Dynamic Panels with Threshold Effect and Endogeneity

Myung Hwan Seo (LSE)   Yongcheol Shin (University of York)

20th IPDC Tokyo

9-10 July 2014

Introduction

- Hansen (1999) develops a static panel threshold model.
- González *et al.* (2005) develop a panel smooth transition regression model.
- However, these approaches are static, the validity of which has not yet been established in dynamic panels.
- Surprisingly, there has been no rigorous study investigating an important issue of nonlinear asymmetry in dynamic panels, though there is a huge literature on GMM estimation of linear dynamic panels.

Endogeniety Issue

- The main limitation is the assumption of exogeneity of the regressors and/or the threshold variable.
- The standard least squares approach (e.g. Hansen, 2000) requires exogeneity in all the covariates.
- Caner and Hansen (2004) allow for endogenous regressors, still assuming the threshold variable to be exogenous.
- See Hansen (2011) for an extensive survey.

Dynamic Threshold Panel Data Modelling

- Dang et al. (2012) propose the GMM for dynamic threshold models in short panels with unobserved individual effects, which can provide consistent estimates and a valid testing procedure for threshold effects.
- Ramirez-Rondan (2013) extends the Hansen's (1999) work to allow the threshold mechanism in dynamic panels, and proposed the MLE, following Hsiao et al. (2002).
- Kremer et al. (2013) consider a hybrid dynamic version by combining forward orthogonal deviations transformation by Arellano and Bover (1995) and the IV estimation of cross-section model by Caner and Hansen (2004).
- However, the crucial assumption is that regressors or the transition variable or both are exogenous.
- Endogeneity in the threshold variable can occur in various examples, e.g., Kourtellos et al. (2009) and Yu (2013).

Motivations and Contributions

- We fill this gap by explicitly addressing an important issue as how best to model nonlinear asymmetric dynamics and cross-sectional heterogeneity, simultaneously.
- We extend Hansen (1999, 2000) and Caner and Hansen (2004) to dynamic panels with endogenous regressors and threshold variable.
- We propose two estimation methods based on FD transformation and evaluate their properties by the diminishing threshold effect asymptotics of Hansen (2000).
- Our approach will overcome the main limitedness in the existing literature, the assumption of exogeneity of regressors and the transition variable.

Motivations and Contributions

- We first develop the FD-GMM method in which the (time-varying) threshold variable is allowed to be endogenous.
- Next, we propose a more efficient 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.

Asymptotic Theory for FD-GMM

- FD-GMM estimator is shown to be asymptotically normal.
- The standard inference based on the Wald statistic is feasible, though convergence rate is slower than $\sqrt{n}$.
- Importantly, the asymptotic normality holds true irrespective of whether the regression function is continuous or not.
- This is in contrast to the LS approach, where the discontinuity changes the asymptotic distribution in a dramatic way.
- Hence, inference on the threshold parameter can be carried out in the standard manner.

Asymptotic Theory for FD-2SLS

- FD-2SLS estimator satisfies the oracle property where the threshold estimate and the slope estimate are asymptotically independent.
- We allow for general continuous or discontinuous nonlinear regression models for the reduced form, and provide the corrected asymptotic variance formula for the slope coefficients.
- FD-2SLS of the threshold parameter is super-consistent, but its inference is non-standard and can be easily conducted by inverting an LR statistic, which follows a known pivotal asymptotic distribution (Hansen, 2000).

Testing for Linearity and Endogeneity

- We provide testing procedures for identifying the threshold effect, using the supremum type statistics that follow non-standard asymptotic distributions due to the loss of identification under the null of no threshold effect.
- Critical values or the p-values can be easily evaluated by the bootstrap.
- Furthermore, we develop the exogeneity test of the threshold variable, following the general principle of the Hausman (1978) test, e.g. Kapetanios (2010) develops the exogeneity test of the regressors in threshold regression.
- This is a straightforward by-product by combining asymptotic results of FD-GMM and FD-2SLS.

Monte Carlo Studies

- 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.
- As there are many different ways to compute the weight matrix in the first step, we propose an averaging of a class of the two-step FD-GMM estimators obtained by randomising the weight matrix.
- This turns out to be successful in reducing sampling errors, so we recommend the use of the averaging method in practice.

Two Empirical Applications

- We provide 2 empirical applications investigating an asymmetric sensitivity of investment to cash flows and an asymmetric dividend smoothing.
- First, employing a balanced panel dataset of 560 UK firms over 1973-1987, we find that cash flow sensitivity of investment is stronger for cash-constrained, high-growth and high-leveraged firms, a consistent finding with Farazzi *et al.* (1988) that the sensitivity of investment to cash flows is an indicator of the degree of financial constraints.
- Next, using the balanced panel dataset of 246 US firms over 1990 - 2001, we find that dividend smoothing is relatively stronger for firms that tend to pay the higher (target) dividend payout, a finding generally consistent with the survey evidence in Brav *et al.* (2005).

