Dimension heterogeneity for threshold model

Yimeng Xie*

November 9, 2020

For the latest version of my JMP please Click Here

Abstract

This paper is concerned with unobserved heterogeneity between regressors of a panel data threshold model. Previous studies unnecessarily assume coefficients of arbitrary regressors change across the same thresholds which may cause misspecification of model and break down inference and prediction. In this paper we allow coefficient of each regressor to be changed across specific thresholds such that regressors sharing the same thresholds are defined as one dimension while across dimensions thresholds can be different. We proposes a new threshold estimator for this generalized model by exploiting the characteristics of threshold estimation. The estimator is shown to be valid regardless of the numbers of both dimensions and thresholds. It is computationally efficient especially when there is sparse interaction, e.g. one threshold of one dimension is too close to a threshold of another dimension. Small sample properties of the estimator are investigated by Monte Carlo simulations and shown to be satisfactory. Moreover, the estimator is applied to two empirical finance studies and it verifies the importance of discussing dimension heterogeneity in threshold modeling.

Keywords: Panel Data Model, Threshold, Projection Matrix, Sequential Estimation, Model Selection, Firm's Investment Spending, Credit Card

JEL classification: C01, C23, C33

*Department of Economics, University of Southern California, Los Angeles, CA 90089. Email: yimengxi@usc.edu.

1 Introduction

Threshold model has been increasingly popular after first proposed in Tong Tong (1978). The model is usually constructed by presuming economic relationship is changing if the value of an observed variable surpasses a predetermined and unknown threshold. It has been widely used in time series and cross sectional applications to explore economic pattern in different economic regime. One of the most famous time series applications is threshold autoregressive models. For example, Beaudry and Koop (1993) which apply threshold autoregressive model to U.S. GNP and verify asymmetric effects of shocks over business cycle. In addition, as Hansen (2011) discussed, threshold autoregressive modeling has been also involved in macro and finance studies such as interest arbitrage, purchasing power parity, exchange rates, stock returns, and transaction cost effects. Nowadays, as the pandemic of COVID 19 swept the global economy, plenty of researches switch to threshold autoregressive models again to study either the pattern of virus transmission or influence of pandemic on economics development, see Kim et al. (2020) and Chudik et al. (2020) for instance. Threshold model is also popular with cross sectional applications. For example, Card et al. (2008) observe a tipping point of minority share in a neighborhood by threshold model such that white population will drop if the minority share surpasses the tipping point. Another example is Yu and Phillips (2018) which study the threshold in income level and find tax deferred savings policies induce different impacts on savings for people in different income level. Recently researches of threshold model have also been extended to panel data model, of which economics regimes can vary along both time series dimension or cross sectional dimension. One of classic panel data applications is examination of the effect of financial constraints on investment decisions by short panel data of firm level, which has been studied by Hansen (1999) and Seo and Shin (2016).

The asymptotic theory of estimation and inference for threshold model for time series and cross sectional data has been developed well by Chan et al. (1993), Hansen (2000). These works are valid for model with exogenous regressors and threshold parameters. Later, many efforts have been made to generalize the model by including either the regressors or threshold parameters to be endogenous, such as Caner and Hansen (2004), Yu (2013), Kourtellos et al. (2014), Yu and Phillips (2018) and Liao et al. (2015) and Yu et al. (2019). For panel data, Hansen (1999) adapts the traditional threshold estimation to nondynamic model with fixed effect, which is then extended to dynamic model by Ramírez-Rondán (2020) and Seo and Shin (2016).

However, all of the above works make arbitrary implicit assumptions about heterogeneity of regressors in threshold model. For example, most of these papers allow coefficients of all regressors to change when the value of threshold variable surpasses the same threshold while others assume coefficients of some arbitrary regressors change across the same threshold. These implicit assumptions indeed are unnecessary and will induce adverse impacts on both specification and prediction. In this paper, we define "dimension" as regressors which share the same threshold structure including number and values of threshods, and allow threshold structure of different dimensions are different. Thereby we design an estimator try to address this potential problem by taking advantage of structure of threshold model. The essence of our estimator is based on the indicator function and linearity in threshold model such that we can filter out other regressors with different thresholds. Our estimation procedure is quite useful for all kinds of data structure while in this paper we focus on short panel data. More than that, we generalize threshold setups such that for each dimension there can be more than one threshold. We are able to determine the number of thresholds for each dimension by a revision of information criterion suggested by Gonzalo and Pitarakis (2002).

Misspecification of the threshold model with dimension heterogeneity will cause the estimates to be biased and also worsen the prediction performance. For example, in the issue of testing whether the sensitivity of investment spending is changing when financial condition is worse, Hansen (1999) and Gonzalez et al. (2017) restrict that only the variable cash flow is subject to the influence of financial condition, while Seo and Shin (2016) assume all variables switch at the same threshold of financial condition. The former find when firm is financial constrained, the sensitivity of investment to cash flow is lower which is contradicted to the corporate finance theory. The latter is consistent with the theory, but we don't know if such modeling is more reasonable than the former. In the application part, we show evidence to supports the modeling of Seo and Shin (2016). Our results indicate it is more powerful in prediction to add threshold to models reasonably. This sheds some light on the reasons why previous studies always find linear model is better at prediction than threshold model. For instance, Clements and Krolzig (1998) compare time series forecasting ability of linear autoregressive model and threshold autoregressive model only to find accounting for non-linearity decreases forecasting power. Though not the necessary condition for that, the implicit restrictions of dimensional homogeneity in their model can be one possible reason.

One can consider dealing with dimension heterogeneity of threshold model in other ways. One alternative is to estimate one threshold by fixing other thresholds. Despite plausible feasibility, this method will undoubtedly bring out a lot of computation burden. Since there are multiple thresholds, one can estimate the model either by joint estimation or sequential estimation as suggested by Gonzalo and Pitarakis (2002). Generally the former is most computationally costly so that practitioners won't choose it. The latter needs to ensure the consistency of the estimated thresholds as the estimator is derived from a misspecified model. Gonzalo and Pitarakis (2002) prove sequential estimator can be consistent and converge to the true value fast when there is only one dimension in the model. For multiple dimensions, it is not clear that classic sequential estimation can hold consistency as one dimension. More than that, when the number of threshold is unknown, one also needs to do model selection with all dimensions which will not be easy to fix.

As far as we know, no one has touched topic of dimension heterogeneity in threshold model before. A related work to our study is Cheng et al. (2019) who discuss classifications of firms in the term of production function. In particular, they classify firms along two separate diemensions, the elasticities of variable units and elasticities of capital. This offers more flexibility in modeling heterogeneity, for example a firm may have high elasticities of variable units and meanwhile low elasticities of capital simultaneously. To justify their classification, Cheng et al. (2019) apply k-means technique to generate groups of units along different dimensions. Obviously, by estimating threshold for different dimensions, our study also classified units to certain groups. So we are able to find similar economics pattern. Also there are obvious differences in modeling between Cheng et al. (2019) and our work: Cheng et al. (2019) restrict the group membership to be time invariant, which ours does not need as the membership is determined by whether observed threshold variable is above or below the fixed threshold. In that sense, we are able to generate forecast for either in sample or out of sample. Compared to that, our estimator associates classification of economic units to observed variable, which in the context of panel data can be generally time varying. Besides, the fact that the classification by our method is determined by the value of threshold variable can be appreciated more as it is utilized to generate forecast for either in sample or out of sample. By the method of Cheng et al. (2019), only in sample forecasting can be conducted and it has to be under the assumption that those economic units will stay in their dimensions for the whole time periods.

The paper is organized as following. Section 2 begins with the motivations for this paper and presents a basic panel threshold model with dimension heterogeneity. Section 3 gives the estimator and its extensions to more general situations with multiple dimensions and multiple thresholds. Section 4 verifies the finite sample performance of our estimator and information criterion by Monte Carlo simulations. Section 5 applied our methods to two empirical studies: influence of financial constraint on firm's investment decisions and influence of payment ratio on credit card . Finally we conclude and propose several future extensions in section 6.

Throughout this paper, we use the following notations. Let $tr\left(\mathbf{A}\right)$ denotes the trace of matrix $\mathbf{A} = \left(a_{ij}\right)_{m\times m}$, the Frobenius norm of $\mathbf{A}$ is defined as $\|\mathbf{A}\| = \sqrt{tr\left(\mathbf{A}'\mathbf{A}\right)}$, and $\|\mathbf{A}\|_\infty = \max\left(a_{ij}\right)$. For arbitrary variables a and b, $a \vee b = max(a,b)$, $a \wedge b = min(a,b)$,

$a \sim b$ denotes a and b are equivalent. $\to_p$ and $\to_d$ denote convergence in probability and in distribution,respectively. For a variable x_n, $x_n = o_p(a_n)$ denotes x_n/a_n converges to zero in probability as n approaches an appropriate limit, while $x_n = O_p(a_n)$ denotes x_n/a_n is stochastically bounded. For any parameter c, denote c^0 as the true value.

2 Model

2.1 Dimension heterogeneity

To highlight the fact of dimensional heterogeneity, suppose we have the following model with scalar variables (y, x_1, x_2, q),

$$\begin{aligned} y &= x_1\beta_1 + x_2\beta_2 + (x_1\delta_1 + x_2\delta_2)\mathrm{l}\{q \leq \gamma\} + u \\ &= x_1\beta_1 + x_1\delta_1\mathrm{l}\{q \leq \gamma\} + x_2\beta_2 + x_2\delta_2\mathrm{l}\{q \leq \gamma\} + u, \end{aligned} \tag{2.1}$$

then it is obvious that coefficients of x_1 and x_2 switch across the same threshold γ. A more generalized form should allow x_1 to have a different threshold from x_2, such as

$$y = x_1\beta_1 + x_1\delta_1\mathrm{l}\{q \leq \gamma_1\} + x_2\beta_2 + x_2\delta_2\mathrm{l}\{q \leq \gamma_2\} + u. \tag{2.2}$$

We call x_1 as one dimension and x_2 as the other dimension and the example shows one kind of dimension of dimension heterogeneity is defined by $\gamma_1 \neq \gamma_2$. Intuitively we cannot guarantee there is no dimension heterogeneity in a threshold model. On the other hand, it is not necessary to force all regressors to be subject to the threshold structure, and the model can also be

$$y = x_1\beta_1 + x_1\delta_1\mathrm{l}\{q \leq \gamma_1\} + x_2\beta_2 + u. \tag{2.3}$$

which is another kind of dimension heterogeneity. As we explain later, the dimension heterogeneity can be more complicated and generalized by allowing multiple thresholds in different dimensions. Besides, the threshold parameter doesn't need to be the same, which means (2.2) can be generalized as

$$y = x_1\beta_1 + x_1\delta_1\mathrm{l}\{q_1 \leq \gamma_1\} + x_2\beta_2 + x_2\delta_2\mathrm{l}\{q_2 \leq \gamma_2\} + u. \tag{2.4}$$

When dealing with threshold models, traditional literature almost all avoids the discussion of the dimension heterogeneity but arbitrarily decide the threshold structure. For example, about the application example of investment and financial constraint, Seo and Shin (2016) construct model by (2.1). Compared to that, for the same application Hansen (1999) construct model by admitting the second dimension heterogeneity. It is hard to say who should be right without any additional systematic analysis.

2.2 Model and assumption

Throughout this paper, we focus on repeated cross sectional data and short panel data. Consider the following model,

$$y_{it} = \beta^{'}\mathbf{x}_{it} + \delta^{'}\mathbf{x}_{it}1\{q_{it} \leq \gamma\} + u_{it} \tag{2.5}$$

for $i = 1, \cdots, N$ and $t = 1, \cdots, T$, where $N \to \infty$ and T is fixed. $\mathbf{x}_{it}$ is a $k \times 1$ vector while q_{it} is a scalar. $(y_{it}, \mathbf{x}_{it}, q_{it})$ are observed data. (β, δ, γ) are unknown parameters. u_{it} denotes the error term. For this moment we don't consider fixed effect, later we will show under regular conditions a fixed effect model can be transformed to (2.5). It is easy to see such model provides heterogeneity for both time and cross section as long as q_{it} is varying across i and t.

As discussed above, we can construct a more generalized threshold model such as the following

$$y_{it} = \beta_1^{\prime}\mathbf{x}_{1it} + \beta_2^{\prime}\mathbf{x}_{2it} + \delta_1^{\prime}\mathbf{x}_{1it}1\{q_{it} \leq \gamma_1\} + \delta_2^{\prime}\mathbf{x}_{2it}1\{q_{it} \leq \gamma_2\} + u_{it} \tag{2.6}$$

where $\mathbf{x}_{1it}$ and $\mathbf{x}_{2it}$ are respectively $k_1 \times 1$ and $k_2 \times 1$ vectors such that $k_1 + k_2 = k$. $(y_{it}, \mathbf{x}_{1it}, \mathbf{x}_{2it}, q_{it})$ are observed data. $(\beta_1, \beta_2, \delta_1, \delta_2, \gamma_1, \gamma_2)$ are unknown parameters and γ_1 may be or not be different from γ_2. Indeed in empirical data there is no supporting evidence to guarantee the threshold structure should be the same for all elements in regressors. The concern arises that we may misspecify the model (2.6) as the same as (2.5) and estimation of coefficeints would therefore based on a wrong sample.

Define $\theta = (\underline{\theta}, \gamma_1, \gamma_2) = (\beta_1, \beta_2, \delta_1, \delta_2, \gamma_1, \gamma_2)$. Correspondingly the parameter space is defined as $\Theta = (\underline{\Theta}, \Gamma)$. We shall need the following assumptions for estimation.

Assumption A (i) The parameter space Θ is convex and compact.

(ii) For each t, $(\mathbf{x}_{1it}, \mathbf{x}_{2it}, q_{it}, u_{it})$ are i.i.d. across i. For each i, $(\mathbf{x}_{1it}, \mathbf{x}_{2it}, q_{it}, u_{it})$ are strictly stationary, ergodic and ρ-mixing, with ρ-mixing coefficients satisfying $\sum_{m=1}^{\infty} \rho_m^{1/2} < \infty$.

(iii)$\sum_{i=1}^{N}\sum_{t=1}^{T} \mathbf{x}_{1it}\mathbf{x}_{2it}^{'} = O\left(N^{1-\alpha}\right)$ where $\alpha \in (0, 1/2]$.

(iv) $E(u_{it}\mathbf{x}_{1it-\tau}) = \mathbf{0}$ and $E(u_{it}\mathbf{x}_{2it-\tau}) = \mathbf{0}$ for $\tau \geq 0$. Also $E(u_{it}q_{it-\tau}) = 0$ for $\tau \geq 0$.

(v) $E|\mathbf{x}_{it}|^4 < \infty$ and $E|\mathbf{x}_{it}u_{it}|^4 < \infty$ where $\mathbf{x}_{it} = (\mathbf{x}_{1it}, \mathbf{x}_{2it})$.

(vi) $E\left(|\mathbf{x}_{it}|^4 u_{it}^4 | q_{it} = \gamma\right) \leq C$ and $E(|\mathbf{x}_{it}|^4 | q_{it} = \gamma) \leq C$ for some $C < \infty$.

(vii) The probability density function of the threshold variable q_{it}, $f_{it}(\gamma)$ satisfies $0 < f_{it}(\gamma) < \bar{f} < \infty$ for all $\gamma \in \Gamma$

Assumption A (i) (ii) mainly specify the regular conditions for repeated cross sectional data model. In particular, for the error term we exclude the cross sectional dependence but allow serial correlation. (iii) specifies the weak correlations between two dimensions, in particular when the cross section number N is sufficiently large. This is a technique assumption required for the proof. For example, suppose $k_1 = k_2 = 1$, then define

$$\theta_{1,2} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} x_{2it}^2 \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=1}^{T} x_{2it} x_{1it} \right)$$

and by Assumption A (iii) coefficients $\theta_{1,2}$ is uniformly bounded by $N^{-\alpha}$. By symmetry, we can define $\theta_{2,1}$ as coefficient for regression of x_{2it} on x_{it}. Intuitively, if two regressors are highly correlated, it is better to form them in one dimension. However as discussed in the simulation results, such assumption can be released and the properties of the estimator are still kept. On the other hand, in practice we can also transform regressors to independent covariates by, for example, principal component analysis as it generates principal components that are independent from each other as they are essentially the eigenvector of a matrix. (iv) allows data to be weakly exogenous so that the model can contain dynamic structure with lags of dependent variable. (v) and (vi) specifies boundedness of unconditional and conditional moments which are required to form central limit theorem. (vii) states the support of distribution of q_{it} is the subset of that of Γ, and it allows us to estimate value of γ from collections of q_{it}. These assumptions are common in classical panel data researches.

