PERUVIAN ECONOMIC ASSOCIATION

Maximum Likelihood Estimation of Dynamic Panel Threshold Models

Nelson Ramírez-Rondán

Working Paper No. 32, March 2015

Maximum Likelihood Estimation of Dynamic Panel Threshold Models*

Nelson Ramírez-Rondán

February 2015

Abstract

Threshold estimation methods are developed for dynamic panels with individual fixed specific effects covering short time periods. Maximum likelihood estimation of the threshold and the slope parameters is proposed using first difference transformations. Threshold estimate is shown to be consistent and it converges to a double-sided standard Brownian motion distribution, when the number of individuals grows to infinity for a fixed time period; and the slope estimates are consistent and asymptotically normally distributed. The method is applied to a sample of 72 countries and 8 periods of 5-year averages to determine the effect of inflation rate on long-run economic growth.

JEL Classification: C13; C23

Keywords: Threshold Models, Dynamic Panel Data, Maximum Likelihood Estimation, Inflation, Economic Growth.

*I am indebted to Bruce Hansen for his guidance, discussions and suggestions in this project. I also would like to thank Jack Porter, Andrés Aradillas-López, Xiaoxia Shi, Chunming Zhang, Arthur Lewbel, Antonio Galvao, and David Jacho-Chávez as well as the participants of the 2014 Annual Meeting of the Latin American Econometric Society (Sao Paulo, Brazil), 2012 Meeting of the Midwest Econometric Group (Lexington, KY), Central Bank of Peru Research Seminar and the University of Wisconsin-Madison Econometrics Workshop, Lunch Seminar and Reading Group for their discussions and useful comments. We also are grateful to Linda Kaltani and Norman Loayza for providing the dataset. All remaining errors are mine. Nelson Ramírez-Rondán is a researcher in the Research Division at the Central Bank of Peru (Email: nramron@gmail.com).

1 Introduction

One of the most interesting forms of non-linear regression models is the threshold regression model developed by Tong (1983)[1] This model has been enormously influential in economics and has popularity in current applied econometric practice. The model splits the sample into classes based on the value of an observed variable, whether or not it exceeds some threshold. That is, the model internally sorts the data on the basis of some threshold determinant into groups of observations, each of which obeys the same model. Hansen (2011) provides an excellent literature review of these models in econometrics and in empirical economics.

Hansen (1999) extended those models to a static panel data model, who proposes econometric techniques for threshold effects with exogenous regressors and exogenous threshold variable, where least squares estimation is proposed using fixed-effects transformation. A challenging extension of Hansen's (1999) work is considering lags of the dependent variable as regressors in the panel data model, in other words considering a dynamic panel threshold model. Hansen's (1999) techniques cannot be used, because any transformation to eliminate the individual fixed specific effect introduce a correlation between the transformed regressors and the transformed error term in the model.

The most popular approaches to estimate a dynamic panel data linear model are Instrumental Variables (IV) and General Method of Moments (GMM) using a first difference transformation of the model, which uses higher lags as valid instruments. In the context of threshold regression, Caner and Hansen (2002) proposed a two-stage least squares estimation for a model with endogenous regressors and exogenous threshold variable. Nevertheless this approach cannot be applied because it needs an explicit reduce form that relates the endogenous variables with the instruments;[2] reduced form equation which is not available in the dynamic panel data model in first differences.

In the context of the dynamic panel data linear model, Hsiao, Pesaran and Tahmiscioglu (2002) propose a maximum likelihood estimator using a first difference transformation. This approach has the advantage that does not require instruments, but needs assumptions on the initial conditions. Thus, upon Hansen (1999) and Hsiao et al. (2002) works, in this paper we propose a maximum likelihood approach to estimate the threshold and slope parameters in dynamic panel threshold models.

In empirical macroeconomics it is crucial to consider dynamics, because all macroeconomic models exhibit dynamics, for example, one of the main implications of the neoclas-

[1]See Tong (2007) for the birth of the threshold model.

[2]Yu (2013) finds that the two-stage least squares estimator of Caner and Hansen (2002) of the threshold parameter is inconsistent without the stronger assumption of the reduced form equation.

sical growth model and indeed of all models is that exhibit transitional dynamics in which the growth rate relies on the previous position of the economy.[3] Thus, dynamic methods are suitable for empirical macroeconomic models.

The outline of the paper is as follows. Section 2 introduces the dynamic panel threshold model with fixed effects. Section 3 develops the estimation procedure for the threshold and slope parameters via maximum likelihood of the model using first difference transformations. Section 4 extends the model by allowing exogenous regressors. Section 5 discusses the estimation of a multiple threshold model. Section 6 establishes consistency and the asymptotic distribution of the parameter estimates. Section 7 shows the performance of the estimators proposed via a Monte Carlo experiments. Section 8 reports an application to the threshold relationship between inflation and long-run economic growth. Finally, section 9 concludes. Proofs of theorems are provided in the appendix.

2 Model

The observed data are from a balanced panel $\{y_{it}, x_{it} : 1 \leq i \leq n, 1 \leq t \leq T\}$. The subscript i indexes the individual and the subscript t indexes time. The dependent variable y_{it} is scalar. The threshold variable $q_{it} = q(x_{it})$ is an element or function of the vector x_{it} of exogenous variables[4] and is assumed to have a continuous distribution. The structural equation of interest is

$$y_{it} = \alpha_i + \beta_1 y_{it-1} 1(q_{it} \leq \gamma) + \beta_2 y_{it-1} 1(q_{it} > \gamma) + u_{it}, \tag{1}$$

where the threshold parameter is $\gamma \in \Gamma$, and Γ is a strict subset of the support of q_{it}. This parameter is unknown and needs to be estimated. $\beta = (\beta_1, \beta_2)'$ are the slope parameters of interest assumed to be different one to each other; α_i is the individual specific effect assumed to be fixed and u_{it} is the error term, assumed to be independently identically normally distributed with mean 0 and variance σ_u^2. We also assume the initial values, y_{i0} and x_{i0}, are observable.

When the individual specific effects, α_i, are fixed, the least-squared dummy variable (LSDV) estimators of the linear version of model (1) leads to an inconsistency of the

[3]Other macroeconomic models, such the Phillips curve, Taylor rule, aggregate demand and aggregate supply, include the lagged of the dependent variable as a regressor.

[4]Endogeneity of the threshold variable is an important unsolved topic in such models. Recent efforts working on endogeneity in the threshold variable are Kourtellos et al. (2013), who propose a two stage concentrated least squares method that involves an inverse Mills ratio bias correction term in each regime; Yu and Phillips (2014), who propose an integrated difference kernel estimator; and Seo and Shin (2014), who propose a GMM estimator in a dynamic panel threshold model.

slope parameter estimator as n grows to infinity for a fixed T (Nickell, 1981). If the errors u_{it} are normally distributed, then the LSDV are also the maximum likelihood estimator (MLE), conditional on the initial observation, y_{i0}, the MLE also leads an inconsistency of the slope parameter estimator, due to the classical incidental parameter problem in which the number of parameters increases with the number of observations (Lancaster, 2000).

To address the incidental parameter problem we take the first difference to eliminate the individual specific effect in model (1), and we get

$$\begin{aligned} y_{it} - y_{it-1} = \beta_1 (y_{it-1} 1(q_{it} \leq \gamma) - y_{it-2} 1(q_{it-1} \leq \gamma)) \\ + \beta_2 (y_{it-1} 1(q_{it} > \gamma) - y_{it-2} 1(q_{it-1} > \gamma)) + u_{it} - u_{it-1}. \quad (2) \end{aligned}$$

To simplify notation, let $\Delta y_{it} \equiv y_{it} - y_{it-1}$, $\Delta y^{*}_{it-1}(\gamma) \equiv y_{it-1} 1(q_{it} \leq \gamma) - y_{it-2} 1(q_{it-1} \leq \gamma)$, $\Delta y^{+}_{it-1}(\gamma) \equiv y_{it-1} 1(q_{it} > \gamma) - y_{it-2} 1(q_{it-1} > \gamma)$ and $\Delta u_{it} \equiv u_{it} - u_{it-1}$. Then equation (2) becomes

$$\Delta y_{it} = \beta_1 \Delta y^{*}_{it-1}(\gamma) + \beta_2 \Delta y^{+}_{it-1}(\gamma) + \Delta u_{it}. \quad (3)$$

Notice that for $t = 2, 3, \ldots, T$, (3) is well defined, but not for Δy_{i1} because $\Delta y^{*}_{i0}(\gamma)$ and $\Delta y^{+}_{i0}(\gamma)$ are missing; that is, $y_{i,-1}$ is not available.

When the time period is fixed, or the panel covers only a short period, the MLE of the dynamic panel linear model depends on the initial condition and the assumption on the initial condition plays a crucial role in devising consistent estimates. Anderson and Hsiao (1981) show under which assumptions the MLE leads to consistent or inconsistent estimates of the slope parameter. This problem arises because the covariance matrix depends on the initial conditions; if T grows to infinity the initial condition problem disappears.

By continuous substitution of equation (3) for the first period, Δy_{i1}, the resulting equation has an intractable form and depends on the structural parameters. Also, it is clear that equation (3) does not depend on the individual specific fixed effect for all t. Thus, to address the initial condition problem, we assume the process has started from a finite period in the past, namely for given values of $y_{i,-1}$ such that[5]

$$E(\Delta y_{i1} | x_i) = \delta_1 1(q_{i1} \leq \gamma) + \delta_2 1(q_{i1} > \gamma),$$

where $x_i = (x_{i0}, x_{i1}, ..., x_{iT})'$. This assumption imposes the restriction that the expected

[5]It is enough to assume that $E(\Delta y_{i1}) = \delta$, for empirical applications it simplifies the procedure.

changes in the initial endowments are the same across all individuals in each regime, but does not necessarily require if the process has reached stationarity.[6] In the dynamic panel linear model, Hsiao et al. (2002) assume $E(\Delta y_{i1}) = b$, while Blundell and Smith (1991) assume $E(y_{i0}) = b$; however, Blundell and Smith (1991) assume random effects models, where there is no incidental parameter issue. Then the marginal distribution of Δy_{i1} conditional on x_i can be written as

$$\Delta y_{i1} = \delta_1 1(q_{i1} \leq \gamma) + \delta_2 1(q_{i1} > \gamma) + \upsilon_{i1}. \tag{4}$$

The auxiliary external parameters, $\delta = (\delta_1, \delta_2)'$, can be a function of the structural parameters, but similar to Hsiao et al. (2002) we can treat the external parameters as free parameters in the sense they do not depend on the structural parameters.

3 Estimation

In this section we propose a Maximum Likelihood approach to estimate equations (3) and (4) together.

3.1 Maximum Likelihood Function

Under the strict exogeneity of x_{it} and by construction, $E(\upsilon_{i1}|x_i) = 0$, $E\upsilon_{i1}^2 = \sigma_\upsilon^2$. We also assume $Cov(\upsilon_{i1}, \Delta u_{i2}|x_{it}) = -\sigma_u^2$ and $Cov(\upsilon_{i1}, \Delta u_{it}|x_{it}) = 0$, for $t = 3, ..., T$, $i = 1, ..., n$; that is, we assume homoscedasticity across regimes.

Let $\Delta y_i = (\Delta y_{i1}, \Delta y_{i2}, ..., \Delta y_{iT})'$ and $\Delta u_i = (\upsilon_{i1}, \Delta u_{i2}, ..., \Delta u_{iT})'$. The Jacobian of the transformation from Δu_i to Δy_i is unity and the joint probability distribution function of Δy_i and Δu_i are therefore the same. The covariance matrix of Δu_i has the form

$$\Omega = \sigma_u^2 \begin{bmatrix} \omega & -1 & 0 & \dots & 0 \\ -1 & 2 & -1 & & \\ 0 & -1 & 2 & & \\ \vdots & & & \ddots & -1 \\ 0 & & & -1 & 2 \end{bmatrix} = \sigma_u^2 \Omega^\star, \tag{5}$$

where $\omega = \sigma_\upsilon^2 / \sigma_u^2$.

Under the assumption that u_{it} is independent normal, the joint probability distribution

[6] Sufficient conditions for the stationarity of the process in the autoregressive threshold model are $|\beta_1| < 1$ and $|\beta_2| < 1$ (see Enders and Granger (1998) and Caner and Hansen (2001)).

function of Δy_i conditional on x_i is given by[7]

$$L(\delta,\beta,\gamma,\sigma_u^2,\omega)=\prod_{i=1}^{n}(2\pi)^{-T/2}|\Omega|^{-1/2}\exp\{-\frac{1}{2}\Delta u_i(\delta,\beta,\gamma)'\Omega^{-1}\Delta u_i(\delta,\beta,\gamma)\}, \qquad (6)$$

where we define $\Delta u_i(\delta,\beta,\gamma)=[\Delta y_{i1}-\delta_1 1(q_{i1}\leq\gamma)-\delta_2 1(q_{i1}>\gamma),\Delta y_{i2}-\beta_1\Delta y_{i1}^*(\gamma)-\beta_2\Delta y_{i1}^+(\gamma),\ldots,\Delta y_{iT}-\beta_1\Delta y_{iT-1}^*(\gamma)-\beta_2\Delta y_{iT-1}^+(\gamma)]'$.

The likelihood function (6) is well defined, depends on a fixed number of parameters. Maximizing the likelihood function (6) is equivalent to Maximizing

$$\ln L(\delta,\beta,\gamma,\sigma_u^2,\omega)=-\frac{nT}{2}\ln(2\pi)-\frac{n}{2}\ln|\Omega|-\frac{1}{2}\sum_{i=1}^{n}\Delta u_i(\delta,\beta,\gamma)'\Omega^{-1}\Delta u_i(\delta,\beta,\gamma). \qquad (7)$$

The only unknown element of Ω^* is ω and it can be shown that $|\Omega|=\sigma_u^{2T}[1+T(\omega-1)]$ (see Hsiao et al., 2002).

For this maximization, γ is assumed to be restricted to a bounded set $\Gamma=[\underline{\gamma};\overline{\gamma}]$. Notice that since this set is also closed, it is compact on the real line. Then, the MLE $(\widehat{\delta},\widehat{\beta},\widehat{\gamma},\widehat{\sigma}_u^2,\widehat{\omega})$ are the values that globally maximize $\ln L(\delta,\beta,\gamma,\sigma_u^2,\omega)$.

In the dynamic panel linear case, Hsiao et al. (2002) found the MLE of the slope parameter is consistent and asymptotically normally distributed when n tends to infinity, whether T is fixed or tends to infinity.

3.2 ML Estimators of $\boldsymbol{\delta}$, $\boldsymbol{\beta}$, $\boldsymbol{\sigma_u^2}$ and $\boldsymbol{\omega}$ for a given $\boldsymbol{\gamma}$

Let $\beta_\delta=(\delta',\beta')'$ and define the matrix $\Delta y_{i,-1}(\gamma)$ as follows

$$\Delta y_{i,-1}(\gamma)=\begin{bmatrix} 1(q_{i1}\leq\gamma) & 1(q_{i1}>\gamma) & 0 & 0 \\ 0 & 0 & \Delta y_{i1}^*(\gamma) & \Delta y_{i1}^+(\gamma) \\ 0 & 0 & \Delta y_{i2}^*(\gamma) & \Delta y_{i2}^+(\gamma) \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \Delta y_{iT-1}^*(\gamma) & \Delta y_{iT-1}^+(\gamma) \end{bmatrix}.$$

Using the above definition, the Log-Likelihood function (7) can be written as

[7]See Appendix A for the formal derivation of the likelihood function.

$$\ln L(\beta_\delta, \gamma, \sigma_u^2, \omega) = -\frac{nT}{2}\ln(2\pi) - \frac{nT}{2}\ln(\sigma_u^2) - \frac{n}{2}\ln[1 + T(\omega - 1)]$$
$$- \frac{1}{2}\sum_{i=1}^{n}[(\Delta y_i - \Delta y_{i,-1}(\gamma)\beta_\delta)'\Omega^{-1}(\Delta y_i - \Delta y_{i,-1}(\gamma)\beta_\delta)]. \quad (8)$$

We start the estimation procedure considering a fixed γ. Then for a given γ, the first-order derivative with respect to β_δ is

$$\frac{\partial \ln L}{\partial \beta_\delta} = \sum_{i=1}^{n}[\Delta y_{i,-1}(\gamma)'\Omega^{-1}(\Delta y_i - \Delta y_{i,-1}(\gamma)\beta_\delta)].$$

Setting the partial derivatives equal to zero gives

$$\widehat{\beta}_\delta(\gamma) = \left(\sum_{i=1}^{n}\Delta y_{i,-1}(\gamma)'\widehat{\Omega}^\star(\gamma)^{-1}\Delta y_{i,-1}(\gamma)\right)^{-1}\left(\sum_{i=1}^{n}\Delta y_{i,-1}(\gamma)'\widehat{\Omega}^\star(\gamma)^{-1}\Delta y_i\right). \quad (9)$$

The first order derivatives with respect to σ_u^2 and ω, for a given γ, are given by

$$\frac{\partial \ln L}{\partial \sigma_u^2} = -\frac{nT}{2\sigma_u^2} + \frac{1}{2\sigma_u^4}\sum_{i=1}^{n}[(\Delta y_i - \Delta y_{i,-1}(\gamma)\beta_\delta)'\Omega^{\star -1}(\Delta y_i - \Delta y_{i,-1}(\gamma)\beta_\delta)],$$

and

$$\frac{\partial \ln L}{\partial \omega} = -\frac{nT}{2[1 + T(\omega - 1)]}$$
$$+ \frac{1}{2\sigma_u^2[1 + T(\omega - 1)]^2}\sum_{i=1}^{n}[(\Delta y_i - \Delta y_{i,-1}(\gamma)\beta_\delta)'\Phi(\Delta y_i - \Delta y_{i,-1}(\gamma)\beta_\delta)],$$

where

$$\Phi = \begin{bmatrix} T^2 & T(T-1) & T(T-2) & \dots & T \\ T(T-1) & (T-1)^2 & (T-1)(T-2) & \dots & (T-1) \\ \vdots & \vdots & \vdots & \dots & \vdots \\ T & (T-1) & (T-2) & \dots & 1 \end{bmatrix}.$$

Setting the above first-order conditions equal to zero yields

$$\widehat{\sigma}_u^2(\gamma) = \frac{1}{nT}\sum_{i=1}^{n}[(\Delta y_i - \Delta y_{i,-1}(\gamma)\widehat{\beta}_\delta(\gamma))'\widehat{\Omega}^\star(\gamma)^{-1}(\Delta y_i - \Delta y_{i,-1}(\gamma)\widehat{\beta}_\delta(\gamma))], \tag{10}$$

$$\widehat{\omega}(\gamma) = \frac{T-1}{T} + \frac{1}{\widehat{\sigma}_u^2(\gamma)nT^2}\sum_{i=1}^{n}[(\Delta y_i - \Delta y_{i,-1}(\gamma)\widehat{\beta}_\delta(\gamma))'\Phi(\Delta y_i - \Delta y_{i,-1}(\gamma)\widehat{\beta}_\delta(\gamma))]. \tag{11}$$

Notice the ML slope estimators depend on σ_u^2 and ω, and those depend on the slope parameters; then we propose two methods to find the MLE $\widehat{\delta}_1(\gamma)$, $\widehat{\delta}_2(\gamma)$, $\widehat{\beta}_1(\gamma)$, $\widehat{\beta}_2(\gamma)$, $\widehat{\sigma}_u^2(\gamma)$ and $\widehat{\omega}(\gamma)$ for a given γ.