The Model

- Consider the dynamic panel threshold regression model:

$$y_{it} = \left(1, x_{it}'\right) \phi_1 1 \left(q_{it} \leq \gamma\right) + \left(1, x_{it}'\right) \phi_2 1 \left(q_{it} > \gamma\right) + \varepsilon_{it}, \quad (1)$$

  for $i = 1, ..., n$; $t = 1, ..., T$, where y_{it} is a scalar stochastic variable, x_{it} is the $k_1 \times 1$ vector of time-varying regressors, $1 (\cdot)$ is an indicator function, and q_{it} is the transition variable.
- γ is the threshold parameter, and ϕ_1 and ϕ_2 regime-dependent slope parameters.
- ε_{it} consists of the error components:

$$\varepsilon_{it} = \alpha_i + v_{it}, \quad (2)$$

  where α_i is an unobserved individual fixed effect and v_{it} is a zero mean idiosyncratic random disturbance.

- 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.
- We do not assume x_{it} or q_{it} to be measurable wrt $\mathcal{F}_{t-1}$, thus allowing endogeneity in both.
- With large n and fixed T, the MDS assumption is just for expositional simplicity.
- A leading example is the SETAR (Tong, 1990), in which case we have x_{it} consisting of the lagged y_{it}'s and $q_{it} = y_{i,t-1}$.
- We allow for both "fixed threshold effect" and "diminishing threshold effect" for γ by defining (e.g. Hansen, 2000):

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

- The fixed effects estimator of the AR parameters is biased downward (e.g. Nickell, 1981).
- To deal with the correlation of regressors with individual effects, we consider the FD transformation of (1):

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

where $\underset{k_1 \times 1}{\beta} = (\phi_{12}, ..., \phi_{1,k_1+1})'$, $\underset{(k_1+1)\times 1}{\delta} = \phi_2 - \phi_1$, 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 $\theta = (\beta', \delta', \gamma)'$ and assume that θ belongs to a compact set, $\Theta = \Phi \times \Gamma \subset \mathbb{R}^k$, with $k = 2k_1 + 2$.

- The transformed model, (4) consists of 4 regimes, which are generated by two threshold variables, q_{it} and q_{it-1}.
- OLS from (4) is not unbiased since transformed regressors are correlated with $\Delta\varepsilon_{it}$.
- To fix this problem we need to find an $l \times 1$ vector of IVs, $\left(z'_{it_0}, ..., z'_{iT}\right)'$ for $2 < t_0 \leq T$, such that either

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

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

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

- z_{it} may include lagged values of (x_{it}, q_{it}) and lagged dependent variables.
- The number of IVs may be different for each time t.

Two Step FD-GMM for endogenous q

- We consider the $l \times 1$ vector of the sample moment conditions:

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

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

- Let $g_i = g_i(\theta_0) = \left(z'_{it_0}\Delta\varepsilon_{it_0}, ..., z'_{iT}\Delta\varepsilon_{iT}\right)'$ and $\Omega = \mathrm{E}(g_i g_i')$.
- For a pd matrix, W_n such that $W_n \xrightarrow{p} \Omega^{-1}$, let

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

- The GMM estimator of θ is given by

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

- The model is linear in ϕ for each γ and $\bar{J}_n(\theta)$ is not continuous in γ, so the grid search algorithm is practical. Let

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

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

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

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

- 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 two-step optimal GMM estimator

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

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

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

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

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

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

FD-2SLS

- Consider the case where q_{it} are exogenous and the conditional moment restriction (6) holds.
- The threshold estimate, $\hat{\gamma}$ can achieve efficient rate of convergence, as in the classical regression (Hansen, 2000), and the slope estimate, $\hat{\phi}$ can achieve the semi-parametric efficiency bound under conditional homoskedasticity.
- This strong result can be obtained since the two sets of estimators are asymptotically independent.

- We consider two cases for the RF regression:
- The first is a general non-linear regression where unknown parameters can be estimated by the standard $\sqrt{n}$ rate;
- The second is the threshold regression with a common threshold. This was also considered by Caner and Hansen $(2004,\ CH)$, albeit in the single equation.
- CH consists of three steps; the first two steps yield the threshold estimate and the third step performs the GMM within each subsample.
- This split-sample GMM approach does not work with panels with a time varying threshold variable, q_{it}, because it generates multiple regimes with cross regime restrictions.
- Their approach is not fully efficient. We will develop a more efficient estimation algorithm.

