Factor-Driven Two-Regime Regression

Sokbae Lee,[1] Yuan Liao,[2] Myung Hwan Seo[3] & Youngki Shin[4]

[1]Columbia Univ. [2]Rutgers Univ. [3]Seoul National Univ. [4]McMaster Univ.

ASSA 2018

Models with Multiple Regimes in Economics

1. Markov Switching Model
   - Latent regime switching: flexible
   - Exogenous regime switching
   - The dynamic structure is stable and stationary
2. Structural Breaks
   - Change at unknown times
   - Easy to model
   - Break is exogenous and unpredictable
3. Threshold Regression or Smooth Transition Regression
   - Regime is determined by an observable scalar covariate
   - Regime is correlated with other observables
   - This paper extends it to a more general regime switching governed by a vector of possibly unobserved factors.

Outline

Factor-Driven Two-Regime Regression

- We propose the following model

$$y_t = x_t'\beta_0 + x_t'\delta_0 1\{f_t'\gamma_0 > 0\} + \varepsilon_t, \quad (1)$$

$$E(\varepsilon_t | \mathcal{F}_{t-1}) = 0, \quad (2)$$

  where

  - f_t is vector-valued and its last element is fixed as -1. And x_t and f_t are adapted to the filtration $\mathcal{F}_{t-1}$,
  - $(\beta_0, \delta_0, \gamma_0)$ is a vector of unknown parameters,
  - The factors f_t may be latent but can be consistently estimable.
  - We explicitly treat the case where the factors are estimated by the principal component analysis (PCA).

- We call the regression model in (1) and (2) a factor-driven two-regime regression model and $f_t'\gamma_0$ the threshold index.

Least Squares Estimation

- Assume the following scale normalization

$$\gamma_0 \in \Gamma \equiv \{(1, \gamma_2')' : \gamma_2 \in \Gamma_2 \subset \mathbb{R}^{d_f - 1}\}.$$

- Under further regularity conditions on the distribution, we can show that (α_0', γ_0') is the unique solution to

$$\min_{(\alpha', \gamma')' \in \mathbb{R}^{2d_x} \times \Gamma} \mathbb{E}(y_t - x_t'\beta - x_t'\delta 1\{f_t'\gamma > 0\})^2.$$

Least Squares Estimation

- Assume the following scale normalization

$$\gamma_0 \in \Gamma \equiv \{(1, \gamma_2')' : \gamma_2 \in \Gamma_2 \subset \mathbb{R}^{d_f - 1}\}.$$

- Under further regularity conditions on the distribution, we can show that (α_0', γ_0') is the unique solution to

$$\min_{(\alpha', \gamma')' \in \mathbb{R}^{2d_x} \times \Gamma} \mathbb{E}(y_t - x_t'\beta - x_t'\delta 1\{f_t'\gamma > 0\})^2.$$

- Thus, the unknown parameters can be estimated by the least squares, that is,

$$(\widehat{\alpha}, \widehat{\gamma}) = \arg \min_{(\alpha', \gamma')' \in \mathcal{A} \times \Gamma} \mathbb{S}_T(\alpha, \gamma), \tag{3}$$

where

$$\mathbb{S}_T(\alpha, \gamma) \equiv \frac{1}{T} \sum_{t=1}^{T} (y_t - x_t'\beta - x_t'\delta 1\{f_t'\gamma > 0\})^2. \tag{4}$$

Optimization

Optimization Algorithm

- The least squares problem is computationally challenging (NP hard) and there is no readily available computational algorithm other than brute-force multi-dimensional grid search.
- We reformulate the problem as a Mixed Integer Optimization (MIO) problem.
- Define

$$M_t \equiv \max_{\gamma \in \Gamma} |f_t' \gamma|$$

  for each $t = 1, \ldots, T$. One can compute M_t easily for each t using linear programming.
- Let δ_j denote the j-th element of δ, where $j = 1, \ldots, d_x$.

Mixed Integer Optimization (MIO)

- Then, rewrite (4) as

$$\min_{\beta,\delta,\gamma,d_1,\ldots,d_T} \frac{1}{T}\sum_{t=1}^{T}\left(y_t - x_t'\beta - x_t'\delta d_t\right)^2 \tag{5}$$

  subject to

$$\begin{aligned} &(\beta,\delta)\in\mathcal{A},\ \gamma\in\Gamma,\\ &L_j \le \delta_j \le U_j,\\ &(d_t-1)(M_t+\epsilon) < f_t'\gamma \le d_t M_t,\\ &d_t\in\{0,1\}\end{aligned} \tag{6}$$

  for each $t = 1,\ldots,T$ and each $j = 1,\ldots,d_x$. Here, $\epsilon > 0$ is a small predetermined constant (e.g. $\epsilon = 10^{-6}$).
- Observe that (5) adds new integer variables $d_1,\ldots,d_T$, but new constraints (6) ensure that the reformulated problem (5) is the same as the original problem.
- Note that each element in δd_t is a product term. Hence, in (5), we have bilinear terms. This creates a difficult optimization problem.

Refinement I

- We propose two solutions:
  1. Reformulate it as a quadratic optimization problem by introducing additional variables
  $$\ell_{j,t} = \delta_j \boldsymbol{d}_t.$$
  2. Iterative algorithm, where we estimate α for a given γ and estimate γ for a given α, iteratively.

