Statistical Inference in Regression Models with Possibly Integrated Processes（未定稿）

千木良 弘朗　山本 拓

東北大学　　一橋大学

研究集会「経済統計・政府統計の理論と応用からの提言」

2019年1月31日 東京大学

1 はじめに

モチベーション

フィリップス曲線の推定・検定:

$$inf_t = \alpha + \beta unem_t + \varepsilon_t,\ t = 1, \cdots, T$$

データ:

$$\begin{cases} inf_t \cdots \text{percentage change in CPI} \\ unem_t \cdots \text{civilian unemployment rate, \%} \\ \text{年次データで}1948 \sim 1996\text{年、}T = 49 \end{cases}$$

⇒ OLS推定の結果（括弧内はt-値）：

$$inf_t = \underset{(0.85)}{1.42} + \underset{(1.65)}{0.47}\, unem_t\text{、}\bar{R}^2 = 0.03$$

⇒ 負のはずの$\hat{\beta}$が正で、しかもあまり有意でない

何かおかしい?

ε_tの系列相関?
ε_tのAR(1)係数は0.57、t-値は5.03で強く有意

$unem_t$の非定常性?
AR(1)係数は0.73、有意水準5%のADF検定でH_0採択

inf_tの非定常性?
AR(1)係数は0.67、有意水準5%のADF検定でH_0棄却

inf_tと$unem_t$で和分の次数が違う?ADF検定の低検出力のせい?

⇒ 非定常性の問題に、単位根・共和分検定せずに対処したい

フィリップス曲線の成立条件

$$inf_t = \alpha + \beta unem_t + \varepsilon_t$$

$\beta < 0$の場合

		ε_tの次数 $I(1)$	ε_tの次数 $I(0)$
$unem_t$の次数	$I(1)$	$inf_t \sim \begin{cases} I(1): \text{見せかけの回帰} \\ I(0): - \end{cases}$ 不成立	$inf_t \sim \begin{cases} I(1): \text{共和分} \\ I(0): - \end{cases}$ 成立
	$I(0)$	$inf_t \sim \begin{cases} I(1): \text{和分の次数が異なる} \\ I(0): - \end{cases}$ 不成立	$inf_t \sim \begin{cases} I(1): - \\ I(0): \text{定常で関係あり} \end{cases}$ 成立

$\beta = 0$の場合

		ε_tの次数 $I(1)$	ε_tの次数 $I(0)$
$unem_t$の次数	$I(1)$	$inf_t \sim \begin{cases} I(1): 見せかけの回帰 \\ I(0): - \end{cases}$ 不成立	$inf_t \sim \begin{cases} I(1): - \\ I(0): 和分の次数が異なる \end{cases}$ 不成立
	$I(0)$	$inf_t \sim \begin{cases} I(1): 和分の次数が異なる \\ I(0): - \end{cases}$ 不成立	$inf_t \sim \begin{cases} I(1): - \\ I(0): 定常で関係無い \end{cases}$ 不成立

⇒ H_0：不成立 H_1：成立なる検定問題を上手く作り、H_0の棄却を以ってフィリップス曲線を立証

階差を取ればいいのか?

$$inf_t = \alpha + \beta unem_t + \varepsilon_t,\ \varepsilon_t = \rho\varepsilon_{t-1} + u_t,\ u_t \sim I(0)$$

$$\xrightarrow[\text{掛けて整理}]{(1-\rho L)\text{を}} inf_t = (1-\rho)\alpha + \beta unem_t - \beta\rho unem_{t-1} + \rho inf_{t-1} + u_t$$

$$\xrightarrow{\rho = 1\text{と仮定}} \Delta inf_t = \beta\Delta unem_t + u_t$$

⇒
- 階差モデルは、$\rho = 1$なら正しいが、$\rho < 1$ではmis-specification
- もし$\rho = 1$でβが非ゼロでも、$\varepsilon_t \sim I(1)$の意味で「見せかけの回帰」

$$inf_t = \alpha + \beta unem_t + \varepsilon_t,\ \varepsilon_t = \rho\varepsilon_{t-1} + u_t,\ u_t \sim I(0),$$
$inf_t \sim I(1),\ unem_t \sim I(1),\ \rho < 1$ ···共和分

$\rightarrow \Delta inf_t = \beta \Delta unem_t + u_t$

⇒ 階差モデルは、共和分の時にはmisspecification

$$inf_t = \alpha + \beta unem_t + \varepsilon_t,\ \varepsilon_t = \rho\varepsilon_{t-1} + u_t,\ u_t \sim I(0),$$
$$\begin{cases} inf_t \sim I(1),\ unem_t \sim I(0) \\ inf_t \sim I(0),\ unem_t \sim I(1) \end{cases}$$

