

# Threshold Regression in Heterogeneous Panel Data with Interactive Fixed Effects\*

Marco Barassi, Yiannis Karavias\*, Chongxian Zhu

*Department of Economics, University of Birmingham*

---

## Abstract

This paper introduces unit-specific heterogeneity in panel data threshold regression. Both slope coefficients and threshold parameters are allowed to vary by unit. The heterogeneous threshold parameters manifest via a unit-specific empirical quantile transformation of a common underlying threshold parameter which is estimated efficiently from the whole panel. In the errors, the unobserved heterogeneity of the panel takes the general form of interactive fixed effects. The newly introduced parameter heterogeneity has implications for model identification, estimation, interpretation, and asymptotic inference. The assumption of a shrinking threshold magnitude now implies shrinking heterogeneity and leads to faster estimator rates of convergence than previously encountered. The asymptotic theory for the proposed estimators is derived and Monte Carlo simulations demonstrate its usefulness in small samples. The new model is employed to examine the Feldstein-Horioka puzzle and it is found that the trade liberalization policies of the 80's significantly impacted cross-country capital mobility.

*Keywords:* Panel Data; Threshold Regression; Heterogeneity; Interactive Fixed Effects; Regime Switching; Feldstein-Horioka Puzzle.

*JEL classification:* C23; C24; F32; F41.

---

## 1. Introduction

Threshold regression was introduced in the seminal paper of Tong (1978) and has since become one of the few key non-linear models widely employed by econometricians. Tsay (1989, 1998) characterizes threshold models as “practical non-linear models” which stand out in the vast expanse of the non-linear world. Indeed, threshold regression can capture complex properties of the data like limit cycles, which in economics correspond to the existence of multiple equilibria often predicted by economic theory, and yet still maintain intuitive interpretation and tractability. Threshold

---

\*Acknowledgements: The authors would like to thank Otilia Boldea, Chihwa Kao, Oliver Linton, Esfandiar Maa-soumi, Hashem Pesaran, Liangjun Su, Lorenzo Trapani, and seminar participants at the University of Birmingham, at the 2023 Italian Congress of Econometrics and Empirical Economics, and at the 28th International Panel Data Conference for their valuable comments and suggestions.

\*Corresponding Author. Email: i.karavias@bham.ac.uk, University of Birmingham, Birmingham, B15 2TT, UK.

regression was introduced in panel data in Hansen (1999) and has since been an active field of theoretical and empirical research. This is because the pooled information across units helps with the identification of the different regimes, as well as the precise estimation of the threshold parameter.

In this paper, we provide a comprehensive asymptotic theory for estimation and testing in panel threshold regression model with two key features: i) heterogeneous parameters across units and ii) interactive fixed effects. We show how to estimate the heterogeneous slopes and thresholds and prove estimator consistency. We derive novel estimator rates of convergence and offer asymptotically valid confidence intervals. Finally, we propose a test for the null hypothesis of linearity against the alternative of a threshold regression.

Parameter heterogeneity arises naturally in panel data which consists of many cross-sectional units. The greater the number of units, the more likely it is that the model parameters will vary across them. This is a fact that is well recognized in the panel data literature, see e.g. the contributions in Swamy (1970), Pesaran and Smith (1995), Pesaran et al. (1999), Pesaran (2006), Fernández-Val and Lee (2013), Bonhomme and Manresa (2015), Gao et al. (2020), Norkutė et al. (2021), Trapani (2021), and Chen et al. (2022), *inter alia*. In the heterogeneous panel model, the parameters of interest are the individual unit slope coefficients and also their average. Therefore, only Mean Group (MG hereafter) estimators can be used here (Pesaran and Smith, 1995). If the model is static and linear, pooled estimators assuming homogeneity across units can consistently estimate the average slope across units (see e.g. Pesaran (2006)) but this property does not extend to non-linear models like the panel threshold regression.

The second key feature of the threshold model considered here is interactive fixed effects in the error term. Interactive fixed effects represent a generalization of the standard two-way fixed effects model. As their name suggests, interactive fixed effects are the inner product of a vector individual effect with a vector time effect, which is a more flexible and empirically relevant way of capturing unobserved heterogeneity. Interactive fixed effects have been widely used in microeconomics, macroeconomics and finance. In microeconomics, they are able to capture the time-varying effect of unobserved unit-invariant characteristics like ability and soft skills, see e.g. Kejriwal et al. (2020). In macroeconomics, they can be interpreted as a set of unobserved multiple common factors useful in capturing cross-section dependence (Pesaran, 2006) and in finance as unobserved factors for asset returns (Giglio and Xiu, 2021).

Both parameter heterogeneity and the existence of interactive fixed effects have important implications for model estimation. Parameter heterogeneity reduces the estimator rate of convergence and interactive fixed effects cannot be removed by the standard fixed effects transformations. There are several types of estimators that can deal with interactive fixed effects, such as the quasi-difference approach (Holtz-Eakin et al., 1988), the generalized method of moments estimator (Ahn et al., 2001, 2013; Juodis and Sarafidis, 2022), the Common Correlated Effects (CCE) estimator (Pesaran, 2006), the Principal Components (PC) estimator (Bai, 2009), the two-stage instrumental variables estimator (Norkutė et al., 2021), the Post-Nuclear Norm Regularized estimator of Moon and Weidner (2018), and the Lasso type shrinkage methods on fixed effects (Lu and Su, 2016; Su

et al., 2016). Yet, not all of the above estimators can deal with heterogeneous coefficients. This paper employs the CCE method for estimation because it is general enough to allow for heterogeneous coefficients, is analytically tractable, has excellent small sample properties, and is computationally fast which is important in threshold regression estimation.

The model estimation by CCE is not straightforward and presents several technical difficulties. First, the identification of individual slope coefficients requires that the threshold variable has a common support across units, something which is unlikely in practice. This is a serious problem that can only be avoided if the threshold parameters are also allowed to be heterogeneous across units. Threshold parameter heterogeneity, however, removes the appeal of panel data, as it the problem becomes equivalent to running separate time series regressions. To break this impasse, we assume that there is a homogeneous threshold parameter across units and propose transforming the threshold variable by the monotonic percentile function. The transformation results in all panel units having a common support for the transformed threshold variable and a transformed common threshold parameter. The common threshold parameter (now a threshold in terms of percentiles) can be estimated precisely using information from both dimensions of the panel as it is desirable. At the same time, the threshold percentile, which splits the model coefficient regimes, translates to heterogeneous thresholds in the original threshold variable. Therefore, we estimate a common parameter across units, whose projected interpretation in the original threshold variable is that of a heterogeneous threshold parameter. This modelling choice gives us the best of both worlds; efficient panel data estimation of the common threshold percentile parameter, no identification issues for the heterogeneous coefficient slopes, and heterogeneous interpretation of each unit's threshold. The percentile function transformation and the impact of parameter heterogeneity on marginal effects are presented in Section 2 below.

Another key obstacle is that the asymptotic distribution of the threshold parameter depends on nuisance parameters related to the distribution of the error term, see e.g. Chan (1993). To deal with this problem, Hansen (2000) derives a nuisance parameter-free asymptotic distribution, under the assumption of a “shrinking threshold”. The shrinking threshold assumption is maintained here, as in almost all other papers in the literature, however, what is new here is that it interacts with the slope coefficient heterogeneity. The shrinking threshold assumption in this model contemporaneously implies shrinking heterogeneity. As the threshold magnitude shrinks, the relevant slope coefficients become less heterogeneous and the estimator rate of convergence becomes faster, reaching in the limit that of a pooled estimator in a homogeneous panel. To the best of our knowledge, this is the first time this effect is observed.

In terms of threshold regression in heterogeneous panels, this paper is closer to that of Chudik et al. (2017), which considers almost the same framework, however, it is mostly focused on the specific on an empirical application and does not provide any asymptotic theory supporting the estimation methodology. We fill this gap with the current contribution. Another paper that is very close is that of Miao et al. (2020b) which assumes that units belong to a small number of groups and parameters vary across groups. This paper allows only for fixed effects, and not for the full

interactive fixed effects model considered in the current contribution. Furthermore, the empirical application is focused on estimating the number of underlying groups and group membership, which is different from the form of parameter heterogeneity considered here. The two papers are therefore clearly distinct. Miao et al. (2020a) consider panel threshold regression with interactive fixed effects but restrict the parameters to be homogeneous across units. Hacıoğlu Hoke and Kapetanios (2021) consider a smooth transition model with heterogeneous coefficients and interactive fixed effects. Massacci (2017) and Massacci et al. (2021) consider threshold regression in the heterogeneous loadings of pure factor models, without additional regressions. Other contributions in the area are those of Yu et al. (2023) which considers threshold regressions with endogenous regressors, Chen et al. (2012) which considers panels with two threshold variables, and Seo and Shin (2016) which considers threshold regression in dynamic panel data models with a short time dimension. All these contributions assume homogeneous coefficients and standard fixed effects.

We conclude this paper by applying the new methodology to examine one of the most important problems of macroeconomics, that of the Feldstein-Horioka puzzle (Feldstein and Horioka, 1980). The puzzling fact is that domestic savings in a country are highly correlated with domestic investment, meaning that savers disregard potential opportunities for higher returns in other countries. The new threshold model studied in this paper allows for heterogeneous coefficients and cross-section dependence which have both been documented previously in country-level data (Chudik et al., 2017). Our results confirm the existence of the puzzle but show that higher trade openness results in greater international capital mobility. We find that most countries moved into this “high capital mobility” regime in the 80’s, which is in line with the trade liberalization policies introduced at that time (Faini, 2004).

The remainder of the paper is organised as follows. Section 2 describes the model we study. Section 3 develops the estimation strategy. Section 4 provides the asymptotic theory. Section 5 introduces a test for the presence of the threshold effects. Section 6 applies the methodology to study the relationship between inflation and economic growth. Section 7 concludes. The supplementary appendix contains the bootstrap algorithm for the test of no nonlinearity, extensive Monte Carlo simulations, estimators for group-specific parameter heterogeneity with known group membership, and all relevant mathematical proofs.

We will use the following notation. The letter  $C$  stands for an universal finite positive constant, while  $I_m$  denotes  $m \times m$  identity matrix. Also, for a real  $m \times n$  matrix  $A$ , the element  $(i, j)$  is denoted by  $A_{ij}$ , while  $\|A\|$  denotes the Frobenius norm.  $1_T = (1, 1, \dots, 1)'$ , a  $T \times 1$  unity vector.  $\lambda_{\min}(A)$  and  $\lambda_{\max}(A)$  denote the smallest and largest eigenvalues of  $A$  respectively. We define the projection matrices  $P_A = A(A'A)^{-1}A'$  and  $M_A = I_m - P_A$ .  $\mathbb{I}(\cdot)$  is the indicator function, and  $\bar{Y}_t = N^{-1} \sum_{i=1}^N Y_{i,t}$  indicates the cross-sectional average of any variable  $Y_{i,t}$ .  $\text{diag}(A)$  denotes a diagonal matrix consisting of main diagonal elements of the matrix  $A$ . The symbol  $\xrightarrow{P}$  denotes convergence in probability,  $\xrightarrow{d}$  convergence in distribution,  $\Rightarrow$  denotes weak convergence with respect to the uniform metric, and  $\text{plim}$  probability limit.  $(N, T) \rightarrow \infty$  denotes that both  $N$  and  $T$  tend to infinity jointly, where  $N$  is the number of units in the panel and  $T$  is the number of time

series observations.