Iterative Procedure Using Initial Estimates

For each γ, we can use lagged $\Delta y^*_{it-2}(\gamma)$ and $\Delta y^+_{it-2}(\gamma)$ as instruments to obtain initial estimates for β_1 and β_2 as in Anderson and Hsiao (1982), respectively. Let

$$\Delta\ddot{y}_i = \begin{bmatrix} \Delta y_{i3} \\ \Delta y_{i4} \\ \vdots \\ \Delta y_{iT} \end{bmatrix}, \quad \Delta\ddot{y}_{i,-1}(\gamma) = \begin{bmatrix} \Delta y^*_{i2}(\gamma) & \Delta y^+_{i2}(\gamma) \\ \Delta y^*_{i3}(\gamma) & \Delta y^+_{i3}(\gamma) \\ \vdots & \vdots \\ \Delta y^*_{iT-1}(\gamma) & \Delta y^+_{iT-1}(\gamma) \end{bmatrix}$$

and

$$\Delta\ddot{y}_{i,-2}(\gamma) = \begin{bmatrix} \Delta y^*_{i1}(\gamma) & \Delta y^+_{i1}(\gamma) \\ \Delta y^*_{i2}(\gamma) & \Delta y^+_{i2}(\gamma) \\ \vdots & \vdots \\ \Delta y^*_{iT-2}(\gamma) & \Delta y^+_{iT-2}(\gamma) \end{bmatrix}.$$

Then, for a given γ the initial estimates of β_1 and β_2 by instrumental variables are

$$\begin{bmatrix} \widetilde{\beta}_1(\gamma) \\ \widetilde{\beta}_2(\gamma) \end{bmatrix} = \left(\sum_{i=1}^{n}\Delta\ddot{y}_{i,-2}(\gamma)'\Delta\ddot{y}_{i,-1}(\gamma)\right)^{-1}\left(\sum_{i=1}^{n}\Delta\ddot{y}_{i,-2}(\gamma)'\Delta\ddot{y}_i\right).$$

Initial estimates $\widetilde{\sigma}_u^2(\gamma)$ and $\widetilde{\omega}(\gamma)$ are given by replacing the initial slope estimates in equations (10) and (11), respectively. And initial estimates of the external parameters, $\widetilde{\delta}_1(\gamma)$ and $\widetilde{\delta}_2(\gamma)$, are given by $(1/n_1)\sum_{i=1}^{n}\Delta y_{i1}1(q_{i1}\leq\gamma)$ and $(1/n_2)\sum_{i=1}^{n}\Delta y_{i1}1(q_{i1}>\gamma)$; where $n_1+n_2=n$, n_1 and n_2 are the subsample in each regimen for $t=1$.

Finally, for a fixed γ, by using those initial estimates, we could use an iterative technique such as the Newton-Raphson procedure. For this purpose the second derivatives of the log-likelihood function are provided in the appendix B.

Grid Search Method

Alternatively, in order to compute the MLE $\widehat{\delta}_1(\gamma)$, $\widehat{\delta}_2(\gamma)$, $\widehat{\beta}_1(\gamma)$, $\widehat{\beta}_2(\gamma)$, $\widehat{\sigma}_u^2(\gamma)$ and $\widehat{\omega}(\gamma)$ for a given γ, we can use a grid search procedure whereby the MLE are computed for a number of values of $\omega(\gamma) > 1 - 1/T$ at a given γ, and then choosing that value of $\omega(\gamma)$, which globally maximizes the log-likelihood function (8).

3.3 ML Estimator for the Threshold Parameter γ

The ML estimators for a given γ are $\widehat{\beta}_\delta(\gamma) = (\widehat{\delta}_1(\gamma), \widehat{\delta}_2(\gamma), \widehat{\beta}_1(\gamma), \widehat{\beta}_2(\gamma))'$, $\widehat{\sigma}_u^2(\gamma)$ and $\widehat{\omega}(\gamma)$.

Therefore the threshold parameter, γ, is estimated by maximizing the concentrated log-likelihood function (12),

$$\begin{aligned} \ln L(\gamma) &= -\frac{nT}{2}\ln(2\pi) - \frac{n}{2}\ln|\widehat{\Omega}(\gamma)| \\ &\quad -\frac{1}{2}\sum_{i=1}^{n}(\Delta y_i - \Delta y_{i,-1}(\gamma)\widehat{\beta}_\delta(\gamma))'\widehat{\Omega}(\gamma)^{-1}(\Delta y_i - \Delta y_{i,-1}(\gamma)\widehat{\beta}_\delta(\gamma)) \\ &= -\frac{nT}{2}\ln(2\pi) - \frac{n}{2}\ln|\widehat{\Omega}(\gamma)| - \frac{1}{2}\sum_{i=1}^{n}\Delta\widehat{u}_i(\gamma)'\widehat{\Omega}(\gamma)^{-1}\Delta\widehat{u}_i(\gamma). \end{aligned} \tag{12}$$

The criterion function (12) is not smooth, so conventional gradient algorithms are not suitable for its maximization. Following Hansen (1999), we suggest using a grid search over the threshold variable space. That is, construct an evenly spaced grid on the empirical support $[\underline{\gamma}; \overline{\gamma}]$ of the threshold variable q_{it}.

Notice the threshold effect only has content if $0 < P(q_{it} \leq \gamma) < 1$. In our environment this constraint is satisfied since we assumed that $\Gamma = [\underline{\gamma}; \overline{\gamma}]$ is a proper subset of the support of the threshold variable q_{it}. Alternatively, Hansen and Seo (2002) impose this constraint by assuming $\pi_0 \leq P(q_{it} \leq \gamma) \leq 1 - \pi_0$, where $\pi_0 > 0$ is a trimming parameter. Using this assumption, similarly we can find the MLE of γ by searching the maximum value of γ on the grid on the support of the threshold variable q_{it}, conditional on $\pi_0 \leq (nT)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T} 1(q_{it} \leq \gamma) \leq 1 - \pi_0$, where we can set $\pi_0 = 0.1$ for empirical applications.

3.4 ML Estimators for the slope Parameters $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$

Once $\widehat{\gamma}$ is obtained by maximizing (12), the ML estimators of slope parameters are $\widehat{\beta}_1 = \widehat{\beta}_1(\widehat{\gamma})$ and $\widehat{\beta}_2 = \widehat{\beta}_2(\widehat{\gamma})$.

Also, the ML estimators of the remaining parameters that involve the estimation method are $\widehat{\delta}_1 = \widehat{\delta}_1(\widehat{\gamma})$, $\widehat{\delta}_2 = \widehat{\delta}_2(\widehat{\gamma})$, $\widehat{\sigma}_u^2 = \widehat{\sigma}_u^2(\widehat{\gamma})$ and $\widehat{\omega} = \widehat{\omega}(\widehat{\gamma})$.

The estimated covariance matrix for the ML slope estimators $\widehat{\beta}_\delta$ is

$$\begin{aligned} \mathrm{Cov}\begin{bmatrix} \widehat{\delta}_1 \\ \widehat{\beta}_1 \\ \widehat{\delta}_2 \\ \widehat{\beta}_2 \end{bmatrix} &= \left(\sum_{i=1}^{n} \Delta y_{i,-1}(\widehat{\gamma})' \Omega^{-1} \Delta y_{i,-1}(\widehat{\gamma}) \right)^{-1} \\ &= \sigma_u^2 \left(\sum_{i=1}^{n} \Delta y_{i,-1}(\widehat{\gamma})' \Omega^{\star -1} \Delta y_{i,-1}(\widehat{\gamma}) \right)^{-1}. \end{aligned}$$

Or, under a suitable partition of the matrix $\Delta y_{i,-1}(\widehat{\gamma})$, the estimated covariance matrix for the ML slope estimators β_1 and β_2 is

$$\begin{aligned} \mathrm{Cov}\begin{bmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{bmatrix} &= \sigma_u^2 \left(\sum_{i=1}^{n} \Delta y^{\circ}_{i,-1}(\widehat{\gamma})' \Omega^{\star -1} \Delta y^{\circ}_{i,-1}(\widehat{\gamma}) \right)^{-1} \\ &\equiv F^{-1}(\widehat{\gamma}), \end{aligned}$$

where,

$$\Delta y^{\circ}_{i,-1}(\gamma) = \begin{bmatrix} 0 & 0 \\ \Delta y^{*}_{i1}(\gamma) & \Delta y^{+}_{i1}(\gamma) \\ \Delta y^{*}_{i2}(\gamma) & \Delta y^{+}_{i2}(\gamma) \\ \vdots & \vdots \\ \Delta y^{*}_{iT-1}(\gamma) & \Delta y^{+}_{iT-1}(\gamma) \end{bmatrix}.$$

In the dynamic panel linear case, Hsiao et al. (2002) discuss a feasible estimator. When a $\sqrt{n}$-consistent estimator of Ω, $\widehat{\Omega}$, instead.

3.5 Summary

In summary, our algorithm has the following procedure:

1. Form a grid on the empirical support, $[\underline{\gamma}; \overline{\gamma}]$, of the threshold variable q_{it}.

2. For each value of γ on this grid, calculate the MLE $\widehat{\delta}_1(\gamma)$, $\widehat{\delta}_2(\gamma)$, $\widehat{\beta}_1(\gamma)$, $\widehat{\beta}_2(\gamma)$, $\widehat{\sigma}_u^2(\gamma)$ and $\widehat{\omega}(\gamma)$ by maximizing the criterion (8), the explicit form for the first four

parameter is in (9), for the fifth parameter is in (10) and for the last one is in (11). We can use either

(a) an iterative technique such as the Newton-Raphson procedure by using initial estimates or

(b) a grid search procedure on $\omega(\gamma)$ at a given γ, and then choosing that value of $\omega(\gamma)$ which globally maximizes the function (8).

3. With the ML estimators computed in part 2 for each γ, find the MLE $\widehat{\gamma}$ as the value of γ on the grid on the empirical support $[\underline{\gamma};\overline{\gamma}]$ which yields the highest value of the concentrated criterion (12).

4. Set $\widehat{\beta}_1 = \widehat{\beta}_1(\widehat{\gamma})$, $\widehat{\beta}_2 = \widehat{\beta}_2(\widehat{\gamma})$, $\widehat{\delta}_1 = \widehat{\delta}_1(\widehat{\gamma})$, $\widehat{\delta}_2 = \widehat{\delta}_2(\widehat{\gamma})$, $\widehat{\sigma}_u^2 = \widehat{\sigma}_u^2(\widehat{\gamma})$, $\widehat{\omega} = \widehat{\omega}(\widehat{\gamma})$ and $\Delta\widehat{u}_i = \Delta\widehat{u}_i(\widehat{\gamma})$.

4 Model with Exogenous Regressors

The model can be extended to allow exogenous regressors as follows

$$y_{it} = \alpha_i + (\beta_1 y_{it-1} + \theta_1' x_{it})1(q_{it} \leq \gamma) + (\beta_2 y_{it-1} + \theta_2' x_{it})1(q_{it} > \gamma) + u_{it}, \tag{13}$$

where x_{it} is a k vector. Again we assume that initial values y_{i0} and x_{i0} are available. By taking first differences to eliminate the individual specific effect in model (13), we get

$$\begin{aligned} y_{it} - y_{it-1} &= \beta_1(y_{it-1}1(q_{it} \leq \gamma) - y_{it-2}1(q_{it-1} \leq \gamma)) + \theta_1(x_{it}1(q_{it} \leq \gamma) \\ &\quad - x_{it-1}1(q_{it-1} \leq \gamma)) + \beta_2(y_{it-1}1(q_{it} > \gamma) - y_{it-2}1(q_{it-1} > \gamma)) \\ &\quad + \theta_2(x_{it}1(q_{it} > \gamma) - x_{it-1}1(q_{it-1} > \gamma)) + u_{it} - u_{it-1}. \end{aligned} \tag{14}$$

To simplify notation let $\Delta y_{it} \equiv y_{it} - y_{it-1}$, $\Delta y^*_{it-1}(\gamma) \equiv y_{it-1}1(q_{it} \leq \gamma) - y_{it-2}1(q_{it-1} \leq \gamma)$, $\Delta x^*_{it}(\gamma) \equiv x_{it}1(q_{it} \leq \gamma) - x_{it-1}1(q_{it-1} \leq \gamma)$, $\Delta y^+_{it-1}(\gamma) \equiv y_{it-1}1(q_{it} > \gamma) - y_{it-2}1(q_{it-1} > \gamma)$, $\Delta x^+_{it}(\gamma) \equiv x_{it}1(q_{it} > \gamma) - x_{it-1}1(q_{it-1} > \gamma)$ and $\Delta u_{it} \equiv u_{it} - u_{it-1}$. Then equation (14) becomes

$$\Delta y_{it} = \beta_1 \Delta y^*_{it-1}(\gamma) + \theta_1' \Delta x^*_{it}(\gamma) + \beta_2 \Delta y^+_{it-1}(\gamma) + \theta_2' \Delta x^+_{it}(\gamma) + \Delta u_{it}. \tag{15}$$

Again, for $t = 2, 3, \ldots, T$, (15) is well defined, but not for Δy_{i1} because $\Delta y^*_{i0}(\gamma)$ and $\Delta y^+_{i0}(\gamma)$ are missing, that is $y_{i,-1}$ is not available. Thus, the mean of Δy_{i1} conditional on $\Delta y^*_{i0}(\gamma)$, $\Delta x^*_{i1}(\gamma)$, $\Delta y^+_{i0}(\gamma)$ and $\Delta x^+_{i1}(\gamma)$ is unknown.

Similar to the model without exogenous regressors and using the exogeneity of x_{it} for all t, we assume the process has started from a finite period in the past, namely for given values of $y_{i,-1}$ such that

$$E(\Delta y_{i1}|x_i) = \delta_1 1(q_{i1} \leq \gamma) + \delta_2 1(q_{i1} > \gamma) + \delta_3' \Delta x_i^*(\gamma) + \delta_4' \Delta x_i^+(\gamma),$$

where $\Delta x_i^*(\gamma) = (\Delta x_{i1}^*(\gamma), \ldots, \Delta x_{iT}^*(\gamma))'$ and $\Delta x_i^+(\gamma) = (\Delta x_{i1}^+(\gamma), \ldots, \Delta x_{iT}^+(\gamma))'$. Notice that we take the conditional expectation given the observables x_{it}.

Thus the marginal distribution of Δy_{i1} conditional on x_i can be written as

$$\Delta y_{i1} = \delta_1 1(q_{i1} \leq \gamma) + \delta_2 1(q_{i1} > \gamma) + \delta_3' \Delta x_i^*(\gamma) + \delta_4' \Delta x_i^+(\gamma) + \upsilon_{i1}, \tag{16}$$

under the exogeneity of x_{it} and by construction $E(\upsilon_{i1}|x_i) = 0$, $E\upsilon_{i1}^2 = \sigma_\upsilon^2$, and we assume $Cov(\upsilon_{i1}, \Delta u_{i2}|x_{it}) = -\sigma_u^2$ and $Cov(\upsilon_{i1}, \Delta u_{it}|x_{it}) = 0$ for $t = 3, ..., T$, $i = 1, ..., n$.

There is no guarantee that the error υ_{i1} is normally distributed, though Hsiao et al. (2002) show that if the exogenous variables are driven by a normal distribution, the error υ_{i1} will be normally distributed.

Therefore, in order to estimate this model for a given γ, we maximize the criterion (7), where $\Delta u_i(\delta, \beta, \theta, \gamma) = [\Delta y_{i1} - \delta_1 1(q_{i1} \leq \gamma) - \delta_2 1(q_{i1} > \gamma) - \delta_3' \Delta x_i^*(\gamma) - \delta_4' \Delta x_i^+(\gamma), \Delta y_{i2} - \beta_1 \Delta y_{i1}^*(\gamma) - \theta_1' \Delta x_{i2}^*(\gamma) - \beta_2 \Delta y_{i1}^+(\gamma) - \theta_2' \Delta x_{i2}^+(\gamma), \ldots, \Delta y_{iT} - \beta_1 \Delta y_{iT-1}^*(\gamma) - \theta_1' \Delta x_{iT}^*(\gamma) - \beta_2 \Delta y_{iT-1}^+(\gamma) - \theta_2' \Delta x_{iT}^+(\gamma)]'$,
where $\delta = (\delta_1, \delta_2, \delta_3', \delta_4')'$ and $\theta = (\theta_1, \theta_2)'$. Let $\theta_\delta = (\delta, \beta_1, \theta_1, \beta_2, \theta_2)'$ and define the matrix $\Delta y_{i,-1}(\gamma)$ as follows

$$\Delta y_{i,-1}(\gamma) = \left[\begin{array}{cccccccc} 1(q_{i1} \leq \gamma) & 1(q_{i1} > \gamma) & \Delta x_i^*(\gamma)' & \Delta x_i^+(\gamma)' & 0 & 0 & 0 & 0 \\ 0 & 0 & 0_{1\times T} & 0_{1\times T} & \Delta y_{i1}^*(\gamma) & \Delta x_{i1}^*(\gamma) & \Delta y_{i1}^+(\gamma) & \Delta x_{i1}^+(\gamma) \\ 0 & 0 & 0_{1\times T} & 0_{1\times T} & \Delta y_{i2}^*(\gamma) & \Delta x_{i2}^*(\gamma) & \Delta y_{i2}^+(\gamma) & \Delta x_{i2}^+(\gamma) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0_{1\times T} & 0_{1\times T} & \Delta y_{iT-1}^*(\gamma) & \Delta x_{iT}^*(\gamma) & \Delta y_{iT-1}^+(\gamma) & \Delta x_{iT}^+(\gamma) \end{array}\right].$$

Then, with our new definition of $\Delta y_{i,-1}(\gamma)$, we can estimate based on the criterion (8).