$\rightarrow \Delta inf_t = \beta \Delta unem_t + u_t$

⇒ 階差を取ると、inf_tと$unem_t$で和分の次数が異なっていても「フィリップス曲線が成立」となりかねない

$$\begin{cases} inf_t\text{と}unem_t\text{の関係式} \cdots \text{フィリップス曲線} \\ \\ \Delta inf_t\text{と}\Delta unem_t\text{の関係式} \\ (inf_t - inf_t^e\text{と}unem_t - \mu_0) \end{cases} \cdots \text{expectations augmented フィリップス曲線}$$

⇒ 階差を取ると経済学的インプリケーションが異なる

分析の目的

levelの単純回帰モデル（共和分回帰の最も簡単なケース）

$$y_t = \alpha + \beta x_t + \varepsilon_t,\ \varepsilon_t = \rho\varepsilon_{t-1} + u_t,\ u_t \sim i.i.d.(0, \sigma_u^2),$$
$$-1 < \rho \leq 1 \tag{1}$$
$$x_t = c + \xi_t,\ \xi_t = \phi\xi_{t-1} + v_t,\ v_t \sim i.i.d.(0, \sigma_v^2),$$
$$-1 < \phi \leq 1,\ u_t \perp\!\!\!\perp v_s,\ \forall t, s$$

⇒ 単位根検定・共和分検定・βの有意性検定の3ステップ

→ {単位根・共和分検定が煩雑 / 単位根・共和分検定の誤りにより有意性検定が歪む}

→ 単位根・共和分検定無しで1ステップで有意性検定

発表の構成

2. アイデア
3. 検定統計量の漸近分布
4. モンテカルロ実験
5. まとめと今後の課題

2 アイデア

(1) 式に $(1-\rho L)$ を掛けて整理

$$y_t = (1-\rho)\alpha + \beta x_t - \beta\rho x_{t-1} + \rho y_{t-1} + u_t,\ u_t \sim i.i.d.(0, \sigma_u^2) \tag{2}$$

⇒ (2) 式で次の検定問題考える。

$$H_0 : \beta + (-\beta\rho) = 0 \ \ H_1 : \beta + (-\beta\rho) \neq 0 \tag{3}$$

→ この H_0 は、次の表の ◯ で正しく、× で正しくない

$$H_0 : \beta + (-\beta\rho) = 0 \quad H_1 : \beta + (-\beta\rho) \neq 0$$

$\beta \neq 0$の場合

ϕ（x_tの次数）		ρ（ε_tの次数）	
		1（$I(1)$）	$-1 \sim 1$（$I(0)$）
	1（$I(1)$）	見せかけの回帰 ○	共和分 ×
	$-1 \sim 1$（$I(0)$）	和分の次数が異なる ○	定常で関係あり ×

$\beta = 0$の場合

ϕ（x_tの次数）		ρ（ε_tの次数）	
		1（$I(1)$）	$-1 \sim 1$（$I(0)$）
	1（$I(1)$）	見せかけの回帰 ○	和分の次数が異なる ○
	$-1 \sim 1$（$I(0)$）	和分の次数が異なる ○	定常で関係無い ○

$\Rightarrow$
- ○は「xとyに経済学的に意味のある関係が無い場合」
- ×は「xとyに経済学的に意味のある関係がある場合」

→ (2)式をOLSで推定し、t-検定統計量

$$t = \frac{\hat{\beta} + \widehat{-\beta\rho}}{\sqrt{\widehat{Var(\hat{\beta})} + 2\widehat{Cov(\hat{\beta}, \widehat{-\beta\rho})} + \widehat{Var(\widehat{-\beta\rho})}}} \tag{4}$$

でH_0を検定

⇒ ×に検出力を持つ
　 ○には検出力を持たない

⇒ 予備的な単位根・共和分検定無しで×と○を判別できる

⇒ 単位根・共和分検定は不要で1ステップで「経済学的に意味のある関係」を検出できる

3 検定統計量の漸近分布

(4)式のtの分布は?

- $\rho = \phi = 1$ではnon-standardな分布（Hamilton (1994)）
- ρとϕの一方だけが1でもnon-standard（モンテカルロ実験）

Toda and Yamamoto (1995)のLA（lag-augment）法

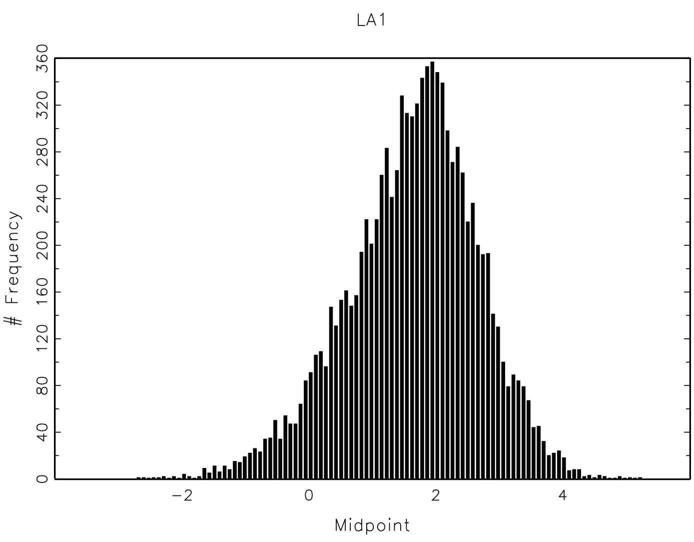
(2)式に、x_{t-2}とy_{t-2}という本来は不要なラグ変数を追加

$$y_t = (1-\rho)\alpha + \beta x_t - \beta\rho x_{t-1} + \rho y_{t-1} + \boxed{0 \times x_{t-2}} + \boxed{0 \times y_{t-2}} + u_t,\ u_t \sim i.i.d.(0, \sigma_u^2) \quad (5)$$