## 2. Model

Consider the following model, in which the response variable of the  $i^{\text{th}}$  unit, observed at time  $t$ ,  $y_{i,t}$  is given by the model:

$$y_{i,t} = \beta_i' x_{i,t} + \delta_i' w_{i,t} \mathbb{I}\{q_{i,t} \leq \gamma\} + e_{i,t} \quad i = 1, \dots, N, t = 1, \dots, T \quad (1)$$

where  $x_{i,t}$  is  $K \times 1$  vector of observable regressors, and  $\beta_i$  is a  $K \times 1$  vector of heterogeneous slope coefficients, which can be different for each unit  $i$ . Additionally, let:

$$w_{i,t} = R' x_{i,t}, \quad (2)$$

be an  $r \times 1$  subset of the regressors in  $x_{i,t}$ . The matrix  $R$  is an  $K \times r$  selection matrix of zeros and ones with full column rank  $r$  that picks elements of  $x_{i,t}$  whose coefficients are subject to the threshold effect.  $R$  is known to the researcher. If  $R = I_K$ , then all  $K$  regressors in  $x_{i,t}$  have the threshold effect and the model is called a pure threshold model, while if  $R = (I_r, 0_{r \times (K-r)})'$ , then only the first  $r$  regressors in  $x_{i,t}$  are affected by the threshold, and the model is called a partial threshold model.

A key characteristic of (1) is that the effect of the regressors on the dependent variable is allowed to vary across two regimes which are identified by  $\mathbb{I}$ , an indicator function which takes value 1 when  $\{q_{i,t} \leq \gamma\}$  and 0 otherwise. The variable  $q_{i,t}$  is a scalar that may belong in  $x_{i,t}$ , and  $\gamma$  is the threshold parameter that defines the two different regimes in the model. When  $q_{i,t} > \gamma$  the effect of  $x_{i,t}$  on  $y_{i,t}$  is  $\beta_i$ . This is frequently called the “high regime”. There is also a “low regime” with  $q_{i,t} \leq \gamma$ , where the coefficient of the variables is  $\beta_i + \delta_i$ . The model is suitable to identify the different equilibria which can arise when the response of the dependent variable is different across periods. Examples of the “low regime” include periods of economic and financial turmoil and distress, such as poor stock market performance or economic crises, or periods of unfavorable economic outlook and low sentiment when compared to normal times, namely the “high regime”.<sup>1</sup>

The model in (1) is sufficiently general to render existing panel models as special cases; (i) the simple linear heterogeneous panel model of Swamy (1970) corresponds to the case where  $\delta_i = 0$  for all  $i$ s and (ii) the homogeneous threshold panel models of Hansen (1999) and Miao et al. (2020a) correspond to the case where  $\beta_i = \beta$ , and  $\delta_i = \delta$ .

The threshold parameter  $\gamma$  is assumed to be common across units, in line with all the pre-existing literature on panel data threshold regression. The  $\gamma$  parameter can be seen as the average of heterogeneous  $\gamma_i$  which vary across units, in the same way that a common break can be seen as the

---

<sup>1</sup>Similarly, it is possible to define  $\mathbb{I}\{q_{i,t} > \gamma\}$  in (1) to capture regimes where inflation or interest rates are above a certain threshold in monetary policy models (high inflation regimes). These cases are formally equivalent and the only difference between them is the interpretations of the  $\delta_i$  coefficients.

average of heterogeneous breaks across units see Bai (2010) and Karavias et al. (2022). Assuming homogeneity in  $\gamma$  allows estimation by pooling the information across  $N$ . Observations on many units allows observing many limit cycles and this pooling results in more precise estimation of the threshold parameter. Contrary to the previous literature, in model (1) the threshold homogeneity assumption creates complications given that the  $\beta_i$  and  $\delta_i$  are now allowed to vary across units.

When  $\gamma$  is homogeneous and the slopes are heterogeneous across units, the model in (1) is hard to estimate in most empirical applications due its stringent identification requirements. The identification of  $\delta_i$  necessitates that the threshold  $\gamma$  is such that there are threshold variable observations in both regimes defined by  $\gamma$ ; explicitly  $\sum_{t=1}^T \mathbb{I}\{q_{i,t} \leq \gamma\} > 0$  and  $\sum_{t=1}^T \mathbb{I}\{q_{i,t} > \gamma\} > 0$ , for every unit  $i$ . If however the supports of  $q_{i,t}$  and  $q_{j,t}$ , for  $i \neq j$ , are disjoint, then the aforementioned condition will not hold and either  $\delta_i$  or  $\delta_j$  will not be identified. To see this, consider the impact of government expenditure on economic growth. The UK's General government final consumption expenditure (% of GDP) varies between 16% to 22% from 1973 to 2021, while that of Mexico in the same period varies between 8% to 12%. Therefore, there is no common  $\gamma$  that creates two regimes in both countries.

To bypass this challenge, we introduce a variation in the functional form of the threshold variable:

$$y_{i,t} = \beta'_i x_{i,t} + \delta'_i w_{i,t} \mathbb{I}\{p_i(q_{i,t}) \leq \gamma\} + e_{i,t}. \quad (3)$$

The function  $p_i(\cdot)$  transforms the threshold variable into an empirical quantile. The benefit of the transformation is that it ensures that  $p_i(q_{i,t})$  has common support across all units and that all the  $\delta_i$ s are identified.

The introduction of the empirical quantile transformation has a significant implication for the interpretation of the model; the threshold parameter  $\gamma$  is now interpreted as a threshold percentile value, and not in the original threshold variable units. To translate the threshold parameter in the original units, the inverse transformation of  $p_i(\cdot)$  must be employed. Because  $p_i(\cdot)$  is unit-specific, the inverse transformation is also unit specific and therefore, the retrieved threshold in the original units will also be heterogeneous and unit-specific:  $\gamma_i = p_i(\gamma)^{-1}$ .

Threshold heterogeneity is a desirable generalization when it does not come at the expense of estimation accuracy. Efficient estimation is achieved by maintaining a common threshold parameter  $\gamma$ . The key additional assumption in model (3) is the specific form of the empirical quantile function. Let for example  $p_i(q_{i,t}) = \text{Rank}_i(q_{i,t})/T$ , where  $\text{Rank}_i(q_{i,t})$  is the rank of  $q_{i,t}$  across the time series dimension  $\{q_{i,t}, t = 1, \dots, T\}$ . This formula transforms  $q_{i,t}$  into percentiles. Notice that the transformed variable  $p_i(q_{i,t})$  is in percentiles and thus has  $T$  unique observations which coincide across all units.

Using the ranks of  $q_{i,t}$  does not change the interpretation of the slope coefficients in the baseline model (1). Let  $q_{i,t}^* = p_i(q_{i,t})$ . The marginal effect of  $x_{i,t}$  is still:

$$\frac{\partial E(y_{i,t}|x_{i,t}, q_{i,t}^*)}{\partial x_{i,t}} = \begin{cases} \beta_i, & \text{if } q_{i,t}^* > \gamma, \\ \beta_i + \delta_i & \text{if } q_{i,t}^* \leq \gamma. \end{cases} \quad (4)$$

The main focus of this paper is to estimate the marginal effects in (4). To this end, we provide consistent estimators for  $\beta_i$ ,  $\delta_i$  and also  $\gamma$ . Additionally, the average marginal effects are of interest:

$$\frac{1}{N} \sum_{i=1}^N \frac{\partial E(y_{i,t}|x_{i,t}, q_{i,t}^*)}{\partial x_{i,t}} = \begin{cases} \bar{\beta}, & \text{if } q_{i,t}^* > \gamma, \\ \bar{\beta} + \bar{\delta} & \text{if } q_{i,t}^* \leq \gamma, \end{cases} \quad (5)$$

where  $\bar{\beta} = N^{-1} \sum_{i=1}^N \beta_i$ , and  $\bar{\delta} = N^{-1} \sum_{i=1}^N \delta_i$ . Following Pesaran and Smith (1995) the average marginal effects are  $\bar{\beta}$  and  $\bar{\delta}$  or the  $E(\beta_i)$  and  $E(\delta_i)$  as  $N \rightarrow \infty$ .

The errors  $e_{it}$  contain the unobserved heterogeneity, which has the general form of interactive fixed effects or a multi-factor error structure as it is sometimes called in the literature (Pesaran, 2006):

$$e_{i,t} = \lambda_i' f_t + \varepsilon_{i,t} \quad (6)$$

where  $f_t$  is a  $m \times 1$  vector containing  $m$  unobserved factors which are common to all units,  $\lambda_i$  are the heterogeneous common factor loadings, and  $\varepsilon_{i,t}$  are the remaining idiosyncratic errors. The unobserved heterogeneity formulation in (6), encompasses all standard panel data models. If  $f_t = 1$  then we have the standard fixed effect (FE) model, while for  $f_t = (1, \tau_t)'$  and  $\lambda_i = (\alpha_i, 1)'$  we have the two-way fixed effect (TWFE). The unobserved common factors  $f_t$  can capture the price of an unobserved skill which changes in time or the price of an input for production, or even aggregate shocks to a particular market or economy. Because all units are impacted by these factors, the factor structure is the source of the cross-sectional dependence across the units.

The interactive fixed effects are allowed to be correlated with the regressors. We follow Pesaran (2006) and assume that the correlation is linear and can be modelled as:

$$x_{i,t} = \Pi_i' f_t + \xi_{i,t} \quad (7)$$

where  $\Pi_i$  is  $m \times K$  fixed factor loading matrix and  $\xi_{i,t}$  is a  $K \times 1$  is the idiosyncratic part. This is a Mundlak-style assumption and is well accepted in the literature.

We further assume that the threshold variable  $q_{i,t}$  is one of the regressors in  $x_{i,t}$  and therefore satisfies (7). This is the most empirically relevant scenario. Alternatively, it is also possible that  $q_{i,t}$  is not included in  $x_{i,t}$ , in which case, the theory below still holds but collapses to that of a structural break model that can be estimated by arranged regression (Tsay, 1998) as in Karavias et al. (2022). The assumption that  $q_{i,t}$  satisfies (7) does not interfere with the empirical quantile transformation of  $q^*$  because the latter appears only in the indicator function  $\mathbb{I}\{q_{i,t}^* \leq \gamma\}$ , and there are no assumptions in the threshold literature on the distribution of  $\mathbb{I}\{q_{i,t}^* \leq \gamma\}$ .

### 3. Estimation

To present the estimators we stack the model in (3) across the time dimension. Letting  $w_{i,t}(\gamma) = w_{i,t}\mathbb{1}\{q_{i,t}^* \leq \gamma\}$ , the stacked model becomes:

$$y_i = X_i\beta_i + W_i(\gamma)\delta_i + e_i, \quad (8)$$

where,  $y_i = (y_{i,1}, y_{i,2}, \dots, y_{i,T})'$  is a  $T \times 1$  vector,  $X_i = (x_{i,1}, x_{i,2}, \dots, x_{i,T})'$  is a  $T \times K$  matrix,  $W_i(\gamma) = (w_{i,1}(\gamma), w_{i,2}(\gamma), \dots, w_{i,T}(\gamma))'$  is a  $T \times r$  matrix and,  $e_i = (e_{i,1}, e_{i,2}, \dots, e_{i,T})'$  is a  $T \times 1$  vector. The interactive effects from (6) in matrix form become:

$$e_i = F\lambda_i + \varepsilon_i, \quad (9)$$

where  $F = (f_1, f_2, \dots, f_T)'$  is a  $T \times m$  matrix, and  $\varepsilon_i = (\varepsilon_{i,1}, \varepsilon_{i,2}, \dots, \varepsilon_{i,T})'$  is a  $T \times 1$  vector. Finally, we express (7) in matrix form as:

$$X_i = F\Pi_i + \xi_i, \quad (10)$$

where,  $\xi_i = (\xi_{i,1}, \xi_{i,2}, \dots, \xi_{i,T})'$  is a  $T \times k$  matrix.