3 Projection estimator

3.1 Intuition behind the estimator

Before our estimator is formally presented, we will show the special facts of the model and intuition that our estimator relies on. Note in the classical threshold model with just one dimension such as (2.5), it is usually impossible to locate the exact true threshold parameter. This is because of the limit of information of q_{it} and the fact that

$$q_{it} \leq \gamma_0 \Rightarrow q_{it} \leq \gamma_0 + \Delta$$

as long as Δ is a small positive constant. Indeed in classical threshold model, we can only identify γ_0 in an interval between two close values of q_{it}. This is the reason why the estimation of γ is usually obtained as q_{it} for some i and t (left-hand estimator) or the middle point value of two close values of q_{it} (middle point estimator), see Hansen (1999) and Yu (2013) for instance. In this paper, we choose the left hand estimator. On the other hand, this identification shortcoming can help us to filter out the term that consists of the indicator function. The essence is that the true threshold can only be identified from $\{q_{i1}, \cdots, q_{iT}\}$ for $i = 1, \cdots, N$ and $t = 1, \cdots, T$. Suppose there is one threshold in the model and the true value is γ_0, then for variable x_{it} and q_{it} we have

$$x_{it}\mathrm{l}\{q_{it} \leq \gamma_0\} = x_{it}\mathrm{l}\{q_{it} \leq \check{\gamma}\}$$

where $\check{\gamma} \in \{q_{11}, \cdots, q_{1T}, q_{21}, \cdots, q_{NT}\}$ despite the fact that we know the exact value of neither γ_0 nor $\check{\gamma}$. So ideally we can construct a large $NT \times NT$ matrix $\mathbf{F} = \left(\mathbf{F}_1', \mathbf{F}_2', \cdots \mathbf{F}_N'\right)'$, where

$$\mathbf{F}_i = \begin{pmatrix} x_{i1}\mathrm{l}\{q_{i1} \leq q_{11}\} & \cdots & x_{i1}\mathrm{l}\{q_{i1} \leq q_{1T}\} & x_{i1}\mathrm{l}\{q_{i1} \leq q_{21}\} & \cdots & x_{i1}\mathrm{l}\{q_{i1} \leq q_{NT}\} \\ x_{i2}\mathrm{l}\{q_{i2} \leq q_{11}\} & \cdots & x_{i2}\mathrm{l}\{q_{i2} \leq q_{1T}\} & x_{i2}\mathrm{l}\{q_{i2} \leq q_{22}\} & \cdots & x_{i2}\mathrm{l}\{q_{i2} \leq q_{NT}\} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{iT}\mathrm{l}\{q_{iT} \leq q_{11}\} & \cdots & x_{iT}\mathrm{l}\{q_{iT} \leq q_{1T}\} & x_{iT}\mathrm{l}\{q_{iT} \leq q_{22}\} & \cdots & x_{iT}\mathrm{l}\{q_{iT} \leq q_{NT}\} \end{pmatrix}_{T \times NT}$$

for $i = 1, \cdots, N$ such that one column in $\mathbf{F}_i$ is exactly equal to $(x_{i1}\mathrm{l}\{q_{i1} \leq \gamma_0\}, x_{i2}\mathrm{l}\{q_{i2} \leq \gamma_0\}, \cdots, x_{iT}\mathrm{l}\{q_{iT} \leq \gamma_0\})$. However N can be large and we may have many values either repeated or close to each other so that we don't need to represent threshold by q_{it} for all $i = 1, \cdots, N$ and $t = 1, \cdots, T$. Instead we can use sufficient number of quantiles of empirical distribution of q_{it} to build up a $NT \times S$ matrix $\tilde{\mathbf{F}} = \left(\tilde{\mathbf{F}}_1', \tilde{\mathbf{F}}_2', \cdots \tilde{\mathbf{F}}_N'\right)'$ where S is the number of quantiles we refer to and

$$\tilde{\mathbf{F}}_i = \begin{pmatrix} x_{i1}\mathrm{l}\{q_{i1} \leq q_{\{1\}}\} & x_{i1}\mathrm{l}\{q_{i1} \leq q_{\{2\}}\} & \cdots & x_{i1}\mathrm{l}\{q_{i1} \leq q_{\{S\}}\} \\ x_{i2}\mathrm{l}\{q_{i2} \leq q_{\{1\}}\} & x_{i2}\mathrm{l}\{q_{i2} \leq q_{\{2\}}\} & \cdots & x_{i2}\mathrm{l}\{q_{i2} \leq q_{\{S\}}\} \\ \vdots & \vdots & \vdots & \vdots \\ x_{iT}\mathrm{l}\{q_{iT} \leq q_{\{1\}}\} & x_{iT}\mathrm{l}\{q_{iT} \leq q_{\{2\}}\} & \cdots & x_{iT}\mathrm{l}\{q_{iT} \leq q_{\{S\}}\} \end{pmatrix}_{T \times S}$$

with $q_{\{s\}}$ for $s = 1, 2, \cdots, S$ as quantiles of empirical distribution of q_{it}. Apparently when S is sufficiently small, we can construct a orthogonal projection matrix

$$\mathbf{M}_0 = I_{NT} - \tilde{\mathbf{F}}(\tilde{\mathbf{F}}'\tilde{\mathbf{F}})^{-1}\tilde{\mathbf{F}}'.$$

such that for

$$\mathbf{x}(\gamma_0) = (x_{11}\mathrm{l}\{q_{11} \leq \gamma_0\}, \cdots, x_{1T}\mathrm{l}\{q_{1T} \leq \gamma_0\}, x_{21}\mathrm{l}\{q_{21} \leq \gamma_0\}, \cdots, x_{iT}\mathrm{l}\{q_{NT} \leq \gamma_0\})',$$

it is spontaneous to get

$$\mathbf{M}_0\mathbf{x}(\gamma_0) = \mathbf{0}$$

which suggests a way to filter out the regressors with threshold.

3.2 A simple example

We consider the data generating process that doesn't have fixed effect. In other words, we can treat constant 1 as a dimension. We can write down the above model in a vector form, basically

$$\mathbf{y}_i = \mathbf{x}_{1i}\beta_1 + \mathbf{x}_{2i}\beta_2 + \mathbf{x}^*_{1i}(\gamma_1)\delta_1 + \mathbf{x}^*_{2i}(\gamma_2)\delta_2 + \mathbf{u}_i \tag{3.1}$$

for $i = 1, \cdots, N$, where $\mathbf{x}_{1i} = (x_{1i1}, \cdots, x_{1iT})'$ and $\mathbf{x}_{2i} = (x_{2i1}, \cdots, x_{2iT})'$, $\mathbf{x}^*_{1i}(\gamma_1) = (x_{1i1}\mathbf{l}\{q_{i1} \leq \gamma_1\}, \cdots, x_{1iT}\mathbf{l}\{q_{iT} \leq \gamma_1\})'$ and $\mathbf{x}^*_{2i}(\gamma_2) = (x_{2i1}\mathbf{l}\{q_{i1} \leq \gamma_2\}, \cdots, x_{2iT}\mathbf{l}\{q_{iT} \leq \gamma_2\})'$. Therefore the criterion function is constructed as following,

$$S_{NT}(\theta) = \sum_{i=1}^{N} (\mathbf{y}_i - \mathbf{x}_{1i}\beta_1 - \mathbf{x}_{2i}\beta_2 - \mathbf{x}^*_{1i}(\gamma_1)\delta_1 - \mathbf{x}^*_{2i}(\gamma_2)\delta_2)' (\mathbf{y}_i - \mathbf{x}_{1i}\beta_1 - \mathbf{x}_{2i}\beta_2 - \mathbf{x}_{1i}(\gamma_1)\delta_1 - \mathbf{x}_{2i}(\gamma_2)\delta_2) \tag{3.2}$$

which leads to the estimator of coefficients and threshold parameter as

$$\hat{\theta} = \arg\min_{\theta \in \Theta} S_{NT}(\theta),$$

We can rewrite the model as

$$\mathbf{Y} = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \mathbf{X}^*_1(\gamma_1)\delta_1 + \mathbf{X}^*_2(\gamma_2)\delta_2 + \mathbf{U} \tag{3.3}$$

where $\mathbf{Y} = (\mathbf{y}'_1, \cdots, \mathbf{y}'_N)'$, $\mathbf{X}_j = (\mathbf{x}'_{j1}, \cdots, \mathbf{x}'_{jN})'$ for $j = 1, 2$, and $\mathbf{X}^*_j(\gamma_j) = (\mathbf{x}^*_{j1}(\gamma_j)', \cdots, \mathbf{x}^*_{jN}(\gamma_j)')'$ for $j = 1, 2$. And correspondingly the criterion function is equivalently written as

$$S_{NT}(\theta) = (\mathbf{Y} - \mathbf{X}_1\beta_1 - \mathbf{X}_2\beta_2 - \mathbf{X}^*_1(\gamma_1)\delta_1 - \mathbf{X}^*_2(\gamma_2)\delta_2)' (\mathbf{Y} - \mathbf{X}_1\beta_1 - \mathbf{X}_2\beta_2 - \mathbf{X}^*_1(\gamma_1)\delta_1 - \mathbf{X}^*_2(\gamma_2)\delta_2) \tag{3.4}$$

For the parameter space Γ, ideally it is represented by the value of q_{it} for $i = 1, 2, \cdots, N$ and $t = 1, \cdots, T$. When the sample size is large, a good approximation can be some quantiles of the empirical distribution of q_{it}. Suppose we can divide the real line between 1% and 99% into m parts and define $v = 98/m$, then the quantiles we consider are

$$q_{\{l\}} = \{\gamma : Pr(q_{it} < \gamma) \leq l\%\}$$

for $l = 1, 1 + v, 1 + 2v, \cdots, 99 - 2v, 99 - v, 99$, where $v = 98/m$. By increasing the value of m, we are able to provide more quantiles of distribution of q_{it} and the estimation of threshold would be more accurate. However, taking too many quantiles will add the computation cost, therefore we will choose the value m as a tradeoff of efficiency and computation cost. The reason we exclude the possibility that true threshold is less than 1% quantile or larger than 99% is because estimation of threshold may be unavailable due to small subsample size. Then we can use the following estimation procedure.

- Step 1: For an arbitrary $\gamma \in \Gamma$, we are going to filter $(\mathbf{Y}, \mathbf{X}_1, \mathbf{X}_1^*(\gamma))$.

  - For each i, let
$$\mathbf{P}_{-1,i} = [\mathbf{x}_{2i}^*(q_{\{1\}}), \cdots, \mathbf{x}_{2i}^*(q_{\{99\}})]$$
where $\mathbf{x}_{2i}^*(q_{\{l\}}) = (x_{2i1}1\{q_{i1} \leq q_{\{l\}}\}, \cdots, x_{2iT}1\{q_{iT} \leq q_{\{l\}}\})'$ for $l = 1, 1+v, \cdots, 99-v, 99$ and the size of $\mathbf{P}_{-1,i}$ is $m+1$. And we can thereby construct the orthogonal projection matrix,
$$\mathbf{M}_{-1} = I_{NT} - \mathbf{P}_{-1}\left(\mathbf{P}_{-1}'\mathbf{P}_{-1}\right)^{-1}\mathbf{P}_{-1}'$$
where
$$\mathbf{P}_{-1} = (\mathbf{P}_{-1,1}', \cdots, \mathbf{P}_{-1,N}')'$$
of which the dimension is $NT \times (m+1)$.
  - Now we filter $(\mathbf{Y}, \mathbf{X}_1, \mathbf{X}_1^*(\gamma))$ such that
$$\begin{aligned}\tilde{\mathbf{Y}}_1 &= \mathbf{M}_{-1}\mathbf{Y}\\ \tilde{\mathbf{X}}_1 &= \mathbf{M}_{-1}\mathbf{X}_1\\ \tilde{\mathbf{X}}_1^*(\gamma) &= \mathbf{M}_{-1}\mathbf{X}_1^*(\gamma)\end{aligned}$$

- Step 2: Define
$$S_N^1(\gamma) = \tilde{\mathbf{Y}}_1'\mathbf{M}_1(\gamma)\tilde{\mathbf{Y}}_1 \tag{3.5}$$
where the orthogonal projection matrix $\mathbf{M}_1(\gamma)$ is defined as
$$\mathbf{M}_1(\gamma) = I_{NT} - \mathbf{P}_1(\gamma)\left(\mathbf{P}_1(\gamma)'\mathbf{P}_1(\gamma)\right)^{-1}\mathbf{P}_1(\gamma)'$$
and
$$\mathbf{P}_1(\gamma) = [\tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_1^*(\gamma)]$$
And we can estimate γ_1 by solving
$$\hat{\gamma}_1 = \arg\min_{\gamma\in\Gamma} S_N^1(\gamma). \tag{3.6}$$
Then we can also get estimation the coefficients $(\hat{\beta}_1, \hat{\delta}_1)$ of this dimension by running the least squares regression of $\tilde{\mathbf{Y}}_1$ on $\tilde{\mathbf{X}}_1$ and $\tilde{\mathbf{X}}_1(\hat{\gamma}_1)$.

- Step 3: For an arbitrary $\gamma \in \Gamma$, we are going to filter $(\mathbf{Y}, \mathbf{X}_2, \mathbf{X}_2^*(\gamma))$.

– For each i, let

$$\mathbf{P}_{-2,i} = [\mathbf{x}^*_{1i}(q_{\{1\}}), \cdots, \mathbf{x}^*_{2i}(q_{\{99\}})]$$

where $\mathbf{x}^*_{1i}(q_{\{l\}}) = (x_{1i1}1\{q_{i1} \leq q_{\{l\}}\}, \cdots, x_{1iT}1\{q_{iT} \leq q_{\{l\}}\})'$ for $l = 1\%, 1\% + v, \cdots, 99\% - v, 99\%$ and the size of $\mathbf{P}_{-2,i}$ is $T \times (m+1)$. And we can thereby construct the orthogonal projection matrix,

$$\mathbf{M}_{-2} = I_{NT} - \mathbf{P}_{-2}\left(\mathbf{P}'_{-2}\mathbf{P}_{-2}\right)^{-1}\mathbf{P}'_{-2}$$

where

$$\mathbf{P}_{-2} = (\mathbf{P}'_{-2,1}, \cdots, \mathbf{P}'_{-2,N})' \tag{3.7}$$

of which the dimension is $NT \times (m+1)$.

– Now we filter $(\mathbf{Y}, \mathbf{X}_2, \mathbf{X}^*_2(\gamma))$ such that

$$\begin{aligned}\tilde{\mathbf{Y}}_2 &= \mathbf{M}_{-2}\mathbf{Y}\\ \tilde{\mathbf{X}}_2 &= \mathbf{M}_{-2}\mathbf{X}_1\\ \tilde{\mathbf{X}}^*_2(\gamma) &= \mathbf{M}_{-2}\mathbf{X}^*_1(\gamma)\end{aligned}$$

- Step 4: And we can estimate γ_2 by solving

$$\hat{\gamma}_2 = \arg\min_{\gamma\in\Gamma} \tilde{\mathbf{Y}}'_2\mathbf{M}_2(\gamma)\tilde{\mathbf{Y}}_2 \tag{3.8}$$

where the orthogonal projection matrix $\mathbf{M}_2(\gamma)$ is defined as

$$\mathbf{M}_2(\gamma) = I_{NT} - \mathbf{P}_2(\gamma)\left(\mathbf{P}_2(\gamma)'\mathbf{P}_2(\gamma)\right)^{-1}\mathbf{P}_2(\gamma)'$$

and

$$\mathbf{P}_2(\gamma) = [\tilde{\mathbf{X}}_2, \tilde{\mathbf{X}}^*_2(\gamma)]$$

Then we can also get estimation the coefficients $(\hat{\beta}_2, \hat{\delta}_2)$ of this dimension by running the least squares regression of $\tilde{\mathbf{Y}}_2$ on $\tilde{\mathbf{X}}_2$ and $\tilde{\mathbf{X}}_2(\hat{\gamma}_2)$.

3.2.1 Asymptotic theory

Notice here we assume there is only one threshold for each dimension. With this prior information, we will show the consistency of estimators for both threshold parameters and coefficients. Indeed just the analysis of one dimension would be enough as the model is symmetric and the same way can be applied to analyze the other dimension.

Define

$$\mathbf{A}(\gamma_1,\gamma_2) = E\begin{pmatrix} \mathbf{X}_1\mathbf{X}_1' & \mathbf{X}_1\mathbf{X}_2' & \mathbf{X}_1\mathbf{X}_1^*(\gamma_1)' & \mathbf{X}_1\mathbf{X}_2^*(\gamma_2)' \\ \mathbf{X}_2\mathbf{X}_1' & \mathbf{X}_2\mathbf{X}_2' & \mathbf{X}_2\mathbf{X}_1^*(\gamma_1)' & \mathbf{X}_2\mathbf{X}_2^*(\gamma_2)' \\ \mathbf{X}_1^*(\gamma_1)\mathbf{X}_1' & \mathbf{X}_1^*(\gamma_1)\mathbf{X}_2' & \mathbf{X}_1^*(\gamma_1)\mathbf{X}_1^*(\gamma_1)' & \mathbf{X}_1^*(\gamma_1)\mathbf{X}_2^*(\gamma_2)' \\ \mathbf{X}_2^*(\gamma_2)\mathbf{X}_1' & \mathbf{X}_2^*(\gamma_2)\mathbf{X}_2' & \mathbf{X}_2^*(\gamma_2)\mathbf{X}_1^*(\gamma_1)' & \mathbf{X}_2^*(\gamma_2)\mathbf{X}_2^*(\gamma_2)' \end{pmatrix}$$

and

$$\mathbf{V}(\gamma_1,\gamma_2) = \mathbf{A}(\gamma_1,\gamma_2)^{-1}E(\mathbf{X}'\boldsymbol{\Omega}\mathbf{X})\mathbf{A}(\gamma_1,\gamma_2)^{-1}$$

where

$$\begin{aligned} \mathbf{X} &= (\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_1^*(\gamma_1), \mathbf{X}_2^*(\gamma_2)) \\ \boldsymbol{\Omega} &= E(\mathbf{U}\mathbf{U}'). \end{aligned}$$

Then we need more assumption about the projection matrix.