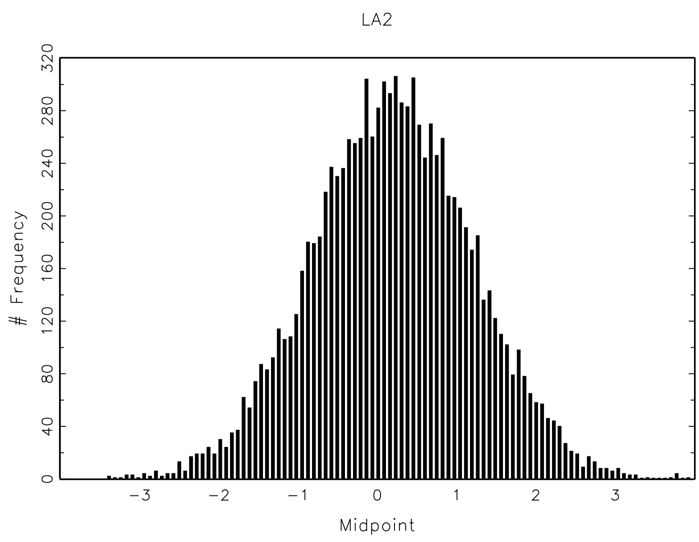
命題（まだ証明はしていない）
(5)式のOLSに基づくt-検定統計量(4)は、(3)のH_0の下で、$\rho = 1$ and/or $\phi = 1$でも、漸近的に標準正規分布に従う。

$\beta = 2$、$\rho = 1$、$\phi = 1$でのtの分布（$T = 1000$、$R = 10000$）

LA無し：(2)式



LAあり：(5)式



4 モンテカルロ実験

実験の設定

DGP:

$$y_t = 10 + \beta x_t + \varepsilon_t,\ \varepsilon_t = \rho\varepsilon_{t-1} + u_t,\ u_t \sim NID(0,1),\ t = 1, \cdots, T$$
$$x_t = 5 + \xi_t,\ \xi_t = \phi\xi_{t-1} + v_t,\ v_t \sim NID(0,1),\ u_t \perp\!\!\!\perp v_s,\ \forall t, s$$

$\beta = \{0, 2\}$、$\rho = \{1, 0.9\}$、$\phi = \{1, 0.9\}$、$T = \{50, 1000\}$、
全ての検定のnominal sizeは5%、実験回数1万回

通常法:	xにDF検定・yにPP検定・y on xの回帰でEG検定 or コクラン-オーカット法
LA法:	(5)式のOLSに基づき、(3)のH_0をt-検定統計量(4)で$N(0,1)$で検定

⇒ 通常法とLA法の「成功率」を比較する。「成功率」の定義は脚注1~8

実験結果

$\beta = 2$の場合

	ρ（ε_tの次数）					
	1			0.9		
ϕ（x_tの次数） 1		$T = 50$	$T = 1000$		$T = 50$	$T = 1000$
	通常法[1]	0.86	0.87	通常法[6]	0.10	0.92
	LA法[2]	0.85	0.95	LA法[7]	0.25	0.83
	（見せかけの回帰）			（共和分）		
0.9		$T = 50$	$T = 1000$		$T = 50$	$T = 1000$
	通常法[3]	0.08	0.49	通常法[8]	0.07	1.00
	LA法[2]	0.85	0.95	LA法[7]	0.23	0.83
	（和分の次数が異なる）			（定常で関係あり）		

[1] x の DF 検定で採択・y の PP 検定で採択・EG 検定で採択で「成功」と判定。
[2] H_0 を採択で「成功」と判定。LA 法の $1-$size のこと。
[3] x の DF 検定で棄却・y の PP 検定で採択で「成功」と判定。
[4] x の DF 検定で採択・y の PP 検定で棄却で「成功」と判定。
[5] x の DF 検定で棄却・y の PP 検定で棄却・コクラン-オーカット法で $\beta = 0$ を採択で「成功」と判定。
[6] x の DF 検定で採択・y の PP 検定で採択・EG 検定で棄却で「成功」と判定。
[7] H_0 を棄却で「成功」と判定。LA 法の power のこと。
[8] x の DF 検定で棄却・y の PP 検定で棄却・コクラン-オーカット法で $\beta = 0$ を棄却で「成功」と判定。

