Linear Control Function Approach in Endogenous Kink Threshold Regression Models

Jianhan Zhang* Chaoyi Chen†‡ Yiguo Sun§
Thanasis Stengos ¶

December 1, 2023

Abstract

This paper considers an endogenous kink threshold regression model with an unknown threshold value in a time series as well as a panel data framework, where both the threshold variable and regressors are allowed to be endogenous. We construct our estimators from a linear control function approach and derive the consistency and asymptotic distribution of our proposed estimators. Monte Carlo simulations are used to assess the finite sample performance of our proposed estimators.

Keywords: Control function approach; Endogeneity; Kink regression model;

*Department of Economics and Finance, University of Guelph, Guelph, Ontario, N1G 2W1, Canada; Email: jzhang56@uoguelph.ca.

†Magyar Nemzeti Bank (Central Bank of Hungary), Budapest, 1054, Hungary; Email: chenc@mnb.hu.

‡MNB Institute, John von Neumann University, Kecskemét, 6000, Hungary.

§Department of Economics and Finance, University of Guelph, Guelph, Ontario, N1G 2W1, Canada; Email: yisun@uoguelph.ca.

¶Department of Economics and Finance, University of Guelph, Guelph, Ontario, N1G 2W1, Canada; Email: tstengos@uoguelph.ca.

1 Introduction

The threshold regression (TR) model is popularly used to capture potential shifts in economic relationships; e.g., Tong (1990) and Hansen (2000). Notwithstanding, the conventional TR model requires the regression function is discontinuous at the true threshold level. But in many empirical applications, there is no reason to expect a discontinuous regression model. As a modification, Chan and Tsay (1998) propose a continuous threshold autoregressive model to allow for a piece-wise linear function of the threshold variable. Notably, this model allows the threshold regression to be continuous. Still, the slope has a discontinuity at the true threshold level and, thus, is widely regarded as a special case of the broad class of threshold autoregressive models. Extending Chan and Tsay (1998), Hansen (2017) provides testing for a threshold effect and inference on the regression parameters for a continuous threshold model with an unknown threshold parameter value (hereinafter, kink threshold regression (KTR) model). It is well established the limiting distribution of the least-squares estimator for the TR model is nonstandard and the estimator is super consistent. For example, with a "fixed threshold effect" assumption, Chan (1993) establishes that the threshold parameter estimator converges to a function of a compound Poisson process. Adopting a "diminishing threshold effect" assumption, Hansen (2000) shows the limiting distribution involves two independent Brownian motions. By contrast, as shown in Hansen (2017), the limiting distribution of the least-squares estimator for the KTR model is normal, and the convergence rate is standard root-n due to the nature of continuity.

All the above studies assume strict exogeneity in both slope regressors and the threshold variable. As many practical issues of nonlinear asymmetric mechanisms are endogenously determined, a growing body of literature has developed for the TR model to allow for endogeneity. Under Hansen (2000)'s diminishing threshold effect framework, Caner and Hansen (2004) allow the slope regressors to be endogenous by using the generalized method of moments (GMM) and two-stage least squares (2SLS) method to estimate the slope parameters and the threshold parameter, respectively. Inspired by the sample selection method of Heckman (1979), Kourtellos et al. (2016) employ a control function (CF) approach to estimate the TR model with endogeneity, where they introduce an inverse Mills ratio as a bias correction term into the regression. Following Kourtellos et al. (2016), Christopoulos et al. (2021) use a copula method to deal with the endogenous threshold variable. Yu et al. (2020) generalize the CF approach of Kourtellos et al. (2016) and classify two groups of CF methods for the TR model with endogeneity based on the choice of variables in the conditional set. Specifically, the first group is an extension of the 2SLS method proposed by Caner and Hansen (2004), while another is a natural extension to the conventional CF approach documented in Newey et al. (1999). It is worth noticing that both CF methods cannot be directly used to estimate the KTR model with endogeneity. In fact, the continuity makes the inference of the least squares estimator for the KTR model quite different from the conventional TR model even without the endogeneity. Hidalgo et al. (2019) underscore that if we wrongly estimate a KTR model with a TR framework of Hansen (2000), ignoring the continuity of the true model, the Hessian matrix becomes irregular[1]. This causes the least squares estimator of the threshold parameter converges at a cube root-n convergence rate, slower than the root-n convergence rate for KTR model as shown by Hansen (2017). As a result, both CF methods proposed by Yu et al. (2020), designing for the TR framework, cannot apply to the KTR model without deviation.[2] More recently, Kourtellos et al. (2022) extend Yu et al. (2020) to allow for the unknown endogenous form by introducing a nonparametric bias correction term into the model. The proposed semiparametric model bypasses any misspecification problem, but is still in the framework of the TR model. Seo and Shin (2016) consider a dynamic panel TR model with endogeneity and develop a first-differenced GMM estimator, which allows both threshold variable and regressors to be endogenous. Yet, the GMM method is notorious for its poor small sample performance. Under a fixed threshold effect framework of Chan (1993) and assuming i.i.d. sample, Yu and Phillips (2018) construct an integrated difference kernel estimator(IDKE) for the threshold parameter. The most appealing feature of the IDKE is the consistency of the estimator holds without requiring the instrumental variables. Also, the IDKE is super-consistent for the TR model with endogenous threshold variable and exogenous slope regressors. Nevertheless, the i.i.d. assumption broadly limits the scope of applications for this method.

In contrast to the proliferate studies on the TR model, surprisingly, to our knowl-

[1]Note that estimating the KTR model under the TR model framework violates the full rank condition that is required for a non-degenerated asymptotic distribution of threshold estimator, see, e.g., Hansen (2000) Assumption 1.7.

[2]The KTR model violates the Assumption I.9 for CF-I and II.9 for CF-II in the Yu et al. (2020).

edge, no estimation and asymptotic result has been developed for the least squares estimator of the KTR model with endogeneity.[3] Thus, this paper aims to fill this gap in the literature. Following Yu et al. (2020) and Kourtellos et al. (2022), our paper employs the CF approach to correct the endogeneity in a KTR model. Our proposed method allows both slope regressors and the threshold variable to be endogenous. Compared with Yu et al. (2020) and Kourtellos et al. (2022), our proposed estimator exhibits a joint normal distribution with a standard root-n convergence rate, similar to Hansen (2017). Also, the method works for both diminishing and fixed threshold effects, which fail to be shown under the TR framework. We develop the model both in a time series and a panel data context. Specifically, we explore the estimation and study the asymptotic properties of the least squares estimator for the time series model with weakly dependent data. For the panel, we eliminate the time-invariant fixed effects by using the first-differencing (FD) method and derive the asymptotic results of our proposed estimator with both large numbers of cross-section (N) and time series (T) observations.