Refinement II

- Moreover, we will consider an additional restriction:

$$\tau_1 \leq \frac{1}{T}\sum_{t=1}^{T} d_t \leq \tau_2 \tag{7}$$

  where $0 < \tau_1 < \tau_2 < 1$.

  - It prevents the proportion of any one regime from getting too small.
  - In the special case that $1\{f_t'\gamma_0 > 0\} = 1\{q_t > \gamma_{0,2}\}$ with a scalar variable q_t and a parameter $\gamma_{0,2}$, it is standard to assume that the parameter space for $\gamma_{0,2}$ is between the α and $(1-\alpha)$ quantiles of q_t for some $0 < \alpha < 1$. We can interpret (7) as a natural generalization of this type of restriction so that the proportion of one regime is never too close to 0 or 1.

Mixed Integer Quadratic Programming

$$\min_{\beta,\delta,\gamma,\boldsymbol{d},\boldsymbol{\ell}} \frac{1}{T}\sum_{t=1}^{T}\left(y_t - x_t'\beta - \sum_{j=1} x_{j,t}\ell_{j,t}\right)^2$$

subject to: $\forall t, j,$

1. $(\beta, \delta) \in \mathcal{A},\ \gamma \in \Gamma,$
2. $L_j \leq \delta_j \leq U_j,$
3. $(d_t - 1)(M_t + \epsilon) < f_t'\gamma \leq d_t M_t,$
4. $d_t \in \{0, 1\},$
5. $d_t L_j \leq \ell_{j,t} \leq d_t U_j,$
6. $L_j(1 - d_t) \leq \delta_j - \ell_{j,t} \leq U_j(1 - d_t),$

$\star$ We show that this is equivalent to the original least squares problem.

Iterative Algorithm I

- Let $\Gamma_T = \{\gamma_j\}_{j=1}^{M_T}$ be a grid on Γ such that $\max_{\gamma \in \Gamma} \min_j |\gamma - \gamma_j| \leq \psi_T \to 0$ as $T \to \infty$ and $M_T = O\left(\psi_T^{1-d_f}\right)$.
- The number M_T signifies the computational burden. We only require $\psi_T \to 0$.
- We do not have to try the initial values exhaustively in our iterative algorithm. $M_T = T^{-c}$ will do for any $c > 0$. We can construct such a grid easily.

Iterative Algorithm II

- For instance, we may consider a unit hypercube Γ and a grid

$$\Gamma_T = \left\{ \left(a_{1i_1}, \cdots, a_{d_f-1, i_{d_f-1}} \right) : i_1, i_2, \ldots = 1, \ldots, m \right\},$$

  where $a_{i,j} - a_{i,j-1} = \zeta_T$ for all i and j.
- Then, $\psi_T = \zeta_T \sqrt{d_f - 1}$ and $M_T = O\left(\left(\zeta_T^{-1} \right)^{d_f - 1} \right)$.

Iterative Algorithm III

1. Obtain an initial consistent estimate

$$\left(\hat{\alpha}^0, \hat{\gamma}^0\right) = \underset{\alpha \in \mathcal{A}, \gamma \in \Gamma_T}{\operatorname{argmin}} \frac{1}{T} \sum_{t=1}^{T} \left(y_t - Z_t \left(\gamma\right)' \alpha\right)^2 .$$

2. Iterate the following steps 3-5, beginning with $i = 1$ and finishing at a prespecified number or until convergence.
3. For the given $\hat{\alpha}^{i-1}$, obtain an estimate $\hat{\gamma}^i$ via the mixed integer linear optimization algorithm

$$\min_{\gamma \in \Gamma, d_1, \dots, d_T} \frac{1}{T} \sum_{t=1}^{T} \left(y_t - x_t' \hat{\beta}^{i-1} - x_t' \hat{\delta}^{i-1} d_t\right)^2$$

subject to

$$\begin{aligned} &(d_t - 1)(M_t + \epsilon) < f_t' \gamma \leq d_t M_t, \\ &d_t \in \{0, 1\}. \end{aligned}$$

Iterative Algorithm IV

4. For the given $\hat{\gamma}^i$, obtain

$$\hat{\alpha}^i = \operatorname*{argmin}_{\alpha \in \mathcal{A}} \frac{1}{T} \sum_{t=1}^{T} \left(y_t - Z_t \left(\hat{\gamma}^i \right)' \alpha \right)^2$$

5. Let $i = i + 1$.

Monte Carlo For Computation

DGPs for Monte Carlo Simulations

We consider the following DGPs for simulation studies:

$$y_t = x_t'\beta_0 + x_t'\delta_0 1\{f_t'\gamma_0 > 0\} + \varepsilon_t$$

- $x_t \equiv (1, \tilde{x}_t)$ with $\tilde{x}_t \sim N(0, \Sigma_x)$
- $f_t \equiv (\tilde{f}_t, -1)$ with $\tilde{f}_t \sim N(0, \Sigma_f)$
- Σ_x and Σ_f are either identity matrices or $(\Sigma)_{ij} = 0.5^{|i-j|}$
- $\epsilon_t \sim N(0, 0.1^2)$
- $\beta_0 = \delta_0 = (1, \ldots, 1)$ and $\gamma_0 = (1, 0.5716, \ldots, 0.5716)$
- The sample size is set to $T = 50, 100, 200,$ and 400
- The number of replications is set to 400

DGPs for Monte Carlo Simulations (cont.)