$\beta = 0$の場合

	ρ（ε_tの次数）					
	1			0.9		
ϕ（x_tの次数） 1		$T = 50$	$T = 1000$		$T = 50$	$T = 1000$
	通常法[1]	0.87	0.87	通常法[4]	0.14	0.95
	LA法[2]	0.91	0.95	LA法[2]	0.92	0.95
	（見せかけの回帰）			（和分の次数が異なる）		
0.9		$T = 50$	$T = 1000$		$T = 50$	$T = 1000$
	通常法[3]	0.11	0.95	通常法[5]	0.02	0.95
	LA法[2]	0.92	0.95	LA法[2]	0.91	0.95
	（和分の次数が異なる）			（定常で関係無し）		

$\Rightarrow$ LA法は、　　　では通常法より「成功率」が低いが、他ではほぼ同等か上回る

モンテカルロ実験から解る「通常法」の大きな欠点

実験結果の［　］のように、$T = 1000$でも「成功率」が49%と極めて低い

$$y_t = 10 + 2x_t + \varepsilon_t,\ \varepsilon_t = \varepsilon_{t-1} + u_t \to y_t \sim I(1)$$

→ PP検定でH_0を51%reject（size distortionが46%）

$$x_t = 5 + \xi_t,\ \xi_t = 0.9\xi_{t-1} + v_t \to x_t \sim I(0)$$

→ DF検定でH_0を100%reject（powerは100%）

⇒ なぜy_tへのPP検定が極端に悪いのか?

(2)式より、

$$y_t = (1-1)10 + 2x_t - 2 \times 1x_{t-1} + y_{t-1} + u_t$$
$$= 2\Delta x_t + y_{t-1} + u_t = 2\Delta x_t + y_{t-1} + u_t$$

⇒ PP検定における誤差項は$2\Delta x_t + u_t$であり、Δx_tには過剰階差によりMA単位根が発生

⇓

$2\Delta x_t + u_t$のMAの根は約0.956（u_tが混じるため完全なMA単位根にはならない）

⇒ PP検定はMA単位根が無いのが前提のため、極端にパフォーマンスが悪化

5 まとめと今後の課題

本稿の貢献

元のモデル: $y_t = \alpha + \beta x_t +$ 誤差

$$\Downarrow \text{2期分LA}$$

$$y_t = \beta_1 + \beta_2 x_t + \beta_3 x_{t-1} + \beta_4 y_{t-1} + \beta_5 x_{t-2} + \beta_6 y_{t-2} + \text{誤差}$$

$$\Downarrow$$

$H_0 : \beta_2 + \beta_3 = 0$ をOLSでt-検定

⇒ 誤差項系列相関・単位根・共和分・見せかけの回帰を気にせずに「元のモデルの β が有意か」解る

課題

・LA法では、H_0 を棄却した際に「x と y が共和分」なのか「x も y も定常で関係がある」のか解らない

・実験結果のように、LA法はpowerの上昇が鈍い

・共和分回帰の枠組みでは $u_t \sim MA(\infty)$ とできるが、LA法では $u_t \sim AR(p)$ としてADFタイプの拡張しかできない[9]

・いわゆる共和分回帰の枠組みでは u_t と v_s に相関を許せる（FM-OLS）が、LA法では相関を許すことはできない[10]

・(1)式の重回帰への拡張については、LA法は計量経済学の方法論的には問題無く行えるだろうが、H_0 を棄却した際の経済学的解釈が難しくなる可能性がある

[9] (5)式の誤差項 u_t にMA(∞)の系列相関が残ると、y_{t-1} や y_{t-2} が説明変数にあるため $I(0)$ の世界では一致推定ができなくなるということ。

[10] u_t と v_s の相関は(5)式での説明変数 x_t と誤差項 u_t ということになり、$I(0)$ の世界では一致推定ができなくなるということ。

フィリップス曲線の推定・検定

$$inf_t = \alpha + \beta unem_t + \varepsilon_t,\ t = 1948, \cdots, 1996$$

	α	β	$\bar{R}^2$
推定値	1.42	0.47	0.03
t-値	0.85	1.65	

（単位根検定では $unem_t \sim I(1)$・$inf_t \sim I(0)$）

$$inf_t = (1-\rho)\alpha + \beta \times unem_t - \beta\rho \times unem_{t-1} + \rho \times inf_{t-1} + 0 \times unem_{t-2} + 0 \times inf_{t-2} + u_t$$

	$(1-\rho)\alpha$	β	$-\beta\rho$	ρ	0	0	$\bar{R}^2$
推定値	1.67	-0.44	-0.22	0.84	0.51	0.00	0.58
t-値	1.35	-0.35	-0.57	6.44	-	-	

$H_0 : \beta + (-\beta\rho) = 0$ に対するt-検定統計量 ⋯ -1.93

Covariance Structure Analysis of Panel Regression Models

Kazuhiko Hayakawa
(joint with Takashi Yamagata(U. of York))