The rest of the paper is organized as follows. Section 2 introduces the times-series KTR model with endogeneity, presenting our proposed estimators' estimation method and asymptotic properties. Section 3 extends the model to the panel context. Section 4 reports Monte Carlo simulation results, suggesting our proposed estimator has a good small sample performance. Section 5 concludes the paper. We relegate all the

[3]We notice the first-differenced GMM estimator proposed by Seo and Shin (2016) works for the KTR model with endogeneity. However, it is widely regarded that the GMM method has a poorer finite sample performance than the least squares estimator.

mathematical proofs to supplementary material.

To proceed, we adopt the following notation throughout the paper. We use subscript 0 to denote the true parameters and the accent $\hat{\cdot}$ to denote the estimators. We define $\|\cdot\|$ as the Euclidean norm. The operators $\overset{p}{\to}$ and $\overset{d}{\to}$ denote convergence in probability and distribution, respectively. We denote $(N,T)\to\infty$ as the joint convergence of N and T, when N and T pass to infinity simultaneously. $\mathbf{0}_{A\times B}$ denotes a $A\times B$ matrix of ones and I_m denotes identical matrix of size m.

2 Time series model

2.1 Model and estimation

Following Hansen (2017), we consider a KTR model

$$y_t = \beta_{10}(x_t-\gamma_0)I(x_t<\gamma_0)+\beta_{20}(x_t-\gamma_0)I(x_t\geq\gamma_0)+\beta_{30}'z_t+u_t, t=1,...,n, \quad (1)$$

where x_t is the threshold variable, a scalar. $I(\cdot)$ is the indicator function and z_t is an $\ell\times 1$ vector of regressors, including an intercept term. Model (1) has $k=3+\ell$ parameters to be estimated, including an unknown threshold value γ_0, which is an interior point of a compact set, Γ. Denote the true value $\beta_0=(\beta_{10},\beta_{20},\beta_{30}')'$, which is a $(k-1)\times 1$ vector.

In the KTR framework, we allow both an endogenous threshold variable x_t and endogenous regressors z_{1t}, where z_{1t} is a $d_{z1}\times 1$ vector and it is a subset of $z_t=[z_{1t}',z_{2t}']'$.

The reduced form equations of x_t and z_{1t} are

$$x_t = \pi_{x0}' p_{xt} + v_{xt}, \tag{2}$$

$$z_{1t} = \pi_{z0}' p_{zt} + v_{zt}, \tag{3}$$

where p_{xt} and p_{zt} allow to have duplicate variables, p_{xt} is a $d_{px} \times 1$ vector with $d_{px} \geq 1$ and p_{zt} is a $d_{pz} \times 1$ vector with $d_{pz} \geq d_{z1}$. To simplify notation, we denote all instrumental variables as p_t, which includes the non-overlapping terms in p_{xt} and p_{zt} and p_{xt} and p_{zt} are allowed to share common variables. The endogeneity of the threshold variable x_t and regressors z_{1t} come from the contemporaneous correlation between u_t and v_t, where $v_t = [v_{xt}, v_{zt}']'$ is a $(1 + d_{z1}) \times 1$ vector. Note here we allow $\mathrm{Cov}(v_{xt}, v_{zt}) \neq 0$. Using the control function approach, we assume $E(u_t|\mathcal{F}_{t-1}, x_t, z_{1t}) = E(u_t|v_t) = \beta_{40}' v_t$ almost surely, where $\mathcal{F}_t$ is the smallest sigma-field generated from $\{(x_s, z_{1s}, z_{2,s+1}, u_s, p_{s+1}) : 1 \leq s \leq t \leq n\}$ and β_{40} is a $(1 + d_{z1}) \times 1$ vector. Therefore, we have

$$E(y_t|\mathcal{F}_{t-1}, x_t, z_{1t}) = \beta_{10}(x_t - \gamma_0)I(x_t < \gamma_0) + \beta_{20}(x_t - \gamma_0)I(x_t \geq \gamma_0) + \beta_{30}' z_t + \beta_{40}' v_t. \tag{4}$$

Let $\delta_0 = \beta_{20} - \beta_{10}$. We can rewrite model (1) as

$$y_t = \beta_{10}(x_t - \gamma_0) + \delta_0(x_t - \gamma_0)I(x_t \geq \gamma_0) + \beta_{30}' z_t + \beta_{40}' v_t + \varepsilon_t, \tag{5}$$

where $\varepsilon_t = u_t - \beta_{40}' v_t$. Note that, since $E(\varepsilon_t|x_t, z_{1t}, \mathcal{F}_{t-1}) = 0$ almost surely, model (5) is free of the endogeneity problem. Thus, we can estimate model (5) by the least squares method.

Below, we outline the steps that are taken in the estimation procedure for model (5).

First step: Applying the OLS estimation to models (2) and (3), we obtain the least squares estimator $\hat{\pi}_x = (\sum_{t=1}^{n} p_{xt}p'_{xt})^{-1} \sum_{t=1}^{n} p_{xt}x_t, \hat{\pi}_z = (\sum_{t=1}^{n} p_{zt}p'_{zt})^{-1} \sum_{t=1}^{n} p_{zt}z'_t$ and collect the residuals $\hat{v}_{xt} = x_t - \hat{\pi}'_{xt}p_{xt}$, $\hat{v}_{zt} = z_t - \hat{\pi}'_{zt}p_{zt}$. Then we have $\hat{v}_t = [\hat{v}_{xt}, \hat{v}'_{zt}]'$.

Second step: Let $\theta = (\beta_1, \delta, \beta'_3, \beta'_4)'$, which is a $(k + d_{z1}) \times 1$ vector. Then, by replacing v_t with $\hat{v}_t$ in (5), the least squares objective function of model (5) becomes

$$S_n(\theta, \gamma) = \frac{1}{n}\sum_{t=1}^{n}[y_t - \beta_1(x_t - \gamma) - \delta(x_t - \gamma)I(x_t \geq \gamma) - \beta'_3 z_t - \beta'_4 \hat{v}_t]^2, \tag{6}$$

and the least squares estimator of model (5) solves the following optimization problem:

$$(\hat{\theta}, \hat{\gamma}) = \underset{(\theta,\gamma)\in B\times\Gamma}{\operatorname{argmin}} S_n(\theta, \gamma). \tag{7}$$

Note that $S_n(\theta, \gamma)$ is non-smooth in γ. Therefore, we use a grid search method empirically. For a given $\gamma \in \Gamma$, we obtain the conditional least squares estimator of θ

$$\hat{\theta}(\gamma) = [X(\gamma)'X(\gamma)]^{-1} X(\gamma)'y, \tag{8}$$

where $y = [y_1, y_2, \ldots, y_n]'$, $X(\gamma) = [x_1(\gamma), x_2(\gamma), \ldots, x_n(\gamma)]'$, and $x_t(\gamma) = [x_t - \gamma, (x_t - \gamma)I(x_t \geq \gamma), z'_t, \hat{v}_t{}']'$ for $t = 1, \cdots, n$.