Then the ML Estimators of θ_δ for a given γ can be written as

$$\widehat{\theta}_\delta(\gamma) = \left(\sum_{i=1}^{n} \Delta y_{i,-1}(\gamma)'\widehat{\Omega}^\star(\gamma)^{-1}\Delta y_{i,-1}(\gamma)\right)^{-1}\left(\sum_{i=1}^{n} \Delta y_{i,-1}(\gamma)'\widehat{\Omega}^\star(\gamma)^{-1}\Delta y_i\right), \tag{17}$$

and the ML estimators of σ_u^2 and ω for a given γ can be written as

$$\widehat{\sigma}_u^2(\gamma) = \frac{1}{nT}\sum_{i=1}^{n}[(\Delta y_i - \Delta y_{i,-1}(\gamma)\widehat{\theta}_\delta(\gamma))'\widehat{\Omega}^\star(\gamma)^{-1}(\Delta y_i - \Delta y_{i,-1}(\gamma)\widehat{\theta}_\delta(\gamma))], \tag{18}$$

$$\widehat{\omega}(\gamma) = \frac{T}{T-1} + \frac{1}{\widehat{\sigma}_u^2(\gamma)nT^2}\sum_{i=1}^{n}[(\Delta y_i - \Delta y_{i,-1}(\gamma)\widehat{\theta}_\delta(\gamma))'\Phi(\Delta y_i - \Delta y_{i,-1}(\gamma)\widehat{\theta}_\delta(\gamma))], \tag{19}$$

where Φ is defined as before.

In order to get the MLE $\widehat{\delta}(\gamma)$, $\widehat{\beta}(\gamma)$, $\widehat{\theta}(\gamma)$, $\widehat{\sigma}_u^2(\gamma)$ and $\widehat{\omega}(\gamma)$ for a given γ, we can use either an iterative procedure using initial estimates of θ_δ for a given γ, or a grid search method on $\omega(\gamma) > 1 - 1/T$ at a given γ.

For the iterative procedure, initial estimators can be obtained by an instrumental variable estimation. Let

$$\Delta\ddot{y}_i = \begin{bmatrix} \Delta y_{i3} \\ \Delta y_{i4} \\ \vdots \\ \Delta y_{iT} \end{bmatrix}, \quad \Delta\ddot{y}_{i,-1}(\gamma) = \begin{bmatrix} \Delta y_{i2}^*(\gamma) & \Delta x_{i3}^*(\gamma) & \Delta y_{i2}^+(\gamma) & \Delta x_{i3}^+(\gamma) \\ \Delta y_{i3}^*(\gamma) & \Delta x_{i4}^+(\gamma) & \Delta y_{i3}^+(\gamma) & \Delta x_{i4}^+(\gamma) \\ \vdots & \vdots & \vdots & \vdots \\ \Delta y_{iT-1}^*(\gamma) & \Delta x_{iT}^*(\gamma) & \Delta y_{iT-1}^+(\gamma) & \Delta x_{iT}^+(\gamma) \end{bmatrix}$$

and

$$\Delta\ddot{y}_{i,-2}(\gamma) = \begin{bmatrix} \Delta y_{i1}^*(\gamma) & \Delta x_{i3}^*(\gamma) & \Delta y_{i1}^+(\gamma) & \Delta x_{i3}^+(\gamma) \\ \Delta y_{i2}^*(\gamma) & \Delta x_{i4}^*(\gamma) & \Delta y_{i2}^+(\gamma) & \Delta x_{i4}^+(\gamma) \\ \vdots & \vdots & & \\ \Delta y_{iT-2}^*(\gamma) & \Delta x_{iT}^*(\gamma) & \Delta y_{iT-2}^+(\gamma) & \Delta x_{iT}^+(\gamma) \end{bmatrix}.$$

Then, for a given γ the initial estimates of β_1, θ_1, β_2 and θ_2 using instrumental variables are

$$[\widetilde{\beta}_1(\gamma) \;\; \widetilde{\theta}_1(\gamma) \;\; \widetilde{\beta}_2(\gamma) \;\; \widetilde{\theta}_2(\gamma)]' = \left(\sum_{i=1}^{n} \Delta\ddot{y}_{i,-2}(\gamma)' \Delta\ddot{y}_{i,-1}(\gamma) \right)^{-1} \left(\sum_{i=1}^{n} \Delta\ddot{y}_{i,-2}(\gamma)' \Delta\ddot{y}_{i} \right);$$

and, initial estimations $\widetilde{\sigma}_u^2(\gamma)$ and $\widetilde{\omega}(\gamma)$ are given by replacing the initial slope estimates in equations (18) and (19), respectively. Initial estimates of the external parameters, $\widetilde{\delta}_1(\gamma)$, $\widetilde{\delta}_2(\gamma)$, $\widetilde{\delta}_3(\gamma)$ and $\widetilde{\delta}_4(\gamma)$, can be obtained by a least square regression of Δy_{i1} on $1(q_{i1} \le \gamma)$, $1(q_{i1} > \gamma)$, $\Delta x_i^*(\gamma)$ and $\Delta x_i^+(\gamma)$.

Once we get the ML estimators $\widehat{\delta}_1(\gamma)$, $\widehat{\delta}_2(\gamma)$, $\widehat{\delta}_3'(\gamma)$, $\widehat{\delta}_4'(\gamma)$, $\widehat{\beta}_1(\gamma)$, $\widehat{\theta}_1'(\gamma)$, $\widehat{\beta}_2(\gamma)$, $\widehat{\theta}_2'(\gamma)$, $\widehat{\sigma}_u^2(\gamma)$ and $\widehat{\omega}(\gamma)$ for each γ; the threshold parameter γ is estimated by maximizing the concentrated likelihood function (12).

Once the MLE of the threshold parameter $\widehat{\gamma}$ is obtained, then the ML slope parameter estimates are $\widehat{\delta}_1 = \widehat{\delta}_1(\widehat{\gamma})$, $\widehat{\delta}_2 = \widehat{\delta}_2(\widehat{\gamma})$, $\widehat{\delta}_3 = \widehat{\delta}_3(\widehat{\gamma})$, $\widehat{\delta}_4 = \widehat{\delta}_4(\widehat{\gamma})$, $\widehat{\beta}_1 = \widehat{\beta}_1(\widehat{\gamma})$, $\widehat{\theta}_1 = \widehat{\theta}_1(\widehat{\gamma})$, $\widehat{\beta}_2 = \widehat{\beta}_2(\widehat{\gamma})$ and $\widehat{\theta}_2 = \widehat{\theta}_2(\widehat{\gamma})$.

5 Multiple Thresholds

Model (1) has a single threshold. In some applications there may be multiple thresholds. For example, in our case, the double threshold model can take the form

$$y_{it} = \alpha_i + \beta_1 y_{it-1} 1(q_{it} \le \gamma_1) + \beta_2 y_{it-1} 1(\gamma_1 < q_{it} \le \gamma_2) + \beta_3 y_{it-1} 1(q_{it} > \gamma_2) + u_{it}, \quad (20)$$

where the thresholds are ordered so that $\gamma_1 < \gamma_2$.

5.1 Estimation

For given $(\gamma_1, \gamma_2)'$, (20) is linear in the slopes, then the ML estimation is appropriate. Thus for given $(\gamma_1, \gamma_2)'$ the concentrated log-likelihood function $\ln L(\gamma_1, \gamma_2)$ is straightforward to calculate (as in the single threshold model). The joint maximum likelihood estimates of $(\gamma_1, \gamma_2)'$ are by definition the values which jointly maximize $\ln L(\gamma_1, \gamma_2)$. While these estimates might seem desirable, Hansen (1999) argues that they may be quite cumbersome to implement in practice.

Hansen (1999) argues that it has been found (Chong, 1994; Bai, 1997) in the multiple change-point model that sequential estimation is consistent. And then the same logic appears to apply to the multiple threshold model.

Following Hansen (1999) the method works as follows. In the first stage, let $\ln L(\gamma)$ be the single threshold concentrated log-likelihood function as defined in (12) and let $\widehat{\gamma}_1$ be the threshold estimate which maximizes $\ln L(\gamma)$. The analysis of Chong (1994) and Bai (1997) suggests that $\widehat{\gamma}_1$ will be consistent for either γ_1 or γ_2.

Fixing the first-stage estimate $\widehat{\gamma}_1$, the second-stage criterion is

$$\ln L_2^r(\gamma_2) = \begin{cases} \ln L(\widehat{\gamma}_1, \gamma_2) & \text{if} \quad \widehat{\gamma}_1 < \gamma_2 \\ \ln L(\gamma_2, \widehat{\gamma}_1) & \text{if} \quad \gamma_2 < \widehat{\gamma}_1 \end{cases} \tag{21}$$

and the second-stage threshold estimate is

$$\widehat{\gamma}_2^r = \underset{\gamma_2}{\operatorname{argmax}} \ln L_2^r(\gamma_2) \tag{22}$$

Bai (1997) has shown that $\widehat{\gamma}_2^r$ is asymptotically efficient, but $\widehat{\gamma}_1$ is not. Hansen (1999) argues that is because the estimate $\widehat{\gamma}_1$ was obtained from a the concentrated function which was contaminated by the presence of a neglected regime. Hansen (1999) states the asymptotic efficiency of $\widehat{\gamma}_1$ can be improved by a third-stage estimation. Bai (1997) suggests the following refinement estimator. Fixing the second-stage estimate $\widehat{\gamma}_2^r$, define the refinement criterion

$$\ln L_1^r(\gamma_1) = \begin{cases} \ln L(\gamma_1, \widehat{\gamma}_2^r) & \text{if} \quad \gamma_1 < \widehat{\gamma}_2^r \\ \ln L(\widehat{\gamma}_2^r, \gamma_1) & \text{if} \quad \widehat{\gamma}_2^r < \gamma_1 \end{cases} \tag{23}$$

and the refinement threshold estimate is

$$\widehat{\gamma}_1^r = \underset{\gamma_1}{\operatorname{argmax}} \ln L_1^r(\gamma_1) \tag{24}$$

Bai (1997) shows that the refinement estimator $\widehat{\gamma}_1^r$ is asymptotically efficient in change-point estimation, and as Hansen (1999) we expect similar results to hold in threshold regression.

6 Asymptotic Theory

In the context of threshold regression it is known that threshold estimate is super-consistent; and since the objective function (12) is not smooth, it is found that the distribution of the threshold estimate is nonstandard.

Those result lie in the assumption of the exogeneity of the threshold variable and the regressors; even though in the structural model we have the same environment, the first difference transformation introduces a correlation between the lagged regressor and the errors in the model. Thus, a different technique will be developed in order to prove consistency and to establish asymptotic distribution of the threshold estimate.

6.1 Assumptions

Let $\beta_{\delta_1} = (\delta_1, \beta_1)'$ and $\beta_{\delta_2} = (\delta_2, \beta_2)'$. Define a partition of the matrix, $\Delta y_{i,-1}(\gamma)$ as follows

$$\Delta y^*_{i,-1}(\gamma) = \begin{bmatrix} 1(q_{i1} \leq \gamma) & 0 \\ 0 & \Delta y^*_{i1}(\gamma) \\ 0 & \Delta y^*_{i2}(\gamma) \\ \vdots & \vdots \\ 0 & \Delta y^*_{iT-1}(\gamma) \end{bmatrix}; \quad \Delta y^+_{i,-1}(\gamma) = \begin{bmatrix} 1(q_{i1} > \gamma) & 0 \\ 0 & \Delta y^+_{i1}(\gamma) \\ 0 & \Delta y^+_{i2}(\gamma) \\ \vdots & \vdots \\ 0 & \Delta y^+_{iT-1}(\gamma) \end{bmatrix}.$$

Define the moment functionals

$$\begin{aligned} M(\gamma) &= E(C'\Delta y^*_{i,-1}(\gamma)'\Omega^{\star-1}\Delta y^*_{i,-1}(\gamma)C) \\ &= c'E\left(\sum_{t=1}^{T}(a_t a_{t-1})^{-1}1(q_{i1} \leq \gamma) + \sum_{t=2}^{T}(a_t a_{t-1})^{-1}\left(\sum_{s=1}^{t-1} a_s \Delta y^*_{is}(\gamma)\right)^2\right. \\ &\quad \left. + 2\sum_{t=2}^{T}(a_t a_{t-1})^{-1}1(q_{i1} \leq \gamma)\sum_{s=1}^{t-1} a_s \Delta y^*_{is}(\gamma)\right) c, \end{aligned} \tag{25}$$

where $a_0 = 1$, $a_1 = \omega$ and $a_0, \cdots, a_{T-1}$ are constants of the matrix H defined in appendix C; that is, H is the matrix such that $H'\Lambda H = \Omega^{\star-1}$. Let $f_t(\gamma)$ and $f_{t|t-1}(\gamma_1|\gamma_2)$ denote the density function of q_{it} and the conditional density of q_{it} given q_{it-1}, respectively. Let

$$\begin{aligned} M &= E(C'\Delta y'_{i,-1}\Omega^{\star-1}\Delta y_{i,-1}C) \\ &= c'E\left(\sum_{t=1}^{T} a_0(a_t a_{t-1})^{-1} + \sum_{t=2}^{T}(a_t a_{t-1})^{-1}\left(\sum_{s=1}^{t-1} a_s \Delta y_{is}\right)^2\right. \\ &\quad \left. + 2\sum_{t=2}^{T}(a_t a_{t-1})^{-1}\sum_{s=1}^{t-1} a_s \Delta y_{is}\right) c, \end{aligned} \tag{26}$$

where

$$
\Delta y_{i,-1} = \begin{bmatrix} 1 & 0 \\ 0 & \Delta y_{i1} \\ 0 & \Delta y_{i2} \\ \vdots & \vdots \\ 0 & \Delta y_{iT-1} \end{bmatrix}
$$

.

Also, let

$$
\begin{aligned}
D(\gamma) \;=\;& c' \Bigg[\sum_{t=1}^{T} (a_t a_{t-1})^{-1} f_1(\gamma) + \sum_{t=2}^{T} (a_t a_{t-1})^{-1} \Bigg(\sum_{s=1}^{t-1} a_s^2 \big(E(y_{is}^2 | q_{is+1} = \gamma) f_{s+1}(\gamma) \\
& + E(y_{is}^2 | q_{is-1} = \gamma) f_s(\gamma) \big) - \sum_{s=1}^{t-2} 2 a_s a_{s+1} E(y_{is}^2 | q_{is+1} = \gamma) f_{s+1}(\gamma) \\
& - 2 a_1 \nu E(y_{i0} | q_{i1} = \gamma) f_1(\gamma) \Bigg) \Bigg] c. \qquad (27)
\end{aligned}
$$

Finally, let

$$
\begin{aligned}
V_1(\gamma) \;=\;& \left(\sum_{t=1}^{T} (a_t a_{t-1})^{-1} \left(\upsilon_{i1} + \sum_{s=2}^{t} a_{s-1} \Delta u_{is} \right) \right)^2 f_1(\gamma) \\
& \sum_{t=2}^{T-1} (a_t a_{t-1})^{-2} \Bigg\{ \sum_{s=1}^{t-1} a_s^2 E \left(\left[y_{is} \left(\upsilon_{i1} + \sum_{s=2}^{t} a_{s-1} \Delta u_{is} \right) \right]^2 | q_{is+1} = \gamma \right) f_{s+1}(\gamma) \\
& + \sum_{s=1}^{t-1} a_s^2 E \left(\left[y_{is-1} \left(\upsilon_{i1} + \sum_{s=2}^{t} a_{s-1} \Delta u_{is} \right) \right]^2 | q_{is} = \gamma \right) f_s(\gamma) \\
& - \sum_{s=1}^{t-2} 2 a_s a_{s+1} E \left(\left[y_{is} \left(\upsilon_{i1} + \sum_{s=2}^{t} a_{s-1} \Delta u_{is} \right) \right]^2 | q_{is+1} = \gamma \right) f_{s+1}(\gamma) \Bigg\},
\end{aligned}
$$

$$
\begin{aligned}
&V_2(\gamma) = \\
&\sum_{t=2}^{T-1}(a_t a_{t-1})^{-1} \sum_{k=t+1}^{T} (a_k a_{k-1})^{-1} \\
&\left\{\sum_{s=1}^{t-1} a_s^2 E\left(y_{is}^2\left(\upsilon_{i1}+\sum_{s=2}^{t} a_{s-1}\Delta u_{is}\right)\left(\upsilon_{i1}+\sum_{s=2}^{t+1} a_{s-1}\Delta u_{is}\right)|q_{is+1}=\gamma\right) f_{s+1}(\gamma)+\right. \\
&\sum_{s=1}^{t-1} a_s^2 E\left(y_{is-1}^2\left(\upsilon_{i1}+\sum_{s=2}^{t} a_{s-1}\Delta u_{is}\right)\left(\upsilon_{i1}+\sum_{s=2}^{t+1} a_{s-1}\Delta u_{is}\right)|q_{is}=\gamma\right) f_s(\gamma)- \\
&\sum_{s=2}^{t-1} 2a_s a_{s-1} E\left(y_{is-1}^2\left(\upsilon_{i1}+\sum_{s=2}^{t} a_{s-1}\Delta u_{is}\right)\left(\upsilon_{i1}+\sum_{s=2}^{t+1} a_{s-1}\Delta u_{is}\right)|q_{is}=\gamma\right) f_s(\gamma)- \\
&\left. a_t a_{t-1} E\left(y_{it-1}^2\left(\upsilon_{i1}+\sum_{s=2}^{t} a_{s-1}\Delta u_{is}\right)\left(\upsilon_{i1}+\sum_{s=2}^{t+1} a_{s-1}\Delta u_{is}\right)|q_{it}=\gamma\right) f_t(\gamma)\right\},
\end{aligned}
$$

$$
\begin{aligned}
V_3(\gamma) &= -2\sum_{t=1}^{T}(a_t a_{t-1})^{-1}\sum_{k=2}^{T}(a_k a_{k-1})^{-1} a_1 \\
&\quad E\left(y_{i0}\left(\upsilon_{i1}+\sum_{s=2}^{t} a_{s-1}\Delta u_{is}\right)\left(\upsilon_{i1}+\sum_{s=2}^{k} a_{s-1}\Delta u_{is}\right)\right)|q_{i1}=\gamma) f_1(\gamma)
\end{aligned}
$$

and

$$
V(\gamma) = c'\left[V_1(\gamma)V_2(\gamma)+V_3(\gamma)\right]c. \tag{28}
$$

The functions $D(\gamma)$ and $V(\gamma)$ look very complicated, but essentially the first one is the derivative of $M(\gamma) = E(C'\Delta y_{i,-1}^*(\gamma)'\Omega^{\star -1}\Delta y_{i,-1}^*(\gamma)C)$ with respect to γ, and the second one is the derivative of $E(C'\Delta y_{i,-1}^*(\gamma)'\Omega^{\star -1}\Delta u_i(\gamma)C)^2$ with respect to γ, where C is defined in Assumption 1.