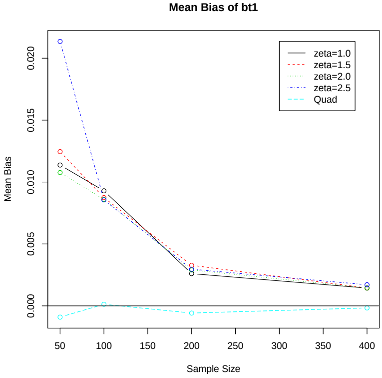
We consider the following DGPs for simulation studies:

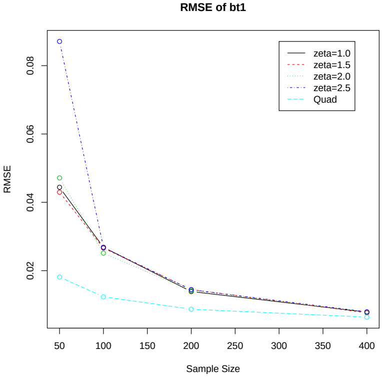
$$y_t = x_t'\beta_0 + x_t'\delta_0 1\{f_t'\gamma_0 > 0\} + \varepsilon_t$$

- The parameter space is set to $[-3, 3]^{2 \cdot dim(x) + dim(f)}$
- For the iteration algorithm, the initial grid size is governed by $\zeta = 1, 1.5, 2, 2.5$
- For the quadratic programming algorithm, the small tuning parameter is set to $\epsilon = 10^{-6}$

Simulation Results

Design: $dim(x) = dim(f) = 3$, and $(\Sigma)_{i,j} = 0.5^{|i-j|}$

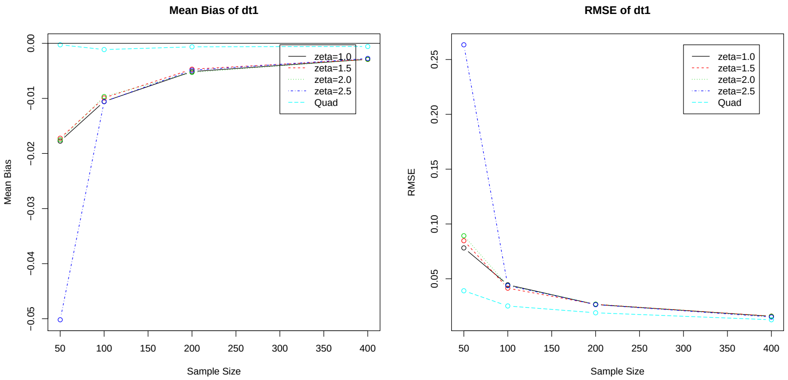




Mean Bias and RMSE for β_1

Simulation Results (cont.)

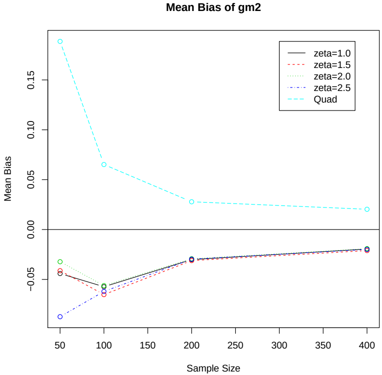
Design: $dim(x) = dim(f) = 3$, and $(\Sigma)_{i,j} = 0.5^{|i-j|}$

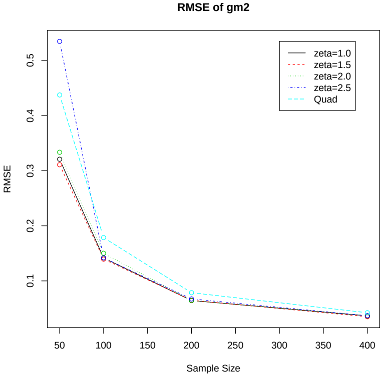


Mean Bias and RMSE for δ_1

Simulation Results (cont.)

Design: $dim(x) = dim(f) = 3$, and $(\Sigma)_{i,j} = 0.5^{|i-j|}$

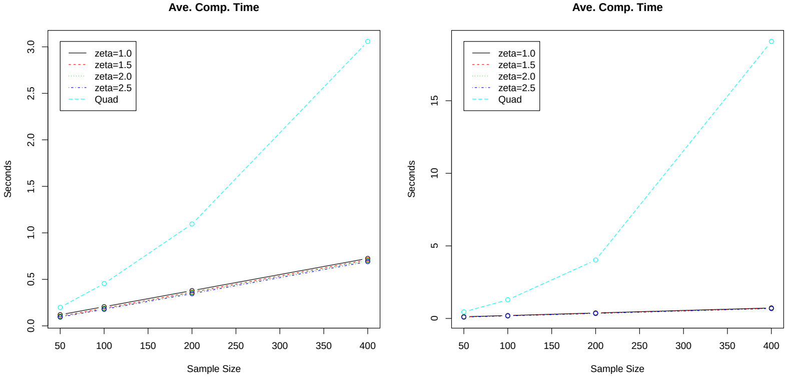




Mean Bias and RMSE for γ_2

Simulation Results (cont.)