Preliminary and incomplete

January 31 2019

Outline

1. Introduction
2. Covariance Structure Analysis of Panel Regression Model
(3. Extension)
4. Monte Carlo simulation
5. Conclusion

1. Introduction

Introduction

- Several panel data models have been proposed in the literature.
- One of the most recent models is the panel regression models with unobserved factor structure or interactive fixed effects.
- The literature can be divided into two categories:
  - Small T and large N
    - GMM: Ahn, Lee and Schmidt (2001, 2013), Robertson and Sarafidis (2015), Hayakawa (2012), etc.
    - ML: Hayakawa, Pesaran and Smith (2018), Bai (2013a,b), **this paper** etc.
  - Large T and large N
    - Pesaran (2006), Bai (2009), Moon and Weidner (2015), Bai (2013a,b), Chudik and Pesaran (2015) etc.
- This paper proposes the maximum likelihood(**ML**) and minimum distance(**MD**) estimators for **small** T and large N panel models with interactive fixed effects based on **covariance structure analysis (CSA)**.

Introduction

- In the literature of panel data analysis, covariance structure analysis has been used in estimation of income process(e.g. Lillard and Willis, 1978; Abowd and Card, 1989)
- However, recently, the covariance structure analysis has been applied to estimation of **panel regression models**.
- Previous studies are
  - Bollen and Brand (2010) :
    - demonstrate that FE/RE panel data model can be estimated by CSA.
    - small T and large N, no theoretical results.
  - Bai (2013a,b):
    - (mainly) panel AR(1) model with standard/interactive FE,
    - (mainly) large T and large N
  - Moral-Benito (2013), Moral-Benito, Allison and Williams (2019) :
    - Dynamic panel model with predetermined regressor and standard FE,
    - small T and large N

Introduction

- This paper proposes a unified approach to estimate for **small** T and large N linear panel regression models that allow
  - **static** / **dynamic** model
  - **endogenous** / **predetermined** / **strictly exogenous** regressors
  - **(known) standard** / **(unknown) interactive** fixed effects
- In this talk, we mainly consider **static**/**dynamic** panel model with **(unknown) interactive** fixed effects and **endogenous**/**strictly exogenous** regressors.
- The novelty of our CSA approach is that we do not need to use instrumental variables even in the presence of endogenous regressor since we use ML and MD.
- Thus, we are free from the weak/many instruments problem.

2. Covariance Structure Analysis of Panel Regression Model

Panel Regression Model

- Let us consider the following model

$$y_{it} = \mu_{y_t} + \beta' \mathbf{x}_{it} + \varepsilon_{it}, \qquad (i = 1, ..., N; t = 1, ..., T) \quad (1)$$

$$\varepsilon_{it} = \mathbf{f}'_t \boldsymbol{\eta}_i + v_{it} \quad (2)$$

  where β and $\mathbf{x}_{it}$ are $K_x \times 1$, unobserved non-random factor $\mathbf{f}_t$ and its random loading $\boldsymbol{\eta}_i$ are both $m \times 1$. $\mu_{y,t}$ denotes the time effects.
- If $\mathbf{f}_t = 1$ with $m = 1$, this model becomes the standard panel data model.
- Thus, this model can be seen as a generalization of standard panel data model.
- When y_{it} is wage, $\mathbf{f}_t$ is business condition and $\boldsymbol{\eta}_i$ is unobserved ability of individual i, this model allows time-varying effect of ability to wage that is driven by business condition.

Panel Regression Model

- In a matrix form, the model can be written as

$$\mathbf{y}_i = \boldsymbol{\mu}_y + \mathbf{X}_i\beta + \varepsilon_i \qquad (i = 1, ..., N) \tag{3}$$

$$\varepsilon_i = \mathbf{F}\boldsymbol{\eta}_i + \mathbf{v}_i \tag{4}$$

where $\mathbf{y}_i = (y_{i1}, ..., y_{iT})'$, $\boldsymbol{\mu}_y = (\mu_{y_1}, ..., \mu_{y_T})'$, $\mathbf{X}_i = (\mathbf{x}_{i1}, ..., \mathbf{x}_{iT})'$, $\mathbf{F} = (\mathbf{f}_1, ..., \mathbf{f}_T)'$, $\mathbf{v}_i = (v_{i1}, ..., v_{iT})'$.
- Since $\mathbf{F}\boldsymbol{\eta}_i$ can be written as

$$\mathbf{F}\boldsymbol{\eta}_i = (\mathbf{FC})(\mathbf{C}^{-1}\boldsymbol{\eta}_i) = \widetilde{\mathbf{F}}\widetilde{\boldsymbol{\eta}}_i \tag{5}$$

for any $m \times m$ invertible matrix, $\mathbf{F}\boldsymbol{\eta}_i$ can not be identified without restriction.
- We use the following normalization for identification in estimation:

$$\mathbf{F} = \begin{bmatrix} \mathbf{F}_1 \\ \mathbf{I}_m \end{bmatrix} \tag{6}$$

- In the following, we provide assumptions for v_{it}, $\boldsymbol{\eta}_i$ and $\mathbf{x}_{it}$.