- We consider general non-linear RF regressions and provide asymptotic variance formula that corrects estimation error stemming from RF.
- This is practically relevant since the linear projection in the RF invalidates consistency of $\hat{\theta}$ when the SF is threshold regression, e.g. Yu (2013).
- The FD model, (4) with conditional moment condition, (6) and exogeneity of q, implies the following regression of Δy_{it} on z_{it}:

$$\mathrm{E}\left(\Delta y_{it}|z_{it}\right) = \beta'\mathrm{E}\left(\Delta x_{it}|z_{it}\right) + \delta'\mathrm{E}\left(X_{it}'|z_{it}\right)\mathbf{1}_{it}\left(\gamma\right). \quad (12)$$

- Assume that the RF regressions are given by

$$\mathrm{E}\left(\begin{array}{c} 1, x_{it}' \\ 1, x_{it-1}' \end{array} \middle| z_{it}\right) = \left(\begin{array}{c} 1, F_{1t}'\left(z_{it}; b_{1t}\right) \\ 1, F_{2t}'\left(z_{it}; b_{2t}\right) \end{array}\right) = \underset{2\times(1+k_1)}{F_t\left(z_{it}; b_t\right)}, \tag{13}$$

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

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

- Caner and Hansen (2004) consider the linear regression and the threshold regression for F_t.
- If $x_{it-1} \in z_{it}$, then $F_{2t} = x_{1t-1}$.
- There are two regressions for x_{it} due to the FD transformation and the possibility that z_{it} varies over time.
- It is not sufficient to consider regression, $\mathrm{E}\left(\Delta x_{it} | z_{it}\right)$ only, due to threshold effect in the SF (12).

(12) and (13) motivates two-step estimation procedure:

1. For each t, estimate the RF, (13) by the LS, and obtain $\hat{b}_t,\ t = t_0, ..., T$, and fitted values, $\hat{F}_{it} = F_t\left(z_{it}; \hat{b}_t\right)$.
2. Estimate θ by

$$\min_{\theta \in \Theta} \hat{\mathbb{M}}_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \sum_{t=t_0}^{T} e_{it}\left(\theta, \hat{b}_t\right)^2, \tag{14}$$

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

3. This step can be done by the grid search. Thus, $\hat{\beta}(\gamma)$ and $\hat{\delta}(\gamma)$ can be obtained from OLS of Δy_{it} on $\hat{H}_{it}$ and $\hat{F}_{it}' \mathbf{1}_{it}(\gamma)$, and $\hat{\gamma}$ is defined as the minimiser of the profiled sum of squared errors, $\mathbb{M}_n(\gamma)$.

- This produces a rate-optimal estimator for γ, implying that β and δ can be estimated as if γ_0 were known.
- The two-step estimation yields the optimal estimate for β and δ if the model is conditionally homoskedastistic, i.e., $\mathrm{E}\left(\Delta\varepsilon_{it}^2|z_{it}\right) = \sigma^2$, see Chamberlain (1987).
- While it requires to estimate the conditional heteroskedasticity to fully exploit the implications of the conditional moment restriction, (6), it is reasonable to employ our two-step estimator and robustify standard errors for heteroskedasticity.

Threshold Regression in Reduced Form

- Suppose that z_{it} includes 1 and x_{it-1}, z_{it}, and

$$x_{it} = \Gamma_{1t} z_{it} 1\{q_{it} \leq \gamma\} + \Gamma_{2t} z_{it} 1\{q_{it} > \gamma\} + \eta_{it}, \ \mathrm{E}(\eta_{it} | z_{it}) = 0.$$

- This implies that

$$\begin{aligned} \Delta y_{it} &= \lambda_{1t}' z_{it} 1\{q_{it} \leq \gamma\} + \lambda_{2t}' z_{it} 1\{q_{it} > \gamma\} \\ &\quad - \lambda_{3t}' z_{it} 1\{q_{it-1} \leq \gamma\} - \lambda_{4t}' z_{it} 1\{q_{it-1} > \gamma\} + e_{it}, \end{aligned} \quad (15)$$

where $\mathrm{E}(e_{it}|z_{it}) = 0$ and $\lambda_{1t}' = (0, \beta'\Gamma_{1t})$, $\lambda_{2t}' = (\delta_1, \phi_{22}'\Gamma_{2t})$, $\lambda_{3t}' z_{it} = \beta' x_{it-1}$, $\lambda_{4t}' z_{it} = \phi_{22}' x_{it-1} - \delta_1$.