Design: $dim(x) = 1$ or 3, $dim(f) = 3$, and $(\Sigma)_{i,j} = 0.5^{|i-j|}$



Average Computation Time

Summary of Simulation Studies

- Results for other parameters are similar across different simulation designs
- Overall, both algorithms show satisfactory computation results and we can confirm the simulation results coincide with the proposed theory
- The iteration algorithm works well with coarse initial grid points
- The quadratic programming works slightly better in a smaller sample at the cost of longer computation time.

Application

Empirical Illustration I

- Growing literature on state-dependent government spending multipliers, e.g. Barro and Redlick (2011), Auerbach and Gorodnichenko (2012,2013), Fazzari et al. (2015), Cogan et al. (2010), Christiano et al. (2011), Coenen et al. (2012), Ramey and Zubairy (2016) etc.
- The states are determined by $I_t = 1\{q_t > \gamma_0\}$.
- The threshold variable q_t and the threshold γ_0 are often chosen by the researchers. (accompanied by various specification checks)

Empirical Illustration II

- The spending multiplier is estimated by the (state dependent) local projection proposed by Jordà (2005). Specifically,

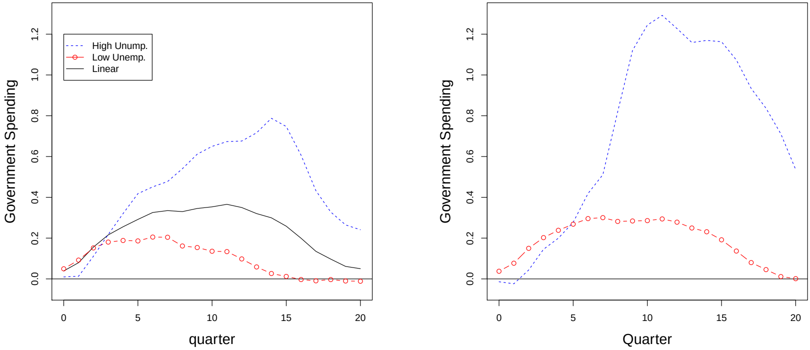
$$\begin{aligned} y_{t+h} &= I_{t-1}\left(x_{t-1}'\beta_{1,h} + \alpha_{1,h} shock_t\right) \\ &\quad + (1 - I_{t-1})\left(x_{t-1}'\beta_{2,h} + \alpha_{2,h} shock_t\right) + \varepsilon_{t+h}, \end{aligned}$$

  where y_{y+h} is either GDP or government spending, x_{t-1} is a vector of lagged variables composed of GDP, government spending, news (shock).
- The impulse response function is hard to construct with nonlinear models and the local projection is a popular alternative, particularly with the threshold model.

Empirical Illustration III

- Data: Ramey and Zubairy's (2016 JPE) historical U.S. data set
  - combined quarterly series for the sample period 1889-2015
  - real GDP, GDP deflator, government purchases, federal government receipts, population, unemployment rate, interest rates, and defense news.
  - includes 3 wars (WWI, WWII, Korean war)
  - For q_t, there are various potential measures of slack, such as output gaps, capacity utilization, or the unemployment rate, moving average of GDP growth, deviation from Hodrick-Prescott trend (time-varying threshold)
  - identification: military news shock and Blanchard-Perotti shock (Cholesky decomposition)

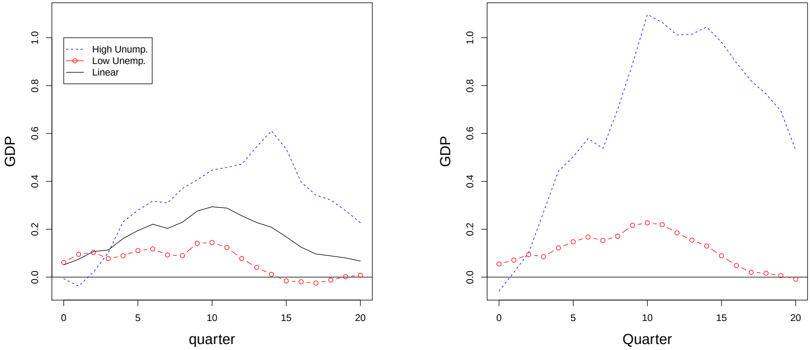
Government spending responses to a news shock



RZ LLSS

Note: RZ used the ad-hoc threshold 6.5% ($F_{unemp}(6.5) = 0.638$). LLSS estimated the threshold with $h = 0$, which is 11.97% ($F_{unemp}(11.97) = 0.918$).