Panel Regression Model

- We assume that v_{it} and $\boldsymbol{\eta}_i$ are independent over i, $E(v_{it}) = 0$, and

$$Cov(v_{is}, v_{it}) = \begin{cases} 0 & s \neq t \\ \sigma^2_{v_t} & s = t \end{cases}, \qquad s, t = 1, ..., T \tag{7}$$

$$Var(\boldsymbol{\eta}_i) = \boldsymbol{\Sigma}_{\eta\eta} > 0 \tag{8}$$

$$Cov(v_{it}, \boldsymbol{\eta}_i) = \mathbf{0}, \qquad t = 1, ..., T \tag{9}$$

- No serial correlation, but time-series heteroskedasticity is allowed.
- Although it is possible to allow for cross-sectional heteroskedasticity such that $\bar{\sigma}^2_{v_t,N} = \frac{1}{N}\sum_{i=1}^{N} \sigma^2_{v_t,i}$, we do not consider this case to simplify the theoretical discussion.

Panel Regression Model

- For the process of $\mathbf{x}_{it}$, we assume

$$\mathbf{x}_{it} = \boldsymbol{\mu}_{x_t} + \boldsymbol{\xi}_{x_t,i}, \qquad (t = 1, 2, ..., T) \tag{10}$$

  where $E(\mathbf{x}_{it}) = \boldsymbol{\mu}_{x_t}$ and $\boldsymbol{\xi}_{x_t,i}$ is a continuous variable which is indepdent over i.

- We also let

$$Cov(\mathbf{x}_{it}, \boldsymbol{\eta}_i) = \boldsymbol{\Sigma}_{x_t\eta}, \qquad t = 1, ..., T \tag{11}$$

$$Cov(\mathbf{x}_{is}, \mathbf{x}_{it}) = E\left(\boldsymbol{\xi}_{x_s,i}\boldsymbol{\xi}'_{x_t,i}\right) = \boldsymbol{\Sigma}_{x_s x_t}, \qquad s, t = 1, ..., T \tag{12}$$

- Note that $\mathbf{x}_{it}$ is correlated with unobserved individual effects $\boldsymbol{\eta}_i$ in an unrestricted way as in FE model.

Panel Regression Model

- For the correlation between $\mathbf{x}_{is}$ and v_{it}, we consider two cases:
  - Strictly exogenous

$$Cov(\mathbf{x}_{is}, v_{it}) = \mathbf{0}, \qquad s, t = 1, ..., T \tag{13}$$

  - Endogenous

$$Cov(\mathbf{x}_{is}, v_{it}) = \begin{cases} \mathbf{0} & s < t \\ \boldsymbol{\sigma}_{x_s v_t} & s \geq t \end{cases}, \qquad s, t = 1, ..., T \tag{14}$$

Panel Regression Model

- When $T = 3$, the general form of covariance between $\mathbf{x}_i = (\mathbf{x}'_{i1}, \mathbf{x}'_{i2}, \mathbf{x}'_{i3})'$ and $\mathbf{v}_i = (v_{i1}, v_{i2}, v_{i3})'$ is given by

$$\boldsymbol{\Sigma}_{xv} = E(\mathbf{x}_i \mathbf{v}'_i) = \begin{bmatrix} E(\mathbf{x}_{i1} v_{i1}) & E(\mathbf{x}_{i1} v_{i2}) & E(\mathbf{x}_{i1} v_{i3}) \\ E(\mathbf{x}_{i2} v_{i1}) & E(\mathbf{x}_{i2} v_{i2}) & E(\mathbf{x}_{i2} v_{i3}) \\ E(\mathbf{x}_{i3} v_{i1}) & E(\mathbf{x}_{i3} v_{i2}) & E(\mathbf{x}_{i3} v_{i3}) \end{bmatrix} \quad (15)$$

- Then, the form of $\boldsymbol{\Sigma}_{xv}$ for each case is given by
  - Strictly exogenous
  - Endogenous

$$\boldsymbol{\Sigma}_{xv} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}, \qquad \boldsymbol{\Sigma}_{xv} = \begin{bmatrix} \boldsymbol{\sigma}_{x_1 v_1} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{\sigma}_{x_2 v_1} & \boldsymbol{\sigma}_{x_2 v_2} & \mathbf{0} \\ \boldsymbol{\sigma}_{x_3 v_1} & \boldsymbol{\sigma}_{x_3 v_2} & \boldsymbol{\sigma}_{x_3 v_3} \end{bmatrix}$$

- Upper triangular part of $\boldsymbol{\Sigma}_{xv}$ is assumed to be zero since it is natural to consider that errors v_{it} are not correlated with past covariates $x_{is}, (s < t)$.
- In the following, we mainly consider the "endogenous" case since it is most general.