- Also, $e_{it} = \Delta\varepsilon_{it} + \eta_{it}' (\beta + 1\{q_{it} > \gamma\} \delta_2)$.
- Since estimates of λ and γ are asymptotically independent, we do not impose these constraints on λ to estimate γ.

Thus, we estimate the model as follows:

1. Estimate γ by the pooled OLS of (15), which can be done by the grid search, and denote the estimate by $\tilde{\gamma}$.
2. Fix γ at $\tilde{\gamma}$ and estimate Γ_{jt}, $j = 1, 2$ by OLS, for each t.
3. Estimate β and δ in (12) by the OLS with γ and the reduced form parameters fixed at the estimates obtained from the preceding steps. Denote these estimates by $\tilde{\beta}$ and $\tilde{\delta}$.

- Our approach is different from Caner and Hansen, who estimate the threshold parameter separately in the reduced and the structural form.
- Their approach introduces dependence between separate threshold estimates, which invalidates their asymptotic distribution.
- 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 reduced and structural forms.

Asymptotic Distributions

- There are two frameworks.
- One is Hansen's (2000) diminishing threshold assumption;
- The other is fixed threshold assumption as in Chan (1993).
- For GMM we present an asymptotics that accommodates both setups;
- For 2SLS we develop the asymptotic distribution only under Hansen's framework.

FD-GMM

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

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

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

where $\mathrm{E}_t[\cdot|\gamma]$ stands for the conditional expectation given $q_{it} = \gamma$ and $p_t(\cdot)$ 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.
- **Assumption 2.**(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) $\mathrm{E}_t\left(z_{it}\left(x_{it}', x_{i,t-1}'\right)|\gamma\right)$ is continuous at γ_0, where $\mathrm{E}_t(\cdot|\gamma) = \mathrm{E}(\cdot|q_{it} = \gamma)$, and $\mathrm{E}_t\left(z_{it}\left(x_{it}', x_{i,t-1}'\right)|\gamma\right)\delta_0 \neq 0$ for some t.
- We do not require the discontinuity of the regression function at the change point. **This is a novel feature of the GMM.**
- **Assumption 3.**Let $G = (G_\beta, G_\delta(\gamma_0), G_\gamma(\gamma_0))$, and G is of the full column rank.

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

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

- The asymptotic variance matrix contains δ_0, and the convergence rate of $\hat{\gamma}$ hinges on unknown α.
- These two 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

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

- G_γ estimated by the Nadaraya-Watson kernel estimator:

$$\hat{G}_\gamma = \begin{bmatrix} \frac{1}{nh}\sum_{i=1}^{n} z_{it_0}\left[(1, x_{it_0-1})' K\left(\frac{\hat{\gamma}-q_{it_0-1}}{h}\right) - (1, x_{it_0})' K\left(\frac{\hat{\gamma}-q_{it_0}}{h}\right.\right. \\ \vdots \\ \frac{1}{nh}\sum_{i=1}^{n} z_{iT}\left[(1, x_{iT-1})' K\left(\frac{\hat{\gamma}-q_{iT-1}}{h}\right) - (1, x_{iT})' K\left(\frac{\hat{\gamma}-q_{iT}}{h}\right)\right. \end{bmatrix} \quad (16)$$

- See Hardle and Linton (1994) for the choice of kernel K and bandwidth h

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

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

- The t-statistic for $\gamma = \gamma_0$ defined by

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

converges to the standard normal distribution.
- Alternatively, nonparametric bootstrap can be employed to construct confidence intervals.

- We collect all distinct RF functions, F_t, $t = t_0, ..., T$, that are not identities, and denote it as $F(z_i, b)$, where z_i and b are the collections of all distinct elements of z_{it} and b_t.
- Denote the collection of the corresponding elements of x_{it}'s by $\S_i$.
- Write the RF as the multivariate regression in the cross section:

$$\S_i = F(z_i, b) + \eta_i, \ \mathrm{E}(\eta_i | z_i) = 0.$$

- Let $\hat{b}$ denote the LSE.
- We follow the convention that $F_i(b) = F(z_i, b)$, $F_i = F(z_i, b_0)$, $\hat{F}_i = F\left(z_i, \hat{b}\right)$, etc, where b_0 is true value.

Reduced Form

- We directly assume asymptotic normality of $\hat{b}$ and existence of a matrix-valued influence function, $\mathbb{F}$ below.
- **Assumption 4.** There exists a matrix-valued function $\mathbb{F}(z_i, b)$ such that $\mathrm{E}\left|\mathbb{F}_i\right|^2 < \infty$ and

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

- We begin with this high-level assumption because our goal is to illustrate how the estimation error in the first step affects asymptotic distribution of the estimator of β, δ and γ in the second step.