Assumption B (i) $\mathbf{P}_{-j}'\mathbf{P}_{-j}$ and $\mathbf{P}_j(\gamma)'\mathbf{P}_j(\gamma)$ have full rank for $\gamma \in \Gamma$ and $j = 1, 2$.

(ii) $\mathbf{A}(\gamma_1,\gamma_2)$ has full rank for $\gamma_1, \gamma_2 \in \Gamma$.

Assumption (i) ensures the projection matrix P_{-j} exists for $j = 1, 2$ and (ii) requires $\mathbf{A}(\gamma_1,\gamma_2)$ is positive definite and thus can form inverse matrix.

Now we can discuss the asymptotic properties.

Theorem 3.1 *Under assumptions A and B, as N goes to infinity,*

$$\begin{aligned} \hat{\gamma}_1 &\xrightarrow{p} \gamma_1^0 \quad \textit{and} \quad \hat{\gamma}_2 \xrightarrow{p} \gamma_2^0 \\ \hat{\beta}_1 &\xrightarrow{p} \beta_1^0 \quad \textit{and} \quad \hat{\beta}_2 \xrightarrow{p} \beta_2^0 \\ \hat{\delta}_1 &\xrightarrow{p} \delta_1^0 \quad \textit{and} \quad \hat{\delta}_2 \xrightarrow{p} \delta_2^0 \end{aligned}$$

Proposition 3.1 *Under assumptions A and B, as $N \to \infty$ and T is fixed, $\hat{\gamma}_1 = \gamma_1^0 + O_p(1/N)$ and $\hat{\gamma}_2 = \gamma_2^0 + O_p(1/N)$.*

See Appendix for a proof.

Proposition 3.1 indicates that the estimator of threshold parameters converges very faster than that of coefficients. Therefore the distribution of estimator of coefficients can be formed by treating the estimator of threshold parameters as the true value.

Theorem 3.2 *Under assumptions A and B, as $N \to \infty$ and T is fixed,*

$$\sqrt{N}(\underline{\hat{\theta}} - \underline{\theta}^0) \xrightarrow{d} \mathcal{N}(0, \mathbf{V}(\gamma_1^0, \gamma_2^0))$$

3.3 More generalized case

Indeed we can make model (2.6) even more generalized. Suppose the number of dimensions defined above is $J \geq 2$ and for each dimension $j = 1, 2, \cdots, J$ we have d_j thresholds and the regressor $\mathbf{x}_{jit}$ is a $k_j \times 1$ vector. Therefore by these parameters we can imply there are at most $\sum_{j=1}^{J}(d_j + 1)$ regimes in total[1]. For example, (2.6) is set by letting $J = 2$ and $d_1 = d_2 = 1$, which therefore generates at most 4 regimes. Finally, there also exists situations that $d_j = 0$ which implies there is no threshold in j-th dimension. It can be in the model that only some of units in regressors are subject to threshold effect, for example in Hansen (1999) it is modeled such that only cash flow is subject to threshold effect among all firm characteristics to determine investment spending for next period. Yet the assumptions made above still holds even model (2.6) is more generalized.

In this subsection, we allow there are J dimensions such that J can be larger than 2. Note it is trivial to discuss the case where threshold variables are different among dimensions after we are clear about the case where threshold variable is the same across dimensions, therefore we will only investigate the latter case. We analyze the models in three cases: (i)$d_j = 1$ for all $j = 1, \cdots, J$, (ii) $d_j = 0$ for some j and (iii) $d_j > 1$ for some j.

3.3.1 $d_j = 1$ for all $j = 1, \cdots, J$

Firstly we adjust Assumption A and B accordingly.

Assumption A'

(i) The parameter space Θ is convex and compact.

(ii) For each t, $(\mathbf{x}_{it}, q_{it}, u_{it})$ are i.i.d. across i. For each i, $(\mathbf{x}_{it}, q_{it}, u_{it})$ are strictly stationary, ergodic and ρ-mixing, with ρ-mixing coefficients satisfying $\sum_{m=1}^{\infty} \rho_m^{1/2} < \infty$.

(iii)$\sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{x}_{1it}\mathbf{x}_{2it}' = O\left(N^{1-\alpha}\right)$ where $\alpha \in (0, 1/2]$.

(iv) $E(u_{it}\mathbf{x}_{it-\tau}) = \mathbf{0}$ for $\tau \geq 0$. Also $E(u_{it}q_{it-\tau}) = 0$ for $\tau \geq 0$.

(v) $E|\mathbf{x}_{it}|^4 < \infty$ and $E|\mathbf{x}_{it}u_{it}|^4 < \infty$.

(vi) $E\left(\|\mathbf{x}_{it}u_{it}\|^4 | q_{it} = \gamma\right) \leq C$ and $E(\|\mathbf{x}_{it}\|^4 | q_{it} = \gamma) \leq C$ for some $C < \infty$.

(vii) The probability density function of the threshold variable q_{it}, $f_{it}(\gamma)$ satisfies $0 < f_{it}(\gamma) < \bar{f} < \infty$ for all $\gamma \in \Gamma$.

---

[1] It is possible that different dimension can share some thresholds so there would be fewer regimes. This issue will be resolved by model selection method as shown below

Basically, Assumption A' is the generalization of assumption of A. Also we have a new version of Assumption B. Define γ_j^l as the l-th threshold in the j-th dimension.

Assumption B'

(i)$\mathbf{X}_j$ is weakly dependent on $\mathbf{X}_{j'}$ where $j' \neq j$ and $j' = 1, 2, \cdots, J$.

(ii) $\mathbf{P}'_{-j}\mathbf{P}_{-j}$ has full rank for $\gamma \in \Gamma$; the minimum eigenvalues of

$$\frac{\tilde{\mathbf{X}}_j(\gamma_j^l - \xi, \gamma_j^l)'\tilde{\mathbf{X}}_j(\gamma_j^l - \xi, \gamma_j^l)}{N} \ and \frac{\tilde{\mathbf{X}}_j(\gamma_j^l - \xi, \gamma_j^l)'\tilde{\mathbf{X}}_j(\gamma_j^l, \gamma_j^l + \xi)}{N}$$

are bounded away from zero for all $l = 1, \cdots, m_j$ and $j = 1, \cdots, J$, where ξ is a small positive constant.

Note Assumption B' is indeed generalization of Assumption A1 of Gonzalo and Pitarakis (2002).

Now the size of projection matrix will expand. In particular, the projection matrix $\mathbf{P}_{-1}$ is constructed as

$$\mathbf{P}_{-1} = (\mathbf{P}'_{-1,1}, \cdots, \mathbf{P}'_{-1,N})', \tag{3.9}$$

where

$$\mathbf{P}_{-1,i} = [\mathbf{x}^*_{2i}(q_{\{1\}}), \cdots, \mathbf{x}^*_{2i}(q_{\{99\}}), \mathbf{x}^*_{3i}(q_{\{1\}}), \cdots, \mathbf{x}^*_{3i}(q_{\{i,99\}}), \cdots, \mathbf{x}^*_{Ji}(q_{\{1\}}), \cdots, \mathbf{x}^*_{Ji}(q_{\{99\}})]$$

and

$$\mathbf{x}^*_{ji}(q_{\{l\}}) = (\mathbf{x}^*_{ji1}(q_{\{l\}}), \cdots, \mathbf{x}^*_{jiT}(q_{\{l\}}))'$$

where $\mathbf{x}^*_{jit}\left(q_{\{l\}}\right) = (x^1_{jit}\mathbf{l}\{q_{it} \leq q_{\{l\}}\}, \cdots, (x^{k_j}_{jit}\mathbf{l}\{q_{it} \leq q_{\{l\}}\})'$ for $j = 1, \cdots, N$. Therefore the size of $\mathbf{P}_{-1}$ is $NT \times \left((m+1) \times \sum_{j\neq 1}^J k_j\right)$. Then to avoid the rank deficiency of $\mathbf{P}'_{-1}\mathbf{P}_{-1}$, we need to satisfy

$$NT \geq (m+1) \times \sum_{j\neq 1}^J k_j$$

We can obtain the similar results for $\mathbf{P}_{-j}$ for $j = 2, \cdots, J$. Then with the same way to construct the orthogonal projection matrix, we can estimate all thresholds γ_j.

3.3.2 $d_j > 1$ for some j

When there is more than one threshold for a dimension, we need to test how many thresholds can exist and where they are located after filtering the data by orthogonal projection matrix. When one dimension has more than one threshold our projection estimator is still able to filter out this dimension. To see this, suppose the first regressor has two thresholds, $\gamma_1^{(1)}$ and $\gamma_1^{(2)}$ while the second regressor have one threshold γ_2. Then (3.1) is replaced by

$$\mathbf{y}_i = \mathbf{x}_{1i}\beta_1 + \mathbf{x}_{1i}^* \left(\gamma_1^{(1)}\right) \delta_{11} + \mathbf{x}_{1i} \left(\gamma_1^{(1)}, \gamma_1^{(2)}\right) \delta_{12} + \mathbf{x}_{2i}\beta_2 + \mathbf{x}_{2i}^*(\gamma_2)\delta_2 + \mathbf{u}_i \tag{3.10}$$

for $i = 1, \cdots, N$, where

$$\mathbf{x}_{1i} \left(\gamma_1^{(1)}, \gamma_1^{(2)}\right) = (x_{1i1}(1\{q_{i1} \leq \gamma_1^{(2)}\} - 1\{q_{i1} \leq \gamma_1^{(1)}\}), \cdots, x_{1iT}(1\{q_{iN} \leq \gamma_1^{(2)}\} - 1\{q_{iN} \leq \gamma_1^{(1)}\})',$$

while the remaining parts also have the same definitions as in (3.1). And as we stack the variables, we can form

$$\mathbf{Y} = \mathbf{X}_1\beta_1 + \mathbf{X}_1^* \left(\gamma_1^{(1)}\right) \delta_{11} + \mathbf{X}_1 \left(\gamma_1^{(1)}, \gamma_1^{(2)}\right) \delta_{12} + \mathbf{X}_2\beta_2 + \mathbf{X}_2^*(\gamma_2)\delta_2 + \mathbf{U}$$

Recall the formula of $\mathbf{P}_{-2}$ in (3.7) then it is obvious that $\left[\mathbf{X}_1, \mathbf{X}_1 \left(\gamma_1^{(1)}\right), \mathbf{X}_1 \left(\gamma_1^{(1)}, \gamma_2^{(2)}\right)\right]$ will be on the linear span of matrix $\mathbf{P}_{-2}$ as long as $\left[\mathbf{X}_1 \left(\gamma_1^{(1)}\right), \mathbf{X}_1 \left(\gamma_2^{(2)}\right)\right]$ is included in $\mathbf{P}_{-2}$, namely for each i there are two columns in $\mathbf{P}_{-2,i}$ such that

$$\begin{aligned}
\mathbf{x}_{1i}^* \left(\gamma_1^{(1)}\right) &= \mathbf{x}_{1i}^* \left(q_{\{l\}}\right) \\
\mathbf{x}_{1i}^* \left(\gamma_2^{(2)}\right) &= \mathbf{x}_{2i}^* \left(q_{\{l'\}}\right).
\end{aligned}$$

Therefore $\mathbf{M}_{-2}$ can cancel out all terms that are formed by $\mathbf{x}_{1i}$ regardless how many thresholds exist for $\mathbf{x}_{1i}$. Indeed even when we have no prior information about the number of thresholds for one dimension, the orthogonal projection matrix still manages to filter out information of this dimension.

As we apply projection method to filter unnecessary dimensions, what remains to be done is indeed an estimation problem of multiple thresholds in one dimension. Suppose we know the number of thresholds for each dimension, a sequential estimation approach proposed by Bai (1997) and Hansen (1999) can then be adopted and conducted as following

- After the transformation of orthogonal projection matrix, find γ which can minimize the sum of squared residuals from least regression of $\tilde{\mathbf{Y}}_j$ on $\tilde{\mathbf{X}}_j^*(\gamma)$ and $\tilde{\mathbf{X}}_j^+(\gamma)$. Denote it as $\hat{\gamma}_j^{(1)}$.
- Then fix $\hat{\gamma}_j^{(1)}$, find γ which can minimize the sum of squared residuals from least regression of $\tilde{\mathbf{Y}}_j$ on $\tilde{\mathbf{X}}_j^*(\gamma_j^{[1]})$, $\tilde{\mathbf{X}}_j(\gamma_j^{[1]}, \gamma_j^{[2]})$, and $\tilde{\mathbf{X}}_j^+(\gamma_j^{[2]})$, where the number in brackets means the ordering of smallest value in $(\gamma, \hat{\gamma}_j^{(1)})$. Denote it as $\hat{\gamma}_j^{(2)}$.

- Then find γ which can minimize the sum of squared residuals from least squares of $\tilde{\mathbf{Y}}_j$ on $\tilde{\mathbf{X}}_j^*(\gamma_j^{[1]})$, $\tilde{\mathbf{X}}_j(\gamma_j^{[1]}, \gamma_j^{[2]})$, $\tilde{\mathbf{X}}_j(\gamma_j^{[2]}, \gamma_j^{[3]})$ and $\tilde{\mathbf{X}}_j^+(\gamma_j^{[3]})$ where the number in brackets means the ordering of smallest value in $(\gamma, \hat{\gamma}_j^{(1)}, \hat{\gamma}_j^{(2)})$. Denote it as $\hat{\gamma}_j^{(3)}$.

- Carry on the same procedure for a new threshold until the number of thresholds is reached

Though the above estimation is derived from misspecified model, we can follow Gonzalo and Pitarakis (2002) to verify the estimators still converge to the corresponding true thresholds. To show it, define $\mathbf{x}_{it} = (\mathbf{x}_{1it}, \mathbf{x}_{2it}, \cdots, \mathbf{x}_{Jit})$. Each step in the above procedure will estimate one threshold. In particular, the procedure for the first step is the same as that for single threshold case, therefore let

$$R_{j,\infty}(\gamma) = \lim_{N\to\infty} S_N^j - S_N^j(\gamma)$$

where S_N^j is the sum of squared errors of least squares estimation of

$$\tilde{\mathbf{Y}}_j = \tilde{\mathbf{X}}_j\beta + \mathbf{e}$$

while $S_N^j(\gamma)$ is defined same as for single threshold. Now suppose we have conducted $h-1$ steps and therefore detected $h-1$ thresholds for the jth dimension, thus splitting this dimension into h intervals. Therefore the additional threshold should lie in one of those intervals. Then for each interval, we can construct the model of one more threshold as

$$\mathbf{Q}_{jl}\tilde{\mathbf{Y}}_j = \tilde{\mathbf{Z}}_{jl}^*(\gamma)\beta_{jl1} + \tilde{\mathbf{Z}}_{jl}^+(\gamma)\beta_{jl2} + \mathbf{e}$$

where

$$\mathbf{Q}_{jl} = I_{NT} - \sum_{s=1,s\neq l}^{h} \tilde{\mathbf{Z}}_{js}\left(\tilde{\mathbf{Z}}_{js}'\tilde{\mathbf{Z}}_{js}\right)^{-1}\tilde{\mathbf{Z}}_{js}'$$

where $\tilde{\mathbf{Z}}_{js} = \tilde{\mathbf{X}}_j\left(\hat{\gamma}_j^{[s-1]}, \hat{\gamma}_j^{[s]}\right)$, $\tilde{\mathbf{Z}}_{jl}^*(\gamma) = \tilde{\mathbf{X}}_j\left(\hat{\gamma}_j^{[l]}, \gamma\right)$ and $\tilde{\mathbf{Z}}_{jl}^+(\gamma) = \tilde{\mathbf{X}}_j\left(\gamma, \hat{\gamma}_j^{[l+1]}\right)$ where the number in brackets means the ordering of smallest value in $(\underline{\Gamma}_j, \hat{\gamma}_j^{(1)}, \cdots, \hat{\gamma}_j^{(h-1)}, \bar{\Gamma}_j)$ such that $\hat{\gamma}_j^{[0]} = \underline{\Gamma}_j$ and $\hat{\gamma}_j^{[h]} = \bar{\Gamma}_j$. This design indicates the next threshold may exist between two closest thresholds that have been estimated. Then define $S_N^{jl}(\gamma)$ as the sum of residuals from least squares estimation. Meanwhile we can construct the model of no more threshold as

$$\mathbf{Q}_{jl}\tilde{\mathbf{Y}}_j = \tilde{\mathbf{X}}_{jl}\beta_{jl} + \mathbf{e}$$

and define S_N^{jl} as the sum of squared residuals from the least squares estimation. Therefore, criterion function for the h-th step is defined as following

$$R_j\left(\gamma|\hat{\gamma}_j^{[1]}, \hat{\gamma}_j^{[2]}, \cdots, \hat{\gamma}_j^{[h-1]}\right) = \sum_{l=1}^{h} R_{jl}\left(\gamma|\hat{\gamma}_j^{[1]}, \hat{\gamma}_j^{[2]}, \cdots, \hat{\gamma}_j^{[h-1]}\right) 1\{\hat{\gamma}_j^{[l-1]} < \gamma < \hat{\gamma}_j^{[l]}\}$$

where

$$R_{jl}\left(\gamma|\hat{\gamma}_j^{[1]},\hat{\gamma}_j^{[2]},\cdots,\hat{\gamma}_j^{[h-1]}\right)=S_N^{jl}-S_N^{jl}(\gamma).$$

Now also define

$$R_{j,\infty}\left(\gamma|\gamma_j^1,\gamma_j^2,\cdots,\gamma_j^{h-1}\right)=\lim_{N\to\infty}R_j\left(\gamma|\gamma_j^1,\gamma_j^2,\cdots,\gamma_j^{h-1}\right)$$

We need some more assumptions.