Panel Regression Model

- To introduce the CSA approach, let us stack the observations and latent variables as follows

$$\begin{aligned} \mathbf{z}_i &= (y_{i1}, ... y_{iT}, \mathbf{x}'_{i1}, ..., \mathbf{x}'_{iT},)' = (\mathbf{y}'_i, \mathbf{x}'_i)' \quad (p \times 1) && (16) \\ \mathbf{u}_i &= (\varepsilon_{i1}, ..., \varepsilon_{iT}, \boldsymbol{\xi}'_{x_1,i}, ..., \boldsymbol{\xi}'_{x_T,i}) = (\varepsilon'_i, \boldsymbol{\xi}'_{x,i})' && (17) \end{aligned}$$

where $p = p_1 + p_2$ with $p_1 = T$ and $p_2 = TK_x$.
- The model can be written as

$$\mathbf{B}\mathbf{z}_i = \boldsymbol{\mu} + \mathbf{u}_i \quad (18)$$

where $\boldsymbol{\mu} = (\mu_{y_1}, ..., \mu_{y_T}, \boldsymbol{\mu}'_{x_1}, ..., \boldsymbol{\mu}'_{x_T})'$ and

$$\mathbf{B} = \left[\begin{array}{ccc|ccc} 1 & \cdots & 0 & -\boldsymbol{\beta}' & \cdots & \mathbf{0}' \\ \vdots & \ddots & 0 & \vdots & \ddots & \mathbf{0}' \\ 0 & 0 & 1 & \mathbf{0}' & \cdots & -\boldsymbol{\beta}' \\ \hline \mathbf{0} & \cdots & \mathbf{0} & & & \\ \vdots & & \vdots & & \mathbf{I}_{p_2} & \\ \mathbf{0} & \cdots & \mathbf{0} & & & \end{array}\right] = \begin{bmatrix} \mathbf{I}_T & \mathbf{B}_{12} \\ \mathbf{0} & \mathbf{I}_{p_2} \end{bmatrix}, \quad (19)$$

Panel Regression Model

- Then, we have
$$\mathbf{z}_i = \mathbf{B}^{-1}\boldsymbol{\mu} + \mathbf{B}^{-1}\mathbf{u}_i \tag{20}$$
- The covariance matrix of $\mathbf{z}_i$ can be written as

$$\begin{aligned} Var(\mathbf{z}_i) &= \boldsymbol{\Sigma}_{zz}(\theta_{all}) = \mathbf{B}^{-1}\boldsymbol{\Sigma}_{uu}\mathbf{B}^{-1\prime} = \begin{bmatrix} \boldsymbol{\Sigma}_{yy} & \boldsymbol{\Sigma}_{yx} \\ \boldsymbol{\Sigma}_{xy} & \boldsymbol{\Sigma}_{xx} \end{bmatrix} \qquad (21) \\ &= \begin{bmatrix} \boldsymbol{\Sigma}_{\varepsilon\varepsilon} + \mathbf{B}_{12}\boldsymbol{\Sigma}_{xx}\mathbf{B}_{12}' - \boldsymbol{\Sigma}_{x\varepsilon}'\mathbf{B}_{12}' - \mathbf{B}_{12}\boldsymbol{\Sigma}_{x\varepsilon} & \boldsymbol{\Sigma}_{x\varepsilon}' - \mathbf{B}_{12}\boldsymbol{\Sigma}_{xx} \\ \boldsymbol{\Sigma}_{x\varepsilon} - \boldsymbol{\Sigma}_{xx}\mathbf{B}_{12}' & \boldsymbol{\Sigma}_{xx} \end{bmatrix} \end{aligned}$$

where

$$\boldsymbol{\Sigma}_{uu} = Var(\mathbf{u}_i) = \begin{bmatrix} Var(\varepsilon_i) & Cov(\mathbf{x}_i, \varepsilon_i)' \\ Cov(\mathbf{x}_i, \varepsilon_i) & Var(\mathbf{x}_i) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{\varepsilon\varepsilon} & \boldsymbol{\Sigma}_{x\varepsilon}' \\ \boldsymbol{\Sigma}_{x\varepsilon} & \boldsymbol{\Sigma}_{xx} \end{bmatrix} \tag{22}$$

$$\underset{(T\times T)}{\boldsymbol{\Sigma}_{\varepsilon\varepsilon}} = \boldsymbol{\Sigma}_{vv} + \mathbf{F}\boldsymbol{\Sigma}_{\eta\eta}\mathbf{F}' \tag{23}$$

$$\underset{(T\times T)}{\boldsymbol{\Sigma}_{vv}} = \text{diag}\left[\sigma_{v_1}^2, ..., \sigma_{v_T}^2\right], \qquad \mathbf{F} = (\mathbf{f}_1, ..., \mathbf{f}_T)' \tag{24}$$