Assumption 1

1. y_{i0} and x_{i0} are observable;

2. For each t, $(x_{it}, u_{it}, \upsilon_{i1})$ are independent and identically distributed (iid) across i;

3. For each i, u_{it} is iid over t, is independent of $\{(y_{it-s})_{s=1}^{t-1}, (x_{it})_{s=1}^{T}\}$; and υ_{i1} is independent of $\{(y_{i0}), (x_{it})_{s=1}^{T}\}$;
4. u_{it} normal distributed with mean 0 and variance σ_u^2, by construction $E(\upsilon_{i1}) = 0$, then we also assume that υ_{i1} is normal distributed with variance σ_υ^2 and $\omega > 1-1/T$.
5. $\beta_1 \neq \beta_2$ and $\delta_1 \neq \delta_2$;
6. $E|y_{i0}|^4 < \infty$ and $E|x_{it}|^4 < \infty$;
7. $f_t(\gamma)$, $D(\gamma)$ and $V(\gamma)$ are continuous at $\gamma = \gamma_0$;
8. For some fixed $c < \infty$ and $0 < \alpha < 1/2$, $\beta_2 - \beta_1 = \delta_2 - \delta_1 = n^{-\alpha}c$ and also $C = (c, c)'$;
9. $0 < V(\gamma_0) < \infty$, $0 < D(\gamma_0) < \infty$, $f_t(\gamma_0) < \infty$ and for $k > t$ $f_{k|t}(\gamma_0|\gamma_0) < \infty$;
10. $0 < M(\gamma) < M$ for all $\gamma \in \Gamma$.

Assumptions 1.1-1.4 are similar to the assumptions of Hsiao et al. (2002) for the dynamic fixed effect panel models with strictly exogenous regressors. Assumption 1.5 excludes the possibility that the threshold parameter is not identified. Assumption 1.6 is the conditional moment bound for the variables (notice that 1.4 implies the conditional moment bounds for both errors terms). Assumption 1.7 requires the threshold variable to have a continuous distribution, and essentially requires the variance conditional on the threshold variable to be continuous at γ_0, which excludes regime-dependent heteroskedasticity. Assumption 1.8 is the small threshold effect as in Hansen (2000). Assumption 1.9 requires the threshold variable q_{it} be continuously distributed with positive support at the threshold γ_0, and the variance is finite, also excluding the possibility that $q_{it} = \gamma_0$ for $t = 1, \ldots, T$. Assumption 1.8 is a conventional full-rank condition that excludes multi-collinearity, restricting Γ to a proper subset of the support of q_{it}.

6.2 Threshold Estimate

In the context of the Conditional Least Squares Estimation (CLSE) of a threshold autoregression, Chan (1993) developed the strong consistency of the threshold parameter while Hansen (2000) shows the consistency of it by using the "small threshold effect" assumption. In the context of Maximum Likelihood Estimation, Qian (1998) obtained results similar to those of Chan (1993) in the same model under some regularity conditions on the error density, not necessarily Gaussian. Samia and Chan (2010) derived consistency of the threshold parameter under the assumption that the conditional probability distribution of the response variable belongs to an exponential family.[8]

In all these models the errors are independent of the threshold variable as well as in our model. They are also independent (mean-independent) of the regressors, even though we have the same environment in the structural model. In the model in first differences we introduce a correlation between the first difference of the errors and the first difference of the lagged variable due to the dynamic nature of the model. Then a different technique will be developed in order to prove the consistency of the threshold parameter. For simplicity we consider the model without exogenous regressors in the proof of theorems; the proofs can be extended for the model with exogenous regressors, but at cost of a more cumbersome notation.

Theorem 1. *Under Assumption 1, the Maximum Likelihood Estimator of γ obtained by minimizing the criterion* (12)*, $\widehat{\gamma}$, is such that $\widehat{\gamma} \to_p \gamma_0$, where γ_0 is the true value.*

In the context of threshold autoregression estimation, Chan (1993) establishes the limiting distribution of the threshold parameter estimator. He shows it converges to a functional of a compound Poisson process at a rate n. The distribution is too complicated to be used in practice due to the dependence on the nuisance parameters (including the marginal distribution of the threshold variable and all the regression coefficients). Hansen (2000) developed an asymptotic distribution for the threshold parameter estimate based on the small threshold effect assumption, in which the threshold model becomes the linear model asymptotically. The limiting distribution converges to a functional of a two-sided Brownian motion process at a rate $n^{1-2\alpha}$. The distribution does not depend on the nuisance parameters; thus, the distribution can be available in a simple closed form.

Hence, we adopt Hansen's (2000) approach in our setting. A two sided Brownian motion $W(\nu)$ on the real line is defined as

$$W(\nu) = \begin{cases} W_1(-\nu), & \nu < 0, \\ 0, & \nu = 0, \\ W_2(\nu), & \nu > 0, \end{cases} \tag{29}$$

where $W_1(\nu)$ and $W_2(\nu)$ are independent standard Brownian motions on $[0, \infty)$.

Theorem 2. *Under Assumption 1, $n^{1-2\alpha}(\widehat{\gamma} - \gamma_0) \to_d \varpi U$, where $\varpi = \frac{V(\gamma_0)}{D(\gamma_0)^2}$ and $U =$* $\operatorname{argmax}_{-\infty<\nu<\infty} \left[-\frac{1}{2}|\nu| + W(\nu)\right]$.

[8] Yu (2012) discusses the consistency and asymptotic distribution of the left-endpoint maximum likelihood estimator and the middle-point maximum likelihood estimator. Seo and Linton (2007) developed another different approach by smoothing the indicator function and proposed a Smoothed Least Squares Estimation (SLSE) in the regression context.

The distribution function for U is known (See Hansen (2000) for the exact form) and the asymptotic distribution in Theorem 2 is scaled by the ratio ϖ. This asymptotic distribution yields a computationally attractive method for constructing confidence intervals, and is described in detail in Hansen (1997) in the context of the threshold autoregression and in Hansen (1999) to the threshold panel data models.

Basically, Hansen (2000) argues the best ways to form confidence intervals for the threshold is to form the no-rejection region using the likelihood ratio statistic for testing on $\widehat{\gamma}$. To test hypothesis $H_0 : \gamma = \gamma_0$, the likelihood ratio test is to reject large values of $LR(\gamma_0)$ where

$$LR(\gamma) = nT\frac{Sn(\gamma) - Sn(\widehat{\gamma})}{Sn(\widehat{\gamma})}, \tag{30}$$

where $S_n(\gamma) = \sum_{i=1}^{n} \Delta\widehat{u}_i(\gamma)'\Omega^{\star -1}\Delta\widehat{u}_i(\gamma)$ is the minimum distance estimator. Hansen (1996) shows the $LR(\gamma)$ converges in distribution to ξ as $n \to \infty$, where ξ is a random variable with distribution function $P(\xi \leq z) = (1 - exp(-z/2))^2$. Then, the asymptotic distribution of the likelihood ratio statistic is non-standard, yet free of nuisance parameters. Since the asymptotic distribution is pivotal, it may be used to form valid asymptotic confidence intervals. Furthermore, the distribution function ξ has the inverse

$$c(a) = -2\ln\left(1 - \sqrt{1-a}\right), \tag{31}$$

where a is the significance level. To form an asymptotic confidence interval for γ, the "no-rejection region" of confidence level $1-a$ is the set of values of γ, such that $LR(\gamma) \leq c(a)$, where $LR(\gamma)$ is defined in (30) and $c(a)$ is defined in (31). This is easiest to find by plotting $LR(\gamma)$ against γ and drawing a flat line at $c(a)$.

6.3 Confidence Region Construction in Multiple Thresholds

Bai (1997) showed (for the analogous case of change-point models) that the refinement estimators have the same asymptotic distributions as the threshold estimate in a single threshold model. Upon that finding, Hansen (1999) suggests that we can construct confidence intervals in the same way as in the threshold estimate in a single threshold model.

For $\widehat{\gamma}_2^r$ let

$$LR_2^r(\gamma) = nT\frac{Sn_2^r(\gamma) - Sn_2^r(\widehat{\gamma}_2^r)}{Sn_2^r(\widehat{\gamma}_2^r)}, \tag{32}$$

where the minimum distance estimator $Sn_2^r(\gamma)$ is defined equivalently to (21), that is

$$Sn_2^r(\gamma_2) = \begin{cases} Sn(\widehat{\gamma}_1, \gamma_2) & \text{if } \widehat{\gamma}_1 < \gamma_2 \\ Sn(\gamma_2, \widehat{\gamma}_1) & \text{if } \gamma_2 < \widehat{\gamma}_1 \end{cases} \tag{33}$$

and $Sn(\gamma) = \sum_{i=1}^{n} \Delta\widehat{u}_i(\gamma)'\Omega^{\star-1}\Delta\widehat{u}_i(\gamma)$.

For $\widehat{\gamma}_1^r$ let

$$LR_1^r(\gamma) = nT\frac{Sn_1^r(\gamma) - Sn_1^r(\widehat{\gamma}_1^r)}{Sn_1(\widehat{\gamma}_1^r)}, \tag{34}$$

where the minimum distance estimator $Sn_1^r(\gamma)$ is defined equivalently to (23), that is

$$Sn_1^r(\gamma_1) = \begin{cases} Sn(\gamma_1, \widehat{\gamma}_2^r) & \text{if } \gamma_1 < \widehat{\gamma}_2^r \\ Sn(\widehat{\gamma}_2^r, \gamma_1) & \text{if } \widehat{\gamma}_2^r < \gamma_1 \end{cases} \tag{35}$$

The asymptotic $(1-a)$ percent confidence intervals for γ_1 and γ_2 are the set of values of γ such that $LR_2^r(\gamma) \leq c(a)$ and $LR_1^r(\gamma) \leq c(a)$, respectively.

6.4 Slope Estimates

The likelihood function (7) or (8) is well defined; it depends on a fixed number of parameters, and satisfies the usual regularity conditions conditional on γ. Therefore, the MLE of (8) is consistent and asymptotically normally distributed, when n tends to infinity when T is fixed. In the next theorem, we state the slope parameters are consistent at the true γ_0.

Theorem 3. *Under Assumption 1, the Maximum Likelihood Estimators $\widehat{\beta} = (\widehat{\beta}_1\ \widehat{\beta}_2)'$ are consistent. That is, $\widehat{\beta} \to_p \beta_0$, where $\beta_0 = (\beta_{10}\ \beta_{20})'$ is the true value of β.*

In the next theorem we state the asymptotic distribution of the ML slope estimators is a normal distribution. Let

$$\begin{aligned} F(\gamma_0) &= -E\frac{\partial^2 \ln L}{\partial\beta\partial\beta'} \\ &= E\left(\Delta y_{i,-1}^{\circ}(\gamma_0)'\Omega^{-1}\Delta y_{i,-1}^{\circ}(\gamma_0)\right) \\ &= \frac{1}{\sigma_u^2}E\left(\Delta y_{i,-1}^{\circ}(\gamma_0)'\Omega^{\star-1}\Delta y_{i,-1}^{\circ}(\gamma_0)\right). \end{aligned} \tag{36}$$

Theorem 4. *Under Assumption 1, $\sqrt{n}(\widehat{\beta} - \beta) \to_d N(0, F^{-1}(\gamma_0))$, where $\beta = (\beta_1\ \beta_2)'$.*

7 Monte Carlo Experiments

7.1 Models

We consider two models: one without exogenous regressors and a second with exogenous regressors. Also, we consider in the second model only one exogenous variable, which for simplicity we consider to be the same as the threshold variable. Then we use the following models to generate y_{it},

$$y_{it} = \alpha_i + \beta_1 y_{it-1} 1(q_{it} \leq \gamma) + \beta_2 y_{it-1} 1(q_{it} > \gamma) + u_{it}, \tag{37}$$

and

$$y_{it} = \alpha_i + (\beta_1 y_{it-1} + \theta_1 q_{it}) 1(q_{it} \leq \gamma) + (\beta_2 y_{it-1} + \theta_2 q_{it}) 1(q_{it} > \gamma) + u_{it}. \tag{38}$$

We generate the variables as $q_{it} \sim N(1/2, 1)$ and $u_{it} \sim N(0, 1)$. The variables are generated from $t = -10$ to $t = T$, and then we discard the first 10 observations by using the observations $t = 0$ through T for estimation. In generating y_{it} we also set $y_{i,-10} = 0$.

7.2 Individual Fixed Effect Construction

For each model, we consider 3 designs to construct the individual fixed effect correlated with the exogenous threshold variable; each design considers different sets of the structural parameters.

Design 1

The individual effects, α_i, are generated as

$$\alpha_i = e_i + (T + 11)^{-1} \sum_{t=-10}^{T} q_{it}, \qquad e_i \sim N(2, 3),$$

and consider these structural parameters (γ, β_1, θ_1, β_2, θ_2)=(0, 0.5, 1.5, -0.5, -1.5). Notice that for the model without exogenous regressors we use only the set of parameters (γ, β_1, β_2).

Design 2

The individual effects, α_i, are generated as

$$\alpha_i = e_i + (T+11)^{-1} \sum_{t=-10}^{T} [-0.7q_{it}1(q_{it} \leq \gamma) + 0.4q_{it}1(q_{it} > \gamma)], \quad e_i \sim N(2,3),$$

and consider these structural parameters (γ, β_1, θ_1, β_2, θ_2)=(-0.5, -0.3, 1, -0.7, -1.2).

Design 3

The individual effects, α_i, are generated as

$$\alpha_i = e_i + (T+11)^{-1} \sum_{t=-10}^{T} [-0.3q_{it}1(q_{it} \leq \gamma) - 0.2q_{it}1(q_{it} > \gamma)], \quad e_i \sim N(2,3),$$

and consider these structural parameters (γ, β_1, θ_1, β_2, θ_2)=(1, -0.6, -1, 0.7, 0.5).

The three above designs of generating α_i ensures that the random effects slope estimates are inconsistent because of the correlation that exists between the individual specific effects and the explanatory variables q_{it}.

7.3 Simulation results

Table 1 presents the performance of the estimators where model a refers to the model without exogenous regressors (37), and model b refers to the model with exogenous regressors (38). The bias and root mean square error of the estimators γ, β_1, θ_1, β_2 and θ_2 for different choices of number of individuals n and a fixed time period $T = 4$ are shown. This table shows that as n increases the bias of the threshold parameter γ decreases quickly; also, the bias of the slope parameters β_1, θ_1, β_2 and θ_2 decreases.

Similarly, this table shows in general that as the number of individuals n increases, the Root Mean Square Error (MRSE) of all parameter estimates decreases. Notice this measure considers the second moments of the data.

8 Inflation and Long-Run Economic Growth

In the long-run, the literature has empathized that nominal variables do not have effects on real variables, i.e., inflation does not have effects on economic growth. Nevertheless, the empirical literature presents evidence on the negative inflation-growth relationship for periods of high inflation. Dornbusch and Fischer (1993) present a country case study

Table 1: Performance of estimators

			Bias of estimators		Root mean square error	
			$T = 4$		$T = 4$	
Model	Design	Coefficient	$n = 50$	$n = 500$	$n = 50$	$n = 500$
a	1	γ = 0.0	-0.006	0.000	0.027	0.002
		β_1 = 0.5	-0.002	0.000	0.055	0.018
		β_2 = -0.5	0.001	-0.001	0.039	0.012
a	2	γ = -0.5	0.014	0.000	0.134	0.013
		β_1 = -0.3	0.008	0.002	0.099	0.028
		β_2 = -0.7	0.001	0.000	0.057	0.017
a	3	γ = 1.0	-0.006	0.000	0.025	0.001
		β_1 = -0.6	-0.002	0.000	0.033	0.011
		β_2 = 0.7	-0.001	-0.001	0.056	0.017
b	1	γ = 0.0	-0.007	0.000	0.031	0.002
		β_1 = 0.5	-0.002	0.000	0.055	0.016
		θ_1 = 1.5	0.009	0.002	0.234	0.071
		β_2 = -0.5	0.001	-0.001	0.037	0.010
		θ_2 = -1.5	0.001	-0.001	0.123	0.035
b	2	γ = -0.5	0.022	0.000	0.170	0.009
		β_1 = -0.3	-0.003	0.001	0.099	0.024
		θ_1 = 1.0	0.020	0.004	0.233	0.061
		β_2 = -0.7	0.000	0.000	0.046	0.012
		θ_2 = -1.2	-0.001	-0.001	0.114	0.032
b	3	γ = 1.0	-0.007	0.000	0.031	0.002
		β_1 = -0.6	-0.002	0.000	0.055	0.018
		θ_1 = -1.0	0.009	0.002	0.234	0.071
		β_2 = 0.7	0.001	-0.001	0.037	0.011
		θ_2 = 0.5	0.001	0.000	0.123	0.040

Note: All results are based on 1000 replications.

in which the range 15 to 30 percent defines a regime of moderate inflation. Bruno and Easterly (1998) find that the negative correlation is robust only if countries with high inflation are included and propose that high inflation occurs when inflation is above 40 percent in annual terms.

Another group of the empirical literature focuses on a formal test for these breaking points that finally defines the threshold under which inflation has effects on growth. Khan and Senhadji (2001) estimate a threshold level of inflation of 1 percent for industrialized countries and 11 percent in a sample of developing countries. Using nonparametric methods and the specification of Khan and Senhadji (2001), Vaona and Schiavo (2007) find a threshold inflation level around 12 percent.

We estimate a threshold level of inflation above which inflation significantly slows down economic growth. The results are based on a dynamic panel threshold model developed in this work that extends the static version of Hansen (1999). We argue that the dynamic properties of our estimations allow us to capture long-run components in the behavior of the variables under analysis, since all economic growth models exhibit dynamics.

8.1 Data and Specification

The period of study spans from 1961 to 2000, over five-year average periods (in order to avoid capturing relations of cyclical type between involved variables) for a sample of 72 countries. The data corresponds primarily to the data in the work of Chang, Kaltani and Loayza (2009). See Chang et al. (2009) for details on the constructions and definitions of the variables used in the estimation analysis.