Next, we substitute θ by $\hat{\theta}(\gamma)$ into $S_n(\theta, \gamma)$ and obtain the least squares estimator of γ_0 as

$$\hat{\gamma} = \underset{\gamma\in\Gamma}{\operatorname{argmin}}\, S_n(\hat{\theta}(\gamma), \gamma) = \underset{\gamma\in\Gamma}{\operatorname{argmin}} \frac{1}{n}[y - X(\gamma)\hat{\theta}(\gamma)]'[y - X(\gamma)\hat{\theta}(\gamma)]. \tag{9}$$

Then, the least squares estimator for θ_0 is given by $\hat{\theta} = \hat{\theta}(\hat{\gamma})$.

2.2 Assumptions and limiting results

Below, we list regularity assumptions used to derive the consistency and asymptotic distribution of our proposed estimators.

Assumptions-time series. For some $r > 1$,

T1. $\{(y_t, x_t, z_t, p_t)\}$ is a strictly stationary, ergodic, and absolutely regular sequence with mixing coefficients $\alpha(m) = O(m^{-\xi})$ for some $\xi > r/(r-1)$;

T2. (a) $E|y_t|^{4r} < \infty$, $E|x_t|^{4r} < \infty$, $E\|z_t\|^{4r} < \infty$; (b) $E\|v_t\|^{4r} < \infty$, and $E\|p_t\|^{4r} < \infty$, $E(p_t p_t')$ is nonsingular;

T3. $\inf_{r\in\Gamma} \det Q(\gamma) > 0$, where $Q(\gamma) = E[x_t^*(\gamma)x_t^{*\prime}(\gamma)]$, and $x_t^*(\gamma)$ equals $x_t(\gamma)$ with $\hat{v}_t$ being replaced with v_t;

T4. x_t has a density function $f(x)$ and $f(x) \le \bar{f} < \infty$ over its domain for some finite constant $\bar{f}$;

T5. (a) $E(u_t|\mathcal{F}_{t-1}, x_t, z_t) = E(u_t|v_t) = \beta_{40}' v_t$ almost surely for all t, where $\mathcal{F}_t$ is the smallest sigma-field generated from $\{(x_s, z_s, u_s, p_{s+1}) : 1 \le s \le t \le n\}$; (b) $\{(v_t, \mathcal{F}_{t-1})\}$ is a martingale difference sequence with $E(v_t|\mathcal{F}_{t-1}) = 0$ almost surely;

T6. $\delta_0 \ne 0$ and $\theta_0 \in B \subset \mathcal{R}^{k+d_{z1}}$, where B is compact;

T7. $\gamma_0 = \underset{\gamma\in\Gamma}{\operatorname{argmin}}\, L^*(\theta^*(\gamma), \gamma)$ is unique, where $\theta^*(\gamma) = E[x_t^*(\gamma)x_t^{*\prime}(\gamma)]^{-1}E[x_t^*(\gamma)y_t]$, $L^*(\theta, \gamma) = E[S_n^*(\theta, \gamma)]$, $S_n^*(\theta, \gamma)$ equals $S_n(\theta, \gamma)$ with $\hat{v}_t$ being replaced with v_t, and Γ is compact.

In Assumptions T1, we assume a β-mixing sequence, where the choice of r involves a trade-off between the allowable degree of serial dependence and the number of finite moments; see discussions given in Remark 2.3 of Chan and Tsay (1998) and

Assumption 1.1 of Hansen (2017). Assumption T2 contains unconditional moment conditions. Assumption T2(a) is the regular moment conditions required and T2(b) and T5(b) ensure that the OLS estimators of the reduced form models (2)-(3) exist and converge to the true parameter vector at the root-n rate. Assumption T3 ensures that the parameter estimation is well defined for all $\gamma \in \Gamma$. Assumption T4 makes sure our x_t has a bounded density function. Assumption T5(a) is the assumption for a linear endogenous structure, which can be easily extended to a non-linear endogenous structure.[4] By Assumption T6, we consider a kink regression model. Assumption T7 is an identification assumption, similar to Assumption 2.1 of Hansen (2017).

Next, denote $\phi = (\theta', \gamma)'$, a $(k+1+d_{z1}) \times 1$ vector, and $H_t = H_t(\phi_0)$ with

$$H_t(\phi) = -\frac{\partial}{\partial \phi}[y_t - x_t^{*\prime}(\gamma)\theta] = \begin{pmatrix} x_t^*(\gamma) \\ -\beta_1 - \delta I(x_t \geq \gamma) \end{pmatrix}. \tag{10}$$

Below, we present the limiting results of our proposed estimator.

Theorem 1-Time Series. Under Assumptions T1-T7, as $n \to \infty$, we have

$$\sqrt{n}\left(\hat{\phi} - \phi_0\right) \xrightarrow{d} N\left(0, V\right), \tag{11}$$

where $V = Q^{-1}SQ^{-1}$, $S = E\left(H_t H_t' \varepsilon_t^2\right)$, and $Q = E(H_t H_t')$ is a $(k+1+d_{z1}) \times (k+1+d_{z1})$ matrix.

Remark 1: The proof of Theorem 1-Time Series is given in the appendix. The slope and threshold estimators converge at the square-root-n rate and are jointly normally distributed with a non-zero asymptotic covariance matrix. By contrast, for

[4] see, e.g, Kourtellos et al. (2022) consider this case in a TR model framework.

the discontinuous TR model, the threshold estimator converges faster than square-root-n, and the limiting distribution is non-standard. Thus, the TR model's threshold estimator is asymptotically independent of the slope estimators, which converge at a standard root-n rate. These stark differences originate from the continuous nature of the KTR function. To make an inference, we suggest using the following as the estimator for the asymptotic variance-covariance matrix

$$\hat{V} = \hat{Q}^{-1}\hat{S}\hat{Q}^{-1},$$

where $\hat{Q} = n^{-1}\sum_{t=1}^{n}\hat{H}_t(\hat{\phi})\hat{H}_t'(\hat{\phi})$, $\hat{S} = n^{-1}\sum_{t=1}^{n}\hat{H}_t(\hat{\phi})\hat{H}_t'(\hat{\phi})\hat{\varepsilon}_t^2(\hat{\phi})$ with $\hat{\varepsilon}_t(\hat{\phi}) = y_t - \hat{\beta}_1(x_t - \hat{\gamma}) - \hat{\delta}(x_t - \hat{\gamma})I(x_t \geq \hat{\gamma}) - \hat{\beta}_3' z_t - \hat{\beta}_4' \hat{v}_t$, and

$$\hat{H}_t(\phi) = -\frac{\partial}{\partial \phi}[y_t - x_t'(\gamma)\theta] = \begin{pmatrix} x_t(\gamma) \\ -\beta_1 - \delta I(x_t \geq \gamma) \end{pmatrix}. \tag{12}$$

3 Panel model extension

Many empirical problems of nonlinear asymmetric mechanisms are explicitly within a panel data context, including but not limited to the potential threshold effect of COVID-19 on the unemployment rate that we will discuss more in section 5. Therefore, we extend our baseline time series model to an endogenous kink threshold panel model with unknown fixed effects and cross-sectional independence. Below, we present our model, the estimation strategy, and the asymptotic results.