We are interested in three sets of parameters: the threshold parameter  $\gamma$ , the average effect of the slope coefficients  $\beta$  and  $\delta$ , and the heterogeneous slope coefficients  $\beta_i$  and  $\delta_i$ , which may be different across units  $i \in \{1, 2, \dots, N\}$ . The model in (8), is linear in the parameters  $\beta_i$  and  $\delta_i$  and non-linear in the parameter  $\gamma$ . For now, let  $\gamma$  be known. In this case the model is linear in the parameters and can be estimated by a variant of the Mean Group CCE estimator in Pesaran (2006), adapted as in Karavias et al. (2022). The key idea is that cross-section averages of  $X_i$  can be used to consistently estimate the space spanned by the unknown factors in (10). This is an alternative to using principal components, see e.g. Westerlund and Urbain (2015). The first step to estimation then involves pre-multiplying (8) by  $M_{\bar{X}} = I_T - \bar{X}(\bar{X}'\bar{X})^{-1}\bar{X}'$ , where  $I_T$  is a  $T$ -order identity matrix. The transformed model becomes:

$$\tilde{y}_i = \tilde{X}_i\beta_i + \tilde{W}_i(\gamma)\delta_i + \tilde{e}_i, \quad (11)$$

where  $\tilde{y}_i = M_{\bar{X}}y_i$ ,  $\tilde{X}_i = M_{\bar{X}}X_i$ ,  $\tilde{W}_i(\gamma) = M_{\bar{X}}W_i(\gamma)$  and  $\tilde{e}_i = M_{\bar{X}}e_i$ . We will show later that  $M_{\bar{X}}e_i = M_{\bar{X}}F\lambda_i + M_{\bar{X}}\varepsilon_i = M_{\bar{X}}\varepsilon_i + o_p(1)$ , asymptotically removing the  $m$  common factors.<sup>2</sup> The model in (11) can be rewritten in a more compact form as:

$$\tilde{y}_i = \tilde{Z}_i(\gamma)\theta_i + \tilde{e}_i, \quad (12)$$

where  $\tilde{Z}_i(\gamma) = \begin{pmatrix} \tilde{X}_i & \tilde{W}_i(\gamma) \end{pmatrix}$  and  $\theta_i = (\beta_i', \delta_i')'$ .

---

<sup>2</sup>Here we only use  $\bar{X}$  to remove the interactive fixed effects, which is different to using both  $\bar{Y}$  and  $\bar{X}$  as in the original CCE estimator of Pesaran (2006). Karavias et al. (2022) show that  $\bar{Y}$  is not rotationally consistent for  $f_t$  in the model with structural breaks and the same applies here.



Assuming  $\gamma$  is known, the CCE estimators are:

$$\hat{\theta}_i(\gamma) = \left( \tilde{Z}_i(\gamma)' \tilde{Z}_i(\gamma) \right)^{-1} \tilde{Z}_i(\gamma)' \tilde{y}_i \quad \hat{\theta}(\gamma) = \frac{1}{N} \sum_{i=1}^N \hat{\theta}_i(\gamma) \quad (13)$$

and more explicitly:

$$\hat{\beta}_i(\gamma) = \left( \tilde{X}_i' M_{\tilde{W}_i(\gamma)} \tilde{X}_i \right)^{-1} \tilde{X}_i' M_{\tilde{W}_i(\gamma)} \tilde{y}_i, \quad \hat{\beta}(\gamma) = \frac{1}{N} \sum_{i=1}^N \hat{\beta}_i(\gamma), \quad (14)$$

$$\hat{\delta}_i(\gamma) = \left( \tilde{W}_i(\gamma)' M_{\tilde{X}_i} \tilde{W}_i(\gamma) \right)^{-1} \tilde{W}_i(\gamma)' M_{\tilde{X}_i} \tilde{y}_i, \quad \hat{\delta}(\gamma) = \frac{1}{N} \sum_{i=1}^N \hat{\delta}_i(\gamma). \quad (15)$$

If  $\gamma$  is unknown, as it is usually the case in practice, we follow Hansen (1999) and estimate it with the  $\gamma$  which minimizes the CCE sum of squared residuals:

$$\hat{\gamma} = \underset{\gamma \in \Gamma}{\operatorname{argmin}} \sum_{i=1}^N \left[ \tilde{y}_i - \tilde{Z}_i(\gamma) \hat{\theta}_i \right]' \left[ \tilde{y}_i - \tilde{Z}_i(\gamma) \hat{\theta}_i \right] \quad (16)$$

$$= \underset{\gamma \in \Gamma}{\operatorname{argmin}} \sum_{i=1}^N \left[ \tilde{y}_i - \tilde{X}_i \hat{\beta}_i(\gamma) - \tilde{W}_i(\gamma) \hat{\delta}_i(\gamma) \right]' \left[ \tilde{y}_i - \tilde{X}_i \hat{\beta}_i(\gamma) - \tilde{W}_i(\gamma) \hat{\delta}_i(\gamma) \right]. \quad (17)$$

The above sum of squared residuals is a step function for  $\gamma$  having only  $O(T)$  distinct values. When  $T$  is large, Hansen (1999) suggests approximating  $\Gamma$  by a grid search method to save computational time, searching in  $\Gamma \cap \{q_{1,t}^*, 1 \leq t \leq T\}$ .<sup>3</sup> First, sort the distinct values of the observations on the threshold variable  $q_{1,t}^*$ , and next, trim the top and bottom 1%, 5%, 10%, or any other specific percentiles of  $q_{1,t}^*$ . Finally, search for  $\hat{\gamma}$  over the remaining values of  $q_{1,t}^*$ .<sup>4</sup> Once  $\hat{\gamma}$  has been obtained, the estimators for  $\hat{\beta}_i$ ,  $\hat{\delta}_i$ ,  $\hat{\beta}$  and  $\hat{\delta}$  can be obtained by substituting  $\hat{\gamma}$  for  $\gamma$  in (14), and (15).

The estimate  $\hat{\gamma}$  will be a quantile, and to interpret it in the original variable  $q_{i,t}$  for each unit  $i$  the inverse transformation  $\hat{\gamma}_i = p_i(\hat{\gamma})^{-1}$  must be applied. We will henceforth denote  $\tilde{\beta}_i = \hat{\beta}_i(\hat{\gamma})$ ,  $\tilde{\delta}_i = \hat{\delta}_i(\hat{\gamma})$ ,  $\tilde{\beta} = \hat{\beta}(\hat{\gamma})$ , and  $\tilde{\delta} = \hat{\delta}(\hat{\gamma})$ . Similarly we define  $\tilde{\theta}_i = (\tilde{\beta}_i, \tilde{\delta}_i)$  and  $\tilde{\theta} = (\tilde{\beta}, \tilde{\delta})$ .

#### 4. Assumptions

This section presents the main assumptions under which we develop the asymptotic theory of the estimators. Since the empirical quantile transformation is a one-to-one and onto map, we only

<sup>3</sup>Notice that for each unit  $i$ , the unique values in  $\{q_{i,t}^*, 1 \leq t \leq T\}$  are the same because they are empirical quantiles based on the same number of time series observations.

<sup>4</sup>This problem is computationally faster to solve because there are only  $O(T)$  unique values, instead of  $O(NT)$  in the case without the quantile transformation. The transformation to quantiles is necessary only when the support of the threshold variable is not common across units. If this is not the case, then the quantile transformation is not necessary and the estimation proceeds in the same way except that now there will be  $O(NT)$  distinct threshold values.

study asymptotic properties with  $q_{i,t}$  to simplify notations. Similar results can be obtained by replacing  $q_{i,t}$  by  $q_{i,t}^*$ . Let  $\mathcal{D} = \sigma(F)$  be the minimal sigma-field generated from the factor structure  $F$ . Let  $\mathbb{P}_{\mathcal{D}}(A) = \mathbb{P}(A|\mathcal{D})$  and  $E_{\mathcal{D}}(A) = E(A|\mathcal{D})$ .

We use the superscript 0 to denote the true parameter values. In particular, the true regression coefficients are denoted by  $\theta_i^0 = (\beta_i^{0'}, \delta_i^{0'})'$  for  $i = 1, \dots, N$ , and the true threshold parameter by  $\gamma^0$ .<sup>5</sup>

**Assumption A.1** (Common Factors). *i)  $f_t$  is strictly stationary and ergodic, distributed independently of  $\varepsilon_{i,t'}$  and  $\xi_{i,t'}$  for all  $i, t, t'$ ; ii)  $E\|f_t\|^{4+\epsilon} < \infty$  for some  $\epsilon > 0$ ; iii)  $\frac{1}{T} \sum_{t=1}^T f_t f_t' \xrightarrow{p} \Sigma_f > 0$  for some  $m \times m$  matrix  $\Sigma_f$  as  $T \rightarrow \infty$ .*

**Assumption A.2** (Fixed Factor Loadings). *i)  $\text{Rank}(\bar{\Pi}) = m \leq k$  for all  $N$ , including  $N \rightarrow \infty$ ; ii)  $\|\bar{\Pi}\| < \infty$ ; iii)  $\|\lambda_i\| < \infty$  for  $i = 1, 2, \dots, N$ .*

**Assumption A.3** (Heterogeneity). *The slopes  $\theta_i$  follow the random coefficient model:*

$$\theta_i^0 = \theta^0 + v_i, \quad v_i \sim \text{iid}(0, \Omega_v), \quad i = 1, 2, \dots, N, \quad (18)$$

where  $\|\theta^0\| < C$ ,  $\Omega_v$  is a  $(k+r) \times (k+r)$  symmetric non-negative definite matrix such that  $\|\Omega_v\| < C$ , and the  $v_i$  are distributed independently of  $\lambda_j$ ,  $\Pi_j$ ,  $\xi_{j,t}$ ,  $\varepsilon_{j,t}$  and  $f_t \forall i, j$  and  $t$ .

