_n^2 1\left(|G_n| > \eta\sqrt{n}\right)\right) \\
\leq\ & \mathrm{E} 2 n^{1-2\alpha} |\delta_0|^2 \sum_{t=t_0-1}^{T} 1\left(|q_{it} - \gamma_0| \leq h_\gamma n^{-1+2\alpha}\right) \\
& \times \left(\sup_{|b-b_0|<\epsilon} |\Delta\varepsilon_i|^2 |f(z_i, b)|^2\right) 1\left(\sup_{|b-b_0|<\epsilon} |\Delta\varepsilon_i| |f(z_i, b)| > \frac{\eta n^\alpha}{2(T+1)|\delta_0|}\right) \\
\leq\ & O\left(n^{-\alpha\zeta}\right) = o(1).
\end{aligned}
$$

due to Assumption 7. In view of the differentiability of f in square mean in Assumption 11, the uniform continuity of $g_n(b, h_\gamma)$ in square mean is obvious. Thus, the second part in (34) is also stochastically equicontinuous. An obvious consequence is that

$$
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} m_{ni}\left(h, \hat{b}\right)' \Delta\varepsilon_i = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} m_{ni}(h, b_0)' \Delta\varepsilon_i + o_p(1) \tag{35}
$$

and the first term converges weakly to a Gaussian process, whose covariance kernel is specified later. Thus, it follows that the second term in (32) is

$$
\begin{aligned}
& \frac{2}{\sqrt{n}} \sum_{i=1}^{n} \hat{e}_i' \hat{m}_{ni}(h) \\
=\ & \left(I - \mathrm{E} m_{ni}(h) \left[I_T \otimes (\iota \otimes \beta_0)\right]' \dot{f}_i' \mathrm{E}\left(\dot{f}_i \dot{f}_i'\right)^{-1}\right) \frac{2}{\sqrt{n}} \sum_{i=1}^{n} \begin{bmatrix} m_{ni}(h)' e_i \\ \dot{f}_i \eta_i \end{bmatrix} + o_p(1).
\end{aligned}
$$

Recall that $m_{ni}(h_1)$ is the sum of a linear function of h_c and $g_n(h_\gamma)$ apart from the negligible term and thus the covariance terms between h_c and h_γ vanish 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,

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

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

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

and the limit of $\frac{1}{n}\sum_{i=1}^{n} |\hat{m}_{ni}(h)|^2$ 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 $\hat{\theta}_1$ and $\hat{\gamma}$.

Turning to the asymptotic distribution of $\hat{\gamma}$, redefine

$$g_n(h_\gamma) = n^{\frac{1}{2}-\alpha}\delta_0'\left(\Xi_{2i}\left(h_\gamma n^{2\alpha-1}\right) - \Xi_{2i}\right)e_i,$$

and 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 $h_\gamma > \acute{h}_\gamma \geq 0$,

$$\begin{aligned} & n^{-1+2\alpha}\mathrm{E}\left(g_n(h_\gamma)\, g_n\left(\acute{h}_\gamma\right)\right) \\ = \; & \delta_0' \sum_{r,t=t_0}^{T} \mathrm{E}\left[e_{it}e_{ir}F_{it}'\left[\mathbf{1}_{it}\left(\gamma_0 + h_\gamma n^{2\alpha-1}\right) - \mathbf{1}_{it}\right]\left[\mathbf{1}_{ir}\left(\gamma_0 + \acute{h}_\gamma n^{2\alpha-1}\right) - \mathbf{1}_{ir}\right]' F_{ir}\right]\delta_0 \end{aligned} \quad (36)$$

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 $t \neq t'$ will vanish. After some algebra, we can show the limit of (36) 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 $\frac{1}{n}\sum_{i=1}^{n}\sum_{t=t_0}^{T} m_{it}(h,b)^2$, the consistency of $\hat{b}$, the same line of argument as above, we may conclude that

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

Given the structure of the weak limit of (31), the minimizer $\hat{h}_c$ is normally distributed and the argmin $\hat{h}_\gamma$ 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 that for a two-sided standard Brownian motion W and for any positive constants c_1 and c_2

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

Furthermore, the same line of proof as that of Theorem 2 of Hansen (2000) goes through for the convergence of $LR_n(\gamma_0)$ given the results obtained above about $\hat{\theta}_1$ and $\hat{\gamma}$. This completes the proof. ■

Proof of Corollary 3. This corollary is a direct consequence of Theorem 2. ■

A.3 Testing

Proof of Theorem 4. (i) GMM case. Recall (24), and apply the standard ULLN and the continuous mapping theorem to conclude that

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

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

(ii) 2SLS case. As the model is linear for each γ, the marginal convergence of $\hat{\delta}(\gamma)$ is standard. And the tightness of the process can be checked following the same line of argument as in the derivation of the asymptotic distribution of the 2SLS estimator. In fact, the current case is simpler as the re-centering of the process at $\gamma = \gamma_0$ is not necessary and $\delta_0 = 0$. Details are omitted. ■

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: 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	26.4	58.9
J-test	60.1 (0.004)	33.3 (0.185)	45.4 (0.091)
No. of IVs	36	36	43

Table 8: A dynamic threshold panel data model of dividend smoothing

$\mathbf{x}_{it} \setminus q_{it}$	ROA	EPS
	Lower Regime ($\boldsymbol{\phi}_1$)	
DPS_{-1}	0.804 (0.030)	0.625 (0.108)
EPS	0.005 (0.005)	-0.021 (0.019)
	Upper Regime ($\boldsymbol{\phi}_2$)	
DPS_{-1}	0.905 (0.029)	0.771 (0.071)
EPS	0.038 (0.008)	0.054 (0.026)
	Difference ($\boldsymbol{\delta}$)	
DPS_{-1}	0.105 (0.026)	0.147 (0.086)
EPS	0.033 (0.009)	0.054 (0.026)
Threshold	0.148 (0.022)	0.605 (0.511)
Upper Regime (%)	61.0	64.2
J-test	47.4 (0.078)	35.6 (0.122)
No. of IVs	40	32