Assumption C

(i) There exists a single threshold parameter $\gamma_j^{[1]}\in\{\gamma_j^1,\gamma_j^2,\cdots,\gamma_j^m\}$ such that $R_{j,\infty}(\gamma_j^{[1]})>R_{j,\infty}(\gamma_j^k)\ \forall\gamma_j^k\neq\gamma_j^{[1]}$ and $k=1,2,\cdots,m$.

(ii) For $j=1,\cdots,J$, there exists a configuration $\left(\gamma_j^1,\gamma_j^2,\cdots,\gamma_j^m\right)$ of the m true threshold parameters such that

$$R_{j,\infty}\left(\gamma_j^h|\gamma_j^1,\gamma_j^2,\cdots,\gamma_j^{h-1}\right)>R_{j,\infty}\left(\gamma|\gamma_j^1,\gamma_j^2,\cdots,\gamma_j^{h-1}\right)$$

$\forall\gamma\in\{\gamma_j^{h+1},\gamma_j^{h+2},\cdots,\gamma_j^m\}$ and $h=1,\cdots,m$.

Proposition 3.2 *Under assumptions A', B' and C, as N goes to infinity and T is fixed, $\hat{\gamma}_j^h=\gamma_j^h+o_p(1)$ and $N|\hat{\gamma}_j^h-\gamma_j^h|=O_p(1)$ for $h=1,\cdots,m$ and $j=1,\cdots,J$.*

Proposition 3.2 is easy to verify as extensions of Proposition 2.4 of Gonzalo and Pitarakis (2002). In particular, the main difference is here the Assumption C states properties for models that have filtered out redundant dimensions while Gonzalo and Pitarakis (2002) focus on just one dimension and therefore make the corresponding assumption.

Since the sequential estimation is conducted from misspecified model, Bai (1997) proposes a refinement estimation procedure. Suppose the jth dimension has only two thresholds, then one can follow our above procedure to obtain corresponding estimators. Next one can fix the second threshold, and estimate the first threshold by grid search. In this way, the estimator of first threshold can be more accurate. As a matter of fact, more efficiency can be achieved by keeping refining $\hat{\gamma}_j^{[1]}$ and $\hat{\gamma}_j^{[2]}$.

3.3.3 Unknown d_j

The number of thresholds in one dimension is generally unknown. One special case is that there is no threshold in some dimension. To test if there is threshold in one dimension, the null hypothesis is usually set such as the coefficient would not change even if regressors are multiplied by indicator function of threshold. Therefore the threshold effect will be taken into model only under alternative hypothesis, and that caused the inferences problem of test statistic, which is well known as Davis' problem. When there is only one dimension, Hansen (1996) has proposed the test for this case where a bootstrapping method is used to calculate the P-value. Alternatively, Gonzalo and Pitarakis (2002) treat it as a model selection problem to see if there is a need to include the term with indicator function. After we filter the data by orthogonal projection matrix, what is left is only one dimension. Thus here we are able to simply adopt the method in Gonzalo and Pitarakis (2002). More specifically, this model selection approach can work for multiple thresholds. Namely we can construct the following information criterion:

$$IC_j(\gamma_j^1, \cdots, \gamma_j^m) = \log S_N^j(\gamma_j^1, \cdots, \gamma_j^m) + \frac{c_N}{N}[k_j(m+1)] \tag{3.11}$$

where $S_N^j(\gamma_j^1, \cdots, \gamma_j^m)$ is the sum of squared errors of regression $\tilde{\mathbf{Y}}_j$ on $\tilde{\mathbf{X}}_j^*(\gamma_j^1), \tilde{\mathbf{X}}_j(\gamma_j^1, \gamma_j^2), \cdots, \tilde{\mathbf{X}}_j^+(\gamma_j^m)$ and c_N is a deterministic function of sample size such that $c_N \to \infty$ and $c_N/N \to 0$ as $N \to \infty$. Then define $IC_j(0) = \log S_N + (c_N/N)k_j$, where S_N^j is the sum of squared error of regression $\tilde{\mathbf{Y}}_j$ on $\tilde{\mathbf{X}}_j$. The model selection based estimator of the number of unknown threshold parameters can then be defined as

$$\hat{m} = \arg \max_{0 \le m \le M_{max}} Q_j(m) \tag{3.12}$$

where

$$Q_j(m) = IC_j(0) - \min_{\gamma_j^1, \cdots, \gamma_j^m} IC_j(\gamma_j^1, \cdots, \gamma_j^m) \tag{3.13}$$

In particular, we can compare $Q_j(0)$ and $Q_j(1)$ to detect whether there is a threshold in the jth dimension. Then we have the following proposition.

Proposition 3.3 *Under assumptions A', B' and C, as N goes to infinity, if $c_N \to \infty$ and $c_N/N \to 0$, $P(\hat{m}_j = m_j) \to 1$.*

3.4 Remarks

Remark 3.1 *It is obvious that we can estimate (3.1) by an alternative. Firstly for an arbitrary $\gamma_1 \in \Gamma$ find its corresponding γ_2 that can minimize the criterion function (3.2). This implies $\hat{\gamma}_2$ is indeed a function of γ_1. Then we can generate the criterion function as a function of γ_1, and thus compute its estimator by finding γ_1 that minimizes the criterion function. This method, however, will add computing cost as in our case there can be $393 \times 393 = 15449$ combinations of (γ_1, γ_2). And it is expected to induce bias of estimator of threshold parameters due to the two-step procedure.*

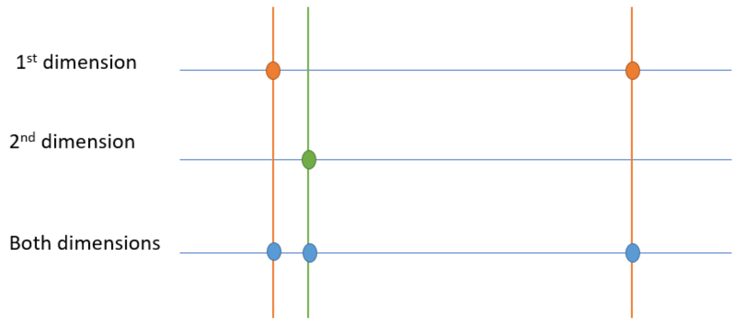

Figure 1: Interactive sparsity problem

On the other hand, when the situation is more complicated, for example there are more than one threshold for one regressor, there can be more computation cost. Compared to that, our method won't include unnecessary computation cost in multiple thresholds case. This is the advantage of our estimator over previous studies (e.g. regression trees) which usually need sacrificing sample size as the tipping point needs to be found from subsample generated from last layer.

Remark 3.2 *When the threshold variable is the same for all dimensions, one is able to model by one dimension with multiple thresholds. For example, (3.10) can be rewritten as a model with two thresholds. Suppose* $\gamma_1 \leq \gamma_2$*,*

$$\mathbf{Y} = \mathbf{X}^*(\gamma_1)\rho_1 + \mathbf{X}(\gamma_1, \gamma_2)\rho_2 + \mathbf{X}^+(\gamma_2)\rho_3 + \mathbf{U} \tag{3.14}$$

where $\rho_1 = (\beta_1, \beta_2)$*,* $\rho_2 = (\beta_1 + \delta_1, \beta_2)$ *and* $\rho_3 = (\beta_1 + \delta_1, \beta_2 + \delta_2)$*. One can apply the sequential estimation by Gonzalo and Pitarakis (2002) to find the number of thresholds and then derive the estimation of the model. However, considering each regime requires sufficiently many observations to be estimated, the threshold estimation may suffer in efficiency when there are two few observations in one regime. Even we make the assumption that the regimes in each dimension have enough observations, two thresholds from two different dimensions may still be too close which makes the associated regime narrow and lack observations. This fact is depicted by Figure 1 where the three horizontal lines are all possible values of threshold variables. The upper line corresponds to the 1st dimension of (3.14), and the middle line the 2nd dimension. The bottom line is for all dimensions. From this figure, we can infer the the first threshold of 1st dimension is very close to the threshold of 2nd dimension. So if we estimate all thresholds by just one dimension, the subsample formed by value of threshold variable between the two thresholds will be too small and it will deteriorate the estimation efficiency. Indeed when we* *have multiple thresholds in different dimensions, it is not surprising such phenomenon occurs frequently. This is equivalent to the sparse interaction problem in multi-dimensional clustering literature. Therefore sequential estimation by Gonzalo and Pitarakis (2002) are unable to assure the efficiency. modeling with dimensional homogeneity is not sufficiently efficient. Compared to that, our orthogonal projection method is not subject to the sparse interaction problem as it analyzes each dimension separately. More than that, when the threshold variable is different across dimensions, one dimension is not able to account for the model.*

Remark 3.3 *Indeed model (2.6) can be furthermore generalized. Instead of restricting transition variable for regressors is just one variable q_i, we can set q_{i1} as the transition variable for $\mathbf{x}_{i1}$ and q_{i2} for $\mathbf{x}_{i2}$. It is obvious that the above estimation procedure can still work for this model. Indeed one example is when $x_{i1} = q_{i1}$ and $x_{i2} = q_{i2}$.*

Remark 3.4 *When the regressors are endogenous, notice there is no need to adjust step 1 and 3. However, we need to modify step 2, 4 and 5. Suppose we have valid instrumental variable $\mathbf{z}_i$, then the 2SLS proposed by Caner and Hansen (2004) can be used to estimate γ_1 and γ_2, namely we need to replace regressors by the predicted value from regressions on instrumental variables and then apply them to step 2, 4 and 5.*

Remark 3.5 *Now we can consider the panel data model with both fixed effect and dimension heterogeneity of thresholds, so the model is*

$$\begin{aligned} y_{it} &= \alpha_{i1} + \alpha_{i2} 1\{q_{it} \le \gamma_\alpha\} + \beta_1' \mathbf{x}_{1it} + \delta_1' \mathbf{x}_{1it} 1\{q_{it} \le \gamma_1\} \\ &\quad + \beta_2' \mathbf{x}_{2it} + \delta_2' \mathbf{x}_{2it} 1\{q_{it} \le \gamma_2\} + u_{it} \end{aligned} \tag{3.15}$$

We can assume the fixed effect is modeled as a function of regressors by Mundlak specification, namely

$$\begin{aligned} \alpha_{i1} &= E\left(\alpha_{i1} | \mathbf{x}_{i1}, \cdots, \mathbf{x}_{iT}\right) + v_{i1} \\ \alpha_{i2} &= E\left(\alpha_{i2} | \mathbf{x}_{i1}, \cdots, \mathbf{x}_{iT}\right) + v_{i2} \end{aligned}$$

In particular, we can model the fixed effect by a linear function of time series average,

$$\begin{aligned} \alpha_{i1} &= \rho_1 \bar{\mathbf{x}}_i + v_{i1} \\ \alpha_{i2} &= \rho_2 \bar{\mathbf{x}}_i + v_{i2} \end{aligned}$$

where $\bar{\mathbf{x}}_i = T^{-1} \sum_{t=1}^{T} \mathbf{x}_{it}$ and therefore (3.15) can be rewritten as

$$\begin{aligned} y_{it} &= \rho_1 \bar{\mathbf{x}}_i + \rho_2 \bar{\mathbf{x}}_i 1\{q_{it} \le \gamma_\alpha\} + \beta_1' \mathbf{x}_{1it} + \delta_1' \mathbf{x}_{1it} 1\{q_{it} \le \gamma_1\} + \\ &\quad \beta_2' \mathbf{x}_{2it} + \delta_2' \mathbf{x}_{2it} 1\{q_{it} \le \gamma_2\} + v_{i1} + v_{i2} 1\{q_{it} \le \gamma_\alpha\} + u_{it} \end{aligned}$$

This indeed satisfies the characteristics of a threshold model with three dimensions except the error term is now serially correlated. Alternatively there is no harm for us to allow dimension heterogeneity in the Mundlak specification such as

$$
\begin{aligned}
y_{it} &= \rho_{11}\bar{\mathbf{x}}_{1i} + \rho_{12}\bar{\mathbf{x}}_{1i}1\{q_{it} \leq \gamma_{\alpha 1}\} + \rho_{21}\bar{\mathbf{x}}_{2i} + \rho_{22}\bar{\mathbf{x}}_{2i}1\{q_{it} \leq \gamma_{\alpha 2}\} + \beta_1'\mathbf{x}_{1it} \\
&\quad + \delta_1'\mathbf{x}_{1it}1\{q_{it} \leq \gamma_1\} + \beta_2'\mathbf{x}_{2it} + \delta_2'\mathbf{x}_{2it}1\{q_{it} \leq \gamma_2\} + v_{i1} + v_{i2}1\{q_{it} \leq \gamma_\alpha\} + u_{it}
\end{aligned}
$$

where $\bar{\mathbf{x}}_{ji} = T^{-1}\sum_{t=1}^{T}\mathbf{x}_{jit}$ *for* $j = 1, 2$. *that is we allow the fixed effect to be formed by two additional dimensions. Note even when the model has a dynamic structure, for instance, the lag of dependent variable is also included as a regressor, the above Mundlak specification still holds if we assume the initial observation* y_{i0} *is uncorrelated with the fixed effect, see Bai (2013) and Hsiao and Zhou (2018). If such correlation is not true, then we may need to add the conditions accounting for the initial observation* y_{i0}. *We will leave that for future work.*

4 Monte carlo simulations

In this section we are going to verify the finite sample performance of the orthogonal projection estimator. In particular, we will focus on a panel threshold model with two dimensions, and two sets of experiments are designed, namely A: each dimension has only one threshold, and B: one dimension has two thresholds and the other has one threshold. We will examine the consistency and normality of estimator by experiment A and model selection by experiment B.

4.1 DGP

Consider the following data generating processes for experiment A,

- DGP 1:

$$y_{it} = \beta_1 x_{1it} + \delta_1 x_{1it}1\{q_{it} \leq \gamma_1\} + \beta_2 x_{2it} + \delta_2 x_{2it}1\{q_{it} \leq \gamma_2\} + u_{it}, \tag{4.1}$$

  where $q_{it} \sim_{iid} \mathcal{N}(1,1)$ and the error $u_{it} \sim_{iid} N\left(0, \sigma_{u,i}^2\right)$ where $\sigma_{u,i}^2$ are independent draws from $0.5(1 + 0.5\chi^2(2))$. The two regressors are generated as

$$x_{jit} = \frac{\eta_{jit} + v g_{it}}{\sqrt{1+v^2}}$$

  where $g_{it} \sim_{iid} \mathcal{N}(1,1)$ and η_{jit} follows an AR(1) process as

$$\eta_{jit} = \rho_{ji}\eta_{jit-1} + \varepsilon_{jit}, \tag{4.2}$$

  where ρ_{ji} is i.i.d draws from $U(0, 0.5)$ for $i = 1, \ldots, N$ and $\varepsilon_{jit} \sim_{iid} (0, \sigma_{\varepsilon ji}^2)$ with $\sigma_{\varepsilon ji}^2$ being independent draw from $0.5(1 + 0.5\chi^2(2))$. v is used to control the pairwise correlation between x_{jit}. We will focus on $v = 0.2, 0.5, 0.8$ which corresponds to $corr(x_{1it}, x_{2it})$ with low, middle and high values.

- DGP 2:

$$y_{it} = \beta_1 x_{1it} + \delta_1 x_{1it} 1\{x_{1it} \leq \gamma_1\} + \beta_2 x_{2it} + \delta_2 x_{2it} 1\{x_{2it} \leq \gamma_2\} + u_{it}, \tag{4.3}$$

  where all variables on the right hand side are generated in the same way as DGP 1.

- DGP 3: Same as DGP 1 except x_{1it} is replaced by lag of dependent variable y_{it-1}, namely

$$y_{it} = \beta_1 y_{it-1} + \delta_1 y_{it-1} 1\{q_{it} \leq \gamma_1\} + \beta_2 x_{2it} + \delta_2 x_{2it} 1\{q_{it} \leq \gamma_2\} + u_{it} \tag{4.4}$$

- DGP 4: Same as DGP 3 except q_{it} is replaced by lag of dependent variable, namely

$$y_{it} = \beta_1 y_{it-1} + \delta_1 y_{it-1} 1\{y_{it-1} \leq \gamma_1\} + \beta_2 x_{2it} + \delta_2 x_{2it} 1\{y_{it-1} \leq \gamma_2\} + u_{it}. \tag{4.5}$$

DGP 1 is designed for strictly exogenous threshold variables and the two dimensions share the same threshold variable. In DGP 2 the threshold variable is replaced by regressor for each dimension. DGP 3 is designed based on DGP 1 but including the the lag of dependent variable y_{it-1} and keeping the threshold variable strictly exogenous. Compared to that, DGP 4 furthermore allows the threshold variable to be weakly exogenous by using the lag of dependent variable as the threshold variable. As our main goal is to compare the difference between dimensions, let $\beta_1 = \beta_2 = 0$. Also let $\delta_1 = 0.3$, $\delta_2 = 0.7$ and $\gamma_1 = 1$ and $\gamma_2 = 0.5$. The error u_{it} is generated with heterogeneity such that $u_{it} \sim \mathcal{N}\left(0, \sigma_{ui}^2\right)$ where $\sigma_{ui} \sim 0.5(1 + 0.5\chi^2(2))$. In the appendix we also discuss simulations where errors are allowed to have either serial correlation and cross sectional dependence. It is shown that the properties of estimation of coefficients and threshold parameters hold regardless of three types of errors.