The database used in Chang et al. (2009) include 82 countries, but the method developed in this work assumes balanced panel data; for that reason, the panel in the estimation analysis has 72 countries, since some countries in Chang et al.'s (2009) database have missing values. The countries that are considered in this estimation are: Argentina, Australia, Austria, Belgium, Burkina Faso, Bolivia, Brazil, Canada, Switzerland, Chile, Cote d'Ivoire, Rep. Congo, Colombia, Costa Rica, Denmark, Dominican Republic, Ecuador, Egypt, Spain, Finland, France, United Kingdom, Ghana, Gambia, Greece, Guatemala, Honduras, Indonesia, India, Ireland, Iran, Iceland, Israel, Italy, Jamaica, Japan, Kenya, Rep Korea, Sri Lanka, Morocco, Madagascar, Mexico, Malaysia, Niger, Nigeria, Nicaragua, Netherlands, Norway, New Zealand, Pakistan, Panama, Peru, Philippines, Portugal, Paraguay, Senegal, Singapore, Sierra Leone, El Salvador, Sweden, Syrian Arab Republic, Togo, Thailand, Trinidad and Tobago, Tunisia, Turkey, Uruguay, United States, Venezuela, South Africa, Dem. Rep. Congo and Zambia.

In order to consider the threshold level of inflation effects on economic growth, we estimate the following variation of an economic growth model with a threshold variable,

$$y_{it} - y_{it-1} = \alpha_i + (\beta_1 y_{it-1} + \theta_1 \log(x_{it}))1(x_{it} \leq \gamma) + (\beta_2 y_{it-1} + \theta_2 \log(x_{it}))1(x_{it} > \gamma) + \pi' z_{it} + u_{it}, \quad (39)$$

where y_{it} is the log of real Gross Domestic Product (GDP) per capita, x_{it} is the inflation rate; α_i is an unobserved country specific-effect; γ is a threshold level of inflation rate; $1(.)$ is an indicator variable; z_{it} is a set of other determinants of economic growth; i refers to country; and t refers to time period.

As is standard in the literature, the dependent variable is the average rate of real per capita GDP growth over 5 years (i.e., the log difference of GDP per capita normalized by the length of the period). The equation is dynamic in that it includes the level of GDP per capita at the previous period in the set of explanatory variables. We include the previous period of the level of GDP per capita to account for transitional convergence because one of the main implications of the neoclassical growth model and indeed of all models is that exhibit transitional dynamics in which the growth rate relies on the initial position of the economy. The hypothesis of transitional convergence posits that in ceteris paribus, poor countries may grow faster than rich ones due to decreasing returns to scale in output. To account for the initial position of the economy we include the level of GDP per capita at the previous period.

All the explanatory variables are in logarithms, including the level of inflation as explanatory variable in equation (39). However, the log function does not exist for negative inflation rates. Thus, following Khan and Senhadji (2001) we use the semi-log transform; that is, $\log(x_{it}) = x_{it} - 1$ if $x_{it} \leq 1$ and it is the usual natural logarithm when $x_{it} > 1$.

We allow having threshold effects in the transitional convergence variable, since the conditional convergence hypothesis suggests countries converge to their own steady state. This hypothesis argues that differences in economic performances are caused by differences in both infrastructure and institutions. Acemoglu et al. (2003) state that countries pursuing poor macroeconomic policies (high inflation, large budget deficits, and a misaligned exchange rate) also have weak institutions. Thus, the periods of inflation crises reflect a country's poor institutions, and then it converges to a different steady state.

A great number of economic and social variables can be posited as determinants of economic growth. Following Loayza et al. (2005) the other determinants of economic growth that we consider are financial depth, governance, public infrastructure, human capital investment, trade openness, and economic instability.

8.2 Estimation and Inference Results

We estimate via maximum likelihood threshold model (39), and we add period-specific dummy variables.

Threshold parameter estimation

The point estimate of the threshold and its asymptotic 99 percent confidence interval are reported in Table 2. The estimate of the threshold level of inflation parameter is 13 percent; thus, the two classes of regimes indicated by the point estimate are those with "high inflation" for inflation rates higher than 13 percent, and "low inflation" for inflation rates lower than 13 percent.

Table 2: Asymptotic confidence interval in threshold model

	Threshold estimate (%)	99% confidence interval lower bound	99% confidence interval upper bound
$\widehat{\gamma}$	13	12	47

Note: Asymptotic critical values are reported in Hansen (2000).

The asymptotic confidence interval for the threshold level of inflation is not tight, indicating an important uncertainty about the nature of this division. This is not surprising since Khan and Senhadji (2001) documented many studies in which the threshold level of the inflation rate ranges from 1 percent to 40 percent using different specifications and samples.

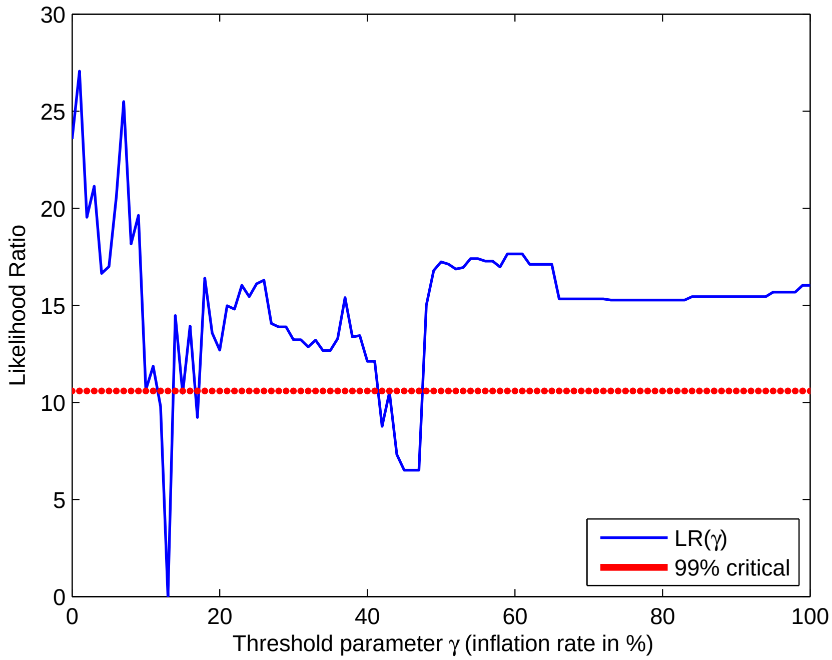
More information can be learned about the threshold estimates from plots of the concentrated likelihood ratio function $LR(\gamma)$. Figure 1 shows the likelihood ratio function, which is computed when estimating a threshold model. The threshold estimate is the point where the $LR(\gamma)$ equals zero, which occurs at $\hat{\gamma} = 13$ percent.

Figure 1 suggests there seems to be a second threshold around 45 percent. Bruno and Easterly (1998) found that the GDP per capita growth decreases dramatically for inflation rates above 40 percent, and they defined inflation above that value as periods of inflation crisis. They do not claim great precision for the 40 percent breakpoint, since there is some arbitrariness in the choice of a threshold.

Slope parameters estimation

Table 3 shows the estimation results of equation (39) for the worldwide sample. The coefficients of primary interest are those on the inflation rate. The point estimates suggest that inflation rates lower than 13 percent have no effect on GDP growth, as the classical

Figure 1: Confidence interval construction in threshold model



dichotomy hypothesis in the long-run suggests. However, inflation rates higher than 13 percent have a negative effect on economic growth at a 1 percent level of significance. The parameter estimate is different from zero at that significance level.

At a level of 1 percent of significance, the initial GDP per capita (as transitional convergence control, because less favored countries often grow faster due to a higher capital return) is significant and negative in both regimes as the neoclassic model predicts.

Trade openness has the expected positive effect on GDP growth because countries take advantage of the international competition and the international specialization. Public infrastructure has a positive effect on economic growth, indicating a significant input of the production function, or it increases total factor productivity by making other economic sectors more competitive.

Financial depth has the expected positive sign. This would mean that a better financial development would stimulate even higher development in the long term because it would facilitate risk diversification in the financial market, identify profitable investment projects, and mobilize savings; however, it is not statistically significant. Governance has a positive impact on GDP growth, indicating that a higher institutional quality of government would increase GDP growth, nevertheless, this variable is not significant. Similarly, the coefficient of human capital investment is not statistically significant.

Economic instability, as predicted, has a negative effect on economic growth. This

Table 3: Estimation results in threshold model

Dependent variable: GDP per capita growth			
Explanatory variables:	Estimates	ML SE	White SE
Inflation rate (Inflation rate < 13%) (Average of annual inflation rate, in semi-logs)	0.05	0.16	0.18
Inflation rate (Inflation rate > 13%) (Average of annual inflation rate, in semi-logs)	-0.92	0.25	0.35
Transitional Convergence (Inflation rate < 13%) (GDP per capita in previous period, in logs)	-3.60	0.47	0.36
Transitional Convergence (Inflation rate > 13%) (GDP per capita in previous period, in logs)	-3.35	0.49	0.41
Governance (Index from ICRG, 0 - 1)	0.19	1.15	0.99
Trade Openness (Structure-adjusted trade volume/GDP, in logs)	1.10	0.36	0.48
Human capital investment (secondary enrollment, in logs)	-0.75	0.40	0.61
Financial Depth (Private domestic credit/GDP, in logs)	0.38	0.26	0.22
Public Infrastructure (Main telephone lines per capita, in logs)	1.00	0.32	0.39
Economic Instability (St.Dev. of the annual GDP per capita growth)	-0.31	0.04	0.09
66-70 period shift	0.68	0.37	0.41
71-75 period shift	0.83	0.28	0.32
76-80 period shift	0.59	0.22	0.30
81-85 period shift	-0.89	0.21	0.33
86-90 period shift	0.05	0.23	0.25
91-95 period shift	-0.55	0.26	0.32
96-00 period shift	-0.71	0.36	0.35
Number of countries (n)	72		
Number of periods, five-year average (T)	8		
Observations used in the estimation (n*(T-1))	504		
Negative log-likelihood	999.9		

leads one to conclude that higher macroeconomic instability has a negative effect on the per capita GDP growth rate.

The period shifts had a negative effect on economic growth in the 1980s and 1990s, indicating that the international trend in economic growth experienced a declining trend from 1960 to 2000, resulting in less favorable external conditions in the 1980s and 1990s than in the previous decades.

9 Conclusion and Possible Extensions

In this paper we introduced econometric techniques for dynamic panel threshold models. Basically, we follow the approach of Hsiao et al. (2002) in the dynamic panel linear model, where we estimate a dynamic panel threshold model in differences using a maximum likelihood estimator.

We show the Maximum Likelihood estimation of the threshold parameter is consistent and converges to a double-sided standard Brownian motion distribution as in Hansen (2000) when the number of individuals grows to infinite for a fixed time period.

Since the likelihood function proposed is well defined and satisfies the usual regularity conditions at a known γ, then we show also the Maximum Likelihood estimation of the slope parameters are consistent and converge to a normal distribution when the number of individuals grows to infinite for a fixed time period.

Also, we evaluate the performance of the estimators in a Monte Carlo simulation for 1000 replications of the data; for a small sample size of number of individuals $n = 50$ and time periods $T = 4$, the estimators show a relatively small bias. When we increase the number of individuals to $n = 500$ for the same time periods $T = 4$, the estimators show practically zero bias. The RMSE also decreases quickly as the number of individuals increases for a fixed time period.

We apply the method to a sample of 72 countries and eight periods of five-year averages from 1961 to 2000, where the threshold level of inflation is estimated. We find a threshold inflation level at 13 percent. This result suggests that there are two regimes: (i) "high inflation" for those countries with an inflation rate higher than 13 percent; and (ii) "low inflation" for countries with inflation rates lower than 13 percent. Countries with inflation rates below 13 percent have no effects on long-run economic growth, while countries with inflation rates above such threshold have negative effects.

Several extensions would be desirable. One important extension is to develop another technique to allow for the lag of the dependent variable to be the threshold variable. Other extensions include allowing for heteroskedasticity, endogenous variables, random effects, non-balanced data, testing threshold effects, testing the number of thresholds, and testing random against fixed effects. Also, it would be interesting to compare our results with alternative approximations based on smoothing the indicator variable $1(q_{it} \leq \gamma)$. These would be interesting subjects for future research.

References

Acemoglu, D., Johnson, S., Robinson, J. and Y. Thaicharoen (2003): "Institutional Causes, Macroeconomic Symptoms: Volatility, Crises and Growth." *Journal of Monetary Economics* 50(1), 49-126.

Anderson, T. and C. Hsiao (1981): "Estimation of Dynamic Models With Error Components." *Journal of the American Statistical Association* 76(375), 598-606.

Anderson, T. and C. Hsiao (1982): "Formulation and Estimation of Dynamics Models Using Panel Data." *Journal of Econometrics* 18(1), 47-82.

Bai, J. (1997): "Estimating Multiple Breaks One at a Time." *Econometric Theory* 13(3), 315-352.

Blundell, R. and R.J. Smith (1991): "Conditions Initiales et Estimation Efficace dans les Modeles Dynamiques sur Donnees de Panel: une Application au Comportement d'Invetissement des Enterprises." *Annales d'Economie et de Statistique* 20/21, 109-123.

Bruno, M. and W. Easterly (1998): "Inflation Crises and Long-Run Growth." *Journal of Monetary Economics* 41(1), 3-26.

Caner, M. and B.E. Hansen (2001): "Threshold Autoregression with a Unit Root." *Econometrica* 69(6), 1555-1596.

Caner, M. and B.E. Hansen (2004): "Instrumental Variable Estimation of a Threshold Model." *Econometric Theory* 20(5), 813-843.

Chan, K.S. (1993): "Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model." *Annals of Statistics* 21(1), 520-533.

Chang, R., Kaltani, L. and N. Loayza (2009): "Openness Can be Good for Growth: The Role of Policy Complementarities." *Journal of Development Economics* 90(1), 33-49.

Chong, T. T-L. (1994): " Consistency of Change-Point Estimators when the Number of Change-Points in Structural Change Models is Underspecified." Working paper, Chinese University of Hong Kong.

Dornbusch, R. and S. Fischer (1993): "Moderate Inflation." *The World Bank Economic Review* 7(1), 1-44.

Enders, W. and C.W.J. Granger (1998): "Unit-Root Tests and Asymmetric Adjustment with an Example Using the Term Structure of Interest Rates." *Journal of Business and Economic Statistics* 16(13), 304-311.

Hansen, B.E. (1996): "Inference When a Nuisance Parameter is not Identified Under the Null Hypothesis." *Econometrica* 64(2), 413-430.

Hansen, B.E. (1997): "Inference in TAR Models." *Studies in Nonlinear Dynamics and Econometrics* 2(1), 1-14.

Hansen, B.E. (1999): "Threshold Effects in Non-Dynamic Panels: Estimation, Testing and Inference." *Journal of Econometrics* 93(2), 345-368.

Hansen, B.E. (2000): "Sample Splitting and Threshold Estimation." *Econometrica* 68(3), 575-603.

Hansen, B.E. (2011): "Threshold Autoregression in Economics." *Statistics and Its Interface* 4, 123-127.

Hansen, B.E. and B. Seo (2002): "Testing for Two-Regime Threshold Cointegration in Vector Error-Correction Models." *Journal of Econometrics* 110(2), 293-318.

Hsiao, C, M. Pesaran and K. Tahmiscioglu (2002): "Maximum Likelihood Estimation of Fixed Effects Dynamic Panel Data Models Covering Short Time Periods." *Journal of Econometrics* 109(1), 107-150.

Khan, M and A. Senhadji (2001): "Threshold Effects in the Relationship Between Inflation and Growth." *IMF Staff Papers* 48(1), 1-21.

Kourtellos, A., T. Stengos and C.M. Tan (2013): "Structural Threshold Regression." *Mimeo.*

Lancaster, T. (2000): "The Incidental Parameter Problem Since 1948." *Journal of Econometrics* 95(2), 391-413.

Loayza, N., P. Fajnzylber and C. Calderon (2005): *Economic Growth in Latin America and the Caribbean : Stylized Facts, Explanations, and Forecasts.* The World Bank.

Newey, W. and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing." *Handbook of Econometrics* IV, ed. by R. Engle and D. McFadden. New York: Elsevier, 2113-2245.

Nickell, S.J. (1981): "Biases in Dynamic Models with Fixed Effects." *Econometrica* 49(6), 1417- 1426.

Qian, L. (1998): "On Maximum Likelihood Estimators for a Threshold Autoregression." *Journal of Statistical Planning and Inference* 75(1), 21-46.

Samia N. and K.S. Chan (2011): "Maximum Likelihood Estimation of a Generalized Threshold Stochastic Regression Model." *Biometrika* 98(2), 433- 448.

Seo, M. and O. Linton (2007): "A Smoothed Least Squares Estimator for Threshold Regression Models." *Journal of Econometrics* 141(2), 704-735.

Seo, M.H. and Y. Shin (2014): "Dynamic Panels with Threshold Effect and Endogeneity." *Mimeo.*

Tong, H. (1983): *Threshold Models in Nonlinear Time Seties Analysis: Lecture Notes in Statistics* 21. Berlin: Springer.

Tong, H. (2007): "Birth of the Threshold Time Series Model." *Statistica Sinica* 17(1), 8-14.

Vaona, A. and S. Schiavo (2007): "Nonparametric and Semiparametric Evidence On the Long-Run Effects of Inflation on Growth." *Economics Letters* 94(3), 452-458.

Yu, P. (2013): "Inconsistency of 2SLS Estimators in Threshold Regression with Endogeneity." *Economics Letters* 120(3), 532-536.

Yu, P. (2012): "Likelihood Estimation and Inference in Threshold Regression." *Journal of Econometrics* 167(1), 274-294.

Yu, P. and P. Phillips (2014): "Threshold Regression with Endogeneity." *Mimeo.*

Appendix A. Derivation of the Maximum Likelihood Function

In the linear case we can derive directly the conditional likelihood function since the regressor in the model in difference is the lag of the dependent variable, but in the differentiated panel threshold model that is not exactly the case.

Let $\mathcal{F}_{it}$ denote the σ-algebra generated by y_{is-1} and x_i for $s \leq t$; and $\mathcal{F}_{i0}$ is the σ-algebra generated by x_i. Thus, The ML function is given by

$$\prod_{i=1}^{n} \prod_{t=1}^{T} f(\Delta y_{it} | \mathcal{F}_{it-1}), \tag{A.1}$$

where $f(\Delta y_{it}|\mathcal{F}_{it-1})$ for $t = 2, 3, \ldots, T$ are fully specified by (3) in the model without exogenous regressors and by (15) in the model with exogenous regressors, and $f(\Delta y_{i1}|\mathcal{F}_{i0})$ can be derived by (4) and (16) in the model without and with exogenous regressors, respectively.