- For each t, let

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

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

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

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

$$\underset{((2k_1+1)+k_b)\times((2k_1+1)+k_b)}{\Omega(\gamma_1,\gamma_2)} = E\left[\begin{pmatrix} \Xi_i(\gamma_1)e_i, \\ \mathbb{F}_i\eta_i \end{pmatrix}\left(e_i'\Xi_i'(\gamma_2), \eta_i'\mathbb{F}_i'\right)\right]$$

$$\underset{(2k_1+1)\times((2k_1+1)+k_b)}{A(\gamma)} = \left(I_{(2k_1+1)}, -\mathrm{E}\left[\frac{\partial}{\partial b'}\sum_{t=t_0}^{T}\left(H_{it}'\beta_0\right)\Xi_{it}(\gamma)\right]\left(\mathrm{E}\mathbb{F}_i\mathbb{F}\right.\right.$$

- For the asymptotic distribution of $\hat{\gamma}$, introduce

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

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

- We write $V_j = V_j(\gamma_0)$ and $M_j = M_j(\gamma_0)$ for $j = 1, 2$.

- **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 $\delta_0 \neq 0$.
- **Assumption 6.** (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) $\mathrm{E}_t\left(w_{it}|\gamma\right)$ 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 $\zeta > 0$, $\mathrm{E}\left(\sup_{t \le T, |b - b_0| < \epsilon}\left|e_{it}F_t\left(z_{it}, b_t\right)\right|^{2+\zeta}\right) < \infty$ and for all $\epsilon > 0$ $\mathrm{E}\left(\sup_{t \le T, |b - b_0| < \epsilon}\left|e_{it}\left(F_t\left(z_{it}, b_t\right) - F_t\left(z_{it}\right)\right)\right|^{2+\zeta}\right) = O\left(\epsilon^{2+\zeta}\right)$.
- **Assumption 8.** The minimum eigenvalue of the matrix $\mathrm{E}\Xi_{it}\left(\gamma\right)\Xi_{it}'\left(\gamma\right)$ is bounded below by a positive value for all $\gamma \in \Gamma$ and $t = 1, ..., T$.

- Asymptotic confidence intervals for γ_0 can be constructed by inverting an LR test statistic (Hansen, 2000):

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

- **Theorem 2.** Let Assumptions 4-8 hold. Then,

$$\sqrt{n} \begin{pmatrix} \hat{\beta} - \beta_0 \\ \hat{\delta} - \delta_n \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0, M_1^{-1} V_1 M_1^{-1}\right), \tag{17}$$

$$n^{1-2\alpha} \frac{M_2^2}{V_2} (\hat{\gamma} - \gamma_0) \xrightarrow{d} \underset{r \in \mathbb{R}}{\operatorname{argmin}} \left(\frac{|r|}{2} - W(r) \right), \tag{18}$$

where $W(r)$ is a two-sided standard Brownian motion and it is independent of the limit variate in (17). Furthermore,

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

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

- The first step estimation error does not affect the asymptotic distribution of $\hat{\gamma}$, while it contributes the asymptotic variance of $\hat{\beta}$ and $\hat{\delta}$ through Ω.
- Estimation of the asymptotic variances of $\hat{\beta}$ and $\hat{\delta}$ is standard due to the asymptotic independence.
- The asymptotic distribution for $\hat{\gamma}$ is symmetric and has a known distribution function,

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

for $x \geq 0$, where Φ is the standard normal distribution.

- 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

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

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

- The normalization 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 $\hat{\sigma}_e^2 = (n(T - t_0 + 1))^{-1} \sum_{i=1}^{n} \sum_{t=t_0}^{T} \hat{e}_{it}^2$.
- Notice that it becomes 1 under the leading case of conditional homoskedasticity and the MDS assumption for e_{it}.
- Hansen (2000) provides the distribution function of the asymptotic distribution of the LR_n statistic, which is $\left(1 - e^{-x/2}\right)^2$.

Threshold Regression in Reduced Form

- $\hat{\theta}$ is obtained from the three-step procedure.
- **Corollary 3**.Let Assumption 5 hold and let $\lambda_j = \left(\lambda'_{jt_0}, ..., \lambda'_{jT}\right)'$, $j = 1, ..., 4$, and assume that $\lambda_1 - \lambda_2 = n^{-\alpha}\delta_1$ for some non-zero vector δ_1 and that Assumption 6 and 8 hold with $F_{1,it} = \Gamma_{1t} z_{it} 1\left\{q_{it} \leq \gamma\right\} + \Gamma_{2t} z_{it} 1\left\{q_{it} > \gamma\right\}$ and $F_{2,it} = x_{it-1}$. Furthermore, assume that $\mathrm{E}\left|z_{it}\right|^4 < \infty$ and $\mathrm{E}e_{it}^4 < \infty$ and M_1, M_2, V_1 and V_2 are defined with $\mathbb{F}_i = \left(z'_{it_0}, ..., z'_{iT}\right)'$. Then, the asymptotic distribution of $\tilde{\theta}$ is the same as in Theorem 2.