3.1 Model and estimation

We consider the panel data with both sufficiently large numbers of cross-sectional units N and time periods T.[5] To remove the time-invariant fixed effects, we apply the first-differencing method, deviating from the within-transformation that is used in Zhang et al. (2017). Our panel kink threshold regression model is as follows

$$y_{it} = \beta_{10}(x_{it} - \gamma_0)I(x_{it} < \gamma_0) + \beta_{20}(x_{it} - \gamma_0)I(x_{it} \geq \gamma_0) + \beta_{30}' z_{it} + b_i + u_{it}, \quad (13)$$

for $i = 1, ..., N, t = 1, ..., T$, where y_{it} is the dependent variable, x_{it} is a scalar threshold variable, z_{it} is an $\ell \times 1$ vector of time-varying regressors, which may include the time-variant fixed effect. b_i is the i^{th} unobserved individual fixed effect, which is independent of the errors u_{it} for all t. We denote $\beta_0 = (\beta_{10}, \beta_{20}, \beta_{30}')' \in R^{k-1}$, where $k = 3 + \ell$. The unknown threshold value γ_0 is an interior point of a compact set, Γ. Again, we have the endogenous threshold variable and endogenous regressors $z_{1,it}$, where $z_{1,it}$ is a $d_{z1} \times 1$ vector and it is a subset of $z_{it} = [z_{1,it}', z_{2,it}']'$. The reduced form equations of x_{it} and $z_{1,it}$ are given by

$$x_{it} = \Pi_{x0}' p_{x,it} + b_{x,i} + v_{x,it}, \quad (14)$$

$$z_{1,it} = \Pi_{z0}' p_{z,it} + b_{z,i} + v_{z,it}, \quad (15)$$

[5] Here we let both N and T go to infinity in order to adapt to our empirical applications. Another interesting setup is to allow only N to go to infinity while taking T fixed as discussed in Zhang et al. (2017). Under this case, our proposed estimator will be root-N consistent instead of root-NT consistent.

where $p_{x,it}$ and $p_{z,it}$ allow to have common variables, $p_{x,it}$ is a $d_{px} \times 1$ vector with $d_{px} \geq 1$ and $p_{z,it}$ is a $d_{pz} \times 1$ vector with $d_{pz} \geq d_{z1}$, $b_{x,i}$ and $b_{z,i}$ are the unknown fixed effects, which are independent of $v_{x,it}$ and $v_{z,it}$, respectively. To simplify notation, we denote all instrumental variables by p_{it}, including $p_{x,it}$ and $p_{z,it}$, and $v_{it} = [v_{x,it}, v'_{z,it}]'$, a $(1 + d_{z1}) \times 1$ vector. In addition, we allow $\text{Cov}(v_{x,it}, v_{z,it}) \neq 0$. Using the control function approach, for each i, we assume $E\left(u_{it}|\mathcal{F}_{i,t-1}, x_{it}, z_{1,it}\right) = E\left(u_{it}|v_{it}\right) = \beta'_{40} v_{it}$ almost surely, where $\mathcal{F}_{it}$ is the smallest sigma-field generated from $\{(x_{is}, z_{1,is}, z_{2,i,s+1}, u_{is}, p_{i,s+1}, b_{x,i}, b_{z,i}) : 1 \leq s \leq t \leq T\}$ and β_{40} is a $(1 + d_{z1}) \times 1$ vector. The endogeneity of the threshold variable x_{it} and regressors $z_{1,it}$ come from the contemporaneous correlation between u_{it} and v_{it}.

Applying the first-differencing to model (13) and denoting $\delta_0 = \beta_{20} - \beta_{10}$ yields

$$\Delta y_{it} = \beta_{10} \Delta x_{it} + \delta_0 (X_{it} - \gamma_0 \tau_2)' \mathbf{I}_{it}(\gamma_0) + \beta'_{30} \Delta z_{it} + \Delta u_{it}, \tag{16}$$

where $\Delta a_{it} = a_{it} - a_{i,t-1}$ denotes the first difference of variable a, τ_m is an $m \times 1$ vector of ones, and

$$X_{it} - \gamma_0 \tau_2 = \begin{pmatrix} x_{it} - \gamma_0 \\ x_{i,t-1} - \gamma_0 \end{pmatrix} \quad \text{and} \quad \mathbf{I}_{it}(\gamma_0) = \begin{pmatrix} I(x_{it} \geq \gamma_0) \\ -I(x_{i,t-1} \geq \gamma_0) \end{pmatrix}.$$

Next, we have

$$E(\Delta y_{it}|\mathcal{F}_{i,t-2}, x_{it}, x_{i,t-1}, z_{it}, z_{1,i,t-1}, p_{it}) = \beta_{10} \Delta x_{it} + \delta_0 (X_{it} - \gamma_0 \tau_2)' \mathbf{I}_{it}(\gamma_0) + \beta'_{30} \Delta z_{it} + \beta'_{40} \Delta v_{it}, \tag{17}$$

where applying the law of iterative expectation and using the reduced form equations

(14) and (15) gives

$$
\begin{aligned}
&E(u_{it}|\mathcal{F}_{i,t-2}, x_{it}, x_{i,t-1}, z_{it}, z_{1,i,t-1}, p_{it}) \\
&= E\left[E(u_{it}|\mathcal{F}_{i,t-1}, x_{it}, z_{1,it})|\mathcal{F}_{i,t-2}, x_{it}, x_{i,t-1}, z_{it}, z_{1,i,t-1}, p_{it}\right] \\
&= \beta_{40}' E\left(v_{it}|\mathcal{F}_{i,t-2}, x_{it}, x_{i,t-1}, z_{it}, z_{1,i,t-1}, p_{it}\right) = \beta_{40}' v_{it},
\end{aligned} \tag{18}
$$

and $E(u_{i,t-1}|\mathcal{F}_{i,t-2}, x_{it}, x_{i,t-1}, z_{it}, z_{1,i,t-1}, p_{it}) = E\left(u_{i,t-1}|\mathcal{F}_{i,t-2}, x_{i,t-1}, z_{1,i,t-1}\right) = \beta_{40}' v_{i,t-1}$, since future information does not affect past information.

Thus, combining (16) with (17) gives

$$
\Delta y_{it} = \beta_{10}\Delta x_{it} + \delta_0(X_{it} - \gamma_0\tau_2)\mathbf{I}_{it}(\gamma_0) + \beta_{30}'\Delta z_{it} + \beta_{40}'\Delta v_{it} + \Delta\varepsilon_{it}, \tag{19}
$$

where $\Delta\varepsilon_{it} = \Delta u_{it} - \beta_{40}'\Delta v_{it}$ and $E(\Delta\varepsilon_{it}|\mathcal{F}_{i,t-2}, x_{it}, x_{i,t-1}, z_{it}, z_{1,i,t-1}, p_{it}) = 0$. Hence, by including the auxiliary regressor Δv_{it} into the regression, our model (19) has no endogeneity issue.