**Assumption A.4** (Regressors). *i)  $\xi_{i,t}$  are independent and identically distributed (i.i.d.) across  $i$ ; ii) For each  $i$ ,  $\xi_{i,t}$  is strictly stationary and ergodic with  $\rho$ -mixing coefficients over  $t$ , such that  $\sum_{m=1}^{\infty} \rho_m^{1/2} < \infty$ ; iii)  $E(\xi_{i,t} | \mathcal{F}_{NT,t-1}^{(\xi)}) = 0$ , where  $\mathcal{F}_{NT,t-1}^{(\xi)}$  is the sigma field such that  $\mathcal{F}_{NT,t-1}^{(\xi)} = \sigma(\{\xi_{i,t-1}\}_{i=1}^N)$ ; iv)  $E|\xi_{i,t}|^{4+\epsilon} < \infty$  for some  $\epsilon > 0$ .*

**Assumption A.5** (Errors). *i)  $\varepsilon_{i,t}$  are independent and identically distributed (i.i.d.) across  $i$ ; ii) For each  $i$ ,  $\varepsilon_{i,t}$  is strictly stationary and ergodic with  $\rho$ -mixing coefficients over  $t$ , such that  $\sum_{m=1}^{\infty} \rho_m^{1/2} < \infty$ ; iii)  $E(\varepsilon_{i,t} | \mathcal{F}_{NT,t-1}^{(\varepsilon)}) = 0$ , where  $\mathcal{F}_{NT,t-1}^{(\varepsilon)}$  is the sigma field such that  $\mathcal{F}_{NT,t-1}^{(\varepsilon)} = \sigma(\{\varepsilon_{i,t-1}\}_{i=1}^N)$ ; iv)  $E|\varepsilon_{i,t}|^{4+\epsilon} < \infty$ , for some  $\epsilon > 0$  v)  $\varepsilon_{i,t}$  and  $\xi_{j,t'}$  are distributed independently for all  $i, j, t, t'$ ; vi)  $E(\xi_{i,t} \xi_{i,t}') = \Sigma_{\xi}$ , and  $E(\varepsilon_i \varepsilon_i') = \Sigma_{\varepsilon_i}$ .*

**Assumption A.6** (Threshold Parameter Space). *The threshold parameter  $\gamma^0 \in \Gamma = [\underline{\gamma}, \bar{\gamma}]$ , where  $\Gamma$  is a compact set.*

**Assumption A.7** (Continuous Threshold Variable). *i)  $q_{i,t}$  is a continuous threshold variable. Conditionally on  $\mathcal{D}$ ,  $q_{i,t}$  has a conditional probability density function,  $f_{i,t,\mathcal{D}}(\gamma)$ . ii)  $f_{i,t,\mathcal{D}}(\gamma)$  is continuous at  $\gamma = \gamma^0$  and it is uniformly bounded, and thus  $\sup_{i,t} \sup_{\gamma \in \Gamma} f_{i,t,\mathcal{D}}(\gamma) \leq C < \infty$ .*

**Assumption A.8** (Full Rank Conditions). *Conditional on  $\mathcal{D}$ , we require that*

*i)  $M_{T,i}^* > M_{T,i}^*(\gamma^1, \gamma^2) > 0$  for all  $i$ , and  $\gamma^1, \gamma^2 \in \Gamma$ , where  $M_{T,i}^* = E_{\mathcal{D}}(T^{-1} X_i' M_F X_i)$ ,  $M_{T,i}^*(\gamma^1, \gamma^2) = E_{\mathcal{D}}(T^{-1} X_i(\gamma^1)' M_F X_i(\gamma^2))$ ; Particularly, if  $\gamma^1 = \gamma^2 = \gamma$ , then  $M_{T,i}^*(\gamma) = E_{\mathcal{D}}(T^{-1} X_i(\gamma)' M_F X_i(\gamma))$ ;*

<sup>5</sup>If  $q_{i,t}$  has common support across units then the quantile transformation may not be necessary. The asymptotic theory presented below applies to this section as well.

- ii)  $M_{T,i}^* > M_{T,i}^{**}(\gamma) > 0$  for all  $i$  and  $\gamma \in \Gamma$ , where  $M_{T,i}^{**}(\gamma) = E_{\mathcal{D}}(T^{-1}X_i' M_F X_i(\gamma))$ ;
- iii)  $M_{NT}^* > M_{NT}^*(\gamma^1, \gamma^2) > 0$  for all  $\gamma^1, \gamma^2 \in \Gamma$ , where  $M_{NT}^* = E_{\mathcal{D}}\left((NT)^{-1} \sum_{i=1}^N X_i' M_F X_i\right)$ , and  $M_{NT}^*(\gamma^1, \gamma^2) = E_{\mathcal{D}}\left((NT)^{-1} \sum_{i=1}^N X_i(\gamma^1)' M_F X_i(\gamma^2)\right)$ . If  $\gamma^1 = \gamma^2 = \gamma$ , then  $M_{NT}^*(\gamma) = E_{\mathcal{D}}\left((NT)^{-1} \sum_{i=1}^N X_i(\gamma)' M_F X_i(\gamma)\right)$ ;
- iv)  $M_{NT}^* > M_{NT}^{**}(\gamma) > 0$  for all  $\gamma \in \Gamma$ , where  $M_{NT}^{**}(\gamma) = E_{\mathcal{D}}\left((NT)^{-1} \sum_{i=1}^N X_i' M_F X_i(\gamma)\right)$ .

**Assumption A.9** (Identification Condition). *Consider the cross-product terms  $\mathcal{S}_{1i}(\gamma) = \tilde{Z}_i(\gamma)' \tilde{Z}_i(\gamma)$  with  $\mathcal{S}_{1i}(\gamma) > 0$ ,  $\mathcal{S}_{2i}(\gamma) = \tilde{Z}_i(\gamma)' \tilde{W}_i(\gamma^0, \gamma)$ ,  $\mathcal{S}_{3i}(\gamma) = \tilde{W}_i(\gamma^0, \gamma)' \tilde{W}_i(\gamma^0, \gamma)$ , and  $\tilde{W}_i(\gamma^1, \gamma^2) = \tilde{W}_i(\gamma^1) - \tilde{W}_i(\gamma^2)$  for any  $\gamma^1, \gamma^2 \in \Gamma$ . Then, i)  $N^{-1} \sum_{i=1}^N \|\delta_i^0\|^2 > 0$ , including  $N \rightarrow \infty$ ; ii) There exists some constant  $\tau > 0$ , as  $T \rightarrow \infty$  for all  $i$  uniformly, such that:*

$$\mathbb{P} \left\{ \min_{\gamma \in \Gamma} \lambda_{\min} \left[ \frac{1}{T} (\mathcal{S}_{3i}(\gamma) - \mathcal{S}_{2i}(\gamma)' \mathcal{S}_{1i}^{-1}(\gamma) \mathcal{S}_{2i}(\gamma)) \right] \geq \tau \min(1, |\gamma - \gamma^0|) \right\} \xrightarrow{p} 1; \quad (19)$$

- iii)  $\limsup_{N,T} (NT)^{-1} \sum_{i=1}^N E \|(\tilde{Z}_i(\gamma)' e_i)' \mathcal{S}_{1i}^{-1}(\gamma) \tilde{Z}_i(\gamma)' e_i\| < \infty$ , uniformly for  $\gamma \in \Gamma$ ; and iv)  $\limsup_{N,T} (NT)^{-1} \sum_{i=1}^N E \|\delta_i' (\mathcal{S}_{3i}(\gamma) - \mathcal{S}_{2i}(\gamma)' \mathcal{S}_{1i}^{-1}(\gamma) \mathcal{S}_{2i}(\gamma)) \delta_i\| < \infty$  uniformly for  $\gamma \in \Gamma$ .

Assumption A.1 is standard in the literature and it is similar to Assumption 1 in Pesaran (2006) which excludes non-stationary factors and trends. Assumption A.2 is the so-called rank condition and states that the number of factors must be smaller or at most equal to the number of regressors. This assumption is standard in the CCE literature and appears in Pesaran (2006) and in Karavias et al. (2022). Applied research typically finds a small number of factors in the error term, see e.g. Juodis and Sarafidis (2022). To further relieve the strain of the rank condition on the number of regressors, notice that some factors in  $f_t$  may be observable. Observed factors such as the intercept, time effects, seasonal dummies, and other common variables like index stock returns, central bank interest rates and oil prices should be included in  $X_i$ . Any factor included in  $X_i$  does not count towards the dimension  $m$  of  $F$ . When it comes to the factor loadings, we follow Westerlund and Kaddoura (2022) and assume that  $\lambda_i$  and  $\Pi_i$  are fixed in our setting, unlike Pesaran (2006), which assumes that they are random and independent of each other, which is much stronger.

Assumption A.3 states that the heterogeneous coefficients are randomly distributed across units and are independent of any other random elements in the model. This is the prevalent assumption in the heterogeneous coefficients literature, see e.g. Pesaran (2006) and Chudik and Pesaran (2013). Assumption A.4 is similar to Assumption 2 of Pesaran (2006) and states that the series must be stationary and that the cross-sectional dependence across units is fully captured by the factor structure. The  $\rho$  mixing assumption controls the degree of time series dependence and is weaker than uniform mixing, yet stronger than strong mixing. This assumption also implies that  $\rho_m = O(\rho^m)$  with  $|\rho| < 1$ .

Assumption A.5 is the typical zero-mean assumption made for errors, similar to Pesaran (2006). We require the errors to be a martingale difference sequence to avoid bias in  $\hat{\theta}_i(\gamma)$ . The series  $(x_{i,t}, q_{i,t})$  are treated as strictly exogenous variables conditional on  $\mathcal{D}$ . Furthermore,  $\xi_{i,t}$  is mean independent with respect to both future and past  $x_{i,t}, q_{i,t}$  conditional on  $\mathcal{D}$ . Finally,  $\varepsilon_{i,t}$  and  $\xi_{j,t}$

are not correlated. All these ensure the correct specification of the conditional mean. The higher order moments are as in Miao et al. (2020a).

Assumptions A.6 and A.7, which are also standard in the literature, impose that the threshold parameter belongs to a compact set (A.6), and also (A.7) that it has a conditional probability density function and is uniformly bounded (Hansen, 2000).

Assumption A.8 contains full rank assumptions that ensure matrix invertibility wherever necessary, as in Hansen (2000), and assumption A.9 ensures the identification of parameter  $\gamma^0$ .<sup>6</sup> Assumption A.9 is not restrictive, and the case where some units do not have threshold effects is permitted if the order of the number of threshold-affected units is  $N$ .

## 5. Asymptotic Theory

In this section, we derive the asymptotic properties of the proposed estimators and related tests by letting  $N$  and  $T$  tend to infinity. Specifically we shall begin with Theorems 1 and 2 which prove the consistency of  $\hat{\gamma}$  and derive its rate of converge. Theorems 3 and 4 derive the asymptotic distributions of the individual  $\tilde{\theta}_i$  and the MG  $\tilde{\theta}$  estimators. Theorem 5 derives the asymptotic distribution of  $\hat{\gamma}$ , while Theorem 6 provides the likelihood ratio statistic for testing hypotheses about  $\gamma$ . Finally, Theorem 7 derives a hypothesis test for the null of no threshold. All proofs are relegated to Appendix A.

**Theorem 1.** *Under Assumptions A.1-A.9, and as  $(N, T) \rightarrow \infty$ ,  $\hat{\gamma} \xrightarrow{P} \gamma^0$ .*

Theorem 1 establishes the consistency for  $\hat{\gamma}$  under weaker conditions than Miao et al. (2020a), in that it does not require the shrinking threshold assumption, or any restriction on the relative magnitude of  $N$  and  $T$ .

While the shrinking threshold assumption is not necessary to prove consistency of the estimator for  $\gamma$ , it is needed to establish the convergence rate of the estimators.