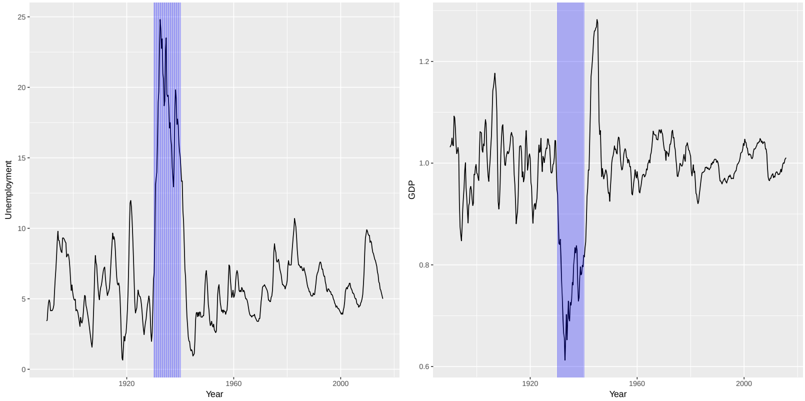
GDP responses to a news shock



RZ LLSS

Note: RZ used the ad-hoc threshold 6.5% ($F_{unemp}(6.5) = 0.638$). LLSS estimated the threshold with $h = 0$, which is 11.97% ($F_{unemp}(11.97) = 0.918$).

Figure: Recession Periods



Shaded periods are High Unemployment periods ($f'\gamma > 0$).

Government Spending Multipliers

Table: Multiplier Estimates

	Low Unemployment	High Unemployment	P-value for difference in multipliers across states
RZ ($\gamma = 6.5$)			
2 year integral	0.59	0.60	0.954
	(0.091)	(0.095)	
4 year integral	0.67	0.68	0.924
	(0.052)	(0.121)	
LLSS ($\hat{\gamma} = 11.97$)			
2 year integral	0.59	1.49	1.23×10^{-13}
	(0.058)	(0.118)	
4 year integral	0.64	0.94	4.72×10^{-6}
	(0.062)	(0.013)	

Adding Irrelevant Factors

We generate some independent (irrelevant) factors and check the robustness of the results by adding $f_{j,t} \sim N\left(0, \widehat{Var}(unem_t)\right)$

Table: Multiplier Estimates

	baseline	$(f_{2,t})$	$(f_{2,t}, f_{3,t})$	$(f_{2,t}, \ldots, f_{4,t})$
$\hat{\gamma}_0$	11.97	12.05	12.95	12.52
$\hat{\gamma}_2$		-0.17	-0.34	-0.04
$\hat{\gamma}_3$			0.10	-0.21
$\hat{\gamma}_4$				-0.01
$F(\hat{\gamma}_0)$	0.918	0.918	0.920	0.918

$F(\cdot)$ is an empirical cdf of $unemp_t$.

Adding Irrelevant Factors (cont.)

Table: Multiplier Estimates

	Low Unemployment	High Unemployment	P-value for difference in multipliers across states
Base			
2 year integral	0.59	1.49	1.23×10^{-13}
	(0.058)	(0.118)	
4 year integral	0.64	0.94	4.72×10^{-6}
	(0.062)	(0.013)	
$(f_{2,t})$			
2 year integral	0.60	1.50	1.03×10^{-14}
	(0.060)	(0.112)	
4 year integral	0.65	0.94	4.76×10^{-5}
	(0.067)	(0.014)	
$(f_{2,t}, f_{3,t})$			
2 year integral	0.60	1.50	1.03×10^{-14}
	(0.060)	(0.113)	
4 year integral	0.65	0.94	4.76×10^{-5}
	(0.052)	(0.014)	
$(f_{2,t}, \ldots, f_{4,t})$			
2 year integral	0.60	1.50	1.03×10^{-14}
	(0.060)	(0.013)	
4 year integral	0.65	0.94	4.76×10^{-5}
	(0.067)	(0.014)	

Asymptotic Theory

Summary of Theoretical Findings

1. Asymptotic Distribution when the factors f_t are directly observed. (Extension of Hansen (2000) from a single threshold value to threshold index)
2. Condition on N/T, that yields the same asymptotic distribution as above (the Oracle result), when f_t is **estimated by PCA** from N separate series.
3. Precise **Phase Transition**, which shows how the asymptotic distribution changes **continuously** as a function of N/T.
4. Validity of the Iterative Algorithm.
5. Test for the Presence of Threshold Effect and its Bootstrap

Generated Regressor Issue

1. Make it Oracle with large N
   - Factor-Augmented Linear Regressions (Bai and Ng 2006) with $\sqrt{T} = o(N)$
   - Factor-Augmented Non-linear GMM and extremum estimator (Bai and Ng 2008) with $T^{5/8} = o(N)$
2. **Phase Transition**
   - In the Factor-Driven Two-Regime Regression, we establish the precise phase transition mechanism, under which the estimation error in the factors disappears from the asymptotic distribution of the parameter estimates as N/T ratio grows.

Challenges

1. No expansion due to the discontinuous transformation of the estimated factors
2. We cannot estimate the factors consistently but the factor space only. The estimated factors are linear combinations of the true factors.
   - random centering of the estimates.
   - uniform control of estimation error in the sequence $\widetilde{f}_t$ (uniformly approximate it by an uncorrelated sequence)
3. Simultaneous asymptotics where both N and T grow to ∞.