Next, we proceed to show the estimation strategy.

First step: Taking the first difference transformation of (14) and (15), we eliminate the unknown fixed effects and obtain

$$
\Delta x_{it} = \Pi_{x0}'\Delta p_{x,it} + \Delta v_{x,it}, \tag{20}
$$

$$
\Delta z_{1,it} = \Pi_{z0}'\Delta p_{z,it} + \Delta v_{z,it}, \tag{21}
$$

Then, we apply the OLS estimation to the first differenced control functions (20) and (21) to acquire the least squares estimators:

$$
\begin{aligned}
\hat{\Pi}_x &= (\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta p_{x,it}\Delta p_{x,it}')^{-1}(\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta p_{x,it}\Delta x_{it}), \\
\hat{\Pi}_z &= (\sum_{i=2}^{N}\sum_{t=1}^{T}\Delta p_{z,it}\Delta p_{z,it}')^{-1}(\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta p_{z,it}\Delta z_{1,it}').
\end{aligned}
$$

We collect the residuals $\Delta\hat{v}_{x,it} = \Delta x_{it} - \hat{\Pi}'_x \Delta p_{x,it}$ and $\Delta\hat{v}_{z,it} = \Delta z_{1,it} - \hat{\Pi}'_z \Delta p_{z,it}$. Let $\Delta\hat{v}_{it} = [\Delta\hat{v}_{x,it}, \Delta\hat{v}'_{z,it}]'$.

Second step: Let $\theta = (\beta_1, \delta, \beta'_3, \beta'_4)' \in R^{k+d_{x1}}$, which is a $(k + d_{z1}) \times 1$ vector. Replacing Δv_{it} by $\Delta\hat{v}_{it}$ in (17), we obtain the least squares criterion function

$$S_{NT}(\theta,\gamma) = \frac{1}{N(T-1)} \sum_{i=1}^{N} \sum_{t=2}^{T} [\Delta y_{it} - \beta_1 \Delta x_{it} - \delta(X_{it} - \gamma\tau_2)\mathbf{I}_{it}(\gamma) - \beta'_3 \Delta z_{it} - \beta'_4 \Delta\hat{v}_{it}]^2. \quad (22)$$

Our least squares estimator is the minimizer of $S_{NT}(\theta,\gamma)$; i.e.,

$$(\hat{\theta},\hat{\gamma}) = \underset{(\theta,\gamma)\in B\times\Gamma}{\operatorname{argmin}} S_{NT}(\theta,\gamma). \quad (23)$$

For a given $\gamma \in \Gamma$, we get the conditional least squares estimator of θ; i.e.,

$$\hat{\theta}(\gamma) = \underset{\theta\in B}{\operatorname{argmin}} \frac{1}{N(T-1)} \sum_{i=1}^{N} \sum_{t=2}^{T} (\Delta y_{it} - \Delta x'_{it}(\gamma)\theta)^2, \quad (24)$$

where $\Delta x_{it}(\gamma) = [\Delta x_{it}, (X_{it} - \gamma\tau_2)'\mathbf{I}_{it}(\gamma), \Delta z'_{it}, \Delta\hat{v}'_{it}]'$.

By solving (24), we have

$$\hat{\theta}(\gamma) = \Big[\sum_{i=1}^{N} \sum_{t=2}^{T} \Delta x_{it}(\gamma)\Delta x'_{it}(\gamma)\Big]^{-1} \Big[\sum_{i=1}^{N} \sum_{t=2}^{T} \Delta x_{it}(\gamma)\Delta y_{it}\Big]. \quad (25)$$

Empirically, we can use a grid search method to obtain $\hat{\gamma}$ by minimizing the sum of squared errors criterion function

$$\hat{\gamma} = \underset{\gamma\in\Gamma}{\operatorname{argmin}} S_{NT}(\hat{\theta}(\gamma),\gamma) = \underset{\gamma\in\Gamma}{\operatorname{argmin}} \frac{1}{N(T-1)} \sum_{i=1}^{N} \sum_{t=2}^{T} \big[\Delta y_{it} - \Delta x'_{it}(\gamma)\hat{\theta}(\gamma)\big]^2. \quad (26)$$

Then, we obtain the estimator of θ_0 with $\hat{\theta} = \hat{\theta}(\hat{\gamma})$.

3.2 Assumptions and limiting results

The assumptions needed for the panel model and its asymptotic theory are collected below.

Assumptions-panel. For some $r > 1$,

P1. (a)$\{(y_{it}, x_{it}, z_{it}, p_{it}) : t = 1, 2, \dots\}$ are independently identically distributed (i.i.d.) across index i; (b) for each i, $\{(y_{it}, x_{it}, z_{it}, p_{it}) : t = 1, 2, \dots\}$ is a strictly stationary, ergodic, and absolutely regular sequence with mixing coefficients $\alpha(m) = O(m^{-\xi})$ for some $\xi > r/(r-1)$;

P2. (a) $E|\Delta y_{it}|^{4r} < \infty$, $E|\Delta x_{it}|^{4r} < \infty$, $E\left\|\Delta z_{it}\right\|^{4r} < \infty$; (b)$E\left\|\Delta v_{it}\right\|^{4r} < \infty$, and $E\left\|\Delta p_{it}\right\|^{4r} < \infty$, $E(\Delta p_{it}\Delta p'_{it})$ is non-singular;

P3. $\inf_{\gamma\in\Gamma} \det Q(\gamma) > 0$, where $Q(\gamma) = E[\Delta x^*_{it}(\gamma)\Delta x^{*\prime}_{it}(\gamma)]$ and Δx^*_{it} equals $\Delta x_{it}(\gamma)$ with $\Delta\hat{v}_{it}$ being replaced with Δv_{it};

P4. x_{it} has a density function $f(x)$ and $f(x) \leq \bar{f} < \infty$ over its domain for a finite real number $\bar{f}$;

P5. For each i, (a) $\{v_{it}, \mathcal{F}_{i,t-1}\}$ is a martingale difference sequence with $E(v_{it}|\mathcal{F}_{i,t-1}) = 0$, where $\mathcal{F}_{it}$ is the smallest sigma-field generated from $\{(x_{is}, z_{1,is}, z_{2,i,s+1}, u_{is}, p_{i,s+1}, b_{x,i}, b_{z,i}) : 1 \leq s \leq t \leq T\}$; (b) $E(u_{it}|\mathcal{F}_{i,t-1}, x_{it}, z_{1,it}) = E(u_{it}|v_{it}) = \beta'_{40}v_{it}$ almost surely; (c) b_i is independent of the error term u_{it} for all t;

P6. $\delta_0 \neq 0$ and $\theta_0 \in B \subset \mathcal{R}^{k+d_{z1}}$, where B is compact;

P7. $\gamma_0 = \underset{\gamma\in\Gamma}{\operatorname{argmin}}\, L^*(\theta^*(\gamma), \gamma)$ is unique, where $\theta^*(\gamma) = E[\Delta x^*_{it}(\gamma)\Delta x^{*\prime}_{it}(\gamma)]^{-1}E[\Delta x^*_{it}(\gamma)\Delta y_{it}]$, $L^*(\theta^*(\gamma), \gamma) = E[S^*_{NT}(\theta^*(\gamma), \gamma)]$, $S^*_{NT}(\theta, \gamma)$ equals $S_{NT}(\theta, \gamma)$ with $\Delta\hat{v}_{it}$ being replaced with Δv_{it}, and Γ is compact.