**Assumption A.10** (Shrinking Threshold Assumption). *Let  $C_{0i} = C_0 + C_{v_i}$ , where  $C_0$  is fixed with  $C_0 \in \mathbb{R}$ , and  $C_{v_i}$  is such that  $C_{v_i} \sim iid(0, \Omega_{v_i})$ , with  $\Omega_{v_i}$  is a  $r \times r$  symmetric non-negative definite matrix and  $\|\Omega_{v_i}\| < C$ . Additionally,  $\sup_{i=1,2,\dots,N} C_{0i} < \infty$ . For  $0 < \alpha < 1/2$ , the threshold effect  $\delta_i^0$  satisfies that  $\delta_i^0 = C_{0i}T^{-\alpha}$  for  $i = 1, 2, \dots, N$ , and therefore, by Assumption A.3,  $\delta^0 = C_0T^{-\alpha}$ .*

Assumption A.10 is called the “shrinking threshold” assumption for the threshold parameter, introduced in the threshold literature in Hansen (1999) and in the structural breaks literature in Bai (1997). This assumption is similar to the idea of local to zero approximations in hypothesis testing, and it is used to derive a pivotal asymptotic distribution of  $\hat{\gamma}$  in our model, such that the critical values can be taken from a table. Without this assumption, the asymptotic distribution is

---

<sup>6</sup>Hansen (2000) suggests an identification condition such as  $\tilde{Z}_i(\gamma)' \tilde{Z}_i(\gamma^0) = \tilde{Z}_i(\gamma^0)' \tilde{Z}_i(\gamma^0)$  for  $\gamma > \gamma^0$ ; yet this equality does not hold in the presence of interactive fixed effects. Instead, we rely on the more complex identification condition proposed by Miao et al. (2020a).

ridden with nuisance parameters arising from the underlying error distribution (see Chan (1993)). However, the assumption implies that the derived asymptotic distribution is a better approximation of the sampling distribution when  $\delta_i^0$  is small. As will be shown below however, when  $\delta_i^0$  is large, it is also easier to estimate.

Hansen (1999) and Miao et al. (2020a) assume that  $\delta_i^0 = O[(NT)^{-\alpha}]$  as opposed to  $\delta_i^0 = O(T)^{-\alpha}$  as in part (ii) of our Assumption A.10. We use this alternative because of the heterogenous coefficients in (1). For a given  $\gamma$ , the estimator for the heterogeneous slope  $\delta_i$  only depends on each unit's information.

Additionally, we require the usual condition on higher conditional moments. Let  $\gamma^1, \gamma^2 \in \Gamma$ , the parameter space and define  $W_i(\gamma^1, \gamma^2) = W_i(\gamma^1) - W_i(\gamma^2)$ . Let  $M_{\mathcal{D}}(\gamma) = (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T E_{\mathcal{D}}(C'_{0i} x_{i,t} x'_{i,t} C_{0i} | q_{i,t} = \gamma) f_{i,t, \mathcal{D}}(\gamma)$ .

**Assumption A.11** (Higher Order Moments). *i) There exist  $\mathcal{D}$ -dependent variables  $C_{i,t}^{\mathcal{D}}$  such that  $\sup_{\gamma \in \Gamma} E_{\mathcal{D}}(\|x_{i,t}\|^4 | q_{i,t} = \gamma) \leq C_{i,t}^{\mathcal{D}}$ ,  $\sup_{\gamma \in \Gamma} E_{\mathcal{D}}(\|x_{i,t} e_{i,t}\|^4 | q_{i,t} = \gamma) \leq C_{i,t}^{\mathcal{D}}$ , and  $\mathbb{P}\left((NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T C_{i,t}^{\mathcal{D}} \leq C\right) = 1$  as  $(N, T) \rightarrow \infty$ .*  
*ii)  $M_{\mathcal{D}}(\gamma)$  is continuous at  $\gamma = \gamma^0$ .*  
*iii) For all  $\epsilon > 0$ , there exist constants  $B > 0$  and  $C_1 > 0$ , for enough large  $N, T$ , we have:*

$$\mathbb{P}\left(\inf_{|\gamma - \gamma^0| < B} \frac{M_{\mathcal{D}}(\gamma)}{|\gamma - \gamma^0|} > C_1\right) > 1 - \epsilon. \quad (20)$$

Assumption A.11 i) assumes that the fourth-order conditional moments of  $x_{i,t}$  and  $x_{i,t} e_{i,t}$  exist and are bounded. Assumption A.11 ii) and iii) are similar to Assumption A.5 of Miao et al. (2020a), and are made to ensure that the square matrix,  $M_{\mathcal{D}}(\gamma)$ , is well-behaved in the neighbourhood of  $\gamma^0$ .

**Theorem 2.** *Under Assumptions A.1-A.11,  $T^\alpha/\sqrt{N} \rightarrow 0$ , as  $(N, T) \rightarrow \infty$ ,  $NT^{1-2\alpha}(\hat{\gamma} - \gamma^0) = O_p(1)$ .*

Theorem 2 shows that the rate of convergence of  $\hat{\gamma}$  is  $NT^{1-2\alpha}$ , thus depending on the magnitude of threshold effects. Smaller  $\alpha$  implies that the threshold effect is larger and in turn the convergence rate is faster. On the other hand, if  $\alpha$  is close to 1/2, the rate of convergence becomes slower since the magnitude of the threshold is smaller. A fixed-magnitude threshold effect, is equivalent to  $\alpha \rightarrow 0$  and generates  $\hat{\gamma} - \gamma^0 = O_p[(NT)^{-1}]$ . However, under this rate of convergence, Chan (1993) finds that the threshold estimator converges in distribution to a functional of a compound Poisson process with unknown nuisance parameters. This distribution can alternatively be simulated by a numerical approach as in Li and Ling (2012). The CCE literature typically requires  $T/N \rightarrow 0$  for pooled estimators (Pesaran, 2006). If this holds, then we have  $T^\alpha/\sqrt{N} \rightarrow 0$  immediately. Therefore, the relative rate of divergence between  $T$  and  $N$  in Theorem 2 above is weaker than what is necessary elsewhere.

Below, we derive the asymptotic distributions of  $\tilde{\theta}_i$ , the MG estimator  $\tilde{\theta}_{MG}$ , and  $\hat{\gamma}$ .

**Theorem 3.** Under Assumptions A.1-A.11,  $T^\alpha/\sqrt{N} \rightarrow 0$ , as  $(N, T) \rightarrow \infty$ , we have:

i)  $\sqrt{T}(\tilde{\theta}_i - \hat{\theta}_i(\gamma^0)) \xrightarrow{p} 0$ .

ii) If we further require that  $\sqrt{T}/N \rightarrow 0$ , and conditional on  $\mathcal{D}$ , then the limiting distribution is  $\sqrt{T}(\tilde{\theta}_i - \theta_i^0) \xrightarrow{d} N(0, V_{\theta_i})$ , where  $V_{\theta_i}$  is the standard asymptotic covariance matrix defined as  $V_{\theta_i} = \Sigma_i^{-1} S_{ie} \Sigma_i^{-1}$ ,  $\Sigma_i = \text{plim}_{T \rightarrow \infty} T^{-1} E_{\mathcal{D}}(Z_i(\gamma^0)' M_F Z_i(\gamma^0))$ ,  $\Sigma_{\varepsilon_i} = T^{-1} E(\varepsilon_i \varepsilon_i')$ , and  $S_{ie} = \text{plim}_{T \rightarrow \infty} T^{-1} E_{\mathcal{D}}(Z_i(\gamma^0)' M_F \Sigma_{\varepsilon_i} M_F Z_i(\gamma^0))$ .

Theorem 3 shows that the difference between  $\tilde{\theta}_i$  and  $\hat{\theta}_i(\gamma^0)$  is  $o_p(1/\sqrt{T})$ , which implies that the distribution of  $\tilde{\theta}_i$  can be approximated by the conventional normal distribution as if  $\gamma$  is known with certainty.

Inference on  $\theta_i^0$  requires a consistent estimator of  $V_{\theta_i}$ . Such an estimator is given by  $\hat{V}_{\theta_i} = \hat{\Sigma}_i^{-1} \hat{S}_{ie} \hat{\Sigma}_i^{-1}$ , where:

$$\hat{\Sigma}_i = \frac{1}{T} \tilde{Z}_i(\hat{\gamma})' \tilde{Z}_i(\hat{\gamma}), \quad \hat{S}_{ie} = \frac{1}{T} \tilde{Z}_i(\hat{\gamma})' \text{diag}(\hat{\varepsilon}_i \hat{\varepsilon}_i') \tilde{Z}_i(\hat{\gamma}), \quad \hat{\varepsilon}_i = \hat{e}_i = \tilde{y}_i - \tilde{Z}_i(\hat{\gamma}) \hat{\theta}_i(\hat{\gamma}). \quad (21)$$

Theorem 3 is useful for testing hypotheses on individual units' parameters. As an example, consider testing the statistical significance of the regression coefficients:  $H_{0i} : \theta_{i_k}^0 = 0$  v.s.  $H_{1i} : \theta_{i_k}^0 \neq 0$ , for some  $i = 1, \dots, N$ , where  $i_k$  denotes  $k$ th element in  $\theta_i^0$ .

Turning to the asymptotic distribution for the MG estimator (13).

**Theorem 4.** Under Assumptions A.1-A.11, as  $(N, T) \rightarrow \infty$  and  $T^\alpha/\sqrt{N} \rightarrow 0$ , we have:

$$\begin{bmatrix} \sqrt{N} I_K & 0_{K \times r} \\ 0_{r \times K} & \sqrt{N} T^\alpha I_r \end{bmatrix} (\tilde{\theta} - \theta^0) \xrightarrow{d} N(0, \Sigma_{MG}), \quad (22)$$

where  $\Sigma_{MG} = \Omega_v$ . An estimator for the variance is:

$$\hat{\Sigma}_{MG} = \begin{bmatrix} I_K & 0_{K \times r} \\ 0_{r \times K} & T^{2\alpha} I_r \end{bmatrix} \frac{1}{N-1} \sum_{i=1}^N (\tilde{\theta}_i - \tilde{\theta})(\tilde{\theta}_i - \tilde{\theta})'$$

Theorem 4 derives the rate of convergence and the asymptotic distribution of the MG estimator. The theorem demonstrates that slope coefficient heterogeneity has unique implications for the asymptotic theory, as  $\tilde{\delta}$  has a rate of convergence that is much faster than that of the MG estimator in linear regression, see e.g., Juodis et al. (2021). This arises due to the shrinking threshold assumption A.10:

$$\delta_i^0 = \frac{C_{0i}}{T^\alpha} = \frac{C_0}{T^\alpha} + \frac{C_{v_i}}{T^\alpha}, \quad (23)$$

which shows that the error term which drives unit heterogeneity  $C_{v_i}/T^\alpha$ , is  $O(T^{-\alpha})$ . This implies that in the limit, heterogeneity vanishes and we have a homogeneous model. The closer the  $\delta_i^0$  are to zero, the closer they are to each other, In other words, they become more homogeneous. This is reflected in the estimator rate of convergence, which increases to reach the standard  $\sqrt{NT}$  found in pooled estimators of homogeneous panel models, see e.g. Pesaran (2006). If the threshold is

large, this can be captured by  $\alpha \rightarrow 0$  so that  $\sqrt{N}(\tilde{\theta} - \theta^0) \xrightarrow{d} N(0, \Sigma_{MG})$ , which is the same result as Pesaran (2006). If the threshold is small, however, then we are closer to the homogeneous case where  $\alpha \rightarrow 1/2$  so that  $\sqrt{NT}(\tilde{\theta} - \theta^0) \xrightarrow{d} N(0, \Sigma_{MG})$ .

The result of Theorem 4 cannot be used because it contains the unknown parameter  $\alpha$ . Dividing (22) by  $T^\alpha$  gives the following usable corollary.