Asymptotic Distribution with Estimated f_t

Suppose $T = O(N)$ and $\delta_0 = d_0 \cdot T^{-\varphi}$. Let

$$\begin{aligned} r_{NT} &= \left(NT^{1-2\varphi}\right)^{1/3} \wedge T^{1-2\varphi}, \\ k &= \lim_{N,T\to\infty} \frac{\sqrt{N}}{T^{1-2\varphi}} \in [0,\infty]. \end{aligned}$$

Theorem

Then as $N, T \to \infty$

$$\begin{aligned} \sqrt{T}(\widehat{\alpha} - \alpha_0) &\xrightarrow{d} \mathcal{N}(0, V) \\ r_{NT}\left(\widehat{\gamma} - \gamma_0\right) &\xrightarrow{d} \arg\min_{g\in\mathcal{G}} A(k,g) + 2W(g), \end{aligned}$$

where W is a centered Gaussian process and $A(k,g)$ is given shortly. Furthermore, the two estimators are asymptotically independent.

Remarks I

- The relative size of N over T affects the shape of the limit criterion function in a way that it is approximated by a quadratic function (adjusted by $\sqrt{N}T^{-2\varphi}$) in a neighborhood of γ_0 when $N = o\left(T^{2-4\varphi}\right)$, while it is better approxiated by a linear function with a kink at γ_0 when $T^{2-4\varphi} = o\left(N\right)$.
- Certainly, it is easier to identify the minimum when the function has a kink at the minimum than when it is a quadratic function, making itself smooth at the minimum. This results in the slower rate of convergence of $\left(NT^{1-2\varphi}\right)^{-1/3}$.

Remarks II

- Intuitively, a bigger N makes the estimated factors $\widehat{f}_t$ more precise. Together, $T^{1-2\varphi} = o\left(\sqrt{N}\right)$ yields the oracle results for both $\widehat{\alpha}$ and $\widehat{\gamma}$ while the smller N in the sense that $N = O\left(T^{2-4\varphi}\right)$ fails to do so. When $N = O\left(T^{2-4\varphi}\right)$, such an effect is not negligible, and thus plays an essential role in the limiting distribution of $\widehat{\gamma}$.
- Bai and Ng (2006, 2008) have shown that the oracle property (with regard to the estimation of the factors) holds for the linear regression if $T^{1/2} = o(N)$ and for the extremum estimation and GMM estimation if $T^{5/8} = o(N)$, when the estimated factors are included in the model. Thus, it appears that the oracle property demands bigger N as the nonlinearity of the estimating equation rises.

Remarks III

- While the rotation matrix H_T converges to a deterministic matrix H in probability, the speed of convergence is unknown. Thus, we should carry this random rotation of the true factors along our derivation of the asymptotic distribution, which adds another complication to our task.
- Nonetheless, the index structure allows us to determine the regime consistently by introducing the cancelling rotation in the parameter values. Our proofs of theorems employ different parametrizations ϕ and γ to handle this random rotation properly.

Drift A(k,g) I

- Let $p(\cdot)$ denote the density function of the standard normal and $\sigma^2_{h,x_t,g_t} := p\lim_{N\to\infty} \mathbb{E}[(h_t'\phi_0)^2 | x_t, g_t, g_t'\phi_0 = 0]$, where h_t is a leading term of an asymptotic expansion of estimated factors.
- Then, for $k \in [0,1]$,

$$A(k,g) = 2\mathbb{E}\left[(x_t'd_0)^2 \int_0^{|f_t'g|} (|f_t'g| - w)\, p\left(\frac{k^{1/3}w}{\sigma_{h,x_t,g_t}}\right) dw \,\middle|\, u_t = 0 \right],$$

and, for $k \in [1,\infty]$,

$$A(k,g) = 2\mathbb{E}\left[(x_t'd_0)^2 \int_0^{k|f_t'g|} \left(|f_t'g| - \frac{w}{k}\right) p\left(\frac{w}{\sigma_{h,x_t,g_t}}\right) dw \,\middle|\, u_t = 0 \right]$$

with the convention that $w/k = 0$ for $k = \infty$.

Drift A(k,g) II

- For a sequence of random variables $\mathcal{Z}_t$, whose conditional distribution given $(x_t, g_t, g_t'\phi_0 = 0)$ is $\mathcal{N}(0, \sigma^2_{h,x_t,g_t})$,

$$A(0,g) = \mathbb{E}\left[(x_t' d_0)^2 (f_t' g)^2 \middle| u_t, \mathcal{Z}_t = 0 \right] p_{u_t, \mathcal{Z}_t}(0,0),$$
$$A(\infty, g) = \mathbb{E}\left[(x_t' d_0)^2 |f_t' g| \middle| u_t = 0 \right] p_{u_t}(0).$$

- To appreciate our asymptotic results, we consider the simple case that $g_t = (q_t, -1)'$, $g = (0, g_2)'$, $x_t = 1$, $d_0 = 1$, and h_t and q_t are independent of each other. Then $A(k,g)$ reduces to (while writing $g_2 = g$ for simplicity)

$$A(k,g) = 2\int_0^{|g|} (|g| - w)\, p\left(k^{1/3} w\right) dw,$$

for $k \in [0,1]$,

$$A(k,g) = 2\int_0^{k|g|} \left(|g| - \frac{w}{k}\right) p(w)\, dw,$$

for $k \in [1,\infty]$.