In Assumption P1(a), we assume cross-sectional independence. Assumption P1(b) assumes a β-mixing sequence across index t, where the choice of r involves a trade-off between the allowable degree of serial dependence and the number of finite moments.

And the asymptotics are taken in large N and large T. Note Zhang et al. (2017) only allow N to go to infinity and treat T as fixed. Assumption P2 contains unconditional moment conditions. Assumption P2(a) is the regularity moment conditions required and P2(b) and P5(b) ensure that the OLS estimators of the first differencing reduced form models (20)-(21) exist and converge to the true parameter vector at the root-NT rate. Assumption P3 ensures that the parameter estimation is well defined for all $\gamma \in \Gamma$. Assumption P4 makes sure our x_{it} has a bounded density function. Assumption P5(a) is the assumption for a linear endogenous structure, which can be easily extended to a non-linear endogenous structure. Assumption P5(c) assumes the unobserved individual effect b_i is independent of the errors u_{it} for all t, which is a standard assumption in the panel data model. By Assumption P6, we consider a kink regression model. Assumption P7 is an identification assumption, similar to Assumption 2.1 of Hansen (2017).

Denote $\phi = (\theta', \gamma)'$ and let

$$\Delta H_{it}(\phi) = -\frac{\partial}{\partial \phi}[\Delta y_{it} - \Delta x_{it}^{*\prime}(\gamma)\theta] = \begin{pmatrix} \Delta x_{it}^{*}(\gamma) \\ -\delta[I(x_{it} \geq \gamma) - I(x_{i,t-1} \geq \gamma)] \end{pmatrix}, \tag{27}$$

and $\Delta H_{it} = \Delta H_{it}(\phi_0)$.

Theorem 2-panel. Under Assumptions P1-P7, as $(N,T) \to \infty$, we have

$$\sqrt{NT}\left(\hat{\phi} - \phi_0\right) \xrightarrow{d} N\left(0, \mathcal{V}\right), \tag{28}$$

where $\mathcal{V} = \mathcal{Q}^{-1}\mathcal{S}\mathcal{Q}^{-1}$, $\mathcal{S} = E\left(\Delta H_{it}\Delta H_{it}'\Delta\varepsilon_{it}^2\right)$, and $\mathcal{Q} = E(\Delta H_{it}\Delta H_{it}')$, is a $(k+1+d_{z1}) \times (k+1+d_{z1})$ matrix.

Remark 2: The proof is provided in the appendix. Similar to the time series model, our slope and threshold estimators are jointly normally distributed with root-NT convergence rate and they have a non-zero asymptotic covariance matrix. To make an inference, we estimate the asymptotic variance-covariance matrix by

$$\hat{\mathcal{V}} = \hat{\mathcal{Q}}^{-1}\hat{\mathcal{S}}\hat{\mathcal{Q}}^{-1},$$

where $\hat{\mathcal{Q}} = \frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta\hat{H}_{it}(\hat{\phi})\Delta\hat{H}'_{it}(\hat{\phi})$ and $\hat{\mathcal{S}} = \frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta\hat{H}_{it}(\hat{\phi})\Delta\hat{H}'_{it}(\hat{\phi})\Delta\hat{\varepsilon}^2_{it}(\hat{\phi})$. Here $\Delta\hat{\varepsilon}_{it}(\hat{\phi}) = \Delta y_{it} - \hat{\beta}_1\Delta x_{it} - \hat{\delta}(X_{it} - \hat{\gamma}\tau_2)\mathbf{I}_{it}(\hat{\gamma}) - \hat{\beta}'_3\Delta z_{it} - \hat{\beta}'_4\Delta\hat{v}_{it}$ and

$$\Delta\hat{H}_{it}(\phi) = -\frac{\partial}{\partial\phi}[\Delta y_{it} - \Delta x'_{it}(\gamma)\theta] = \begin{pmatrix} \Delta x_{it}(\gamma) \\ -\delta[I(x_{it} \geq \gamma) - I(x_{i,t-1} \geq \gamma)] \end{pmatrix}. \quad (29)$$

4 Monte Carlo simulations

This section contains Monte Carlo simulations to evaluate the finite sample performance of our proposed estimator. Below, we list the four data generating processes (DGPs)– two give time series data and two yield panel data.

DGP1:

$$y_t = c_0 + \beta_{10}x_t + \delta_0(x_t - \gamma_0)I(x_t \geq \gamma_0) + u_t, \quad u_t = 0.2u_{t-1} + \kappa v_t, \quad (30)$$

$$x_t = 2 + v_t + p_t, \quad t = 1, ..., n. \quad (31)$$

DGP2:

$$y_t = c_0 + \beta_{10}x_t + \delta_0(x_t - \gamma_0)I(x_t \geq \gamma_0) + \beta_{30}z_t + u_t, \quad u_t = 0.2u_{t-1} + \kappa(v_{1t} + v_{2t}), \quad (32)$$

$$x_t = 2 + (0.9v_{1t} + 0.1v_{2t}) + p_{1t}, \quad z_t = 2 + (0.1v_{1t} + 0.9v_{2t}) + p_{2t}, \tag{33}$$

$$t = 1, ..., n.$$

DGP3:

$$y_{it} = c_0 + \beta_{10}x_{it} + \delta(x_{it} - \gamma_0)I(x_{it} \geq \gamma_0) + b_i + u_{it}, \quad u_{it} = 0.2u_{i,t-1} + \kappa v_{it}, \tag{34}$$

$$x_{it} = 2 + v_{it} + p_{it}, \quad i = 1, ..., N, t = 1, ..., T. \tag{35}$$

DGP4:

$$y_{it} = c_0 + \beta_{10}x_{it} + \delta_0(x_{it} - \gamma_0)I(x_{it} \geq \gamma_0) + \beta_{30}z_{it} + b_i + u_{it}, \quad u_{it} = 0.2u_{i,t-1} + \kappa(v_{1,it} + v_{2,it}), \tag{36}$$

$$x_{it} = 2 + (0.9v_{1,it} + 0.1v_{2,it}) + p_{1t} \quad z_{it} = 2 + (0.1v_{1,it} + 0.9v_{2,it}) + p_{2t} \tag{37}$$

$$i = 1, ..., N, t = 1, ..., T.$$

In the time series setup, we consider two different data-generating processes, DGP1, and DGP2. In DGP1, we only allow the threshold variable to be endogenous, while in DGP2, we allow both the threshold variable x_t and slope regressor z_t to be endogenous. The endogeneity of x_t in DGP1 comes from the common factor v_t between x_t and u_t. In DGP2, the endogeneity of (x_t, z_t) comes from the common factors v_{1t} and v_{2t} shared with u_t. DGP3 and DGP4 are designed for the panel KTR context. Specifically, DGP3 allows the threshold variable x_t to be endogenous, and DGP4 allows both the threshold variable and regressors to be endogenous. In DGP3, the endogeneity of x_{it} comes from the common factor v_{it} between x_{it} and u_{it}. In DGP4, the endogeneity of (x_{it}, z_{it}) comes from the common factors $v_{1,it}$ and