- The test for threshold effects requires to develop the different asymptotic theory due to the presence of unidentified parameters under the null hypothesis (e.g. Davies, 1977).
- Specifically, we consider the null hypothesis:

$$\mathcal{H}_0 : \delta_0 = 0, \quad \text{for any } \gamma \in \Gamma, \tag{19}$$

against the alternative

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

- Then, a natural test statistic for $\mathcal{H}_0$ is

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

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

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

where $\hat{\delta}(\gamma)$ is the estimate of δ, given γ by FD-GMM or FD-2SLS.

- $\hat{\Sigma}_{\delta}(\gamma)$ is consistent asymptotic variance estimator for $\hat{\delta}(\gamma)$.
- For FD-GMM we 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 $R = \left(\mathbf{0}_{(k_1+1)\times k_1}, I_{k_1+1}\right)$.
- For FD-2SLS we can 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.
- For FD-2SLS, the limit is the supremum of the square of a Gaussian process with some unknown covariance kernel, yielding non-pivotal asymptotic distribution.
- In case of the 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 $\inf_{\gamma\in\Gamma} \det(D(\gamma)) > 0$ and Assumption 2(i) holds. Then, under the null (19),

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

where $Z \sim \mathcal{N}\left(0, \Omega^{-1}\right)$.
(ii) Consider the 2SLS estimation. Suppose that Assumptions, 10, 11, 6(i), 7, and 8, hold. Then, under the null (19),

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

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

- When the RF is a threshold regression, our test can be performed based on the model, (15).
- A null model might be that both RF and SF are linear for all t,

$$\mathcal{H}_0' : \lambda_{1t} - \lambda_{2t} = \lambda_{3t} - \lambda_{4t} = 0, \quad \text{for all } \gamma \in \Gamma \text{ and } t = t_0, ..., T. \tag{20}$$

- The model, (15) is estimated by the pooled OLS and the construction of supW statistic is standard.
- These limiting distributions are not asymptotically pivotal. We bootstrap or simulate the asymptotic critical values or p-values.
- **Bootstrap procedure in details:**

Let $\hat{\theta}$ be FD-GMM or 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 from bootstrap samples.
4. Repeat steps 1-3 B times, and evaluate the bootstrap p-values by the frequency of supW* that exceeds supW.

Test for the exogeneity of the threshold variable.

- Kapetanios (2010) develops the exogeneity test of the regressors in threshold regression models, following the principle of the Hausman (1978) test.
- Similarly, we can develop the Hausman type testing procedure for the validity of the null that the threshold variable is exogenous.
- This is a straightforward by-product by combining FD-GMM and FD-2SLS estimation methods and their asymptotic results.

- We propose the t-statistic for the null that GMM estimate of threshold, $\hat{\gamma}_{GMM}$, equals 2SLS estimate, $\hat{\gamma}_{2SLS}$:

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

- Notice that

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

  due to its super-consistency.
- The asymptotic distribution of the t-statistic is the standard normal under the null of strict exogeneity of the threshold variable, q_{it}.

Monte Carlo Experiments

- We explore finite sample performance of FD-GMM.
- The GMM estimator is first to be examined in this general context.
- We thus focus on the GMM estimator.

- We consider the following two models:

$$\begin{aligned} y_{it} &= (0.7 - 0.5 y_{it-1}) 1\{y_{it-1} \leq 0\} \\ &\quad + (-1.8 + 0.7 y_{it-1}) 1\{y_{it-1} > 0\} + \sigma_1 u_{it}, \end{aligned}$$

$$\begin{aligned} y_{it} &= (0.52 + 0.6 y_{it-1}) 1\{y_{it-1} \leq 0.8\} \\ &\quad + (1.48 - 0.6 y_{it-1}) 1\{y_{it-1} > 0.8\} + \sigma_2 u_{it}, \end{aligned}$$

  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 at the threshold point. The second is the continuous model by Chan and Tsay (1998).
- The threshold is located around the center of the distribution.
- Unknown true parameter values: $\beta = -0.5$ and $\delta = (-2.5, 1.2)'$ in the first 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.