Secondly, we consider the following DGPs for experiment B to show the model selection of threshold works after filtering out other dimensions. We have

- DGP 5:

$$y_{it} = \rho_{11} x_{1it} 1\{q_{it} \leq \gamma_{11}\} + \rho_{12} x_{1it} 1\{\gamma_{11} < q_{it} \leq \gamma_{12}\} + \rho_2 x_{2it} 1\{q_{it} \leq \gamma_2\} + u_{it} \tag{4.6}$$

  where all the variables on the right hand side are generated in the same way as DGP 1.

- DGP 6:

$$y_{it} = \rho_{11} x_{1it} 1\{x_{1it} \leq \gamma_{11}\} + \rho_{12} x_{1it} 1\{\gamma_{11} < x_{1it} \leq \gamma_{12}\} + \rho_2 x_{2it} 1\{x_{2it} \leq \gamma_2\} + u_{it} \tag{4.7}$$

  where all the variables on the right hand side are generated in the same way as DGP 2.

- DGP 7:

$$y_{it} = \rho_{11} y_{it-1} 1\{q_{it} \leq \gamma_{11}\} + \rho_{12} y_{it-1} 1\{\gamma_{11} < q_{it} \leq \gamma_{12}\} + \rho_2 x_{2it} 1\{q_{it} \leq \gamma_2\} + u_{it} \quad (4.8)$$

where all the variables on the right hand side are generated in the same way as DGP 3.

- DGP 8:

$$y_{it} = \rho_{11} y_{it-1} 1\{y_{it-1} \leq \gamma_{11}\} + \rho_{12} y_{it-1} 1\{\gamma_{11} < y_{it-1} \leq \gamma_{12}\} + \rho_2 x_{2it} 1\{y_{it-1} \leq \gamma_2\} + u_{it} \quad (4.9)$$

where all the variables on the right hand side are generated in the same way as DGP 4.

DGP 5-8 are constructed based on DGP 1-4 correspondingly. The difference between these two groups of DGPs is now we allow two thresholds for the first dimension. Let $(\rho_{11}, \rho_{12}, \rho_2) = (0.3, 0.7, 0.5)$ and $(\gamma_{11}, \gamma_{12}, \gamma_2) = (1, 2, 1.5)$. Obviously as in experiment A, now we suppress the coefficients of x_{1it} to be zero when the threshold variable q_{it} is larger than γ_{12} and the coefficient of x_{2it} to be zero when the threshold variable q_{it} is larger than γ_2.

4.2 Simulation results

For each DGP, we do 1000 iterations for samples with $N = 100, 200, 500, 1000$ and $T = 5$. In particular, we generate data with $T = 1005$ and then drop the first 1000 time periods. The simulation results of experiment A are shown in Table 1 and 2 while that of experiment B are shown in Table 3.

4.2.1 Experiment A

For coefficients δ_1 and δ_2, we provide mean estimates, bias, root mean squared error, IQR (inter-quantile range, $75\% - 25\%$ percentile) and $95\% - percentage$ coverage probability (which is equivalent to $1 - size$). While for the threshold parameters γ_1 and γ_2, we provide the same statistics as coefficient estimation but not the coverage probability. Many previous studies have shown that the estimator of threshold parameter has a nonregular distribution, so here we won't discuss it and thus the coverage probability.

The results in Table 1 and 2 indicate estimators of both coefficients and threshold parameters achieve consistency for all DGPS. It is also observed that coverage rate of the coefficient estimation is around 95%. In particular for DGP1 and DGP2 though we vary the correlation between regressors, the simulation results are not affected significantly.

4.2.2 Experiment B

We consider five information criteria, namely BIC, AIC, HQ, BIC2 and BIC3, of which accordingly $\lambda_N = \log N$, $\lambda_N = 2$, $\lambda_N = 2\log\log N$, $\lambda_N = 2\log N$, $\lambda_N = 3\log N$. We select the number of threshold from a pool of $(0, 1, 2, 3)$ such that we also test if the threshold exists and assume the largest number of threshold is up to 3. In Table 3, we report the decision frequencies for true number of thresholds.

The results show five information criteria work well in large sample when regressors share the same exogenous threshold variables, which can be found from the decision frequencies of DGP5. And generally the correlation between regressors does not change the results significantly. However, when we adopt regressor as its corresponding threshold variable, the decision frequencies of 1st dimension by AIC and BIC3 break down as shown by results of DGP6. Also if the threshold variable is weakly exogenous, then the decision frequencies of 1st dimension by BIC, BIC2, BIC3 all break down. Compared to that, HQ penalty function behaves more steadily.

5 Application

To justify the meaning of the dimension heterogeneity in threshold model, here we give two empirical applications. These two applications highlight the importance of dimension heterogeneity in both economics implications and prediction.

5.1 Investment and financial constraint

The influence of financial constraint on investment spending of a firm has been discussed for a long time. In theory of corporate finance, when there is a funding need firm can either finance it from its own assets (internal financing) or borrow money from outside by means of loans, bonds and equities (external financing). However, when firm is under financial constraint, it will be hard to finance externally as the risk of the firm's default is also rising. New investment therefore has to rely more on internal funding, for example the cash in hand. If firm under financial constraint is inclined to use its own cash to fund, it is then spontaneous that cash flow is more important than normal situation to decide the investment spending in the next period. This hypothesis is examined in Hansen (1999) and Gonzalez et al. (2017), where they find two thresholds exist for the debt level such that from the lowest regime to the highest regime of debt level sensitivity of cash flow firstly rises and then drops. In general both papers find evidence that is against the hypothesis above, since the highest regime indeed has the lowest sensitivity

			δ_1					δ_2				
	v	N	mean	bias	rmse	IQR	CP	mean	bias	rmse	IQR	CP
DGP1	0.2	100	0.334	0.034	0.149	0.120	0.984	0.707	0.007	0.101	0.135	0.954
		200	0.316	0.016	0.062	0.085	0.946	0.698	-0.002	0.069	0.090	0.944
		500	0.302	0.002	0.040	0.053	0.945	0.697	-0.003	0.042	0.054	0.948
		1000	0.301	0.001	0.028	0.038	0.951	0.698	-0.003	0.032	0.043	0.957
	0.5	100	0.328	0.028	0.155	0.126	0.986	0.703	0.003	0.103	0.138	0.949
		200	0.315	0.015	0.063	0.085	0.945	0.696	-0.004	0.071	0.096	0.942
		500	0.302	0.002	0.040	0.053	0.941	0.697	-0.003	0.042	0.053	0.941
		1000	0.301	0.001	0.029	0.039	0.948	0.698	-0.003	0.032	0.044	0.953
	0.8	100	0.322	0.022	0.181	0.134	0.979	0.698	-0.002	0.110	0.148	0.944
		200	0.314	0.014	0.067	0.087	0.952	0.693	-0.007	0.075	0.101	0.974
		500	0.301	0.001	0.041	0.054	0.949	0.697	-0.003	0.043	0.056	0.946
		1000	0.301	0.001	0.030	0.039	0.951	0.697	-0.003	0.033	0.046	0.956
DGP2	0.2	100	0.348	0.048	0.134	0.137	0.935	0.697	-0.003	0.105	0.146	0.958
		200	0.318	0.018	0.070	0.091	0.934	0.699	-0.001	0.071	0.094	0.955
		500	0.303	0.003	0.043	0.057	0.944	0.702	0.002	0.044	0.057	0.948
		1000	0.302	0.002	0.030	0.042	0.941	0.699	-0.001	0.031	0.041	0.943
	0.5	100	0.344	0.044	0.145	0.144	0.945	0.697	-0.003	0.108	0.144	0.943
		200	0.317	0.017	0.072	0.092	0.948	0.699	-0.001	0.071	0.095	0.948
		500	0.304	0.004	0.044	0.058	0.955	0.701	0.001	0.045	0.060	0.949
		1000	0.303	0.003	0.030	0.040	0.947	0.699	-0.001	0.031	0.042	0.949
	0.8	100	0.340	0.040	0.158	0.144	0.945	0.697	-0.003	0.112	0.141	0.947
		200	0.317	0.017	0.073	0.094	0.945	0.698	-0.002	0.074	0.100	0.958
		500	0.304	0.004	0.045	0.061	0.957	0.700	0.000	0.047	0.063	0.948
		1000	0.303	0.003	0.030	0.042	0.953	0.699	-0.001	0.032	0.043	0.953
DGP3		100	0.334	0.034	0.136	0.115	0.980	0.707	0.007	0.100	0.133	0.952
		200	0.314	0.014	0.066	0.083	0.945	0.698	-0.002	0.069	0.090	0.952
		500	0.304	0.004	0.040	0.053	0.947	0.697	-0.003	0.042	0.054	0.951
		1000	0.300	0.000	0.028	0.039	0.952	0.698	-0.002	0.032	0.044	0.956
DGP4		100	0.319	0.019	0.088	0.110	0.943	0.704	0.004	0.097	0.121	0.946
		200	0.310	0.010	0.061	0.083	0.954	0.702	0.002	0.066	0.091	0.950
		500	0.303	0.003	0.038	0.052	0.947	0.698	-0.002	0.046	0.060	0.949
		1000	0.301	0.001	0.027	0.035	0.955	0.697	-0.003	0.033	0.045	0.948

Table 1: Simulation results of coefficients for Experiment A

	v	N	γ_1 mean	γ_1 bias	γ_1 rmse	γ_1 IQR	γ_2 mean	γ_2 bias	γ_2 rmse	γ_2 IQR
DGP1	0.2	100	1.030	0.030	0.499	0.246	0.500	0.000	0.101	0.057
		200	1.013	0.013	0.196	0.113	0.500	0.000	0.040	0.031
		500	1.005	0.005	0.092	0.038	0.500	0.000	0.017	0.019
		1000	1.000	0.000	0.031	0.023	0.500	0.000	0.012	0.016
	0.5	100	1.036	0.036	0.506	0.271	0.502	0.002	0.099	0.057
		200	1.018	0.018	0.192	0.116	0.501	0.001	0.045	0.031
		500	1.004	0.004	0.092	0.040	0.500	0.000	0.017	0.019
		1000	0.999	-0.001	0.031	0.024	0.500	0.000	0.012	0.016
	0.8	100	1.025	0.025	0.555	0.315	0.506	0.006	0.115	0.062
		200	1.014	0.014	0.240	0.136	0.502	0.002	0.048	0.033
		500	1.004	0.004	0.098	0.046	0.500	0.000	0.019	0.020
		1000	0.998	-0.002	0.034	0.026	0.500	0.000	0.012	0.016
DGP2	0.2	100	0.614	-0.386	1.106	0.669	0.375	-0.125	0.523	0.223
		200	0.857	-0.143	0.600	0.210	0.440	-0.060	0.271	0.107
		500	0.988	-0.012	0.164	0.064	0.485	-0.015	0.105	0.035
		1000	0.995	-0.005	0.076	0.040	0.496	-0.005	0.040	0.022
	0.5	100	0.605	-0.395	1.094	0.665	0.367	-0.133	0.522	0.250
		200	0.887	-0.113	0.584	0.180	0.440	-0.069	0.256	0.100
		500	0.977	-0.023	0.188	0.071	0.485	-0.015	0.103	0.035
		1000	0.997	-0.003	0.066	0.040	0.497	-0.003	0.029	0.021
	0.8	100	0.634	-0.366	1.060	0.608	0.370	-0.130	0.527	0.216
		200	0.861	-0.139	0.626	0.224	0.446	-0.054	0.260	0.094
		500	0.972	-0.028	0.219	0.074	0.480	-0.020	0.110	0.036
		1000	0.999	-0.001	0.064	0.042	0.497	-0.003	0.036	0.021
DGP3		100	1.029	0.029	0.466	0.256	0.499	-0.001	0.089	0.057
		200	1.002	0.002	0.194	0.095	0.499	-0.001	0.041	0.029
		500	1.005	0.005	0.086	0.040	0.500	0.000	0.017	0.019
		1000	0.998	-0.002	0.031	0.024	0.500	0.000	0.012	0.016
DGP4		100	0.737	-0.263	0.935	0.457	0.511	0.011	0.123	0.067
		200	0.912	-0.088	0.490	0.192	0.501	0.001	0.043	0.034
		500	0.985	-0.015	0.168	0.072	0.500	0.000	0.018	0.020
		1000	0.998	-0.003	0.061	0.042	0.499	-0.001	0.013	0.018

Table 2: Simulation results of threshold parameters for Experiment A

			1st dimension					2nd dimension				
	v	N	BIC	AIC	HQ	BIC2	BIC3	BIC	AIC	HQ	BIC2	BIC3
DGP5	0.2	100	11%	60%	40%	0%	0%	88%	85%	95%	27%	2%
		200	61%	85%	90%	3%	0%	100%	92%	100%	95%	57%
		500	100%	90%	99%	78%	17%	100%	94%	100%	100%	100%
		1000	100%	84%	99%	100%	97%	100%	93%	100%	100%	100%
	0.5	100	10%	57%	34%	0%	0%	85%	84%	94%	21%	1%
		200	53%	85%	87%	2%	0%	100%	91%	99%	90%	42%
		500	100%	91%	100%	66%	10%	100%	94%	100%	100%	100%
		1000	100%	86%	99%	100%	92%	100%	94%	100%	100%	100%
	0.8	100	4%	50%	25%	0%	0%	76%	85%	92%	14%	1%
		200	4%	50%	25%	0%	0%	100%	91%	100%	78%	22%
		500	98%	91%	99%	45%	2%	100%	95%	100%	100%	100%
		1000	100%	89%	100%	99%	68%	100%	94%	100%	100%	100%
DGP6	0.2	100	2%	40%	14%	0%	0%	4%	96%	97%	19%	1%
		200	21%	84%	63%	0%	0%	100%	96%	100%	93%	45%
		500	92%	78%	96%	23%	1%	100%	92%	100%	100%	100%
		1000	97%	63%	80%	89%	36%	100%	87%	99%	100%	100%
	0.5	100	1%	40%	11%	0%	0%	77%	96%	94%	11%	0%
		200	17%	81%	55%	0%	0%	100%	95%	100%	87%	31%
		500	88%	80%	98%	16%	0%	100%	96%	100%	100%	100%
		1000	98%	65%	82%	86%	24%	100%	94%	100%	100%	100%
	0.8	100	1%	36%	10%	0%	0%	70%	94%	91%	8%	0%
		200	13%	79%	51%	0%	0%	94%	96%	100%	77%	19%
		500	82%	77%	95%	9%	0%	100%	96%	100%	100%	100%
		1000	98%	65%	82%	77%	15%	100%	95%	100%	100%	100%
DGP7		100	86%	76%	91%	29%	4%	88%	86%	95%	25%	3%
		200	100%	82%	98%	93%	55%	100%	92%	99%	92%	45%
		500	99%	71%	94%	100%	100%	100%	93%	100%	100%	100%
		1000	95%	55%	81%	100%	100%	100%	92%	100%	100%	100%
DGP8		100	0%	12%	2%	0%	0%	5%	36%	18%	0%	0%
		200	0%	21%	2%	0%	0%	19%	70%	45%	1%	0%
		500	2%	64%	22%	0%	0%	73%	96%	94%	14%	2%
		1000	17%	78%	69%	0%	0%	97%	95%	100%	56%	13%

Table 3: Simulation results of threshold numbers for Experiment B

to cash flow. However, it is suspicious that their estimation is biased because of neglecting the dimension heterogeneity.

5.1.1 Modeling

We employ the same data set as used in Hansen (1999) and Gonzalez et al. (2017). This data set is a balanced panel of 565 US firms over 15 years from 1973 to 1987. Following Gonzalez et al. (2017), we exclude seven companies with extreme data, and consider a final sample of 558 companies with 7812 company year observations. To avoid the use of potentially persistent series, we normalize variables by the book value of assets. Namely, it is measured by investment to the book value of assets, CF_{it} by cash flow to the book value of assets, Q_{it} by the market value to the book value of assets, and D_{it} by the long-term debt to the book value of assets.

Hansen (1999) models the financial constraint's influence on investment spending as following,

$$\begin{aligned} I_{it} &= \mu_i + \theta_1 Q_{it-1} + \theta_2 Q^2_{it-1} + \theta_3 Q^3_{it-1} + \theta_4 D_{it-1} \\ &\quad + \theta_5 Q_{it-1} D_{it-1} + \beta_1 CF_{it-1} 1\{D_{it-1} \leq \gamma_1\} \\ &\quad + \beta_2 CF_{it-1} 1\{\gamma_1 \leq D_{it-1} \leq \gamma_2\} + \beta_3 CF_{it-1} 1\{\gamma_2 < D_{it-1}\} + e_{it} \end{aligned} \tag{5.1}$$

where μ_i is the fixed effect and D_{it-1} is the threshold variable and thus serves as an indicator of financial condition of a firm. He subjectively sets two dimensions, one consisting of $(\mu_i, Q_{it-1}, Q^2_{it-1}, Q^3_{it-1}, D_{it-1}, Q_{it-1}D_{it-1})$ which has no threshold, the other one consisting of CF_{it-1} with threshold as some values of D_{it-1}. The first dimension not only contains two fundamental characteristics of firm, Q_{it-1} and D_{it-1}, but also introduces non-linear terms in order to reduce the possibility of spurious correlations due to omitted variable bias. Hansen (1999) mentions the two dimensions are adopted because his focus is to see investment's sensitivity to cash flow. Under this framework, he designs a likelihood ratio test and determine the number of thresholds is two. Therefore Hansen (1999) indeed admits dimension heterogeneity but chooses to ignore the effect of other dimensions. Compared to that, Gonzalez et al. (2017) do not allow dimension heterogeneity by simply assuming all regressors can be classified as one dimension.