$v_{2,it}$ shared with u_t. For all data-generating processes, the error terms are stationary AR(1) sequences. We use κ to control the severity of endogeneity and we set $c_0 = \beta_{10} = \delta_0 = \beta_{30} = 1$, and $\gamma_0 = 2$.

In DGP1, $(v_t, p_t) \sim i.i.d.N(0, I_2)$, where p_t is our instrumental variable . In DGP2, $(v_{1t}, v_{2t}, p_{1t}, p_{2t}, \varepsilon_t) \sim i.i.d.N(0, I_5)$ and p_{1t} and p_{2t} are the instrumental variable for x_t and z_t, respectively. In DGP3, we generate $(v_{it}, p_{it}) \sim i.i.d.N(0, I_2)$, and the unknown fixed effects, b_i ~i.i.d.$N(0,1)$. In DGP4, we have $(v_{1,it}, v_{2,it}, p_{1,it}, p_{2,it}) \sim i.i.d.N(0, I_4)$ and b_i represents individual fixed effects with distribution $N(0,1)$ across i. With $\kappa \in \{0.05, 0.5, 0.95, 2\}$, we check the performance of our estimator under low, moderate, and high endogenous severity.[6] We set the sample size n = 100, 200, 300, and 400 for GDP1 and DGP2, and N = 10,20,30, and 40 and T = 10,20,30, and 40 for DGP3 and DGP4. The number of Monte Carlo replications is 5,000. Tables 1-2 , 3-4 report the root mean squared errors (RMSEs) for our proposed estimator for DGP1 and DGP3, respectively[7].

[Table 1; Table 2]

Table 1 and2 display the Monte Carlo simulation results for our DGP1. We compare the results of our proposed estimator and the least squares estimator ignoring the endogeneity issue under different sample sizes. We have the following observations.

[6]To save space, we only report the panel model results with severe endogeneity (i.e. $\kappa = 2$).The results for other cases are available from the authors upon request.

[7]The online supplementary material collects the Monte Carlo results of DGP2 and DGP4 in Section ??. The results confirm our reports from DGP1 and DGP3, respectively.

First, we find that, as the number of observations increases, the RMSE without control functions remains large as the endogenous severity rises (κ increases). For example, without using the control function correction approach, the RMSE for β_1 barely decreases, even with a mild degree of endogeneity. By contrast, with the control function correction, the RMSEs for all parameters decrease rapidly as the sample size increases, confirming the validity of our CF approach to tackling endogeneity.

[Table 3; Table 4]

Tables 3 and 4 give the Monte Carlo simulation results for DGP3. With severe endogeneity, the findings are similar to those in the time series model.

5 Conclusion

As in Hansen (2017), we consider a kink threshold model and allow for possible endogenous threshold variables and linear regressors. Following Kourtellos et al. (2016) and Yu et al. (2020), we extend the usage of the control function approach to tackle the problem and propose an estimator that achieves a standard joint normal distribution. Compared with other methods dealing with endogeneity in the context of a threshold regression model our method is easy to apply and more reliable, especially with a small sample size. Also, the continuity of the regression function and the joint normal distribution of our proposed estimator provide more useful possibilities for economic applications. Monte Carlo simulations show that the small sample performance of our proposed estimator is quite satisfactory both for time series and panel cases.

Our method has several possible extensions. Instead of introducing bias correction terms linearly ($\beta_{40}' v_t$ in (5)), one can introduce a nonparametric endogeneity correction term ($g(v_t)$, where $g(\cdot)$ is an unknown function), like Kourtellos et al. (2022) in the threshold regression model. Also, one can relax the linear specification in reduced-form functions (2)-(3) to a more flexible semi-/nonparametric specification. Finally, it is interesting to allow the coefficients of the reduced form equations to be heterogeneous across i. In that case, both N and T going to infinity is not optional but necessary. We leave these for future research.

References

Caner, M. and B. E. Hansen (2004). Instrumental variable estimation of a threshold model. *Econometric Theory 20*(05), 813–843.

Chan, K. S. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *The Annals of Statistics 21*(1), 520–533.

Chan, K. S. and R. S. Tsay (1998, 06). Limiting properties of the least squares estimator of a continuous threshold autoregressive model. *Biometrika 85*(2), 413–426.

Christopoulos, D., P. McAdam, and E. Tzavalis (2021). Dealing with endogeneity in threshold models using copulas. *Journal of Business & Economic Statistics 39*(1), 166–178.

Table 1: Estimation Results for DGP1

	MEAN-β ($\beta_{10}=1$)	RMSE-β	MEAN-δ ($\delta_0=1$)	RMSE-δ	MEAN-γ ($\gamma_0=2$)	RMSE-γ
$\kappa=0.05$	No CF					
T=100	1.0249	0.0255	0.9999	0.009	2	0
T=200	1.025	0.0252	1	0.0063	2	0
T=300	1.025	0.0251	0.9999	0.005	2	0
T=400	1.025	0.0251	1	0.0044	2	0
$\kappa=0.05$	CF					
T=100	0.9969	0.0174	1	0.0025	2	0
T=200	0.9986	0.0093	1	0.0017	2	0
T=300	0.9992	0.0069	1	0.0014	2	0
T=400	0.9992	0.0061	1	0.0012	2	0
$\kappa=0.5$	NO CF					
T=100	1.2459	0.2556	1.0083	0.0915	2.0045	0.1531
T=200	1.2472	0.2515	1.0036	0.0632	1.9975	0.1002
T=300	1.2489	0.2516	1.0017	0.0502	2.0015	0.0822
T=400	1.2488	0.2509	1.0017	0.0444	1.9992	0.0721
$\kappa=0.5$	CF					
T=100	0.9684	0.1744	1.0006	0.0256	2.0006	0.041
T=200	0.9863	0.0936	1.0001	0.0174	2	0.021
T=300	0.992	0.0695	0.9999	0.014	1.9999	0.011
T=400	0.9923	0.0609	0.9998	0.0119	2.0001	0.0068