**Corollary 1.** *Under the Assumptions of of Theorem 4,  $\sqrt{N}(\tilde{\theta} - \theta^0) \xrightarrow{d} N(0, \Sigma_{MG})$ , and a consistent estimator for  $\Sigma_{MG}$  is given by  $\hat{\Sigma}_{MG} = (N - 1)^{-1} \sum_{i=1}^N (\tilde{\theta}_i - \tilde{\theta})(\tilde{\theta}_i - \tilde{\theta})'$ .*

Below, we derive the asymptotic distribution of the threshold parameter estimator. We will need the following two assumptions. Let  $f_{i,t}(\cdot)$  denote the pdf of  $q_{i,t}$ .

**Assumption A.12** (No jumps in threshold variable). *For all  $i$ , if  $s > t$ ,  $f_{i,s|t}(\gamma^0 | \gamma^0) < \infty$ .*

**Assumption A.13** (Well-behaved limits). *Define:*

$$D_{NT}(\gamma) = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T E [(C'_{0i} x_{i,t})^2 | q_{i,t} = \gamma] f_{i,t}(\gamma), \quad (24)$$

$$V_{NT}(\gamma) = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T E [(C'_{0i} x_{i,t})^2 \varepsilon_{i,t}^2 | q_{i,t} = \gamma] f_{i,t}(\gamma), \quad (25)$$

and let  $D(\gamma) = \text{plim}_{(N,T) \rightarrow \infty} D_{NT}(\gamma)$  and  $V(\gamma) = \text{plim}_{(N,T) \rightarrow \infty} V_{NT}(\gamma)$ .

i) *The limits  $D(\gamma)$  and  $V(\gamma)$  exist and are continuous at  $\gamma = \gamma^0$ .*

ii) *There is a constant  $C^*$  such that:*

$$\text{Var} \left[ \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T E_{\mathcal{D}} (\|g_{i,t}^{(1)}(\gamma^1, \gamma^2)\|^2) \right] \leq C^* \sup_{i,t} E (\|g_{i,t}^{(1)}(\gamma^1, \gamma^2)\|^4), \quad (26)$$

$$\text{Var} \left[ \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T E_{\mathcal{D}} (\|g_{i,t}^{(2)}(\gamma^1, \gamma^2)\|^2) \right] \leq C^* \sup_{i,t} E (\|g_{i,t}^{(2)}(\gamma^1, \gamma^2)\|^4), \quad (27)$$

where  $g_{i,t}^{(1)}(\gamma^1, \gamma^2) = x_{i,t} |d_{i,t}(\gamma^1) - d_{i,t}(\gamma^2)|$ ,  $g_{i,t}^{(2)}(\gamma^1, \gamma^2) = x_{i,t} \varepsilon_{i,t} |d_{i,t}(\gamma^1) - d_{i,t}(\gamma^2)|$  and  $d_{i,t}(\gamma) = \mathbb{I}\{q_{i,t}^* \leq \gamma\}$ .

Assumption A.12 excludes the possibility for the unit  $i$ ,  $q_{i,t} = \gamma^0$  holds for all  $t = 1, 2, \dots, T$ , which is similar as Assumption 8 of Hansen (1999). Assumption A.13 provides conditional moment boundedness for all  $\gamma \in \Gamma$ .

**Theorem 5.** *Under Assumptions A.1-A.13, if  $T^\alpha/\sqrt{N} \rightarrow 0$ , and  $\sqrt{T}/N \rightarrow 0$  as  $(N, T) \rightarrow \infty$ , then:*

$$NT^{1-2\alpha}(\hat{\gamma} - \gamma^0) \xrightarrow{d} \phi \zeta, \quad (28)$$

where  $\phi = V(\gamma^0)/D(\gamma^0)^2$ ,  $\zeta = \text{argmax}_{-\infty < r < \infty} \{(-1/2)|r| + W(r)\}$ , and  $W(r)$  is a two-sided standard Brownian motion on the real line.

Theorem 5 gives the asymptotic distribution of the threshold estimate  $\hat{\gamma}$ . The threshold parameter's asymptotic distribution is the same as in Hansen (2000). In the presence of conditional homoskedasticity in  $\varepsilon_{i,t}$ ,  $\phi$  can be further simplified as  $\sigma^2[D(\gamma^0)]^{-1}$ . Hansen (2000) suggests that the asymptotic distribution of Theorem 5 should not be used for the construction of confidence intervals on  $\gamma$ , because the nuisance parameter  $\phi$  is hard to estimate accurately.

We now propose a likelihood ratio test for the null hypothesis  $H_0 : \gamma = \gamma^0$ . This test can be used to get confidence intervals for  $\gamma^0$ , instead of the result in Theorem 5. Let

$$LR(\gamma) = \frac{RSS(\hat{\theta}_1(\gamma), \dots, \hat{\theta}_N(\gamma), \gamma) - RSS(\hat{\theta}_1(\hat{\gamma}), \dots, \hat{\theta}_N(\hat{\gamma}), \hat{\gamma})}{\hat{\sigma}^2},$$

where  $RSS(\hat{\theta}_1(\gamma), \dots, \hat{\theta}_N(\gamma), \gamma) = \sum_{i=1}^N (\tilde{y}_i - \tilde{Z}_i(\gamma)\hat{\theta}_i(\gamma))'(\tilde{y}_i - \tilde{Z}_i(\gamma)\hat{\theta}_i(\gamma))$ , and the error variance estimator is given by  $\hat{\sigma}^2 = (NT)^{-1}RSS(\tilde{\theta}_1, \dots, \tilde{\theta}_N, \hat{\gamma})$ .

**Theorem 6.** Under  $H_0 : \gamma = \gamma^0$ , Assumptions A.1-A.13,  $T^\alpha/\sqrt{N} \rightarrow 0$ , and  $\sqrt{T}/N \rightarrow 0$ , as  $(N, T) \rightarrow \infty$ ,

$$LR(\gamma^0) \xrightarrow{d} \eta^2 \Xi,$$

where  $\eta^2 = V(\gamma^0)/(\sigma^2 D(\gamma^0))$  and  $\sigma^2 = \text{plim}_{(N,T) \rightarrow \infty} (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T \varepsilon_{i,t}^2$ . The random variable  $\Xi = \sup_{s \in \mathbb{R}} [2W(s) - |s|]$  has a distribution function given by  $\mathbb{P}(\Xi \leq x) = (1 - \exp(-x/2))^2$ .

In a special case where the error  $\varepsilon_{i,t}$  is homoskedastic along both the cross-section and time dimension,  $\eta^2 = 1$  and inference can be made based on readily available critical values. The distribution function of  $\Xi$  has the inverse:

$$c(a) = -2\log(1 - \sqrt{1 - a}). \quad (29)$$

where  $c(a)$  is the critical value and  $a$  is the significance level. A test of  $H_0$  rejects at the asymptotic level  $a$  if  $LR(\gamma^0)$  exceeds  $c(a)$ . The critical values appear in Table 1 of Hansen (2000) and are: for  $a = 0.1$  it is 5.94, for  $a = 0.05$  it is 7.35 and for  $a = 0.01$  it is 10.59. In terms of forming an asymptotic confidence interval for  $\gamma$ , the non-rejection region of confidence level  $1 - a$  is the set of values of  $\gamma$  such that  $LR(\gamma) \leq c(a)$ . The confidence interval can be found by plotting  $LR(\gamma)$  against  $\gamma$  and drawing a flat line at  $c(a)$ .

For homoskedastic errors the above Theorem 6 is asymptotically correct under the shrinking threshold assumption  $\delta_i^0 \rightarrow 0$  for all  $i = 1, 2, \dots, N$ . However, if the errors are additionally normal, i.i.d. across both dimensions and independent of the regressors and the threshold variable, we conjecture that the result of Theorem 3 of Hansen (2000) holds, which says that inference based on the LR test is asymptotically valid even if  $\delta_i^0$  does not shrink towards 0.

If the errors are not homoskedastic, however,  $\eta^2$  needs to be estimated consistently. This can be done by noticing that

$$\eta^2 = \frac{\text{plim}_{(N,T) \rightarrow \infty} \sum_{i=1}^N \sum_{t=1}^T E((\delta_i^0 x_{i,t} \varepsilon_{i,t})^2 | q_{i,t} = \gamma^0) f_t(\gamma^0)}{\sigma^2 \text{plim}_{(N,T) \rightarrow \infty} \sum_{i=1}^N \sum_{t=1}^T E((\delta_i^0 x_{i,t})^2 | q_{i,t} = \gamma^0) f_t(\gamma^0)}, \quad (30)$$



which can be estimated nonparametrically by

$$\hat{\eta}^2 = \frac{\sum_{i=1}^N \sum_{t=1}^T K_h(\hat{\gamma} - q_{i,t}) (\tilde{\delta}'_i x_{i,t} \hat{\varepsilon}_{i,t})^2}{\hat{\sigma}^2 \sum_{i=1}^N \sum_{t=1}^T K_h(\hat{\gamma} - q_{i,t}) (\tilde{\delta}'_i x_{i,t})^2}, \quad (31)$$

where  $\hat{\sigma}^2 = (NT)^{-1} RSS(\tilde{\theta}_1, \dots, \tilde{\theta}_N, \hat{\gamma})$ ,  $K_h(u) = h^{-1}K(u/h)$ ,  $K(\cdot)$  is a kernel function, and  $h \rightarrow 0$  is the bandwidth parameter. It is straightforward to show consistency of the above estimator as in Hansen (2000) and Miao et al. (2020a). Given the estimator  $\hat{\eta}^2$ , a normalized  $LR$  statistic is defined as  $NLR(\gamma^0) = LR(\gamma^0)/\hat{\eta}^2$  and its critical values are the same as in the homoskedastic case studied above.

## 6. Testing the threshold effect

Testing the null hypothesis of no threshold regression is challenging for two reasons. First, under  $H_0 : \delta_i^0 = 0$ , for  $i = 1, \dots, N$ , the parameter  $\gamma^0$  disappears and thus can only be identified under the alternative. This is a nonstandard hypothesis testing problem which has been studied by Davies (1977) and further developed by Hansen (1996), among others. To deal with this problem we will follow the previous literature and test the null hypothesis based on a supremum-type Wald statistic whose limiting distribution is non-standard but can be estimated by the bootstrap.

The second challenge, which is new to the threshold literature is that the null hypothesis above is equivalent to testing  $N$  individual hypotheses, where  $N$  goes to infinity. This is a multiple testing problem that can lead to low “power”, where power here has the more loose interpretation of “not the null”. To avoid the multiple testing issue altogether, we employ an approach recently put forward in Juodis et al. (2021), which notices that under the null hypothesis, the model becomes homogeneous in  $\delta_i^0$  because  $\delta_i^0 = \delta^0 = 0$  for all  $i$ . For homogeneous threshold coefficients, the data generating process becomes

$$y_i = X_i \beta_i + W_i(\gamma) \delta + e_i. \quad (32)$$

For a given  $\gamma$ ,  $\delta^0$  can be estimated by the pooled estimator

$$\hat{\delta}_p(\gamma) = \left( \sum_{i=1}^N W_i(\gamma)' M_{Z_i^*(\gamma)} W_i(\gamma) \right)^{-1} \left( \sum_{i=1}^N W_i(\gamma)' M_{Z_i^*(\gamma)} y_i \right), \quad (33)$$

where  $M_{Z_i^*(\gamma)} = I - Z_i^*(\gamma)(Z_i^{*'}(\gamma)Z_i^*(\gamma))^{-1}Z_i^{*'}(\gamma)$ ,  $Z_i^*(\gamma) = (\bar{X}, \bar{W}(\gamma), X_i)$ , and  $\bar{W}(\gamma) = N^{-1} \sum_{i=1}^N W_i(\gamma)$ .