5.1.2 Estimation

We apply our orthogonal projection estimator to reexamine the modeling procedure of Hansen (1999). In terms of modeling, we have two main differences from Hansen (1999). First of all, we would like to show if there is an intercept term that is subject to any threshold structure, so

Table 4: The number and values of thresholds for each dimension

	Intercept	QD	CF
Number of thresholds	1	2	1
1st threshold	0.0059	0.1842	0.7997
2nd threshold	NA	0.3125	NA
3rd threshold	NA	NA	NA

we just replace the fixed effect by an intercept term[2]. Secondly, we classify the regressors that contains Q_{it} and D_{it} into one dimension denoted as QD. This is because we may have correlation between those regressors, and more examinations are needed to ensure such correlation will not affect our estimation on coefficients of CF_{it}. Table 4 displays the estimated threshold numbers and values. The test results show three dimensions are completely different from each other in threshold structures. The intercept dimension and CF dimension has only one threshold while the QD dimension has two thresholds. In terms of value, thresholds of intercept dimension and QD dimension are small and thus remark there is heterogeneity for financially unconstrained firms. Compared to that, the threshold of CF dimension is high and can be regarded as evidence that financially constrained firms are different from financially unconstrained firms in the sensitivity of cash flow to investment. Also note Hansen (1999) observe two thresholds by neglecting the effect of intercept dimension and QD dimension, while by accounting the two dimensions there is only one threshold in CF dimension.

Next we adopt the threshold estimated above to generate threshold regressors and analyze by least squares estimation. The estimation results are shown in table 5. Note most of the regressors in in intercept and QD dimensions are significant. Despite that, the coefficient of debt level D_{it} is insignificant when the level of D_{it} is higher than the lower threshold. It means the influence of debt on investment spending is not through itself but rather other firm characteristics. This fact can also be verified by the insignificance of coefficients for the variable $Q_{it}D_{it}$. Now for the cash flow dimension, the value of coefficient decreases and becomes insignificant after the debt level is too high. It indicates when a firm's debt level is too high, it will indeed not rely on its own cash flow as generally firms that have high debt are not rich in cash flows.This finding about high debt level regime is consistent with results for previous work such as Hansen (1999) and Gonzalez et al. (2017). Overall, we have shown that modeling investment spending should account for the dimension heterogeneity and our robust examination of this issue indicates the sensitivity of cash flow will not rise when the debt level is high.

[2]Note that our estimator is also able to analyze model with fixed effect if Mundlak specification is accepted/

Table 5: Estimation results of investment spending model

Dimension	Status	Value	t-statistic	P-value
Intercept	$D_{it} \leq 0.0059$	0.038	12.247	0.000
	$D_{it} > 0.0059$	0.050	18.985	0.000
Q_{it}	$D_{it} \leq 0.1842$	0.020	8.131	0.000
	$0.1842 < D_{it} \leq 0.3125$	0.044	4.696	0.000
	$D_{it} > 0.3125$	0.047	9.248	0.000
Q_{it}^2	$D_{it} \leq 0.1842$	-0.002	-4.570	0.000
	$0.1842 < D_{it} \leq 0.3125$	-0.012	-4.281	0.000
	$D_{it} > 0.3125$	-0.007	-3.522	0.000
Q_{it}^3	$D_{it} \leq 0.1842$	0.000	3.456	0.001
	$0.1842 < D_{it} \leq 0.3125$	0.001	2.539	0.011
	$D_{it} > 0.3125$	0.000	1.886	0.059
D_{it}	$D_{it} \leq 0.1842$	0.049	2.303	0.021
	$0.1842 < D_{it} \leq 0.3125$	0.013	0.912	0.362
	$D_{it} > 0.3125$	0.002	0.377	0.706
$Q_{it}D_{it}$	$D_{it} \leq 0.1842$	0.016	1.396	0.163
	$0.1842 < D_{it} \leq 0.3125$	-0.015	-0.459	0.646
	$D_{it} > 0.3125$	-0.003	-0.651	0.515
CF_{it}	$D_{it} \leq 0.7997$	0.059	13.126	0.000
	$D_{it} > 0.7997$	0.015	0.694	0.488

5.2 Credit card default

The second application we consider is the modeling of credit card default. Credit cards are an important consumer credit product banks and credit card companies (hereafter banks) offer to more than 170 million U.S. consumers (CFPB, 2019). They usually provide high returns to banks but at the same time pose high default risks. According to the Federal Reserve (Fed), credit card has been the single largest potential loss generator in bank stress tests with predicted losses between $100-144 billion during 2017-2020 under the Fed severely adverse scenario. Therefore, it is crucial to build models that predict credit card default well.

Credit card default happens when the card holder (borrower) fails to make the minimum payment, i.e., is "in delinquency", for a period of time. To capture early default risk, we define default as more than 30-day delinquency in this study. Credit card default or delinquency is usually predicted by borrower characteristics such as credit score, macroeconomic variables like unemployment rate, and other factors.

A growing literature documents substantial heterogeneity in how consumers use credit cards, for example, transactors that pay off their credit card debt each month vs. revolvers that carry credit card debt from month to month (see, e.g., Gathergood et al. (2019); Agarwal et al. (2015); Keys and Wang (2019)). Industry research has found that the default risks of transactor and revolvers to be systematically different. Note that transactors and revolvers are defined by payment ratios. Therefore, in that regard, payment ratio could be a threshold variable and that the relation between credit card default and unemployment rate and alike variables could be different for accounts that are above and below the payment ratio threshold.

Existing analyses define transactors as those that pay off their credit card balance in full each month. In other words, the payment ratio threshold is set arbitrarily at 100 percent. However, it is not clear if that is the optimal definition of transactor. For example, we can imagine that borrowers who pay off 90 percent of their monthly balance might be very similar to the narrowly defined "transactors" in terms of their default risk. Also, there might be other thresholds that are important in this context. We thus think the modeling of credit card default is an ideal situation to test if the model developed in this essay fits the data and whether it can help us uncover important thresholds in a credit card default risk model.

5.2.1 Modeling

The data we use to estimate and test our threshold model are a representative sample of general purpose consumer credit card accounts in California. We have nearly 10,000 accounts that we follow from June 2009 to June 2017. For each account, we know the characteristics of the account such as the credit limit, the borrower's credit score, and the location (county) where the borrower resides, We also observe the balance, the purchase amount (using the card), the payment amount, and other account information for each month (billing cycle). Therefore, it is longitudinal data, but it forms an unbalanced panel due to the closure or inactivity of some accounts during our study periods.

The modeling question is how to predict future default, e.g., default in the next 12 months, using current information including borrower and account characteristics and macroeconomic variables we observe in the current month. To minimize data overlap, we select data of each June, e.g., June 2009, June 2010, etc.. Therefore, it is nine years of data and the the number of observations for each account varies from one to nine, depending on the duration of the account in our data. Altogether we have about 34,000 observations. Empirical studies of the default risks of credit cards and other debt like mortgages usually treat the longitudinal data as repeated cross sectional data. We follow the same convention to treat those 34,000 observations as independent, even though we allow for clustered error terms, which we will discuss below.

Let i denote individual account and t denote time (i.e., year). The outcome variable we are trying to model is a default event dummy $default_{it}$, which takes the value of 1 if the borrower defaults during any month of the next year, and 0 otherwise, The key explanatory variables include borrower credit score $score_{it}$, which represents the creditworthiness of an individual, the county-level year-over-year percent change in unemployment rate ue_{it}, and the size of the credit limit $limit_{it}$. Credit limit is used as a categorical variable in the model.

The threshold variables is the payment ratio pr_{it}, which is simply the ratio of current month payment amount to cycle beginning balance. We use recent six months' average to smooth out noises in payment ratios.

The model we estimate is

$$\begin{aligned} default_{i,t+1} \quad = \quad & \sum_{s=1}^{S_1} \beta_{1s} \mathrm{1}\{\gamma_{1,s} < pr_{i,t} \leq \gamma_{1,s+1}\} + \sum_{s=1}^{S_2} \beta_{2s} \times score_{i,t} \mathrm{1}\{\gamma_{2,s} < pr_{i,t} \leq \gamma_{2,s+1}\} \\ & + \sum_{s=1}^{S_3} \beta_{3s} \times ue_{i,t} \mathrm{1}\{\gamma_{3,s} < pr_{i,t} \leq \gamma_{3,s+1}\} + \sum_{s=1}^{4} \beta_{4s} \mathrm{I}\{limit_{i,t} = s\} + u_{i,t}. \quad (5.2) \end{aligned}$$

In this model, we consider thresholds for the credit score variable, the unemployment rate variable and the intercept. The economic meaning of the model is that, among the credit card borrowers, there exists different risk groups in that some groups are more or less sensitive than others to the risk factors (dimensions) such as change in unemployment rate when it comes to default decisions. The different risk groups are identified with the payment ratio variable. In other words, there exist payment ratio thresholds that separate different risk groups. Those thresholds will be estimated along with other model parameters that govern the relation between default risk and the risk factors. We start with the unconstrained model where the thresholds of different dimensions are allowed to vary. We also allow clustered error terms at the county-level to account for possibly correlations of the error terms across different individuals/observations.

Since the size of the data is quite large, when applying the orthogonal projection to filter out dimensions, it would be computationally unrealistic to construct a $NT \times NT$ matrix. Instead we can replace it with multiple small projection matrix. First, we classify all units into G groups where for each group we denote the number of observations as $\#_g$ for $g = 1, \cdots, G$. Then consider a possible threshold γ for dimension j, we can build the projection matrix $\mathbf{P}_{-j,g}$ by using only data in group g and the definition in (3.9). It will generate the orthogonal projection matrix and filter information, namely

$$\mathbf{M}_{j,g} \quad = \quad I_{(nt)_g} - \mathbf{P}_{-j,g} \left(\mathbf{P}'_{-j,g} \mathbf{P}_{-j,g} \right)^{-1} \mathbf{P}'_{-j,g}, \tag{5.3}$$

which can then be used to filter out data in group g. Denote $\mathbf{Y}_{j1}$ and $\tilde{\mathbf{X}}_{j1}$ as the dependent variable and the jth dimension of regressors in group g, we can get $(\tilde{\mathbf{Y}}_{jg}, \tilde{\mathbf{X}}_{jg}, \tilde{\mathbf{X}}_{jg}(\gamma))$ as the following:

$$\begin{aligned} \tilde{\mathbf{Y}}_{j1} &= \mathbf{M}_{j,g} \mathbf{Y}_{j1} \\ \tilde{\mathbf{X}}_{j1} &= \mathbf{M}_{j,g} \mathbf{X}_{j1} \\ \tilde{\mathbf{X}}_{j1}(\gamma) &= \mathbf{M}_{j,g} \mathbf{X}_{j1}(\gamma). \end{aligned}$$

Note though it is not required that $\#_g$ holds the same for all g, we still need to restrict $\#_g$ to be sufficiently large so that the invertible matrix in (5.3) is of full rank.

After filtering for all groups, we can combine the filtered data in each group to get

$$\begin{aligned} \tilde{\mathbf{Y}}_{j} &= \left(\tilde{\mathbf{Y}}'_{j1}, \cdots, \tilde{\mathbf{Y}}'_{jG} \right)' \\ \tilde{\mathbf{X}}_{j} &= \left(\tilde{\mathbf{X}}'_{j1}, \cdots, \tilde{\mathbf{X}}'_{jG} \right)' \\ \tilde{\mathbf{X}}_{j}(\gamma) &= \left(\tilde{\mathbf{X}}'_{j1}(\gamma), \cdots, \tilde{\mathbf{X}}'_{jG}(\gamma) \right)'. \end{aligned}$$

Now we can use $(\tilde{\mathbf{Y}}_j, \tilde{\mathbf{X}}_j, \tilde{\mathbf{X}}_j(\gamma))$ to do least squares estimation, and then construct the sum of squares as a function of γ. Note that this transformation is valid here because we assume weak dependence between dimensions so that the orthogonal projection matrix of one dimension will not change the basic properties of another dimension.

5.2.2 Estimation

We set the maximum number of thresholds to be 3 for each dimension, and then apply the orthogonal projection method in estimating thresholds for each dimension. Considering the performance of different model selection penalty functions in Monte Carlo simulations, we choose HQ that guarantee steadiness in most cases. The threshold estimation results are shown in Table 6. We find there are three thresholds for both the intercept and the credit score variable. In particular, they share two thresholds of relatively small values. The third thresholds in both dimensions are very close.

These threshold results are economically meaningful in that the thresholds of the credit score variable separate borrowers who are transactors and those who are revolvers. Interestingly, the payment ratio threshold for transactors is at about 60 percent, much lower than the usually used 100 percent threshold that is used to define transactors. In addition, the first threshold in payment ratio is at 3 percent, which is likely for minimum payment payers vs. other revolvers. Gathergood et al. (2019) estimates that 28 percent of individuals with exactly two credit cards are minimum payers. Keys and Wang (2019) finds that 29 percent of accounts pay within $50 of the minimum payment, We can imagine that those minimum payers have very different default risks than other borrowers. We also see that there is another threshold at a payment ratio of about 7 percent, which indicates that among revolvers there are two distinct groups that differ in default behavior. Altogether, the credit score and intercept threshold results show that, among credit card borrowers that we study, there are potentially four distinct risk groups: the minimum payers, the group 1 revolvers, the group 2 revolvers, and the transactors.

Our results show only one payment ratio threshold for the unemployment rate variable at 31 percent. In other words, when it comes to default risk sensitivity to change in unemployment rate, there are two distinct risk groups among the credit card borrowers we study. The value of the threshold is different from those of the credit score variable. This difference suggests that if we modelled the data by applying the same thresholds to all explanatory variables we would end up with a misspecified model.

Combining the credit score threshold results and the unemployment rate threshold results, it seems that the group 2 revolvers can be further split into two risk groups due to their different sensitivities to unemployment rate change. For ease of notation, we label all the different risk groups we have identified as: minimum payers, type I revolvers, type II revolvers, type III revolvers, and transactors.

Table 8 shows the coefficient estimates of our threshold model, with the thresholds we just discussed. Starting with the intercept, we see that minimum payers show the highest default risk, followed by type I revolvers. All else equal, the default risk difference between these two groups is almost 25 percentage points. Consider that the average default rate of our sample is 7 percent, this difference is huge. Type II and type III revolvers are about 16 percentage points lower in default risk, compared to type I revolvers. Finally, transactors are the lowest default

Table 6: The number and values of thresholds for each dimension

	$Intercept$	$score_{it}$	ue_{it}
Number of thresholds	3	3	1
1st threshold	0.0299	0.0299	0.3085
2nd threshold	0.0697	0.0697	NA
3rd threshold	0.5871	0.6070	NA

risk group, which is not surprising.

For credit score, the coefficients of all the risk segments are negative, which is conforming to conventional wisdom as higher credit score borrowers are less likely to fall into dedfault, ceteris paribus. What is interesting is that the slopes are significantly different among minimum payers, revolvers and transactors. For minimum payers, a 20 points difference in credit score can lead to 4 percentage points difference in default rate, but for transactors it is less than 1 percentage point. There is also a noticeable difference between type I and type II/III revolvers.

As we discussed previously, borrowers can be split into two groups according to their default risk sensitivity to change in unemployment. Results in Table 8 shows that for lower payment ratio borrowers, including minimum payers, type I and type II revolvers, a increase in unemployment rate can lead to significant increase in default risk. However, type III revolvers and transactors are not sensitive to change in unemployment rate, as the zero coefficient shows. These unemployment results are intuitive as as deterioration in the economy will mostly affect those who are living hand-to-mouth.

Finally, the coefficients of credit limit buckets are as expected – those with smaller than or equal to $500 credit limit have higher default risk, everything else equal.