NOTE: CF=Control function approach

Table 2: Estimation Results for DGP1(continue)

	MEAN-β ($\beta_{10}=1$)	RMSE-β	MEAN-δ ($\delta_0=1$)	RMSE-δ	MEAN-γ ($\gamma_0=2$)	RMSE-γ
$\kappa=0.95$	NO CF					
T=100	1.4547	0.4783	1.0409	0.1853	2.0013	0.3474
T=200	1.4642	0.4736	1.017	0.1204	1.9958	0.213
T=300	1.4708	0.4766	1.0102	0.0994	2.0027	0.165
T=400	1.469	0.4732	1.0116	0.0847	1.9998	0.1379
$\kappa=0.95$	CF					
T=100	0.9342	0.3088	1.0011	0.0483	1.9997	0.0782
T=200	0.9716	0.1756	1.0004	0.0324	1.9982	0.0549
T=300	0.9852	0.1353	1.0003	0.0268	2.0006	0.0443
T=400	0.989	0.1149	1.0001	0.0227	1.9998	0.0368
$\kappa=2$	NO CF					
T=100	1.905	0.9721	1.1914	0.4693	2.0048	0.707
T=200	1.9475	0.9751	1.0986	0.2855	1.9976	0.5218
T=300	1.9729	0.9884	1.0604	0.2234	2.0074	0.414
T=400	1.9741	0.9852	1.0519	0.1873	2.0008	0.3481
$\kappa=2$	CF					
T=100	0.8568	0.6511	1.0092	0.1018	1.9955	0.1711
T=200	0.9389	0.3701	1.0038	0.0683	1.9971	0.1143
T=300	0.9681	0.2846	1.0024	0.0564	2.0022	0.0914
T=400	0.976	0.2421	1.0016	0.0477	1.9994	0.0795

NOTE: CF=Control function approach

Table 3: Estimation Results for DGP3

		MEAN-β_1 ($\beta_{10}=1$)	RMSE-β_1	MEAN-δ ($\delta_0=1$)	RMSE-δ	MEAN-γ ($\gamma_0=2$)	RMSE-γ
NO CF	FD						
T=10	N=10	1.7878	0.8682	1.2221	0.4992	1.9805	0.7298
	N=20	1.8474	0.8819	1.1121	0.3057	2.0028	0.5573
	N=30	1.867	0.8876	1.069	0.2377	2.0002	0.4444
	N=40	1.8725	0.8869	1.0513	0.1993	1.9919	0.3793
T=20	N=10	1.8388	0.8745	1.113	0.3031	1.9792	0.5532
	N=20	1.8786	0.8927	1.0471	0.1941	2.004	0.3813
	N=30	1.8842	0.8929	1.0298	0.1529	1.9958	0.297
	N=40	1.8907	0.8967	1.0193	0.1309	1.9976	0.2401
T=30	N=10	1.867	0.8879	1.0621	0.2335	1.9968	0.442
	N=20	1.8872	0.8954	1.0273	0.154	2.0014	0.2933
	N=30	1.8921	0.8971	1.0181	0.1223	2.0031	0.2202
	N=40	1.8958	0.8996	1.0121	0.1074	2.0031	0.1865
T=40	N=10	1.8779	0.8924	1.0462	0.2001	2.0072	0.3807
	N=20	1.8919	0.8978	1.0193	0.131	2.0052	0.2421
	N=30	1.8954	0.8993	1.0128	0.1061	2.005	0.1903
	N=40	1.8964	0.899	1.0075	0.0912	2.0002	0.155

NOTE: FD=first difference; CF=control function approach;

Table 4: Estimation Results for DGP3(continue)

		MEAN-β_1 ($\beta_{10}=1$)	RMSE-β_1	MEAN-δ ($\delta_0=1$)	RMSE-δ	MEAN-γ ($\gamma_0=2$)	RMSE-γ
CF	FD						
T=10	N=10	0.9656	0.2541	1.0169	0.1148	1.9975	0.192
	N=20	0.9824	0.171	1.006	0.077	1.9991	0.1249
	N=30	0.9905	0.1355	1.0032	0.0605	1.9995	0.0998
	N=40	0.992	0.1192	1.0036	0.0532	2.0016	0.0846
T=20	N=10	0.9867	0.1661	1.005	0.0653	1.9979	0.1055
	N=20	0.9911	0.1138	1.0025	0.0451	1.9998	0.0738
	N=30	0.9942	0.0943	1.0012	0.0365	1.9997	0.061
	N=40	0.995	0.0807	1.0004	0.0315	2.0011	0.0536
T=30	N=10	0.9887	0.131	1.0024	0.0493	2.0005	0.0805
	N=20	0.9985	0.0934	1.0008	0.0341	1.9994	0.0577
	N=30	0.9971	0.0755	1.0008	0.0281	1.9994	0.0472
	N=40	0.9988	0.0649	1.0003	0.0241	1.9997	0.0391
T=40	N=10	0.9932	0.1161	1.0009	0.0413	1.9999	0.0679
	N=20	0.9985	0.0792	1.0007	0.0285	2.0006	0.0477
	N=30	0.9965	0.066	1	0.0232	2.0003	0.038
	N=40	0.9973	0.0563	1	0.0205	2.0005	0.0295

NOTE: FD=first difference; CF=control function approach;

Hansen, B. E. (2000). Sample splitting and threshold estimation. *Econometrica 68*, 575–603.

Hansen, B. E. (2017). Regression kink with an unknown threshold. *Journal of Business & Economic Statistics 35*(2), 228–240.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica 47*(1), 153–161.

Hidalgo, J., J. Lee, and M. H. Seo (2019). Robust inference for threshold regression models. *Journal of Econometrics 210*(2), 291–309.

Kourtellos, A., T. Stengos, and Y. Sun (2022). Endogeneity in semiparametric threshold regression. *Econometric Theory 38*(3), 562–595.

Kourtellos, A., T. Stengos, and C. M. Tan (2016). Structural threshold regression. *Econometric Theory 32*(4), 827–860.

Newey, W. K., J. L. Powell, and F. Vella (1999). Nonparametric estimation of triangular simultaneous equations models. *Econometrica 67*(3), 565–603.

Seo, M. H. and Y. Shin (2016). Dynamic panels with threshold effect and endogeneity. *Journal of Econometrics 195*(2), 169–186.

Tong, H. (1990). *Non-linear time series: a dynamical system approach.* Oxford university press.

Yu, P., Q. Liao, and P. C. B. Phillips (2020). New control function approaches in threshold regression with endogeneity. *mimeo*.

Yu, P. and P. C. Phillips (2018). Threshold regression with endogeneity. *Journal of Econometrics 203*(1), 50–68.

Zhang, Y., Q. Zhou, and L. Jiang (2017). Panel kink regression with an unknown threshold. *Economics Letters 157*, 116–121.