- We also consider an averaging of a class of FD-GMM estimators.
- There are many different ways to compute the weight matrix W_n in the first step, though there is no way to tell which is optimal.
- If the first step estimators are consistent, all the second step estimators are asymptotically equivalent. Thus the averaging does not change the first order asymptotic distribution.
- We propose to randomize the weight matrix, W_n in the first step as follows: We compute W_n in (11) 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 $\mathcal{N}(0,1)$.

- We examine the bias, standard error (s.e.), and mean square error (MSE) of FD-GMM estimator with 1,000 iterations.
- For $n = 50$, 100 and 200, we set $\sigma_1 = 1$ and $\sigma_2 = 0.5$.
- The simulation results are reported in Tables 1-3.
- MSEs of FD-GMM 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 which is consistent with our theoretical finding.
- We find that the averaging significantly reduces the MSEs.
- As a rule of thumb, the reduction in MSEs by averaging becomes larger when the original MSEs are rather big.

- Turning to biases and standard errors, 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 FD-GMM estimator is almost the same as that of the averaging whilst the standard deviation of the former is always larger than the latter.
- This implies that the averaging has some positive bias reduction effect on FD-GMM estimator.

Table: 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: Bias of FD-GMM estimators

		FD-GMM				Avera	
DGP	n	γ	β	δ_1	δ_2	γ	β
Jump	50	-0.041	0.005	-0.044	0.100	-0.269	0.199
	100	-0.047	0.007	-0.044	0.095	-0.106	0.073
	200	-0.029	-0.011	-0.018	0.098	-0.060	0.016
Cont.	50	0.057	0.180	-0.288	0.184	0.055	0.105
	100	0.064	0.145	-0.271	0.199	0.057	0.099
	200	0.074	0.190	-0.298	0.162	0.067	0.158

Table: 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

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

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

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

- From Tables 4-6, we find that the averaging reduces MSEs and standard errors even more substantially.
- 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 sampling errors with the two-step FD-GMM

Table: 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: 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: 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

A dynamic threshold panel data model of investment

- Farazzi *et al.* (1988) find that investment spending by firms with low dividend payments is strongly affected by availability of cash flows.
- 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 financial constraints.
- One of the main problems is that the distinction between constrained and unconstrained firms is routinely based on an arbitrary 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 model takes the form of a Tobin's Q model:

$$I_{it} = \beta_1 CF_{it} + \beta_2 Q_{it} + \varepsilon_{it} \tag{21}$$

  where I_{it} is investment, CF_{it} cash flows, Q_{it} Tobin's Q, and ε_{it} consists of the one-way error components, $\varepsilon_{it} = \alpha_i + v_{it}$.
- β_1 represents the cash flow sensitivity of investment.
- If firms are not financially constrained, β_1 is expected to be close to zero.
- 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 (Aivazian *et al.*, 2005).

- Consider the augmented model:

$$I_{it} = \phi I_{it-1} + \theta_1 CF_{it} + \theta_2 Q_{it} + \theta_3 L_{it} + \varepsilon_{it}, \qquad (22)$$

where L_{it} is leverage.

- We extend (22) into the dynamic panel data with threshold effects:

$$\begin{aligned} I_{it} &= \left(\phi_1 I_{it-1} + \theta_{11} CF_{it} + \theta_{21} Q_{it} + \theta_{31} L_{it}\right) 1_{\{q_{it} \leq \gamma\}} \qquad (23) \\ & \left(\phi_2 I_{it-1} + \theta_{12} CF_{it} + \theta_{22} Q_{it} + \theta_{32} L_{it}\right) 1_{\{q_{it} > \gamma\}} + \alpha_i + v_{it}, \end{aligned}$$

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

The data

- We employ the same data set as in Hansen (1999) and González *et al.* (2005).
- It is a balanced panel of 565 US firms over 973-1987.
- Following González *et al.* (2005), we exclude five companies with extreme data values, and consider a final sample of 560 companies with 7840 company-year observations.
- 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.

- Table 7 summarises the results for the dynamic threshold model of investment, (23), 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 leverage and Tobin's Q.
- We only report the FD-GMM results which allow for both (contemporaneous) regressors and the transition variable to be endogenous.
- The estimation results are reported respectively in the low and high regimes.

Cash flow used as the transition variable

- Threshold estimate is 0.36 (about 80% of observations falling 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.
- We find the more negative impacts of the leverage when firms are cash-constrained. This is consistent with our expectations that the leverage should have a stronger negative impact on investment for constrained firms, in line with the overinvestment hypothesis (Jensen, 1986).

- 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).
- This finding supports evidence for the role of financial constraints in the investment-cash flow sensitivity.