5.2.3 Forecasting power

We compare the forecasting power of our threshold model and that of alternative models using out-of-sample tests. The forecast involves using the estimated models and current information of the accounts to predict the probabilities of default for the next year. The data we use in the forecast are a different sample of credit card accounts in California than the one we use in our model estimation. Recall that our estimation sample ends in June 2017. Our forecasting sample include data from September and December of 2017 and March of 2018. The sample contains over 5,000 accounts with about 15,000 observations. These accounts are tracked through March 2019, so we know the actual performance of these accounts. The two benchmark models that

Table 7: Estimation of credit card default model

Dimension	Status	value	t-statistic	P-value
Intercept	$pr_{it} \leq 0.0299$	0.736	14.042	0.000
	$0.0299 < pr_{it} \leq 0.0697$	0.488	20.915	0.000
	$0.0697 < pr_{it} \leq 0.5871$	0.329	20.410	0.000
	$pr_{it} > 0.5871$	0.198	14.319	0.000
$score_{it}/100$	$pr_{it} < 0.0299$	-0.198	-12.331	0.000
	$0.0299 < pr_{it} \leq 0.0697$	-0.126	-17.063	0.000
	$0.0697 < pr_{it} \leq 0.6070$	-0.079	-17.618	0.000
	$pr_{it} > 0.6070$	-0.045	-13.344	0.000
ue_{it}	$pr_{it} \leq 0.3085$	0.034	5.895	0.000
	$pr_{it} > 0.3085$	0.000	0.120	0.905
$\mathrm{l}\{limit_{it} = 1\}$		0.034	3.923	0
$\mathrm{l}\{limit_{it} = 2\}$		-0.013	-2.455	0.014
$\mathrm{l}\{limit_{it} = 3\}$		-0.005	-1.470	0.142
$\mathrm{l}\{limit_{it} = 4\}$		0.004	1.006	0.314

we compare our threshold model to are:

$$default_{i,t+1} = \beta_{11} + \beta_{21} \times score_{i,t} + \beta_{31} \times ue_{i,t} + \sum_{s=1}^{4} \beta_{4s} \mathrm{I}\{limit_{i,t} = s\} + u_{i,t}, \quad (5.4)$$

$$\begin{aligned} default_{i,t+1} &= (\beta_{11} + \beta_{21} \times score_{i,t} + \beta_{31} \times ue_{i,t})\,\mathrm{l}\{pr_{i,t} \leq 1\} \\ &\quad + (\beta_{12} + \beta_{22} \times score_{i,t} + \beta_{32} \times ue_{i,t})\,\mathrm{l}\{pr_{i,t} > 1\} \\ &\quad + \sum_{s=1}^{4} \beta_{4s} \mathrm{I}\{limit_{i,t} = s\} + u_{i,t}. \quad (5.5) \end{aligned}$$

Model (5.4) is a linear model which ignores the threshold effect while (5.5) is a threshold model that neglects the dimension heterogeneity and thus use the same threshold for different explanatory variables. The threshold is also set arbitrarily – we adopt the 100 percent payment ratio widely used in the industry. We use root mean squared error (RMSE) and mean absolute

	no threshold	given threshold	estimated threshold
RMSE	0.2127	0.2122	0.2119
Ratio 1	1	0.9976	0.9964
MAE	0.1081	0.1069	0.1037
Ratio 2	1	0.9888	0.9595

Table 8: Forecasting performance

error (MAE) as our measures of forecasting power, namely

$$
\begin{aligned}
RMSE &= (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(default_{it} - \widehat{default}_{it} \right)^2, and \\
MAE &= (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \left| default_{it} - \widehat{default}_{it} \right|.
\end{aligned}
$$

Results are shown in table 8. As we see, both metrics show that our threshold model is superior in out-of-sample (and out-of-time) forecast. Given that we are running linear probability models, it is not surprising that the RMSE differences of these alternative models are small. However, the MAE metric shows that the difference is sizable between the performance of our threshold model and that of the two benchmark models.

6 Conclusion

In this paper, we study the dimension heterogeneity in threshold model, which indicates the situation where thresholds can be different for different regressors. We provide the projection estimator which can work efficiently compared to traditional threshold models especially when the model is complicated in number of thresholds. By applying our method to credit card default issue, we are able to show two determinant variables credit score and unemployment rate have different thresholds. Involving the heterogeneous thresholds in the model help increase out of sample forecasting power for default case.

In the future it is interesting to extend our method to nonlinear models such as probit and logit model. Another extension is about more complicated structure of error term. For instance, the recently popular research in interactive fixed effect can also be applied which allows the error can be cross sectional correlated.

A Appendix: Mathematical Derivations

Proof of Proposition 3.1. Consider model (3.10). by multiplying both sides of equation by $\mathbf{M}_{-1}$, we can get

$$\mathbf{M}_{-1}\mathbf{Y} = \mathbf{M}_{-1}\mathbf{X}_1\beta_1 + \mathbf{M}_{-1}\mathbf{X}_2\beta_2 + \mathbf{M}_{-1}\mathbf{X}_1^*(\gamma_1)\delta_1 + \mathbf{M}_{-1}\mathbf{X}_2^*(\gamma_2)\delta_2 + \mathbf{M}_{-1}\mathbf{U}$$

Since $\mathbf{M}_{-1} = I_{NT} - \mathbf{P}_{-1}\left(\mathbf{P}'_{-1}\mathbf{P}_{-1}\right)^{-1}\mathbf{P}'_{-1}$ and $\mathbf{P}_{-1}$ spans the linear space that contains $\mathbf{X}_1$ and $\mathbf{X}_1^*(\gamma_1)$, we can therefore filter out model (3.10) as

$$\mathbf{M}_{-1}\mathbf{Y} = \mathbf{M}_{-1}\mathbf{X}_2\beta_2 + \mathbf{M}_{-1}\mathbf{X}_2^*(\gamma_2)\delta_2 + \mathbf{M}_{-1}\mathbf{U}$$

which is indeed

$$\tilde{\mathbf{Y}}_1 = \tilde{\mathbf{X}}_1\beta_1 + \tilde{\mathbf{X}}_1^*(\gamma_1)\delta_1 + \tilde{\mathbf{U}}_1$$

where $\tilde{\mathbf{U}}_1 = \mathbf{M}_{-1}\mathbf{U}$. It is obvious that minimizing $S_N^1(\theta)$ can be equivalent to minimizing the following equation

$$\begin{aligned}
\frac{S_N^1(\theta) - \tilde{\mathbf{U}}_1^{0\prime}\tilde{\mathbf{U}}_1^0}{N} &= \frac{1}{N}\tilde{\mathbf{Y}}_1'\left(I_{NT} - \mathbf{P}_1(\gamma)\right)\tilde{\mathbf{Y}}_1 - \frac{1}{N}\tilde{\mathbf{U}}_1^{0\prime}\tilde{\mathbf{U}}_1^0 \\
&= \frac{1}{N}\delta_1^{0\prime}\tilde{\mathbf{X}}_1^*(\gamma_1^0)'\left(I_{NT} - \mathbf{P}_1(\gamma)\right)\tilde{\mathbf{X}}_1^*(\gamma_1^0)\delta_1^0 + \frac{2}{N}\delta_1^{0\prime}\tilde{\mathbf{X}}_1^*(\gamma_1^0)'\tilde{\mathbf{U}}_1^0 \\
&= \delta_1^{0\prime}\mathbf{R}_1(\gamma)\delta_1^0 + o_p(1)
\end{aligned}$$

where $\mathbf{R}_1(\gamma) = \frac{1}{N}\tilde{\mathbf{X}}_1^*(\gamma_1^0)'\left(I_{NT} - \mathbf{P}_1(\gamma)\right)\tilde{\mathbf{X}}_1^*(\gamma_1^0)$, Consider the case $\gamma \in [\gamma_1^0, \bar{\gamma}]$, then

$$\begin{aligned}
\mathbf{R}_1(\gamma) &= \frac{1}{N}\tilde{\mathbf{X}}_1^*(\gamma_1^0)'\tilde{\mathbf{X}}_1^*(\gamma_1^0) - \frac{1}{N}\tilde{\mathbf{X}}_1^*(\gamma_1^0)'\tilde{\mathbf{X}}_1^*(\gamma)\left(\tilde{\mathbf{X}}_1^*(\gamma)'\tilde{\mathbf{X}}_1^*(\gamma)\right)^{-1}\tilde{\mathbf{X}}_1^*(\gamma)'\tilde{\mathbf{X}}_1^*(\gamma_1^0) \\
&= \frac{1}{N}\tilde{\mathbf{X}}_1^*(\gamma_1^0)'\tilde{\mathbf{X}}_1^*(\gamma_1^0) - \frac{\tilde{\mathbf{X}}_1^*(\gamma_1^0)'\tilde{\mathbf{X}}_1^*(\gamma_1^0)}{N}\left(\frac{\tilde{\mathbf{X}}_1(\gamma)'\tilde{\mathbf{X}}_1^*(\gamma)}{N}\right)^{-1}\frac{\tilde{\mathbf{X}}_1^*(\gamma_1^0)'\tilde{\mathbf{X}}_1^*(\gamma_1^0)}{N} \\
&\xrightarrow{p} \mathbf{D}_1(\gamma_1^0) - \mathbf{D}_1(\gamma_1^0)\mathbf{D}_1(\gamma)^{-1}\mathbf{D}_1(\gamma_1^0)
\end{aligned}$$

And by taking the derivative of $\mathbf{D}_1(\gamma)^{-1}$ with respect to γ, we have

$$\frac{d\mathbf{D}_1(\gamma)^{-1}}{d\gamma} = -\mathbf{D}_1(\gamma)^{-1}\frac{d\mathbf{D}_1(\gamma)}{d\gamma}\mathbf{D}_1(\gamma)^{-1}$$

where

$$\frac{d\mathbf{D}_1(\gamma)}{d\gamma} = \sum_{i=1}^{N}\sum_{t=1}^{T} E\left[\tilde{\mathbf{x}}_{it}^1\tilde{\mathbf{x}}_{it}^{1\prime}|q_{it} = \gamma\right] f_{it}(\gamma)$$

which is positive definite. Therefore for $\gamma \in [\gamma_1^0, \bar{\gamma}]$ the limit of function $\mathbf{R}_1(\gamma)$ is increasing in γ. Similarly we can show the limit of $\mathbf{R}_1(\gamma)$ is decreasing in γ when $\gamma \in [\underline{\gamma}, \gamma_1^0)$. Therefore γ_1^0

minimizes the limit of criterion function $S_1(\theta)$. Since $\hat{\gamma}_1$ is the minimizer of $S_1(\theta)$, by Newey (1994), it indicates $\hat{\gamma}_1 \xrightarrow{p} \gamma_1^0$.

By the symmetry of our model, we can infer the convergence for $\hat{\gamma}_2$ in the same way. Now with the consistency of threshold parameter, we can show the consistency of coefficients straightly as the remaining part is just linear model. ■

Proof of Proposition 3.1. We just show the proof for convergence rate of $\hat{\gamma}_1$ as the procedure is the same for $\hat{\gamma}_2$. The convergence rate of threshold estimator is proved by showing that $S_N^1(\gamma) - S_N^1(\gamma_1^0) \geq 0$ if $\gamma \in \left(\underline{\gamma}, \gamma_0 - \frac{B}{N}\right) \cup \left(\gamma_0 + \frac{B}{N}, \bar{\gamma}\right)$ where B can be any positive constant. This is equivalent to $\gamma \in (\gamma_0 - \frac{B}{N}, \gamma_0 + \frac{B}{N})$ if $S_N^1(\gamma) - S_N^1(\gamma_1^0) < 0$ which is the fact based on our estimator.

To show it holds, suppose $\gamma < \gamma_1^0$, then we can decompose $S_N^1(\gamma) - S_N^1(\gamma_1^0)$ as

$$S_N^1(\gamma) - S_N^1(\gamma_1^0) = (S_N^1(\gamma) - S_N^1(\gamma, \gamma_1^0)) - (S_N^1(\gamma_1^0) - S_N^1(\gamma, \gamma_1^0)) \tag{A.1}$$

where $S_N^1(\gamma, \gamma_1^0)$ is the concentrated sum of squared errors function from the following specification

$$\tilde{\mathbf{Y}}_1 = \tilde{\mathbf{X}}_1^*(\gamma)\rho_1 + \tilde{\mathbf{X}}_1(\gamma, \gamma_1^0)\rho_2 + \tilde{\mathbf{X}}_1^+(\gamma_1^0)\rho_3 + \tilde{U}_1 \tag{A.2}$$

where $\tilde{\mathbf{X}}_1(\gamma, \gamma_1^0) = \tilde{\mathbf{X}}_1^*(\gamma_1^0) - \tilde{\mathbf{X}}_1^*(\gamma)$ and $\tilde{\mathbf{X}}_1^+(\gamma_1^0) = \tilde{\mathbf{X}}_1 - \tilde{\mathbf{X}}_1^*(\gamma_1^0)$. This specification can therefore regard $S_N^1(\gamma)$ as concentrated sum of squared errors function from (A.2) by restricting $\rho_2 = \rho_3$. Then by the fact that $\tilde{\mathbf{X}}_1(\gamma, \gamma_1^0)$ and $\tilde{\mathbf{X}}_1^+(\gamma_1^0)$ are orthogonal and straight algebra the first part of (A.1) can be represented by function of $\hat{\rho}_2$ and $\hat{\rho}_3$, namely

$$S_N^1(\gamma) - S_N^1(\gamma, \gamma_1^0) = (\hat{\rho}_2 - \hat{\rho}_3)'\tilde{\mathbf{X}}_1(\gamma, \gamma_1^0)'\tilde{\mathbf{X}}_1(\gamma, \gamma_1^0)(\mathbf{Z}_1'\mathbf{Z}_1)^{-1}\tilde{\mathbf{X}}_1^+(\gamma_1^0)'\tilde{\mathbf{X}}_1^+(\gamma_1^0)(\hat{\rho}_2 - \hat{\rho}_3) \tag{A.3}$$

where $\mathbf{Z}_1 = \tilde{\mathbf{X}}_1(\gamma, \gamma_1^0) + \tilde{\mathbf{X}}_1^+(\gamma_1^0)$. Similarly $S_N^1(\gamma_1^0)$ can be regarded as concentrated sum of squared errors function from (A.2) by restricting $\rho_1 = \rho_2$. Then by the fact that $\tilde{\mathbf{X}}_1^*(\gamma)$ and $\tilde{\mathbf{X}}_1(\gamma, \gamma_1^0)$ are orthogonal and straight algebra the second part of (A.1) can be represented by function of $\hat{\rho}_1$ and $\hat{\rho}_2$, namely

$$S_N^1(\gamma_1^0) - S_N^1(\gamma, \gamma_1^0) = (\hat{\rho}_1 - \hat{\rho}_2)'\tilde{\mathbf{X}}_1^*(\gamma)'\tilde{\mathbf{X}}_1^*(\gamma)(\mathbf{Z}_2'\mathbf{Z}_2)^{-1}\tilde{\mathbf{X}}_1(\gamma, \gamma_1^0)'\tilde{\mathbf{X}}_1(\gamma, \gamma_1^0)(\hat{\rho}_1 - \hat{\rho}_2) \tag{A.4}$$

where $\mathbf{Z}_2 = \tilde{\mathbf{X}}_1^*(\gamma) + \tilde{\mathbf{X}}_1(\gamma, \gamma_1^0)$.

Now we can divide both sides of (A.1) by $N(\gamma - \gamma_1^0)$ and decompose as

$$\begin{aligned} \frac{S_N^1(\gamma) - S_N^1(\gamma_1^0)}{N(\gamma - \gamma_1^0)} &= (\hat{\rho}_2 - \hat{\rho}_3)' \left[\frac{\tilde{\mathbf{X}}_1(\gamma, \gamma_1^0)'\tilde{\mathbf{X}}_1(\gamma, \gamma_1^0)(\mathbf{Z}_1'\mathbf{Z}_1)^{-1}\tilde{\mathbf{X}}_1^+(\gamma_1^0)'\tilde{\mathbf{X}}_1^+(\gamma_1^0)}{N(\gamma - \gamma_1^0)} \right] (\hat{\rho}_2 - \hat{\rho}_3) \\ &\quad - (\hat{\rho}_1 - \hat{\rho}_2)' \left[\frac{\tilde{\mathbf{X}}_1^*(\gamma)'\tilde{\mathbf{X}}_1^*(\gamma)(\mathbf{Z}_2'\mathbf{Z}_2)^{-1}\tilde{\mathbf{X}}_1(\gamma, \gamma_1^0)'\tilde{\mathbf{X}}_1(\gamma, \gamma_1^0)}{N(\gamma - \gamma_1^0)} \right] (\hat{\rho}_1 - \hat{\rho}_2) \end{aligned} \tag{A.5}$$

Now taking the advantage of orthogonality between $\tilde{\mathbf{X}}_1(\gamma,\gamma_1^0)$ and $\tilde{\mathbf{X}}_1^+(\gamma_1^0)$, we have $\tilde{\mathbf{X}}_1(\gamma,\gamma_1^0)'\mathbf{Z}_1 = \tilde{\mathbf{X}}_1(\gamma,\gamma_1^0)'\tilde{\mathbf{X}}_1(\gamma,\gamma_1^0)$. Then as $\tilde{\mathbf{X}}_1^+(\gamma_1^0) = \mathbf{Z}_1 - \tilde{\mathbf{X}}_1(\gamma,\gamma_1^0)$, we have

$$\tilde{\mathbf{X}}_1(\gamma,\gamma_1^0)'\tilde{\mathbf{X}}_1(\gamma,\gamma_1^0)(\mathbf{Z}_1'\mathbf{Z}_1)^{-1}\tilde{\mathbf{X}}_1^+(\gamma_1^0)'\tilde{\mathbf{X}}_1^+(\gamma_1^0) = \tilde{\mathbf{X}}_1(\gamma,\gamma_1^0)'(I - \mathbf{Z}_1(\mathbf{Z}_1'\mathbf{Z}_1)^{-1}\mathbf{Z}_1')\tilde{\mathbf{X}}_1(\gamma,\gamma_1^0)$$

which allows us to decompose (A.5) as

$$\begin{aligned}\frac{S_N^1(\gamma) - S_N^1(\gamma_1^0)}{N(\gamma-\gamma_1^0)} &= (\hat{\rho}_2-\hat{\rho}_3)'\left[\frac{\tilde{\mathbf{X}}_1(\gamma,\gamma_1^0)'\tilde{\mathbf{X}}_1(\gamma,\gamma_1^0)}{N(\gamma-\gamma_1^0)}\right](\hat{\rho}_2-\hat{\rho}_3) \\ &\quad -(\hat{\rho}_2-\hat{\rho}_3)'\left[\frac{\tilde{\mathbf{X}}_1(\gamma,\gamma_1^0)'(\mathbf{Z}_1(\mathbf{Z}_1'\mathbf{Z}_1)^{-1}\mathbf{Z}_1')\tilde{\mathbf{X}}_1(\gamma,\gamma_1^0)}{N(\gamma-\gamma_1^0)}\right](\hat{\rho}_2-\hat{\rho}_3) \\ &\quad -(\hat{\rho}_1-\hat{\rho}_2)'\left[\frac{\tilde{\mathbf{X}}_1^*(\gamma)'\tilde{\mathbf{X}}_1^*(\gamma)(\mathbf{Z}_2'\mathbf{Z}_2)^{-1}\tilde{\mathbf{X}}_1(\gamma,\gamma_1^0)'\tilde{\mathbf{X}}_1(\gamma,\gamma_1^0)}{N(\gamma-\gamma_1^0)}\right](\hat{\rho}_1-\hat{\rho}_2) \qquad (A.6)\end{aligned}$$