The estimator  $\hat{\delta}_p(\gamma)$  is a pooled estimator with a faster  $\sqrt{NT}$  rate of convergence. It also varies from  $\hat{\delta}(\gamma)$  in terms of the annihilator matrix. In  $M_{Z_i^*(\gamma)}$ ,  $\bar{W}(\gamma)$  is included to remove asymptotic bias from the interactive effects following Karavias et al. (2022), and  $X_i$  is included to project out the variables with the heterogeneous coefficients as in Juodis et al. (2021).

The supremum Wald statistic for the null hypothesis is:

$$\sup W = \sup_{\gamma \in \Gamma} W_{NT}(\gamma), \quad (34)$$

$$W_{NT}(\gamma) = NT \hat{\delta}_p(\gamma)' \hat{V}_{\hat{\delta}_p(\gamma)} \hat{\delta}_p(\gamma), \quad (35)$$

where  $\hat{V}_{\hat{\delta}_p(\gamma)} = \hat{\Sigma}^{-1}(\gamma, \gamma) \hat{K}(\gamma, \gamma) \hat{\Sigma}^{-1}(\gamma, \gamma)$ ,  $\hat{\Sigma}(\gamma, \gamma) = \frac{1}{NT} \sum_{i=1}^N \tilde{W}_i(\gamma)' \tilde{W}_i(\gamma)$ , and where  $\hat{K}(\gamma, \gamma) = \frac{1}{NT} \sum_{i=1}^N \tilde{W}_i(\gamma)' \text{diag}(\hat{\varepsilon}_i(\gamma) \hat{\varepsilon}_i(\gamma)') \tilde{W}_i(\gamma)$ . In the above  $\tilde{y}_i(\gamma) = M_{Z_i^*(\gamma)} y_i$ ,  $\tilde{W}_i(\gamma) = M_{Z_i^*(\gamma)} W_i(\gamma)$ , and  $\hat{\varepsilon}_i(\gamma) = \hat{\tilde{\varepsilon}}_i(\gamma) = \tilde{y}_i(\gamma) - \tilde{W}_i(\gamma) \hat{\delta}_p(\gamma)$ .

The implementation of  $\sup W$  requires an approximation of  $\Gamma$  like the one used in estimating  $\hat{\gamma}$  above. To derive the limiting distribution of  $\sup W$  we will need an additional assumption. Define:

$$S_{NT}(\gamma) = \frac{1}{\sqrt{NT}} \sum_{i=1}^N W_i(\gamma)' M_F \varepsilon_i, \text{ and } \Sigma(\gamma^1, \gamma^2) = \underset{(N,T) \rightarrow \infty}{\text{plim}} \frac{1}{NT} \sum_{i=1}^N E(W_i(\gamma^1)' M_F W_i(\gamma^2) | \mathcal{D}).$$

Assumption A.14 is a high level assumption. It is straight forward to prove its claims from more primitive conditions following Hansen (1996) or Lemma A.9 in the appendix of Miao et al. (2020a).

**Assumption A.14.**  $S_{NT}(\gamma) \Rightarrow S(\gamma)$ , where  $S(\gamma)$  is a mean-zero Gaussian process with covariance kernel  $K(\gamma^1, \gamma^2) = \underset{(N,T) \rightarrow \infty}{\text{plim}} (NT)^{-1} \sum_{i=1}^N E(W_i(\gamma^1)' M_F \Sigma_{\varepsilon_i} M_F W_i(\gamma^2) | \mathcal{D})$ , where  $\Sigma_{\varepsilon_i} = \underset{(N,T) \rightarrow \infty}{\text{plim}} E(\varepsilon_i \varepsilon_i')$ .

**Theorem 7.** Suppose that Assumptions A.1-A.8 and A.14 hold, as  $(N, T) \rightarrow \infty$  and  $T/N \rightarrow 0$ , under  $H_0 : \delta_i^0 = 0$ , for  $i = 1, \dots, N$ , we have:

$$\sup W \xrightarrow{d} \sup_{\gamma \in \Gamma} W_{NT}^c(\gamma) \quad (36)$$

where  $W_{NT}^c(\gamma) = \bar{S}(\gamma)' \bar{K}(\gamma, \gamma)^{-1} \bar{S}(\gamma)$ , and  $\bar{S}(\gamma) = \Sigma(\gamma, \gamma)^{-1} S(\gamma)$  is a mean-zero Gaussian process with covariance kernel  $\bar{K}(\gamma^1, \gamma^2) = \Sigma(\gamma^1, \gamma^1)^{-1} K(\gamma^1, \gamma^2) \Sigma(\gamma^2, \gamma^2)^{-1}$ .

The limiting distribution under the null depends on nuisance parameters. Following Hansen (1996), Chudik et al. (2017) and Giannerini et al. (2023) we find the critical values using the bootstrap. The steps for the bootstrap and Monte Carlo simulations evaluating its performance can be found in the Appendix.

## 7. Empirical Application

As an empirical application of the above theory, we revisit one of the key puzzles in international economics, the Feldstein-Horioka puzzle (Feldstein and Horioka, 1980). In theory, perfect capital mobility should allow savings from one country to be invested in other countries, where better investment opportunities with higher returns are available. Feldstein and Horioka (1980), however, find that this is not the case, and that domestic investments are highly correlated with domestic

savings. Ever since, this has become one of the six major puzzles of international macroeconomics (Obstfeld and Rogoff, 2000).

Feldstein and Horioka (1980) used cross-sectional data to estimate the relationship:

$$\frac{I}{Y} = a + \beta \frac{S}{Y}, \quad (37)$$

where  $Y$  is national income,  $I$  is domestic investment, and  $S$  is domestic savings. They find that  $\beta$  is closer to 1, than to 0, meaning that there is a strong relationship between domestic saving rates and domestic investment.

The closest contribution to ours is that of Hacıoğlu Hoke and Kapetanios (2021) which examine this problem for a panel of OECD countries by introducing a nonlinear relationship between investment and savings, based on the trade openness of the country. The idea is that higher trade openness would imply fewer frictions and therefore higher capital mobility. This would in turn weaken the relationship between domestic savings and investment, given that more investment opportunities are made available worldwide. The nonlinearity is modelled via a smooth transition model based on the logistic function.

Our analysis is different from that in Hacıoğlu Hoke and Kapetanios (2021) in that we employ the discontinuous and heterogeneous threshold model developed previously. The discontinuous threshold model has several advantages over smooth transition models. First, it has fewer parameters to estimate and there is no need to select the continuous transition function. Second, it is more intuitive to interpret as it implies that there are only two regimes, below and above the threshold, while smooth transition models can be viewed as models with a wide continuum of states between two extreme regimes. Finally, the threshold parameter estimator in the discontinuous model is super-consistent and thus has a faster rate of convergence when compared to the smooth transition threshold parameter. The model we consider is the following:

$$\begin{aligned} Investment_{i,t} = & \beta_{1i} Savings_{i,t} + \beta_{2i} Trade\ Openness_{i,t} \\ & + \delta_i Savings_{i,t} \mathbb{I}\{p_i(Trade\ Openness_{i,t}) \geq \gamma\} + \alpha_i + \lambda'_i f_t + \varepsilon_{i,t}, \end{aligned} \quad (38)$$

where  $Investment_{i,t}$  is the investment share of real GDP per capita for country  $i$  at year  $t$ ,  $Savings_{i,t}$  is the percentage share of current savings to GDP per capita for country  $i$  at year  $t$ ,  $Trade\ Openness_{i,t}$  denotes the trade openness for country  $i$  at year  $t$ ,  $p_i(\gamma)$  is the percentile function of the distribution of  $Trade\ Openness_{i,t}$  across  $t = 1, \dots, T$  for country  $i$ ,  $\alpha_i$  is the fixed effects,  $\lambda'_i f_t$  are the interactive fixed effects, and  $\varepsilon_{i,t}$ , are the innovations. The data is taken from the Penn World Tables version 7.1 as in Hacıoğlu Hoke and Kapetanios (2021) and covers the period of 1951-2000 resulting in a balanced panel with  $N = 45$  and  $T = 50$ .

The main results are displayed in Table 1. The top panel of the table presents the results of the test for non-linearity which can be seen to reject the null of no-threshold regression. The middle panel, reveals the threshold parameter estimate, which is 0.685. This means that for country  $i$  if trade openness is below the 68.5 percentile of that country's historical trade openness distribution,

**Table 1:** Main Results

Test for the presence of threshold effects:			
SupW	96.194		
P-value	0.000		
Thereshold estimate:			
Threshold( $\hat{\gamma}$ )	0.685		
95%CI	[0.685, 0.700]		
Coefficient estimates:			
	Investment		
	EST	SE	HC
Savings	0.730***	0.062	0.867
Saving(Trade Openness $\geq$ Threshold Level)	-0.085***	0.024	0.578
Control variables:			
Trade Openness	0.081*	0.049	0.689

Notes: EST denote the coefficient estimates  $\tilde{\beta}$  and  $\tilde{\delta}$ , SE are the corresponding standard errors, and HC indicates the proportion of units that have statistically significant  $\beta_i$  and  $\delta_i$  are defined in the section 5.3. \*, \*\* and \*\*\* denote statistical significance at the 10%, 5% and 1% levels, respectively.

then that country is in the low regime. Beyond that threshold, country  $i$  is in the high regime. The confidence interval is asymmetric because it is created by inverting the normalised LR statistic as suggested in Hansen (2000). The bottom panel of Table 1 reports the estimates for  $\tilde{\beta}$  and  $\tilde{\delta}$  which are the MG estimates based on the estimated threshold values. The coefficient of savings in the low regime is 0.73, which is in line with previous literature. Hacıoğlu Hoke and Kapetanios (2021) report estimates of  $\tilde{\beta}$  of 0.69 (pooled) and 0.61 (MG) for OECD countries when the cross-sectional dependence is accounted for. On the other hand, Feldstein and Horioka (1980) reported the savings retention parameter to be 0.89, while Obstfeld and Rogoff (2000) found a smaller estimate of 0.60. Notice that the forms of non-linearity are modelled in different ways in the above mentioned papers. In the high regime, the savings retention parameter is reduced by 0.085 to  $0.73 - 0.085 = 0.645$  showing that as trade openness increases international capital mobility also increases. The standalone trade openness regressor is included to avoid potential omitted variable bias. However, removing it will not alter the results obtained.

Table 2, translates the 68.5 percentile threshold into actual trade openness threshold values for the countries in our sample. Each country has a different distribution of the trade openness variable and thus the 68.5 percentile translates to a different trade openness threshold. Table 2 presents these thresholds in an increasing order. The country with the lowest threshold is the US. Generally, from the left column of the table, larger economies seem to have lower thresholds, implying that capital flows are easier to leave the domestic market. On the contrary, smaller economies require significant levels of trade openness in order to divert domestic savings toward investment opportunities abroad.

Figure 1 displays the proportion of countries that are in the high regime. This proportion can be seen to increase rapidly in the 1980s, and can be explained by the extensive trade liberalisation policies implemented internationally at the time, see Faini (2004). The results show that higher levels of trade liberalisation increase international capital mobility.