Leverage used as the transition variable

- Threshold estimate is 0.10, lower than the mean (0.24), with more than 73% of observations falling into the high-leverage regime.
- Past investment has a much higher positive impact for highly-levered firms, suggesting that firms with high leverage attempt to respond to growth options quickly.
- The effect of Tobin's Q is higher for lowly-levered firms, providing a support that by lowering the risky "debt overhang" to control underinvestment incentives *ex ante*, firms can take more growth opportunities and make more investments *ex post*, though these impacts are rather small.
- We 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 have a stronger effect on the level of investment for financially constrained firms.
- Notice, however, that non-dynamic threshold model of investment by Hansen (1999) fails to find conclusive evidence in favor of this prediction.

Tobin's Q as the transition variable

- Threshold is estimated at 0.56 with about 59% of observations falling into the higher growth regime.
- Past investment has a slightly stronger positive effect for firms with low Tobin's Q, but the differentials 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 relatively higher for high-growth firms than low-growth firms.
- This supports the hypothesis that cash flow should be more relevant for firms with potentially high financial constraints.
- Our results are qualitatively similar to González et al. (2005) regarding the impacts on investment of 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.

Table: A dynamic threshold panel data model of investment

$x_{it} \setminus q_{it}$	Cash Flow	-Leverage	Tobin Q
	Lower Regime (ϕ_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 (ϕ_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 (δ)		
I_{-1}	−0.796 (0.561)	−0.336 (0.439)	−0.016 (0.325)

- In sum, 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.
- Our results demonstrate the usefulness of the proposed dynamic threshold panel data estimation despite that the transition variables used in the current study may be imperfect measures of financial constraints.

A dynamic threshold panel data model of dividend smoothing

- 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 find evidence in favour of such dividend smoothing at both firm and aggregate levels.
- Adjustment of dividends may be asymmetric as managers react differently to earnings shocks.
- 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 that of the average dividend increase.

- 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.
- 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 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.
- There is a conflict between the results at the disaggregate and the aggregate level, though the micro-evidence in Leary and Michaely (2011) is more consistent with the survey evidence in Brav *et al.* (2005).

- We examine the issue of asymmetric dividend smoothing by extending the Lintner's (1956) partial adjustment model:

$$\Delta d_{it} = (\phi_1 d_{i,t-1} + \theta_1 e_{it})\, 1_{\{q_{it} \le \gamma\}} + (\phi_2 d_{i,t-1} + \theta_2 e_{it})\, 1_{\{q_{it} > \gamma\}} + \alpha_i + v \tag{24}$$

- We construct the annual firm data on dividend per share real price (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 consider $q_{it} = \{ROA_{it}, e_{it}\}$. Both measures are expected to provide a reasonable proxy for market conditions.
- 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, (24).
- When ROA is used as the transition variable, the results show that the threshold estimate is 0.148 such that 61% of observations falling into the higher ROA regime.
- Coefficient on lagged dividend is significantly higher for firms with the higher ROA (0.905 vs 0.804), suggesting that dividend smoothing is stronger for firms with higher ROA.
- As expected, the impact reaction of dividend to earning is stronger for the higher regime at 0.038 than for the lower regime at 0.005, but statistically significant only at the upper.
- Furthermore, the long-run target payout coefficients, $\hat{\beta}_1 = \hat{\theta}_1 / \left(1 - \hat{\phi}_1\right)$ and $\hat{\beta}_2 = \hat{\theta}_2 / \left(1 - \hat{\phi}_2\right)$, are 0.007 and 0.43 for firms with lower and higher ROA.

- Next, when EPS is used as the transition variable, the threshold is estimated at 0.605, lower than median, with more than 64% of observations falling into the high 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.
- These results 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).

Table: A dynamic threshold panel data model of dividend smoothing

$x_{it} \setminus q_{it}$	ROA	EPS
	Lower Regime (ϕ_1)	
DPS_{-1}	0.804 (0.030)	0.625 (0.108)
EPS	0.005 (0.005)	−0.021 (0.019)
	Upper Regime (ϕ_2)	
DPS_{-1}	0.905 (0.029)	0.771 (0.071)
EPS	0.038 (0.008)	0.054 (0.026)
	Difference (δ)	
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
Linearity test	0.002	0.528
J-test	47.4	35.6

Concluding Remarks

- The investigation of nonlinear asymmetric dynamic modelling has recently assumed a prominent role.
- 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.

- Depending upon whether a threshold variable is endogenous or not, we have proposed the two alternative estimation procedures, 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 endogenous.
- 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 and greatly extend the scope of the applicability of dynamic threshold panel data model, as demonstrated in our two empirical applications.

- 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.