Figure: $A(k, g)$

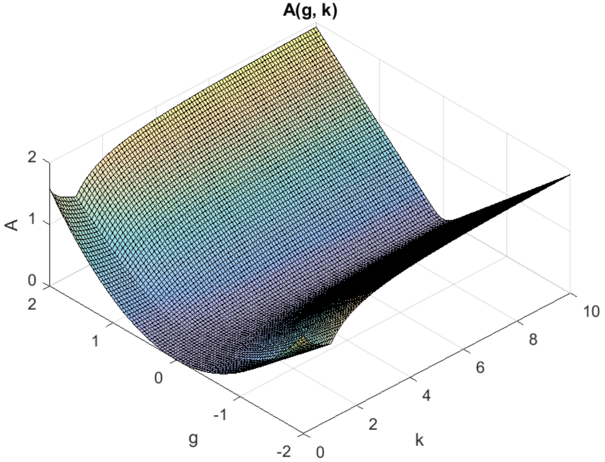


Figure: $A(k, g)$

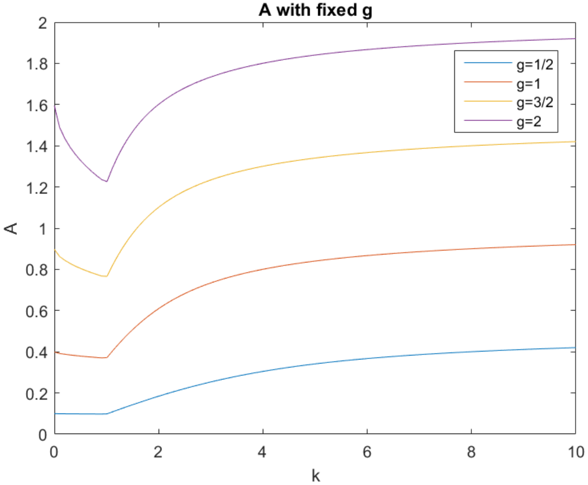
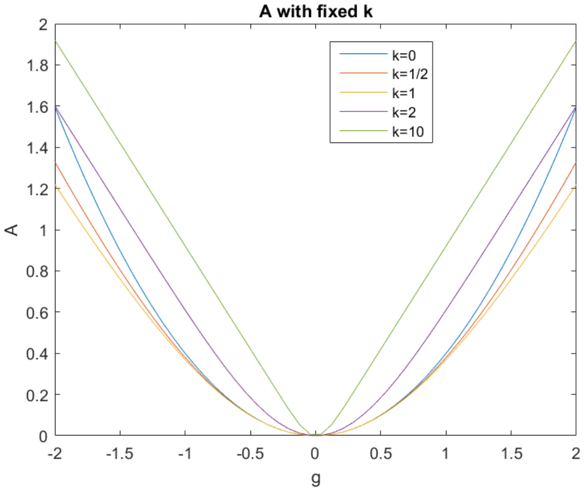


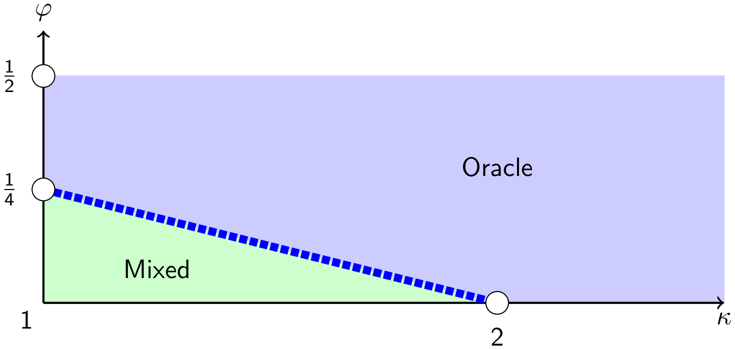
Figure: $A(k, g)$


A with fixed k
k=0
k=1/2
k=1
k=2
k=10
A
g
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2

Phase Transition

- To demonstrate that our asymptotic results are sharp and continuous among three different cases, we consider a special case that $N = T^{\kappa}$ for $\kappa \geq 1$.
- In this case, the asymptotic results can be depicted on the (κ, φ)-space.
  - oracle phase : $T^{1-2\varphi} = o\left((NT^{1-2\varphi})^{1/3}\right) = o(T^{(\kappa+1-2\varphi)/3})$. The resulting convergence rate and asymptotic distribution for $\widehat{\gamma}$ are the same as those when the unknown factors are observed.
  - mixed phase : $T^{(\kappa+1-2\varphi)/3} = o(T^{1-2\varphi})$. The resulting convergence rate and asymptotic distribution for $\widehat{\gamma}$ are different from those under the oracle phase in terms of the convergence rate as well as the drifting term. Even in this case, the convergence rate and asymptotic distribution for $\widehat{\alpha}$ are still the same as those when the unknown factors are observed.
  - critical boundary : $T^{1-2\varphi} = T^{(\kappa+1-2\varphi)/3}$. Changes in the convergence rates and asymptotic distributions are continuous along this boundary.
- As a result, we expect that inference on (α_0, γ_0) can be carried out in a uniform fashion.