**Table 2:** Heterogeneous Threshold Levels

United States	13.005	United Kingdom	32.635	Israel	58.489
India	13.121	Portugal	33.131	Norway	58.580
Brazil	13.248	Pakistan	33.492	Philippines	59.025
Japan	15.252	Peru	35.001	Thailand	60.204
Mexico	15.440	New Zealand	35.463	Ireland	64.784
Turkey	18.903	Finland	40.968	Netherlands	69.857
Argentina	19.025	Canada	43.879	Trinidad and Tobago	70.627
Colombia	20.603	Iceland	45.518	Cyprus	85.200
Congo, Dem. Rep.	20.841	Sweden	48.747	Egypt, Arab Rep.	88.981
Spain	22.683	Bolivia	49.311	Belgium	96.196
Australia	23.453	Denmark	51.214	Sri Lanka	116.871
Uruguay	28.454	Venezuela	51.767	Honduras	126.058
Uganda	29.121	South Africa	55.277	Puerto Rico	129.323
France	29.136	Switzerland	56.129	Panama	174.804
Italy	30.279	Austria	56.526	Luxembourg	200.054
Mean	54.237				

Notes: Mean denotes the average of heterogeneous threshold levels in the sample.

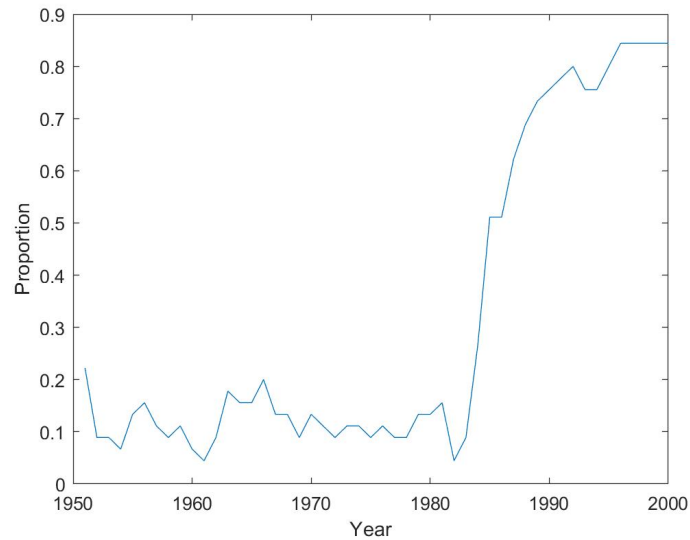
## 8. Conclusion

In this paper, we studied a new panel data threshold regression model with heterogeneous coefficients and interactive fixed effects. We employed a quantile transformation approach to achieve pooled estimation of the threshold parameter and threshold heterogeneity across units. We provided a complete and novel inferential theory for the estimated heterogeneous slope coefficients and the threshold parameter. We also proposed a test for the presence of the threshold effect. The Monte Carlo simulations showed that the new theory works well in practice. The new model was used to examine the Feldstein-Horioka puzzle, where we found that once trade openness is greater than a threshold, capital mobility increases.

There are still several interesting topics for future research. Interactive fixed effects represent a great novelty in panel threshold regression and existing methods could be extended along this direction. Such examples include panel threshold models with endogenous threshold variables as in Seo and Shin (2016) and Yu et al. (2023), multiple-regime threshold models (Yau et al., 2015), binary response models (Gao et al., 2023) and quantile models as in Chen (2019), Harding et al. (2020) and Zhang et al. (2021).

## References

- Ahn, S.C., Lee, Y.H., Schmidt, P., 2001. Gmm estimation of linear panel data models with time-varying individual effects. *Journal of Econometrics* 101, 219–255.
- Ahn, S.C., Lee, Y.H., Schmidt, P., 2013. Panel data models with multiple time-varying individual effects. *Journal of Econometrics* 174, 1–14.
- Bai, J., 1997. Estimation of a change point in multiple regression models. *Review of Economics and Statistics* 79, 551–563.



**Figure 1:** Proportion of countries experiencing the threshold effects

- Bai, J., 2009. Panel data models with interactive fixed effects. *Econometrica* 77, 1229–1279.
- Bai, J., 2010. Common breaks in means and variances for panel data. *Journal of Econometrics* 157, 78–92.
- Bonhomme, S., Manresa, E., 2015. Grouped patterns of heterogeneity in panel data. *Econometrica* 83, 1147–1184.
- Chan, K.S., 1993. Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *The Annals of Statistics* 21, 520–533.
- Chen, H., Chong, T.T.L., Bai, J., 2012. Theory and applications of tar model with two threshold variables. *Econometric Reviews* 31, 142–170.
- Chen, J., Shin, Y., Zheng, C., 2022. Estimation and inference in heterogeneous spatial panels with a multifactor error structure. *Journal of Econometrics* 229, 55–79.
- Chen, L., 2019. Two-step estimation of quantile panel data models with interactive fixed effects. *Econometric Theory* , 1–28.
- Chudik, A., Mohaddes, K., Pesaran, M.H., Raissi, M., 2017. Is there a debt-threshold effect on output growth? *Review of Economics and Statistics* 99, 135–150.
- Chudik, A., Pesaran, M.H., 2013. Large panel data models with cross-sectional dependence: a survey. *CAFE Research Paper* .
- Davies, R.B., 1977. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 64, 247–254.

- Faini, R., 2004. Trade liberalization in a globalizing world. IZA Discussion Papers No. 1406.
- Feldstein, M., Horioka, C., 1980. Domestic saving and international capital flows. *The Economic Journal* 90, 314–329.
- Fernández-Val, I., Lee, J., 2013. Panel data models with nonadditive unobserved heterogeneity: Estimation and inference. *Quantitative Economics* 4, 453–481.
- Gao, J., Liu, F., Peng, B., Yan, Y., 2023. Binary response models for heterogeneous panel data with interactive fixed effects. *Journal of Econometrics* .
- Gao, J., Xia, K., Zhu, H., 2020. Heterogeneous panel data models with cross-sectional dependence. *Journal of Econometrics* 219, 329–353.
- Giannerini, S., Goracci, G., Rahbek, A., 2023. The validity of bootstrap testing for threshold autoregression. *Journal of Econometrics* .
- Giglio, S., Xiu, D., 2021. Asset pricing with omitted factors. *Journal of Political Economy* 129, 1947–1990.
- Hacıoğlu Hoke, S., Kapetanios, G., 2021. Common correlated effect cross-sectional dependence corrections for nonlinear conditional mean panel models. *Journal of Applied Econometrics* 36, 125–150.
- Hansen, B.E., 1996. Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* 64, 413–430.
- Hansen, B.E., 1999. Threshold effects in non-dynamic panels: Estimation, testing, and inference. *Journal of Econometrics* 93, 345–368.
- Hansen, B.E., 2000. Sample splitting and threshold estimation. *Econometrica* 68, 575–603.
- Harding, M., Lamarche, C., Pesaran, M.H., 2020. Common correlated effects estimation of heterogeneous dynamic panel quantile regression models. *Journal of Applied Econometrics* 35, 294–314.
- Holtz-Eakin, D., Newey, W., Rosen, H.S., 1988. Estimating vector autoregressions with panel data. *Econometrica* 56, 1371–1395.
- Juodis, A., Karavias, Y., Sarafidis, V., 2021. A homogeneous approach to testing for granger non-causality in heterogeneous panels. *Empirical Economics* 60, 93–112.
- Juodis, A., Sarafidis, V., 2022. A linear estimator for factor-augmented fixed-t panels with endogenous regressors. *Journal of Business & Economic Statistics* 40, 1–15.
- Karavias, Y., Narayan, P.K., Westerlund, J., 2022. Structural breaks in interactive effects panels and the stock market reaction to covid-19. *Journal of Business & Economic Statistics* , 1–14.

- Kejriwal, M., Li, X., Totty, E., 2020. Multidimensional skills and the returns to schooling: Evidence from an interactive fixed-effects approach and a linked survey-administrative data set. *Journal of Applied Econometrics* 35, 548–566.
- Li, D., Ling, S., 2012. On the least squares estimation of multiple-regime threshold autoregressive models. *Journal of Econometrics* 167, 240–253.
- Lu, X., Su, L., 2016. Shrinkage estimation of dynamic panel data models with interactive fixed effects. *Journal of Econometrics* 190, 148–175.
- Massacci, D., 2017. Least squares estimation of large dimensional threshold factor models. *Journal of Econometrics* 197, 101–129.
- Massacci, D., Sarno, L., Trapani, L., 2021. Factor models with downside risk. Available at SSRN 3937321 .
- Miao, K., Li, K., Su, L., 2020a. Panel threshold models with interactive fixed effects. *Journal of Econometrics* 219, 137–170.
- Miao, K., Su, L., Wang, W., 2020b. Panel threshold regressions with latent group structures. *Journal of Econometrics* 214, 451–481.
- Moon, H.R., Weidner, M., 2018. Nuclear norm regularized estimation of panel regression models. arXiv preprint arXiv:1810.10987 .
- Norkutė, M., Sarafidis, V., Yamagata, T., Cui, G., 2021. Instrumental variable estimation of dynamic linear panel data models with defactored regressors and a multifactor error structure. *Journal of Econometrics* 220, 416–446.
- Obstfeld, M., Rogoff, K., 2000. The six major puzzles in international macroeconomics: is there a common cause? *NBER macroeconomics annual* 15, 339–390.
- Pesaran, M.H., 2006. Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica* 74, 967–1012.
- Pesaran, M.H., Shin, Y., Smith, R.P., 1999. Pooled mean group estimation of dynamic heterogeneous panels. *Journal of the American Statistical Association* 94, 621–634.
- Pesaran, M.H., Smith, R., 1995. Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics* 68, 79–113.
- Seo, M.H., Shin, Y., 2016. Dynamic panels with threshold effect and endogeneity. *Journal of Econometrics* 195, 169–186.
- Su, L., Shi, Z., Phillips, P.C., 2016. Identifying latent structures in panel data. *Econometrica* 84, 2215–2264.



- Swamy, P.A., 1970. Efficient inference in a random coefficient regression model. *Econometrica* 38, 311–323.
- Tong, H., 1978. On a threshold model in pattern recognition and signal processing, ed. CH Chen, Amsterdam: Sijhoff & Noordhoff .
- Trapani, L., 2021. Inferential theory for heterogeneity and cointegration in large panels. *Journal of Econometrics* 220, 474–503.
- Tsay, R.S., 1989. Testing and modeling threshold autoregressive processes. *Journal of the American Statistical Association* 84, 231–240.
- Tsay, R.S., 1998. Testing and modeling multivariate threshold models. *Journal of the American Statistical Association* 93, 1188–1202.
- Westerlund, J., Kaddoura, Y., 2022. Cce in heterogenous fixed-t panels. *The Econometrics Journal* 25, 715–738.
- Westerlund, J., Urbain, J.P., 2015. Cross-sectional averages versus principal components. *Journal of Econometrics* 185, 372–377.
- Yau, C.Y., Tang, C.M., Lee, T.C., 2015. Estimation of multiple-regime threshold autoregressive models with structural breaks. *Journal of the American Statistical Association* 110, 1175–1186.
- Yu, P., Liao, Q., Phillips, P.C.B., 2023. New control function approaches in threshold regression with endogeneity. *Econometric Theory* , 1–55.
- Zhang, Y., Wang, H.J., Zhu, Z., 2021. Single-index thresholding in quantile regression. *Journal of the American Statistical Association* , 1–16.