Using the notation of chapter 3 and 4, the ML function (A.1) is equivalent to

$$\prod_{i=1}^{n} f(\Delta y_i | x_i), \tag{A.2}$$

where we can write the joint density of Δy_i given x_i as

$$\begin{aligned} f(\Delta y_i|x_i) = f(\Delta y_{iT}|\Delta y_{iT-1}, \ldots, \Delta y_{i1}, x_i) f(\Delta y_{iT-1}|\Delta y_{iT-2}, \ldots, \Delta y_{i1}, x_i) \ldots \\ f(\Delta y_{i2}|\Delta y_{i1}, x_i) f(\Delta y_{i1}|x_i). \end{aligned} \tag{A.3}$$

All the terms in (A.3) can be derived using (3) and (4) in the model without exogenous regressors and using (15) and (16) in the model with exogenous regressors.

Appendix B. Second-Order Derivatives of the Log-Likelihood Function

In this appendix we provide the second-order derivatives of function (8) considering the model (1). We can redefine the parameters and the matrix $\Delta y_{i,-1}(\gamma)$ for the model with exogenous regressors (13) and then we can get analogous expressions.

The second-derivatives are:

$$\frac{\partial^2 \ln L}{\partial\beta_\delta\partial\beta_\delta'} = -\frac{1}{\sigma_u^2}\sum_{i=1}^{n}[\Delta y_{i,-1}(\gamma)'\Omega^{\star-1}\Delta y_{i,-1}(\gamma)], \tag{B.1}$$

$$\frac{\partial^2 \ln L}{\partial\beta_\delta\partial\sigma_u^2} = -\frac{1}{\sigma_u^4}\sum_{i=1}^{n}[\Delta y_{i,-1}(\gamma)'\Omega^{\star-1}(\Delta y_i - \Delta y_{i,-1}(\gamma)\beta_\delta)], \tag{B.2}$$

$$\frac{\partial^2 \ln L}{\partial\beta_\delta\partial\omega} = -\frac{1}{\sigma_u^2[1+T(\omega-1)]^2}\sum_{i=1}^{n}[\Delta y_{i,-1}(\gamma)'\Phi(\Delta y_i - \Delta y_{i,-1}(\gamma)\beta_\delta)], \tag{B.3}$$

$$\frac{\partial^2 \ln L}{\partial\sigma_u^2\partial\sigma_u^2} = \frac{nT}{2\sigma_u^4} - \frac{1}{\sigma_u^6}\sum_{i=1}^{n}[(\Delta y_i - \Delta y_{i,-1}(\gamma)\beta_\delta)'\Omega^{\star-1}(\Delta y_i - \Delta y_{i,-1}(\gamma)\beta_\delta)], \tag{B.4}$$

$$\begin{aligned}\frac{\partial \ln L}{\partial\omega\partial\omega} = \frac{nT^2}{2[1+T(\omega-1)]^2} - \frac{1}{\sigma_u^2[1+T(\omega-1)]^3}&\\ \sum_{i=1}^{n}[(\Delta y_i - \Delta y_{i,-1}(\gamma)\beta_\delta)'\Phi(\Delta y_i - \Delta y_{i,-1}(\gamma)\beta_\delta)],&\end{aligned} \tag{B.5}$$

and

$$\frac{\partial \ln L}{\partial\omega\partial\sigma_u^2} = -\frac{1}{\sigma_u^4[1+T(\omega-1)]^2}\sum_{i=1}^{n}[(\Delta y_i - \Delta y_{i,-1}(\gamma)\beta_\delta)'\Phi(\Delta y_i - \Delta y_{i,-1}(\gamma)\beta_\delta)]. \tag{B.6}$$

Appendix C. Proof of Theorems

Lemma 1.

$$E\left(C'\Delta y_{i,-1}^{*}(\gamma)'\Omega^{\star-1}\Delta u_i\right) = 0 \quad \textit{for all } \gamma \in \Gamma. \tag{C.1}$$

Proof. Recall

$$\Omega = \sigma_u^2 \begin{bmatrix} \omega & -1 & 0 & \dots & 0 \\ -1 & 2 & -1 & & \\ 0 & -1 & 2 & & \\ \vdots & & & \ddots & -1 \\ 0 & & & -1 & 2 \end{bmatrix} = \sigma_u^2\Omega^\star,$$

where $\omega = \sigma_v^2/\sigma_u^2$.

We note that the only unknown element of $\Omega^{\star}$ is ω. Then following to Hsiao et al. (2002), let

$$H = \begin{bmatrix} a_0 & 0 & 0 & \dots & 0 \\ a_0 & a_1 & 0 & \dots & \vdots \\ a_0 & a_1 & a_2 & \dots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_0 & a_1 & a_2 & \dots & a_{T-1} \end{bmatrix}, \tag{C.2}$$

where $a_{s+1} - 2a_s + a_{s-1} = 0$, $s = 1, 2, \dots, T$. With $a_0 = 1$, $a_1 = \omega$, which has the solution $a_s = 1 + s(\omega - 1)$ for $s = 1, 2, \dots, T$. Then

$$H\Omega^{\star}H' = diag(a_0 a_1, a_1 a_2, \dots, a_{T-1}a_T) = \Lambda. \tag{C.3}$$

Since $\Omega > 0$ then $\Omega^{\star -1} = \sigma_u^2 \Omega^{-1} = H'\Lambda^{-1}H$ and then there is a unique root $\Lambda^{-1/2}H$ which is also positive definite $\Lambda^{-1/2}H > 0$. Then (C.1) can be written in the form

$$E\left(c' \sum_{t=1}^{T} (a_t a_{t-1})^{-1} 1(q_{i1} \leq \gamma) \left(\upsilon_{i1} + \sum_{s=2}^{t} a_{s-1} \Delta u_{is} \right) + \right.$$
$$\left. c' \sum_{t=2}^{T} (a_t a_{t-1})^{-1} \left(\sum_{s=1}^{t-1} a_s \Delta y_{is}^{*}(\gamma) \right) \left(\upsilon_{i1} + \sum_{s=2}^{t} a_{s-1} \Delta u_{is} \right) \right). \tag{C.4}$$

The first term of (C.4) is the threshold variable, by assumption this variable is exogenous with respect to υ_{i1} and u_{it} for all t. Then we proof that for the second term.

Let $\Delta 1(q_{is} \leq \gamma) = 1(q_{is} \leq \gamma) - 1(q_{is-1} \leq \gamma)$ and $\Delta P(q_{is} \leq \gamma) = P(q_{is} \leq \gamma) - P(q_{is-1} \leq \gamma)$. And compute the expectations

$$E\left[\Delta y_{is}^*(\gamma)\upsilon_{i1}\right]\begin{cases} = \sigma_\upsilon^2 P(q_{i2}\leq\gamma) \text{ for } s=1. \\ = \beta_1\sigma_\upsilon^2 P(q_{i2}\leq\gamma)P(q_{i3}\leq\gamma)+\beta_2\sigma_\upsilon^2 P(q_{i2}>\gamma)P(q_{i3}\leq\gamma)-\sigma_u^2 P(q_{i3}\leq\gamma) \\ +\sigma_\upsilon^2\Delta P(q_{i3}\leq\gamma) \text{ for } s=2. \\ = \beta_1\left(E\left[\Delta y_{is-1}^*(\gamma)\upsilon_{i1}\right]P(q_{is+1}\leq\gamma)+\sum_{j=1}^{s-2}E\left[\Delta y_{ij}^*(\gamma)\upsilon_{i1}\right]\Delta P(q_{is+1}\leq\gamma)\right) \\ +\beta_2\left(E\left[\Delta y_{is-1}^+(\gamma)\upsilon_{i1}\right]P(q_{is+1}\leq\gamma)+\sum_{j=1}^{s-2}E\left[\Delta y_{ij}^+(\gamma)\upsilon_{i1}\right]\Delta P(q_{is+1}\leq\gamma)\right) \\ (-\sigma_u^2+\sigma_\upsilon^2)\Delta P(q_{is+1}\leq\gamma) \text{ for } s=3,\ldots,T. \end{cases} \tag{C.5}$$

$$\begin{aligned} E\left[\Delta y_{is}^*(\gamma)\Delta u_{is}\right] &= E\left[(y_{is}1(q_{is+1}\leq\gamma)-y_{is-1}1(q_{is}\leq\gamma))\Delta u_{is}\right] \\ &= E\left[(\Delta y_{is}1(q_{is+1}\leq\gamma)+y_{is-1}\Delta 1(q_{is+1}\leq\gamma))\Delta u_{is}\right] \\ &= E\big[(\beta_1\Delta y_{is-1}^*(\gamma)+\beta_2\Delta y_{is-1}^+(\gamma)+\Delta u_{is})1(q_{is+1}\leq\gamma)\Delta u_{is} \\ &\quad +y_{is-1}\Delta 1(q_{is+1}\leq\gamma)\Delta u_{is}\big] \\ &= -\beta_1\sigma_u^2 P(q_{is}\leq\gamma)P(q_{is+1}\leq\gamma)-\beta_2\sigma_u^2 P(q_{is}>\gamma))P(q_{is+1}\leq\gamma)) \\ &\quad +\sigma_u^2 P(q_{is+1}\leq\gamma)+\sigma_u^2 P(q_{is}\leq\gamma), \end{aligned} \tag{C.6}$$

$$\begin{aligned} E\left[\Delta y_{is}^*(\gamma)\Delta u_{is+1}\right] &= E\left[(y_{is}1(q_{is+1}\leq\gamma)-y_{is-1}1(q_{is}\leq\gamma))\Delta u_{is+1}\right] \\ &= E\left[(\Delta y_{is}1(q_{is+1}\leq\gamma)+y_{is-1}\Delta 1(q_{is+1}\leq\gamma))\Delta u_{is+1}\right] \\ &= E\big[(\beta_1\Delta y_{is-1}^*(\gamma)+\beta_2\Delta y_{is-1}^+(\gamma)+\Delta u_{is})1(q_{is+1}\leq\gamma)\Delta u_{is+1} \\ &\quad +y_{is-1}\Delta 1(q_{is+1}\leq\gamma)\Delta u_{is+1}\big] \\ &= -\sigma_u^2 P(q_{is+1}\leq\gamma), \end{aligned} \tag{C.7}$$

and

$$E\left[\Delta y_{is}^*(\gamma)\Delta u_{is+j}\right]=0 \quad \text{for } j=2,\ldots,T. \tag{C.8}$$

Notice that it is enough to show that for any γ

$$E\left[\left(\sum_{s=1}^{t-1}a_s\Delta y_{is}^*(\gamma)\right)\left(\upsilon_{i1}+\sum_{s=2}^{t}a_{s-1}\Delta u_{is}\right)\right]=0. \tag{C.9}$$

The analytic form of (C.9) is very complicated, then we proof the lemma by induction. Let

$$E[A_t] = E\left[\left(\sum_{s=1}^{t-1} a_s \Delta y_{is}^*(\gamma)\right)\left(\upsilon_{i1} + \sum_{s=2}^{t} a_{s-1}\Delta u_{is}\right)\right]. \tag{C.10}$$

Using (C.8), we can write

$$E[A_t] = E[A_{t-1}] + a_{t-1}E\left[\Delta y_{it-1}^*(\gamma)\upsilon_{i1}\right] + a_{t-1}\sum_{s=2}^{t} a_{s-1}\left[\Delta y_{it-1}^*(\gamma)\Delta u_{is}\right], \tag{C.11}$$

if $E[A_{t-1}] = 0$, then $a_{t-1}E\left[\Delta y_{it-1}^*(\gamma)\upsilon_{i1}\right] + a_{t-1}\sum_{s=2}^{t} a_{s-1}\left[\Delta y_{it-1}^*(\gamma)\Delta u_{is}\right]$ must be zero. Thus, by using (C.5), (C.6), (C.7) and (C.8) for $t = 2$

$$\begin{aligned}
E[A_2] &= a_1 E\left[\Delta y_{i1}^*(\gamma)\upsilon_{i1}\right] + a_1^2 E\left[\Delta y_{i1}^*(\gamma)\Delta u_{i2}\right] \\
&= a_1\sigma_\upsilon^2 P(q_{i2}\le\gamma) - a_1^2\sigma_u^2 P(q_{i2}\le\gamma) \\
&= a_1^2\sigma_u^2 P(q_{i2}\le\gamma) - a_1^2\sigma_u^2 P(q_{i2}\le\gamma) \\
&= 0,
\end{aligned} \tag{C.12}$$

for $t = 3$

$$\begin{aligned}
E[A_3] &= a_2 E\left[\Delta y_{i2}^*(\gamma)\upsilon_{i1}\right] + a_2(a_1 E\left[\Delta y_{i2}^*(\gamma)\Delta u_{i2}\right] + a_2 E\left[\Delta y_{i2}^*(\gamma)\Delta u_{i3}\right]) \\
&= a_2\beta_1\sigma_\upsilon^2 P(q_{i2}\le\gamma)P(q_{i3}\le\gamma) + a_2\beta_2\sigma_\upsilon^2 P(q_{i2}>\gamma)P(q_{i3}\le\gamma) \\
&\quad -a_2\sigma_u^2 P(q_{i3}\le\gamma) + a_2\sigma_\upsilon^2\Delta P(q_{i3}\le\gamma) - a_1a_2\beta_1\sigma_u^2 P(q_{i2}\le\gamma)P(q_{i3}\le\gamma) \\
&\quad -a_1a_2\beta_2\sigma_u^2 P(q_{i2}>\gamma))P(q_{i3}\le\gamma)) + a_1a_2\sigma_u^2 P(q_{i3}\le\gamma) + a_1a_2\sigma_u^2 P(q_{i2}\le\gamma) \\
&\quad -a_2^2\sigma_u^2 P(q_{i3}\le\gamma) \\
&= (-1+2a_1-a_2)a_2\sigma_u^2 P(q_{i3}\le\gamma) \\
&= 0,
\end{aligned} \tag{C.13}$$

For $t = 4$,

$$
\begin{aligned}
E[A_4] &= a_3E\left[\Delta y^*_{i3}(\gamma)v_{i1}\right] + a_3(a_1E\left[\Delta y^*_{i3}(\gamma)\Delta u_{i2}\right] + a_2E\left[\Delta y^*_{i3}(\gamma)\Delta u_{i3}\right] + \\
&= a_3E\left[\Delta y^*_{i3}(\gamma)\Delta u_{i4}\right]) \\
&= \{(-1+2a_1-a_2)\left[(\beta_1P(q_{i3}\le\gamma)+\beta_2P(q_{i3}>\gamma))P(q_{i4}\le\gamma)+\Delta P(q_{i4}\le\gamma)\right] \\
&\quad +(-a_1+2a_2-a_3)P(q_{i4}\le\gamma)\}a_3\sigma_u^2 \\
&= 0. \qquad \text{(C.14)}
\end{aligned}
$$

We now know that it is true for at least $t=2$ by (C.12), then we can assume that it is true up to some fixed number j, which is at least 3, by proving that it is true for $j+1$ by (C.13); then it is true for at least $t=3$. Thus, now $j=3$, but since the statement is true for $j+1$, t is at least 4 (proved in (C.14)). In this manner, we can repeat this pattern indefinitely. Therefore, the statement holds for $t=2,\ldots,T$ for all γ.

□

Define

$$
J_n(\gamma) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} C'\Delta y^*_{i,-1}(\gamma)'\Omega^{\star-1}\Delta u_i. \qquad \text{(C.15)}
$$

Let "$\Rightarrow$" denote weak convergence with respect to the uniform metric.

Lemma 2. *$J_n(\gamma)\Rightarrow J(\gamma)$, a mean zero Gaussian process with almost surely continuous sample paths.*

Proof. For each γ and for a fix t, $\Delta y^*_{i,-1}(\gamma)$ is iid across i. So $J_n(\gamma)$ converges pointwise to a Gaussian distribution by the Central Limit Theorem. This can be extended to any finite collection of γ to yield a convergence of the finite dimensional distributions. The mean-zero for each γ is guaranteed by Lemma 1.

Then by using Lemma A.3 and Lemma A.4 of Hansen (2000), $J_n(\gamma)$ is tight. Therefore $J_n(\gamma)\Rightarrow J(\gamma)$. □

Lemma 3. *Uniform strong law*

If assumptions 1 holds, then

$$
\sup_{\gamma\in R}\left|\frac{1}{n}\sum_{i=1}^{n}\Delta y^*_{i,-1}(\gamma) - E(\Delta y^*_{i,-1}(\gamma))\right| \to_{a.s.} 0 \qquad \text{(C.16)}
$$

where $|.|$ denotes the absolute value and $\to_{a.s.}$ denotes convergence almost surely.

Proof. For a fix t, $y^*_{i,-1}(\gamma)$ is iid across i, $|y^*_{i,-1}(\gamma)| < \infty$ by assumption 1.6 and x_{it} have a continuous distribution by assumption 1.7. These establishes the conditions of Lemma 1 of Hansen (1996). Hence following the same argument it can proved the uniform strong law.

If the asymptotic theory is when T grows to infinity, and since the process y_{it} is not iid over t, we require to assume that y_{it} is stationary and ergodic, sufficient conditions for the stationarity of y_{it} are $|\beta_1| < 1$ and $|\beta_2| < 1$ (see Enders and Granger (1998) and Caner and Hansen (2001)). $\square$

Proof of Theorem 1. Notice that by Lemma 3, uniformly in $\gamma \in \Gamma$

$$\frac{1}{n}\sum_{i=1}^{n} C'\Delta y^*_{i,-1}(\gamma)'\Omega^{\star-1}\Delta y^*_{i,-1}(\gamma)C \to_p M(\gamma), \tag{C.17}$$

and

$$\frac{1}{n}\sum_{i=1}^{n} C'\Delta y'_{i,-1}\Omega^{\star-1}\Delta y_{i,-1}C \to_p M. \tag{C.18}$$

Let $\nabla y_{it-1}(\gamma) = y_{it-1}[1(q_{it} \leq \gamma) - 1(q_{it} \leq \gamma_0)]$.