By the setups of variables and parameters we assume before, the true model should be

$$\tilde{\mathbf{Y}}_1 = \tilde{\mathbf{X}}_1^*(\gamma_1^0)\phi_1^0 + \tilde{\mathbf{X}}_1^+(\gamma_1^0)\phi_2^0 + \tilde{\mathbf{U}}_1 \qquad (A.7)$$

where $\phi_1^0 = \beta_1^0 + \delta_1^0$ and $\phi_2^0 = \beta_1^0$. we can deduce the closed form value of $(\hat{\rho}_1, \hat{\rho}_2, \hat{\rho}_3)$ as

$$\hat{\rho}_1 = \phi_1^0 + \left(\frac{\tilde{\mathbf{X}}_1^*(\gamma_1^0)'\tilde{\mathbf{X}}_1^*(\gamma_1^0)}{N}\right)^{-1}\frac{\tilde{\mathbf{X}}_1^*(\gamma_1^0)'\tilde{\mathbf{U}}_1}{N} \qquad (A.8)$$

$$\hat{\rho}_2 = \phi_1^0 + \left(\frac{\tilde{\mathbf{X}}_1(\gamma,\gamma_1^0)'\tilde{\mathbf{X}}_1(\gamma,\gamma_1^0)}{N}\right)^{-1}\frac{\tilde{\mathbf{X}}_1(\gamma,\gamma_1^0)'\tilde{\mathbf{U}}_1}{N} \qquad (A.9)$$

$$\hat{\rho}_3 = \phi_2^0 + \left(\frac{\tilde{\mathbf{X}}_1^+(\gamma_1^0)'\tilde{\mathbf{X}}_1^+(\gamma_1^0)}{N}\right)^{-1}\frac{\tilde{\mathbf{X}}_1^+(\gamma_1^0)'\tilde{\mathbf{U}}_1}{N} \qquad (A.10)$$

Under our assumptions,

$$\hat{\rho}_1 - \hat{\rho}_2 = o_p(1)$$

and

$$\hat{\rho}_2 - \hat{\rho}_3 = \delta_1^0 + o_p(1).$$

Also we have

$$\left\|\frac{\tilde{\mathbf{X}}(\gamma,\gamma_1^0)'\tilde{\mathbf{X}}(\gamma,\gamma_1^0)}{N(\gamma-\gamma_1^0)}\right\| = O_p(1)$$

by simple Taylor expansion. Then the third term in (A.6) is arbitrarily small. As for the second term, we can show

$$\frac{\tilde{\mathbf{X}}_1(\gamma,\gamma_1^0)'(\mathbf{Z}_1(\mathbf{Z}_1'\mathbf{Z}_1)^{-1}\mathbf{Z}_1')\tilde{\mathbf{X}}_1(\gamma,\gamma_1^0)}{N(\gamma-\gamma_1^0)} = \frac{\tilde{\mathbf{X}}_1(\gamma,\gamma_1^0)'\mathbf{Z}_1}{N(\gamma-\gamma_1^0)}\left(\frac{\mathbf{Z}_1'\mathbf{Z}_1}{T}\right)^{-1}\frac{\mathbf{Z}_1'\tilde{\mathbf{X}}_1(\gamma,\gamma_1^0)}{N(\gamma-\gamma_1^0)}(\gamma-\gamma_1^0)$$

which therefore indicates

$$\begin{aligned}
& \left\| \frac{\tilde{\mathbf{X}}_1(\gamma,\gamma_1^0)'(\mathbf{Z}_1(\mathbf{Z}_1'\mathbf{Z}_1)^{-1}\mathbf{Z}_1')\tilde{\mathbf{X}}_1(\gamma,\gamma_1^0)}{N(\gamma-\gamma_1^0)} \right\| \\
\leq & \left\| \frac{\tilde{\mathbf{X}}_1(\gamma,\gamma_1^0)'\mathbf{Z}_1}{N(\gamma-\gamma_1^0)} \left(\frac{\mathbf{Z}_1'\mathbf{Z}_1}{T}\right)^{-1} \right\| \left\| \frac{\mathbf{Z}_1'\tilde{\mathbf{X}}_1(\gamma,\gamma_1^0)}{N(\gamma-\gamma_1^0)} \right\| (\gamma-\gamma_1^0),
\end{aligned}$$

therefore the second term is also arbitrarily small. Compared to that, the first term is bounded by some constant multiplied by $\|\delta_1^0\|^2$, therefore $S_N^1(\gamma) - S_N^1(\gamma_1^0) > 0$ when $\gamma \in (\underline{\gamma}, \gamma_1^0 - \frac{B}{N})$. Similarly we can show the inequality also holds when $\gamma \in (\gamma_1^0 + \frac{B}{N})$. Therefore our proof is completed. ■

Proof of Theorem 3.2. The estimator $\hat{\underline{\theta}}$ can be written as

$$\hat{\underline{\theta}} = \left(\bar{\mathbf{X}}(\hat{\gamma}_1,\hat{\gamma}_2)'\bar{\mathbf{X}}(\hat{\gamma}_1,\hat{\gamma}_2)\right)^{-1} \bar{\mathbf{X}}(\hat{\gamma}_1,\hat{\gamma}_2)'\mathbf{Y} \tag{A.11}$$

where

$$\bar{\mathbf{X}}(\hat{\gamma}_1,\hat{\gamma}_2) = \left(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}^*(\hat{\gamma}_1), \mathbf{X}_2^*(\hat{\gamma}_2)\right).$$

Therefore the closed form of the estimator can be rewritten as

$$\begin{aligned}
\hat{\underline{\theta}} &= \left(\bar{\mathbf{X}}(\hat{\gamma}_1,\hat{\gamma}_2)'\bar{\mathbf{X}}(\hat{\gamma}_1,\hat{\gamma}_2)\right)^{-1} \bar{\mathbf{X}}(\hat{\gamma}_1,\hat{\gamma}_2)(\bar{\mathbf{X}}(\gamma_1^0,\gamma_2^0)\underline{\theta}^0 + \mathbf{U}) \\
&= \underline{\theta}^0 + \left(\mathbf{X}(\hat{\gamma}_1,\hat{\gamma}_2)'\mathbf{X}(\hat{\gamma}_1,\hat{\gamma}_2)\right)^{-1} \mathbf{X}(\hat{\gamma}_1,\hat{\gamma}_2)[(\mathbf{X}(\gamma_1^0,\gamma_2^0) - \mathbf{X}(\hat{\gamma}_1,\hat{\gamma}_2))\underline{\theta}^0 + \mathbf{U}]
\end{aligned}$$

As we show the convergence rates of $\hat{\gamma}_1$ and $\hat{\gamma}_2$ are both $O_p(1/N)$, it is straightforward to infer $\left\|\bar{\mathbf{X}}(\gamma_1^0,\gamma_2^0) - \bar{\mathbf{X}}(\hat{\gamma}_1,\hat{\gamma}_2)\right\| = O_p(1/N)$. Then

$$\begin{aligned}
\sqrt{N}(\hat{\underline{\theta}} - \underline{\theta}^0) &= \left(\frac{1}{N}\bar{\mathbf{X}}(\hat{\gamma}_1,\hat{\gamma}_2)'\bar{\mathbf{X}}(\hat{\gamma}_1,\hat{\gamma}_2)\right)^{-1} \frac{1}{\sqrt{N}}\bar{\mathbf{X}}(\hat{\gamma}_1,\hat{\gamma}_2)'[(\bar{\mathbf{X}}(\gamma_1^0,\gamma_2^0) - \bar{\mathbf{X}}(\hat{\gamma}_1,\hat{\gamma}_2))\underline{\theta}^0 + \mathbf{U}] \\
&= \left(\frac{1}{N}\bar{\mathbf{X}}(\gamma_1^0,\gamma_2^0)'\bar{\mathbf{X}}(\gamma_1^0,\gamma_2^0) + o_P(1)\right)^{-1} \frac{1}{\sqrt{N}}\bar{\mathbf{X}}(\gamma_1^0,\gamma_2^0)'\mathbf{U} + o_p(1)
\end{aligned}$$

Then the asymptotic distribution in Theorem 3.2 follows. ■

Lemma A.1

$$\frac{\mathbf{X}_j'\mathbf{M}_{-j}\mathbf{X}_j^*(\gamma)}{N} = \frac{\mathbf{X}_j^*(\gamma)'\mathbf{M}_{-j}\mathbf{X}_j^*(\gamma)}{N} + o_p(1)$$

Proof.

$$\begin{aligned}
\mathbf{X}_j'\mathbf{M}_{-j}\mathbf{X}_j^*(\gamma) &= \left(\mathbf{X}_j^*(\gamma) + \mathbf{X}_j^+(\gamma)\right)'\mathbf{M}_{-j}\mathbf{X}_j^*(\gamma) \\
&= \mathbf{X}_j^*(\gamma)'\mathbf{M}_{-j}\mathbf{X}_j^*(\gamma) + \mathbf{X}_j^+(\gamma)'\mathbf{M}_{-j}\mathbf{X}_j^*(\gamma) \\
&= \mathbf{X}_j^*(\gamma)'\mathbf{M}_{-j}\mathbf{X}_j^*(\gamma) + \mathbf{X}_j^+(\gamma)'\mathbf{P}_{-j}\left(\mathbf{P}_{-j}'\mathbf{P}_{-j}\right)^{-1}\mathbf{P}_{-j}'\mathbf{X}_j^*(\gamma)
\end{aligned}$$

where we take the advantage of fact that $\mathbf{X}_j^+(\gamma)'\mathbf{X}_j^*(\gamma) = 0$. Then,

$$\frac{\mathbf{X}_j'\mathbf{M}_{-j}\mathbf{X}_j^*(\gamma)}{N} = \frac{\mathbf{X}_j^*(\gamma)'\mathbf{M}_{-j}\mathbf{X}_j^*(\gamma)}{N} + \frac{\mathbf{X}_j^+(\gamma)'\mathbf{P}_{-j}}{N}\left(\frac{\mathbf{P}_{-j}'\mathbf{P}_{-j}}{N}\right)^{-1}\frac{\mathbf{P}_{-j}'\mathbf{X}_j^*(\gamma)}{N} \tag{A.12}$$

if $\mathbf{X}_j'$ and $\mathbf{P}_{-j}$ are independent, then $\mathbf{X}_j^+(\gamma)'\mathbf{P}_{-j}\left(\mathbf{P}_{-j}'\mathbf{P}_{-j}\right)^{-1}\mathbf{P}_{-j}'\mathbf{X}_j^*(\gamma)/N = o_p(1)$. If $\mathbf{X}_j'$ and $\mathbf{P}_{-j}$ are weakly dependent, without loss of generality we can set

$$\mathbf{X}_j^*(\gamma) = \mathbf{P}_{-j}\rho + \mathbf{v}_j,$$

then the least squares estimator is

$$\hat{\rho} = \left(\frac{\mathbf{P}_{-j}'\mathbf{P}_{-j}}{N}\right)^{-1}\frac{\mathbf{P}_{-j}'\mathbf{X}_j^*(\gamma)}{N}$$

which indicates

$$\mathbf{X}_j^*(\gamma) - \mathbf{P}_{-j}\hat{\rho} = \mathbf{v}_j - \mathbf{P}_{-j}\left(\mathbf{P}_{-j}'\mathbf{P}_{-j}\right)^{-1}\mathbf{P}_{-j}'\mathbf{v}_j \tag{A.13}$$

Now we have the second term of (A.12) can be decomposed to

$$\frac{\mathbf{X}_j^+(\gamma)\mathbf{X}_j^*(\gamma)}{N} - \frac{\mathbf{X}_j^+(\gamma)'\left(\mathbf{v}_j - \mathbf{P}_{-j}\left(\mathbf{P}_{-j}'\mathbf{P}_{-j}\right)^{-1}\mathbf{P}_{-j}'\mathbf{v}_j\right)}{N} \tag{A.14}$$

The first term is obviously zero. If $\mathbf{v}_j \perp \mathbf{P}_{-j}$, together with $E(v_j) = \mathbf{0}$ it can be indicated that the second term will be $o_p(1)$. If $\mathbf{v}_j \not\perp \mathbf{P}_{-j}$, we can infer

$$\mathbf{v}_j = \mathbf{P}_{-j}'\eta + \mathbf{h}_j$$

such that $\mathbf{h}_j \perp \mathbf{P}_j$ which is usually used in instrumental literature. Therefore the second term of (A.14) is

$$\frac{\mathbf{X}_j^+(\gamma)'\left(\mathbf{h}_j - \mathbf{P}_{-j}\left(\mathbf{P}_{-j}'\mathbf{P}_{-j}\right)^{-1}\mathbf{P}_{-j}'\mathbf{h}_j\right)}{N}$$

which should be $o_p(1)$. $\blacksquare$

References

Agarwal, S., S. Chomsisengphet, N. Mahoney, and J. Stroebel (2015): "Regulating consumer financial products: Evidence from credit cards," *The Quarterly Journal of Economics*, 130, 111–164.

Bai, J. (1997): "Estimation of a change point in multiple regression models," *Review of Economics and Statistics*, 79, 551–563.

——— (2013): "Fixed-effects dynamic panel models, a factor analytical method," *Econometrica*, 81, 285–314.

Beaudry, P. and G. Koop (1993): "Do recessions permanently change output?" *Journal of Monetary economics*, 31, 149–163.

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

Card, D., A. Mas, and J. Rothstein (2008): "Tipping and the Dynamics of Segregation," *The Quarterly Journal of Economics*, 123, 177–218.

Chan, K.-S. et al. (1993): "Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model," *The annals of statistics*, 21, 520–533.

Cheng, X., F. Schorfheide, and P. Shao (2019): "Clustering for Multi-Dimensional Heterogeneity," .

Chudik, A., K. Mohaddes, M. H. Pesaran, M. Raissi, and A. Rebucci (2020): "A counterfactual economic analysis of Covid-19 using a threshold augmented multi-country model," Tech. rep., National Bureau of Economic Research.

Clements, M. P. and H.-M. Krolzig (1998): "A comparison of the forecast performance of Markov-switching and threshold autoregressive models of US GNP," *The Econometrics Journal*, 1, 47–75.

Gathergood, J., N. Mahoney, N. Stewart, and J. Weber (2019): "How do individuals repay their debt? The balance-matching heuristic," *American Economic Review*, 109, 844–75.

Gonzalez, A., T. Teräsvirta, D. Van Dijk, and Y. Yang (2017): *Panel smooth transition regression models*.

Gonzalo, J. and J.-Y. Pitarakis (2002): "Estimation and model selection based inference in single and multiple threshold models," *Journal of Econometrics*, 110, 319–352.

Hansen, B. E. (1996): "Inference when a nuisance parameter is not identified under the null hypothesis," *Econometrica: Journal of the econometric society*, 413–430.

——— (1999): "Threshold effects in non-dynamic panels: Estimation, testing, and inference," *Journal of econometrics*, 93, 345–368.

——— (2000): "Sample splitting and threshold estimation," *Econometrica*, 68, 575–603.

——— (2011): "Threshold autoregression in economics," *Statistics and its Interface*, 4, 123–127.

Hsiao, C. and Q. Zhou (2018): "Incidental parameters, initial conditions and sample size in statistical inference for dynamic panel data models," *Journal of Econometrics*, 207, 114–128.

Keys, B. J. and J. Wang (2019): "Minimum payments and debt paydown in consumer credit cards," *Journal of Financial Economics*, 131, 528–548.

Kim, Y.-J., M. H. Seo, and H.-E. Yeom (2020): "Estimating a breakpoint in the pattern of spread of COVID-19 in South Korea," *International Journal of Infectious Diseases*, 97, 360–364.

Kourtellos, A., T. Stengos, and C. M. Tan (2014): "Structural threshold regression," *Econometric Theory, Forthcoming.*

Liao, Q., P. C. Phillips, and P. Yu (2015): "Inferences and Specification Testing in Threshold Regression with Endogeneity," Tech. rep., mimeo, HKU.

Newey, K. (1994): "Large sample estimation and hypothesis," .

Ramírez-Rondán, N. (2020): "Maximum likelihood estimation of dynamic panel threshold models," *Econometric Reviews*, 39, 260–276.

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

Tong, H. (1978): "On a threshold model," .

Yu, P. (2013): "Inconsistency of 2SLS estimators in threshold regression with endogeneity," *Economics Letters*, 120, 532–536.

Yu, P., Q. Liao, and P. C. Phillips (2019): "New Control Function Approaches in Threshold Regression with Endogeneity," Tech. rep., mimeo.

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