Figure: Phase Diagram



Notes. This figure depicts a phase transition on the (κ, φ)-space. The possible region we consider on the (κ, φ)-space is $0 < \varphi < 1/2$ and $\kappa \geq 1$. The critical boundary ($\varphi = -\kappa/4 + 1/2$) is shown by closely dotted points in the figure. The oracle phase is shaded in blue, wheres the mixed phase is in green.

Testing for Linearity

Testing for Linearity I

Consider

$$\mathcal{H}_0 : \delta = 0 \quad \text{for all } \gamma \in \Gamma.$$

Under this hypothesis the model becomes the linear regression model and thus γ is not identified.

This testing problem has been studied intensively in the literature, see e.g. Andrews and Ploberger (1994), Hansen (1996), Lee et al. (2011) among many others.

Testing for Linearity II

Utilizing our computational algorithm, we consider

$$\begin{aligned}
\text{supQ} &= \sup_{\gamma\in\Gamma} Q_T(\gamma) = \sup_{\gamma\in\Gamma} T\frac{\min_{\alpha:\delta=0}\tilde{\mathbb{S}}_T(\alpha,\gamma) - \min_{\alpha}\tilde{\mathbb{S}}_T(\alpha,\gamma)}{\min_{\alpha}\tilde{\mathbb{S}}_T(\alpha,\gamma)} \\
&= T\frac{\min_{\alpha:\delta=0}\tilde{\mathbb{S}}_T(\alpha,\gamma) - \tilde{\mathbb{S}}_T(\hat{\alpha},\hat{\gamma})}{\tilde{\mathbb{S}}_T(\hat{\alpha},\hat{\gamma})}.
\end{aligned}$$

The statistic $Q_T(\gamma)$ is the likelihood ratio (LR) statistic for $\delta = 0$ when γ is given and the error is Gaussian.

Testing for Linearity III

Theorem

Then, under $\mathcal{H}_0$

$$supQ \xrightarrow{d} \max_{\gamma \in \Gamma} W\left(\gamma\right)' \left(R\left(\mathbb{E}Z_t\left(\gamma\right)Z_t\left(\gamma\right)'\right)^{-1}\mathbb{E}\varepsilon_t^2 R'\right)^{-1} W\left(\gamma\right),$$

where $W\left(\gamma\right)$ *is a vector of centered Gaussian processes with covariance kernel*

$$R\left(\mathbb{E}Z_t\left(\gamma_1\right)Z_t\left(\gamma_1\right)'\right)^{-1}\mathbb{E}Z_t\left(\gamma_2\right)Z_t\left(\gamma_2\right)'\varepsilon_t^2\left(\mathbb{E}Z_t\left(\gamma_2\right)Z_t\left(\gamma_2\right)'\right)^{-1}R'$$

with $R = \left(0, I_{d_x}\right)$ *being the selection matrix.*

Bootstrap Testing I

We proceed as follows.

1. Generate an iid sequence $\{\eta_t\}$ whose mean is zero and variance is one.
2. Construct $\{y_t^*\}$ by
$$y_t^* = x_t'\widehat{\beta} + \eta_t \widehat{\varepsilon}_t.$$
3. Construct the bootstrap statistic*

$$\begin{aligned} \text{supQ}^* &= \sup_{\gamma \in \Gamma} Q_T^*(\gamma) = \sup_{\gamma \in \Gamma} T \frac{\min_{\alpha:\delta=0} \ \tilde{\mathbb{S}}_T^*(\alpha,\gamma) - \min_\alpha \ \tilde{\mathbb{S}}_T^*(\alpha,\gamma)}{\min_\alpha \ \tilde{\mathbb{S}}_T^*(\alpha,\gamma)} \\ &= T \frac{\min_{\alpha:\delta=0} \tilde{\mathbb{S}}_T^*(\alpha,\gamma) - \tilde{\mathbb{S}}_T^*(\hat{\alpha}^*,\hat{\gamma}^*)}{\tilde{\mathbb{S}}_T^*(\hat{\alpha}^*,\hat{\gamma}^*)}, \end{aligned}$$

where $(\hat{\alpha}^*, \hat{\gamma}^*)$ is a minimizer of

$$\tilde{\mathbb{S}}_T^*(\alpha,\gamma) = \frac{1}{T}\sum_{t=1}^{T}\left(y_t^* - x_t'\beta - x_t'\delta 1\left\{\tilde{f}_t'\gamma > 0\right\}\right)^2.$$

Bootstrap Testing II

4. Repeat 1-3 many times and compute the empirical distribution of $\sup Q^*$

Then, with the obtained empirical distribution, say $F_T^*(\cdot)$, one can compute the bootstrap p-value by

$$p^* = 1 - F_T^*(\sup Q),$$

or a-level critical value

$$c_a^* = {F_T^*}^{-1}(1 - a).$$

Conclusion

- We generalize the threshold regression model to allow for *multiple (estimated) factors* as the threshold index.
- We develop efficient computation algorithms by means of the mixed integer optimization (MIO).
- We specifically analyze the model with factors $\hat{f}_t$ estimated by the principal component analysis using a large panel of macro and financial series, such as Stock and Watson's dataset.
- Formal phase transition of the asymptotic distribution to the oracle state is developed as the N/T ratio increases, where N and T are the number of cross sections and the time span, respectively.