The equation (1) holds when $\gamma = \gamma_0$, the true value. For values of $\gamma \neq \gamma_0$, note that (1) can be re-written as

$$\begin{aligned}
y_{it} &= \alpha_i + \beta_1 y_{it-1}1(q_{it} \leq \gamma_0) + \beta_2 y_{it-1}1(q_{it} > \gamma_0) + u_{it} \\
&= \alpha_i + \beta_1 y_{it-1}1(q_{it} \leq \gamma) + \beta_2 y_{it-1}1(q_{it} > \gamma) \\
&\quad -\beta_1 y_{it-1}[1(q_{it} \leq \gamma) - 1(q_{it} \leq \gamma_0)] - \beta_2 y_{it-1}[1(q_{it} > \gamma) - 1(q_{it} > \gamma_0)] + u_{it} \\
&= \alpha_i + \beta_1 y_{it-1}1(q_{it} \leq \gamma) + \beta_2 y_{it-1}1(q_{it} > \gamma) \\
&\quad +(\beta_2 - \beta_1) y_{it-1}[1(q_{it} \leq \gamma) - 1(q_{it} \leq \gamma_0)] + u_{it} \\
&= \alpha_i + \beta_1 y_{it-1}1(q_{it} \leq \gamma) + \beta_2 y_{it-1}1(q_{it} > \gamma) + cn^{-\alpha}\nabla y_{it-1}(\gamma) + u_{it}
\end{aligned} \tag{C.19}$$

The first difference transformation is linear, so can be applied to (C.19) to yield

$$\Delta y_{it} = \beta_1 \Delta y^*_{it-1}(\gamma) + \beta_2 \Delta y^+_{it-1}(\gamma) + cn^{-\alpha}\nabla\Delta y^*_{it-1}(\gamma) + \Delta u_{it}. \tag{C.20}$$

Similarly, (4) can be re-written as

$$\begin{aligned}
\Delta y_{i1} &= \delta_1 1(q_{i1} \leq \gamma_0) + \delta_2 1(q_{i1} > \gamma_0) + \upsilon_{i1} \\
&= \delta_1(1(q_{i1} \leq \gamma) + \delta_2 1(q_{i1} > \gamma) \\
&\quad -\delta_1[1(q_{i1} \leq \gamma) - 1(q_{i1} \leq \gamma_0)] - \delta_2[1(q_{i1} > \gamma) - 1(q_{i1} > \gamma_0)] + \upsilon_{i1} \\
&= \delta_1 1(q_{i1} \leq \gamma) + \delta_2 1(q_{i1} > \gamma) + (\delta_2 - \delta_1)[1(q_{i1} \leq \gamma) - 1(q_{i1} \leq \gamma_0)] + \upsilon_{i1} \\
&= \delta_1 1(q_{i1} \leq \gamma) + \delta_2 1(q_{i1} > \gamma) + cn^{-\alpha}\nabla 1(q_{i1} \leq \gamma) + \upsilon_{i1}. \qquad \text{(C.21)}
\end{aligned}$$

Hansen (2000) shows that the asymptotic distribution of $\widehat{\gamma}$ is not affected by the estimation of the slope parameters and the variance, then this holds in our environment as well. We can thus simplify matters by assuming that δ, β, σ_u^2 and ω are known and only γ is estimated, thus the covariance matrix Ω^* is known and the estimation residual (for fixed γ) is

$$\Delta\widehat{u}_i(\gamma) = \nabla\Delta y^*_{i,-1}(\gamma)Cn^{-\alpha} + \Delta u_i, \qquad \text{(C.22)}$$

where $\nabla\Delta y^*_{i,-1}(\gamma)$ is given by

$$\nabla\Delta y^*_{i,-1}(\gamma) = \begin{bmatrix} \nabla 1(q_{i1} \leq \gamma) & 0 \\ 0 & \nabla\Delta y^*_{i1}(\gamma) \\ 0 & \nabla\Delta y^*_{i2}(\gamma) \\ \vdots & \vdots \\ 0 & \nabla\Delta y^*_{iT-1}(\gamma) \end{bmatrix}.$$

Notice that conditional on $\Omega^\star$, the MLE is asymptotically equivalent to the minimum distance estimator $S_n(\gamma) = \sum_{i=1}^n \Delta\widehat{u}_i(\gamma)'\Omega^{\star-1}\Delta\widehat{u}_i(\gamma)$. Then using (C.22) we have

$$
\begin{aligned}
S_n(\gamma) - S_n(\gamma_0) &= \sum_{i=1}^{n} \Delta\widehat{u}_i(\gamma)'\Omega^{\star-1}\Delta\widehat{u}_i(\gamma) - \sum_{i=1}^{n} \Delta u_i'\Omega^{\star-1}\Delta u_i \\
&= \sum_{i=1}^{n} (\nabla\Delta y_{i,-1}^{*}(\gamma)Cn^{-\alpha} + \Delta u_i)'\Omega^{\star-1}(\nabla\Delta y_{i,-1}^{*}(\gamma)Cn^{-\alpha} + \Delta u_i) \\
&\quad - \sum_{i=1}^{n} \Delta u_i'\Omega^{\star-1}\Delta u_i \\
&= n^{-2\alpha} \sum_{i=1}^{n} C'\nabla\Delta y_{i,-1}^{*}(\gamma)'\Omega^{\star-1}\nabla\Delta y_{i,-1}^{*}(\gamma)C \\
&\quad + 2n^{-\alpha} \sum_{i=1}^{n} C'\nabla\Delta y_{i,-1}^{*}(\gamma)'\Omega^{\star-1}\Delta u_i \qquad \text{(C.23)}
\end{aligned}
$$

Using Lemma 2, we see that uniformly over $\gamma \in \Gamma$

$$
n^{2\alpha-1}(S_n(\gamma) - S_n(\gamma_0)) = n^{-1} \sum_{i=1}^{n} C'\nabla\Delta y_{i,-1}^{*}(\gamma)'\Omega^{\star-1}\nabla\Delta y_{i,-1}^{*}(\gamma)C + o_p(1) \qquad \text{(C.24)}
$$

Then, using Lemma 3 we calculate the uniformly over $\gamma \in \Gamma$

$$
\begin{aligned}
n^{2\alpha-1}(S_n(\gamma) - S_n(\gamma_0)) &\rightarrow_p (M(\gamma)'M(\gamma) + M(\gamma_0)'M(\gamma_0) - 2M(\gamma)'M(\gamma_0)) \\
&\equiv k(\gamma).
\end{aligned}
$$

Note that $k(\gamma)$ is a continuous non-negative real function which achieves its unique minimum of 0 at γ_0. For $\gamma \geq \gamma_0$, $\Delta y_{i,-1}^{*}(\gamma)'\Omega^{\star-1}\Delta y_{i,-1}^{*}(\gamma_0) = \Delta y_{i,-1}^{*}(\gamma_0)'\Omega^{\star-1}\Delta y_{i,-1}^{*}(\gamma_0)$. Thus, uniformly over $\gamma \in [\gamma_0, \bar{\gamma}]$

$$
\begin{aligned}
n^{2\alpha-1}(S_n(\gamma_0) - S_n(\gamma)) &\rightarrow_p (M(\gamma)'M(\gamma) - M(\gamma_0)'M(\gamma_0)) \\
&\equiv k_1(\gamma).
\end{aligned}
$$

The derivative of $k_1(\gamma)$ with respect to γ is given by

$$
\frac{d}{d\gamma}k_1(\gamma) = (M(\gamma)'D(\gamma) + D(\gamma)'M(\gamma)) \geq 0. \qquad \text{(C.25)}
$$

Thus, $k_1(\gamma)$ is continuous and weakly increasing on $[\gamma_0, \bar{\gamma}]$. Additionally, by assumption

(10)

$$\frac{d}{d\gamma}k_1(\gamma_0) = (M(\gamma_0)'D(\gamma_0) + D(\gamma_0)'M(\gamma_0)) > 0. \tag{C.26}$$

Hence, $k(\gamma)$ is uniquely minimized at γ_0 on $[\gamma_0, \bar{\gamma}]$.

Symmetrically, it can be shown that the function is uniquely minimized at γ_0 on $[\underline{\gamma}, \gamma_0]$. Therefore, since $\widehat{\gamma}$ minimizes $S_n(\gamma) - S_n(\gamma_0)$, it follows that $\widehat{\gamma} \to_p \gamma_0$ (see, e.g., Theorem 2.1 of Newey and McFadden (1994)).

□

The proof of the theorem 2 is based on the following lemmas. Let $\lambda_n = n^{1-2\alpha}$.

Lemma 4. *As $n \to \infty$, uniformly over $\nu \in [-\overline{\nu}, \overline{\nu}]$,*

$$n^{-2\alpha} \sum_{i=1}^{n} C'\nabla\Delta y^*_{i,-1}(\gamma_0 + \nu/\lambda_n)'\Omega^{\star-1}\nabla\Delta y^*_{i,-1}(\gamma_0 + \nu/\lambda_n)C \Rightarrow D(\gamma_0)|\nu|. \tag{C.27}$$

Proof. We prove (C.27) for the case $\nu \in [0, \overline{\nu}]$. We will show that for $\gamma = \gamma_0 + \nu/\lambda_n$,

$$\begin{aligned} &E\left(n^{-2\alpha} \sum_{i=1}^{n} C'\nabla\Delta y^*_{i,-1}(\gamma)'\Omega^{\star-1}\nabla\Delta y^*_{i,-1}(\gamma)C\right) \\ &= \lambda_n E\left(C'\nabla\Delta y^*_{i,-1}(\gamma)'\Omega^{\star-1}\nabla\Delta y^*_{i,-1}(\gamma)C\right) \\ &\to D(\gamma_0)|\nu|. \end{aligned} \tag{C.28}$$

Arguments similar to those in the proof of Lemma A.10 of Hansen (2000) and Lemma A.1 of Hansen (1999) show that (C.28) implies (C.27) under the assumptions. Notice that

$$\begin{aligned} &\lambda_n E\left(C'\nabla\Delta y^*_{i,-1}(\gamma)'\Omega^{\star-1}\nabla\Delta y^*_{i,-1}(\gamma)C\right) \\ &= \lambda_n c'E\left(\sum_{t=1}^{T}(a_t a_{t-1})^{-1}\nabla 1(q_{i1} \le \gamma) + \sum_{t=2}^{T}(a_t a_{t-1})^{-1}\left(\sum_{s=1}^{t-1} a_s \nabla\Delta y^*_{is}(\gamma)\right)^2\right. \\ &\left.+ 2\sum_{t=2}^{T}(a_t a_{t-1})^{-1}\nabla 1(q_{i1} \le \gamma)\sum_{s=1}^{t-1} a_s \nabla\Delta y^*_{is}(\gamma)\right) c. \end{aligned} \tag{C.29}$$

Observe that since $\gamma = \gamma_0 + \nu/\lambda_n \to \gamma_0$,

$$\begin{aligned}
\lambda_n P(\gamma_0 < q_{it} \leq \gamma) &= \nu \frac{P(q_{it} \leq \gamma) - P(q_{it} \leq \gamma_0)}{\gamma - \gamma_0} \\
&\to f_t(\gamma_0)\nu
\end{aligned} \tag{C.30}$$

as $n \to \infty$. Thus, the first term on the right-hand-side of (C.29),

$$\begin{aligned}
\lambda_n c' \sum_{t=1}^{T} (a_t a_{t-1})^{-1} P(\gamma_0 < q_{i1} \leq \gamma) c &= c' \sum_{t=1}^{T} (a_t a_{t-1})^{-1} \frac{P(q_{i1} \leq \gamma) - P(q_{i1} \leq \gamma_0)}{\gamma - \gamma_0} \nu c \\
&\to c' \sum_{t=1}^{T} (a_t a_{t-1})^{-1} f_1(\gamma_0) \nu c
\end{aligned} \tag{C.31}$$

Expansion of the quadratic of the expectation of the second term on the right-hand-side of (C.29) yields

$$\begin{aligned}
E\left(\sum_{s=1}^{t-1} a_s \nabla \Delta y_{is}^*(\gamma)\right)^2 = \; & \sum_{s=1}^{t-1} a_s^2 \left(E(\nabla y_{is}(\gamma))^2 + E(\nabla y_{is-1}(\gamma))^2\right. \\
& \left. -2E(\nabla y_{is}(\gamma) \nabla y_{is-1}(\gamma))\right) \\
& + \sum_{s=1}^{t-2} 2 a_s a_{s+1} \left(E(\nabla y_{is+1}(\gamma) \nabla y_{is}(\gamma))\right. \\
& -E(\nabla y_{is+1}(\gamma) \nabla y_{is-1}(\gamma)) - E(\nabla y_{is}(\gamma))^2 \\
& \left. + E(\nabla y_{is}(\gamma) \nabla y_{is-1}(\gamma))\right) \\
& + \sum_{s=1}^{t-3} \sum_{j=s+2}^{t-1} 2 a_s a_j \left(E(\nabla y_{ij}(\gamma) \nabla y_{is}(\gamma))\right. \\
& -E(\nabla y_{ij}(\gamma) \nabla y_{is-1}(\gamma)) \\
& -E(\nabla y_{ij-1}(\gamma) \nabla y_{is}(\gamma)) \\
& \left. + E(\nabla y_{ij-1}(\gamma) \nabla y_{is-1}(\gamma))\right).
\end{aligned} \tag{C.32}$$

Regardless to the constants, the relevant terms in (C.32) are of the form $E(\nabla y_{is}(\gamma))^2$ and $E(\nabla y_{is}(\gamma) \nabla y_{ik}(\gamma))$ for $k \neq s$. Thus,

$$
\begin{aligned}
\lambda_n E(\nabla y_{is}(\gamma))^2 &= \lambda_n E(y_{is}^2 1(\gamma_0 < q_{is+1} \leq \gamma)) \\
&= E(y_{is}^2 | \gamma_0 < q_{is+1} \leq \gamma)\lambda_n P(\gamma_0 < q_{is+1} \leq \gamma)) \\
&\to E(y_{is}^2 | q_{is+1} = \gamma_0) f_{s+1}(\gamma_0)\nu. \qquad (C.33)
\end{aligned}
$$

For the form $E(\nabla y_{is}(\gamma)\nabla y_{ik}(\gamma))$, first notice that Lemma A.1 of Hansen (1999) shows that for $k \neq s$,

$$
\lambda_n E(\nabla y_{is}(\gamma)\nabla y_{ik}(\gamma)) \to 0. \qquad (C.34)
$$

Then, by using (C.33) and (C.34) we have

$$
\begin{aligned}
\lambda_n E\left(\sum_{s=1}^{t-1} a_s \nabla \Delta y_{is}^*(\gamma)\right)^2 \to & \sum_{s=1}^{t-1} a_s^2 \big(E(y_{is}^2 | q_{is+1} = \gamma_0) f_{s+1}(\gamma_0)\nu \\
& + E(y_{is-1}^2 | q_{is} = \gamma_0) f_s(\gamma_0)\nu\big) \\
& - \sum_{s=1}^{t-2} 2 a_s a_{s+1} E(y_{is}^2 | q_{is+1} = \gamma_0) f_{s+1}(\gamma_0)\nu, \qquad (C.35)
\end{aligned}
$$

replacing this result in the the second term of (C.29), we will get

$$
\begin{aligned}
& c' \sum_{t=2}^{T} (a_t a_{t-1})^{-1} \lambda_n E\left(\sum_{s=1}^{t-1} a_s \nabla \Delta y_{is}^*(\gamma)\right)^2 c \\
& \to c' \sum_{t=2}^{T} (a_t a_{t-1})^{-1} \left(\sum_{s=1}^{t-1} a_s^2 \big(E(y_{is}^2 | q_{is+1} = \gamma_0) f_{s+1}(\gamma_0)\right. \\
& \left. + E(y_{is}^2 | q_{is-1} = \gamma_0) f_s(\gamma_0)\big) - \sum_{s=1}^{t-2} 2 a_s a_{s+1} \big(E(y_{is}^2 | q_{is+1} = \gamma_0) f_{s+1}(\gamma_0)\big)\right) \nu c. (C.36)
\end{aligned}
$$

Using a similar argument of (C.34), we have for the third term on the right-hand-side of (C.29)

$$
\begin{aligned}
&2c' \sum_{t=2}^{T} (a_t a_{t-1})^{-1} \sum_{s=1}^{t-1} a_s \lambda_n E(\nabla \Delta y_{is}^*(\gamma) \nabla 1(q_{i1} \leq \gamma)) c \\
&\to -2c' \sum_{t=2}^{T} (a_t a_{t-1})^{-1} a_1 E(y_{i0} | q_{i1} = \gamma_0) f_1(\gamma_0) \nu c.
\end{aligned} \tag{C.37}
$$

Finally, by plugging (C.31), (C.36) and (C.37) in (C.29), we have

$$
\begin{aligned}
&\lambda_n E\left(C' \nabla \Delta y_{i,-1}^*(\gamma)' \Omega^{\star -1} \nabla \Delta y_{i,-1}^*(\gamma) C\right) \\
&\to c' \left[\sum_{t=1}^{T} (a_t a_{t-1})^{-1} f_1(\gamma_0) + \sum_{t=2}^{T} (a_t a_{t-1})^{-1} \left(\sum_{s=1}^{t-1} a_s^2 \left(E(y_{is}^2 | q_{is+1} = \gamma_0) f_{s+1}(\gamma_0) \right.\right.\right. \\
&\left. +E(y_{is}^2 | q_{is-1} = \gamma_0) f_s(\gamma_0) \right) - \sum_{s=1}^{t-2} 2 a_s a_{s+1} E(y_{is}^2 | q_{is+1} = \gamma_0) f_{s+1}(\gamma_0) \\
&\left. -2 a_1 \nu E(y_{i0} | q_{i1} = \gamma_0) f_1(\gamma_0)) \right] c\nu \\
&= D(\gamma_0) \nu.
\end{aligned} \tag{C.38}
$$

A similar argument applies for the case $\nu \in [-\overline{\nu}, 0]$. Then (C.28) is proven and hence (C.27).

$\square$

Lemma 5. *As $n \to \infty$, uniformly over $\nu \in [-\overline{\nu}, \overline{\nu}]$,*

$$
n^{-\alpha} \sum_{i=1}^{n} C' \nabla \Delta y_{i,-1}^*(\gamma_0 + \nu/\lambda_n)' \Omega^{\star -1} \Delta u_i \Rightarrow \sqrt{V(\gamma_0)} W(\nu). \tag{C.39}
$$

Proof. As the previous lemma, we prove (C.39) for the case $\nu \in [0, \overline{\nu}]$. Let

$$
\begin{aligned}
&E\left(n^{-\alpha}\sum_{i=1}^{n}C'\nabla\Delta y_{i,-1}^{*}(\gamma)'\Omega^{\star-1}\Delta u_i\right)^2\\
&=\lambda_n E\left(C'\nabla\Delta y_{i,-1}^{*}(\gamma)'\Omega^{\star-1}\Delta u_i\right)^2\\
&=\lambda_n E\left(c'\sum_{t=1}^{T}(a_t a_{t-1})^{-1}\nabla 1(q_{i1}\le\gamma)\left(\upsilon_{i1}+\sum_{s=2}^{t}a_{s-1}\Delta u_{is}\right)\right.\\
&\left.+c'\sum_{t=2}^{T}(a_t a_{t-1})^{-1}\left(\sum_{s=1}^{t-1}a_s\nabla\Delta y_{is}^{*}(\gamma)\right)\left(\upsilon_{i1}+\sum_{s=2}^{t}a_{s-1}\Delta u_{is}\right)\right)^2\\
&=\lambda_n E\left(c'\sum_{t=1}^{T}(a_t a_{t-1})^{-1}\nabla 1(q_{i1}\le\gamma)\left(\upsilon_{i1}+\sum_{s=2}^{t}a_{s-1}\Delta u_{is}\right)\right)^2\\
&+\lambda_n E\left(c'\sum_{t=2}^{T}(a_t a_{t-1})^{-1}\left(\sum_{s=1}^{t-1}a_s\nabla\Delta y_{is}^{*}(\gamma)\right)\left(\upsilon_{i1}+\sum_{s=2}^{t}a_{s-1}\Delta u_{is}\right)\right)^2\\
&+2\lambda_n E\left(c'\sum_{t=1}^{T}(a_t a_{t-1})^{-1}\nabla 1(q_{i1}\le\gamma)\left(\upsilon_{i1}+\sum_{s=2}^{t}a_{s-1}\Delta u_{is}\right)\right)\times\\
&\left(c'\sum_{t=2}^{T}(a_t a_{t-1})^{-1}\left(\sum_{s=1}^{t-1}a_s\nabla\Delta y_{is}^{*}(\gamma)\right)\left(\upsilon_{i1}+\sum_{s=2}^{t}a_{s-1}\Delta u_{is}\right)\right).
\end{aligned}
\tag{C.40}
$$

The expression (C.40) has three terms, we start with the first term,

$$
\begin{aligned}
&\lambda_n E\left(c'\sum_{t=1}^{T}(a_t a_{t-1})^{-1}\nabla 1(q_{i1}\le\gamma)\left(\upsilon_{i1}+\sum_{s=2}^{t}a_{s-1}\Delta u_{is}\right)\right)^2\\
&=c'^2\left(\sum_{t=1}^{T}(a_t a_{t-1})^{-1}E\left(\upsilon_{i1}+\sum_{s=2}^{t}a_{s-1}\Delta u_{is}\right)\right)^2\lambda_n P(\gamma_0<q_{i1}\le\gamma)\\
&\to c'^2\left(\sum_{t=1}^{T}(a_t a_{t-1})^{-1}E\left(\upsilon_{i1}+\sum_{s=2}^{t}a_{s-1}\Delta u_{is}\right)\right)^2 f_1(\gamma_0)\nu.
\end{aligned}
\tag{C.41}
$$

Notice that by assumption 9, for $k>t$, $f_{k|t}(\gamma_0|\gamma_0)<\infty$ implies that $P(\gamma_0\le q_{ik}<\gamma|\gamma_0\le q_{it}<\gamma)=0$ (see Hansen, 1999). We use this fact to simplify the algebra in the proof of convergence of the second and third terms of (C.40).

The second term of (C.40) is similar to the expression (C.35) multiplied by the random variable $\left(\upsilon_{i1}+\sum_{s=2}^{t}a_{s-1}\Delta u_{is}\right)$, which is independent of q_{it} and it does not depend on γ. thus we have

$$
\begin{aligned}
&\lambda_n E\left(c'\sum_{t=2}^{T}(a_t a_{t-1})^{-1}\left(\sum_{s=1}^{t-1} a_s \nabla\Delta y_{is}^*(\gamma)\right)\left(\upsilon_{i1}+\sum_{s=2}^{t} a_{s-1}\Delta u_{is}\right)\right)^2 \\
&=\lambda_n E\left(c'^2\sum_{t=2}^{T}(a_t a_{t-1})^{-2}\left(\sum_{s=1}^{t-1} a_s \nabla\Delta y_{is}^*(\gamma)\right)^2\left(\upsilon_{i1}+\sum_{s=2}^{t} a_{s-1}\Delta u_{is}\right)^2\right) \\
&+\lambda_n E\left(c'^2\sum_{t=2}^{T-1}(a_t a_{t-1})^{-1}\left[\left(\sum_{s=1}^{t-1} a_s \nabla\Delta y_{is}^*(\gamma)\right)\left(\upsilon_{i1}+\sum_{s=2}^{t} a_{s-1}\Delta u_{is}\right)\right]\right. \\
&\left.\times\left[\sum_{k=t+1}^{T}(a_k a_{k-1})^{-1}\left(\sum_{s=1}^{k-1} a_s \nabla\Delta y_{is}^*(\gamma)\right)\left(\upsilon_{i1}+\sum_{s=2}^{k} a_{s-1}\Delta u_{is}\right)\right]\right). \qquad \text{(C.42)}
\end{aligned}
$$

The first component of equation (C.42) converges to

$$
\begin{aligned}
&\lambda_n E\left(c'^2\sum_{t=2}^{T}(a_t a_{t-1})^{-2}\left(\sum_{s=1}^{t-1} a_s \nabla\Delta y_{is}^*(\gamma)\right)^2\left(\upsilon_{i1}+\sum_{s=2}^{t} a_{s-1}\Delta u_{is}\right)^2\right) \\
&\to c'^2\sum_{t=2}^{T-1}(a_t a_{t-1})^{-2}\left\{\sum_{s=1}^{t-1} a_s^2 E\left(\left[y_{is}\left(\upsilon_{i1}+\sum_{s=2}^{t} a_{s-1}\Delta u_{is}\right)\right]^2 |q_{is+1}=\gamma_0\right) f_{s+1}(\gamma_0)\right. \\
&+\sum_{s=1}^{t-1} a_s^2 E\left(\left[y_{is-1}\left(\upsilon_{i1}+\sum_{s=2}^{t} a_{s-1}\Delta u_{is}\right)\right]^2 |q_{is}=\gamma_0\right) f_s(\gamma_0) \\
&\left.-\sum_{s=1}^{t-2} 2a_s a_{s+1} E\left(\left[y_{is}\left(\upsilon_{i1}+\sum_{s=2}^{t} a_{s-1}\Delta u_{is}\right)\right]^2 |q_{is+1}=\gamma_0\right) f_{s+1}(\gamma_0)\right\}\nu. \qquad \text{(C.43)}
\end{aligned}
$$

The second component of equation (C.42) converges to

$$
\begin{aligned}
&\lambda_n E\left(c'^2\sum_{t=2}^{T-1}(a_t a_{t-1})^{-1}\left[\left(\sum_{s=1}^{t-1}a_s\nabla\Delta y^*_{is}(\gamma)\right)\left(\upsilon_{i1}+\sum_{s=2}^{t}a_{s-1}\Delta u_{is}\right)\right]\right.\\
&\left.\times\left[\sum_{k=t+1}^{T}(a_k a_{k-1})^{-1}\left(\sum_{s=1}^{k-1}a_s\nabla\Delta y^*_{is}(\gamma)\right)\left(\upsilon_{i1}+\sum_{s=2}^{k}a_{s-1}\Delta u_{is}\right)\right]\right)\\
&\to c'^2\sum_{t=2}^{T-1}(a_t a_t)^{-1}\sum_{k=t+1}^{T}(a_k a_{k-1})^{-1}\\
&\left\{\sum_{s=1}^{t-1}a_s^2 E\left(y_{is}^2\left(\upsilon_{i1}+\sum_{s=2}^{t}a_{s-1}\Delta u_{is}\right)\left(\upsilon_{i1}+\sum_{s=2}^{t+1}a_{s-1}\Delta u_{is}\right)|q_{is+1}=\gamma_0\right)f_{s+1}(\gamma_0)+\right.\\
&\sum_{s=1}^{t-1}a_s^2 E\left(y_{is-1}^2\left(\upsilon_{i1}+\sum_{s=2}^{t}a_{s-1}\Delta u_{is}\right)\left(\upsilon_{i1}+\sum_{s=2}^{t+1}a_{s-1}\Delta u_{is}\right)|q_{is}=\gamma_0\right)f_{s}(\gamma_0)-\\
&\sum_{s=2}^{t-1}2a_s a_{s-1} E\left(y_{is-1}^2\left(\upsilon_{i1}+\sum_{s=2}^{t}a_{s-1}\Delta u_{is}\right)\left(\upsilon_{i1}+\sum_{s=2}^{t+1}a_{s-1}\Delta u_{is}\right)|q_{is}=\gamma_0\right)f_{s}(\gamma_0)-\\
&\left.a_t a_{t-1} E\left(y_{it-1}^2\left(\upsilon_{i1}+\sum_{s=2}^{t}a_{s-1}\Delta u_{is}\right)\left(\upsilon_{i1}+\sum_{s=2}^{t+1}a_{s-1}\Delta u_{is}\right)|q_{it}=\gamma_0\right)f_{t}(\gamma_0)\right\}\nu.
\end{aligned}
\tag{C.44}
$$

Thus, using (C.43) and (C.44), the second term (C.42) of (C.40) converges to

$$
\begin{aligned}
&\lambda_n E\left(c'\sum_{t=2}^{T}(a_t a_{t-1})^{-1}\left(\sum_{s=1}^{t-1} a_s \nabla\Delta y_{is}^*(\gamma)\right)\left(\upsilon_{i1}+\sum_{s=2}^{t} a_{s-1}\Delta u_{is}\right)\right)^2 \\
&\to c'^2\sum_{t=2}^{T-1}(a_t a_{t-1})^{-2}\left\{\sum_{s=1}^{t-1} a_s^2 E\left(\left[y_{is}\left(\upsilon_{i1}+\sum_{s=2}^{t} a_{s-1}\Delta u_{is}\right)\right]^2 |q_{is+1}=\gamma_0\right) f_{s+1}(\gamma_0)\right. \\
&+\sum_{s=1}^{t-1} a_s^2 E\left(\left[y_{is-1}\left(\upsilon_{i1}+\sum_{s=2}^{t} a_{s-1}\Delta u_{is}\right)\right]^2 |q_{is}=\gamma_0\right) f_s(\gamma_0) \\
&\left.-\sum_{s=1}^{t-2} 2a_s a_{s+1} E\left(\left[y_{is}\left(\upsilon_{i1}+\sum_{s=2}^{t} a_{s-1}\Delta u_{is}\right)\right]^2 |q_{is+1}=\gamma_0\right) f_{s+1}(\gamma_0)\right\}\nu \\
&+c'^2\sum_{t=2}^{T-1}(a_t a_{t-1})^{-1}\sum_{k=t+1}^{T}(a_k a_{k-1})^{-1} \\
&\left\{\sum_{s=1}^{t-1} a_s^2 E\left(y_{is}^2\left(\upsilon_{i1}+\sum_{s=2}^{t} a_{s-1}\Delta u_{is}\right)\left(\upsilon_{i1}+\sum_{s=2}^{t+1} a_{s-1}\Delta u_{is}\right)|q_{is+1}=\gamma_0\right) f_{s+1}(\gamma_0)+\right. \\
&\sum_{s=1}^{t-1} a_s^2 E\left(y_{is-1}^2\left(\upsilon_{i1}+\sum_{s=2}^{t} a_{s-1}\Delta u_{is}\right)\left(\upsilon_{i1}+\sum_{s=2}^{t+1} a_{s-1}\Delta u_{is}\right)|q_{is}=\gamma_0\right) f_s(\gamma_0)- \\
&\sum_{s=2}^{t-1} 2a_s a_{s-1} E\left(y_{is-1}^2\left(\upsilon_{i1}+\sum_{s=2}^{t} a_{s-1}\Delta u_{is}\right)\left(\upsilon_{i1}+\sum_{s=2}^{t+1} a_{s-1}\Delta u_{is}\right)|q_{is}=\gamma_0\right) f_s(\gamma_0)- \\
&\left. a_t a_{t-1} E\left(y_{it-1}^2\left(\upsilon_{i1}+\sum_{s=2}^{t} a_{s-1}\Delta u_{is}\right)\left(\upsilon_{i1}+\sum_{s=2}^{t+1} a_{s-1}\Delta u_{is}\right)|q_{it}=\gamma_0\right) f_t(\gamma_0)\right\}\nu.
\end{aligned}
\tag{C.45}
$$

The third term of (C.40) converges to

$$
\begin{aligned}
&+2\lambda_n E\left(c'\sum_{t=1}^{T}(a_t a_{t-1})^{-1}\nabla 1(q_{i1}\le\gamma)\left(\upsilon_{i1}+\sum_{s=2}^{t} a_{s-1}\Delta u_{is}\right)\right)\times \\
&\left(c'\sum_{t=2}^{T}(a_t a_{t-1})^{-1}\left(\sum_{s=1}^{t-1} a_s\nabla\Delta y_{is}^*(\gamma)\right)\left(\upsilon_{i1}+\sum_{s=2}^{t} a_{s-1}\Delta u_{is}\right)\right) \\
&\to -c'^2 2\sum_{t=1}^{T}(a_t a_{t-1})^{-1}\sum_{k=2}^{T}(a_k a_{k-1})^{-1} a_1 \\
&E\left(y_{i0}\left(\upsilon_{i1}+\sum_{s=2}^{t} a_{s-1}\Delta u_{is}\right)\left(\upsilon_{i1}+\sum_{s=2}^{k} a_{s-1}\Delta u_{is}\right)\right)|q_{i1}=\gamma_0)f_1(\gamma_0)\nu.
\end{aligned}
\tag{C.46}
$$

Finally, by plugging (C.41), (C.45) and (C.46) into (C.40), we get

$$E\left(n^{-\alpha}\sum_{i=1}^{n}C'\nabla\Delta y^{*}_{i,-1}(\gamma)'\Omega^{\star-1}\Delta u_i\right)^2 \quad\rightarrow\quad V(\gamma_0)\nu. \tag{C.47}$$

By assumption 1.9, this establishes that the finite dimensional distributions of the stochastic process are those of the stated double-sided Brownian motion. By arguments identical to those in the proof of Lemma A.11 of Hansen (2000) and Lemma A.1 of Hansen (1999), and using (C.47), we will have

$$E\left(n^{-\alpha}\sum_{i=1}^{n}C'\nabla\Delta y^{*}_{i,-1}(\gamma)'\Omega^{\star-1}\Delta u_i\right) \quad\rightarrow\quad \sqrt{V(\gamma_0)}W(\nu). \tag{C.48}$$

□

Proof of Theorem 2. By lemma A.9 of Hansen (2000), lemma 4 and lemma 5, where the limit functional $S_n(\gamma_0+\nu/\lambda_n)-S_n(\gamma_0)$ is continuous with a unique minimum almost surely, and following the argument in the proofs of Theorem 1 of Hansen (2000) give the stated result. □

Proof of Theorem 3. The slope estimators are asymptotically equivalent to their ideal counterparts constructed with the unknown true value of γ rather than the estimated value $\widehat{\gamma}$, then we examine the case of known γ and we will prove the theorem for β_1, the argument is similar for β_2.

Define the following vectors $\Delta\widetilde{y}^{*}_{i,-1}(\gamma)=\Lambda^{-1/2}H(0,\Delta y^{*}_{i,1}(\gamma),\cdots,y^{*}_{i,T-1}(\gamma))'$, $\Delta\widetilde{y}_i(\gamma)=\Lambda^{-1/2}H(\Delta y_{i,1}(\gamma),\cdots,y_{i,T}(\gamma))'$ and by using a suitable partition of the form (9), we have

$$\widehat{\beta}_1=\left(\sum_{i=1}^{n}\widetilde{y}^{*}_{i,-1}(\gamma)'\Delta\widetilde{y}^{*}_{i,-1}(\gamma)\right)^{-1}\left(\sum_{i=1}^{n}\Delta\widetilde{y}^{*}_{i,-1}(\gamma)'\Delta\widetilde{y}_i\right), \tag{C.49}$$

and therefore

$$\widehat{\beta}_1=\beta_1+\left(\sum_{i=1}^{n}\widetilde{y}^{*}_{i,-1}(\gamma)'\Delta\widetilde{y}^{*}_{i,-1}(\gamma)\right)^{-1}\left(\sum_{i=1}^{n}\Delta\widetilde{y}^{*}_{i,-1}(\gamma)'(\Lambda^{-1/2}H\Delta u_i)\right). \tag{C.50}$$

The consistency of β_1 is established by a law of large numbers if it is shown that the last term of (C.50) has a zero mean. This term has the form

$$\sum_{i=1}^{n} \Delta\widetilde{y}^{*}_{i,-1}(\gamma)'(\Lambda^{-1/2}H\Delta u_i) =$$

$$\sum_{i=1}^{n}\sum_{t=2}^{T}(a_t a_{t-1})^{-1}\left(\sum_{s=1}^{t-1} a_s \Delta y^{*}_{is}(\gamma)\right)\left(\upsilon_{i1} + \sum_{s=2}^{t} a_{s-1}\Delta u_{is}\right), \quad \text{(C.51)}$$

and using lemma 1

$$E\left[\left(\sum_{s=1}^{t-1} a_s \Delta y^{*}_{is}(\gamma)\right)\left(\upsilon_{i1} + \sum_{s=2}^{t} a_{s-1}\Delta u_{is}\right)\right] = 0. \quad \text{(C.52)}$$

□

Proof of Theorem 4. Recall that

$$F(\widehat{\gamma}) = \left(\sum_{i=1}^{n} \Delta y^{\circ}_{i,-1}(\widehat{\gamma})'\Omega^{-1}\Delta y^{\circ}_{i,-1}(\widehat{\gamma})\right), \quad \text{(C.53)}$$

and

$$F(\gamma_0) = E\left(\Delta y^{\circ}_{i,-1}(\gamma_0)'\Omega^{-1}\Delta y^{\circ}_{i,-1}(\gamma_0)\right). \quad \text{(C.54)}$$

Let

$$F_n(\widehat{\gamma}) = \left(\frac{1}{n}\sum_{i=1}^{n} \Delta y^{\circ}_{i,-1}(\widehat{\gamma})'\Omega^{\star-1}\Delta y^{\circ}_{i,-1}(\widehat{\gamma})\right). \quad \text{(C.55)}$$

By Lemma 3 and the consistency of $\widehat{\gamma}$, we have

$$F_n^{-1}(\widehat{\gamma}) \to_p F^{-1}(\gamma_0). \quad \text{(C.56)}$$

Conditional on γ, the structural model (1) and the differentiated model (2) are linear in β_1 and β_2. The likelihood function (7) or (8) depends on a fixed number of parameters, and by the normality assumption, it satisfies the standard regularity conditions conditional on γ. Therefore, the ML estimators of β_1 and β_2 are asymptotically normally distributed with a covariance matrix given by $F^{-1}(\gamma_0)$ as the number of individuals n grows to infinity whether the time period T is fixed or tends to infinity.

□