Panel Regression Model

$$\underset{(p_2\times T)}{\boldsymbol{\Sigma}_{x\varepsilon}} = Cov(\mathbf{x}_i, \varepsilon_i) = \boldsymbol{\Sigma}_{xv} + \boldsymbol{\Sigma}_{x\eta}\mathbf{F}' \tag{25}$$

$$\underset{(p_2\times T)}{\boldsymbol{\Sigma}_{xv}} = \begin{bmatrix} \boldsymbol{\sigma}_{x_1v_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \boldsymbol{\sigma}_{x_2v_1} & \boldsymbol{\sigma}_{x_2v_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \boldsymbol{\sigma}_{x_Tv_1} & \cdots & \boldsymbol{\sigma}_{x_Tv_{T-1}} & \boldsymbol{\sigma}_{x_Tv_T} \end{bmatrix}, \quad \underset{(p_2\times m)}{\boldsymbol{\Sigma}_{x\eta}} = \begin{bmatrix} \boldsymbol{\Sigma}_{x_1\eta} \\ \vdots \\ \boldsymbol{\Sigma}_{x_T\eta} \end{bmatrix} \tag{26}$$

$$\underset{(p_2\times p_2)}{\boldsymbol{\Sigma}_{xx}} = \begin{bmatrix} \boldsymbol{\Sigma}_{x_1x_1} & \cdots & \boldsymbol{\Sigma}_{x_1x_T} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{x_Tx_1} & \cdots & \boldsymbol{\Sigma}_{x_Tx_T} \end{bmatrix} \tag{27}$$

Panel Regression Model

- By using the sample covariance matrix of $\mathbf{z}_i$, $\mathbf{S}_N$, we estimate the following parameters

$$\boldsymbol{\theta}_{all} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')' = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_{\varepsilon\varepsilon}', \boldsymbol{\theta}_{x\varepsilon}', \boldsymbol{\theta}_{xx}')' \quad (28)$$

where

$$\boldsymbol{\theta}_1 = (\boldsymbol{\beta}', \text{vec}(\mathbf{F}_1)')'$$
$$\boldsymbol{\theta}_2 = (\boldsymbol{\theta}_{\varepsilon\varepsilon}', \boldsymbol{\theta}_{x\varepsilon}', \boldsymbol{\theta}_{xx}')'$$

with

$$\boldsymbol{\theta}_{\varepsilon\varepsilon} = (\sigma_{v_1}^2, ..., \sigma_{v_T}^2, \text{vech}(\boldsymbol{\Sigma}_{\eta\eta}))'$$
$$\boldsymbol{\theta}_{x\varepsilon} = (\text{vec}(\boldsymbol{\Sigma}_{x_1\eta})', ..., \text{vec}(\boldsymbol{\Sigma}_{x_T\eta})', \boldsymbol{\sigma}_{x_1 v_1}', \boldsymbol{\sigma}_{x_2 v_1}', ..., \boldsymbol{\sigma}_{x_T v_T}',)'$$
$$\boldsymbol{\theta}_{xx} = \text{vech}(\boldsymbol{\Sigma}_{xx})$$

Panel Regression Model

- The order condition is given by $p(p+1)/2 \geq \dim(\boldsymbol{\theta}_{all})$.
- The minimum value of T required for order condition depends on m and the exogeneity property of $\mathbf{x}_{it}$.

Minimum number of T required for order condition ($K_x = 1, 2, 3$)

	$m=1$	$m=2$	$m=3$
Endogenous	3	5	6
Strictly exogenous	2	3	4

- Throughout the paper, we assume that the order condition is satisfied.

- Next, we consider estimation of $\boldsymbol{\theta}_{all}$ based on ML and MD.

ML/MD estimators

ML

- The likelihood function associated with $\mathbf{z}_i$ can be written as

$$\log L_{ML}(\boldsymbol{\theta}_{all}) = -\frac{N}{2}\log|\boldsymbol{\Sigma}_{zz}(\boldsymbol{\theta}_{all})| - \frac{N}{2}\operatorname{tr}\left[\mathbf{S}_N \boldsymbol{\Sigma}_{zz}(\boldsymbol{\theta}_{all})^{-1}\right] \quad (29)$$

where $\mathbf{S}_N$ is sample covariance matrix of $\mathbf{z}_i$.
- The (Q)ML estimator is defined as

$$\widehat{\boldsymbol{\theta}}_{all}^{ML} = \underset{\theta_{all}}{\operatorname{argmax}} \log L_{ML}(\boldsymbol{\theta}_{all})$$

- The dimension of $\boldsymbol{\theta}_{xx}$ can be very large$(K_x T(K_x T+1)/2)$.
- Solution for $\boldsymbol{\Sigma}_{xx}$ is given by

$$\boldsymbol{\Sigma}_{xx} = \frac{1}{N}\left(\mathbf{X} - \mathcal{E}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{-1}\boldsymbol{\Sigma}_{\varepsilon x}\right)'\left(\mathbf{X} - \mathcal{E}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{-1}\boldsymbol{\Sigma}_{\varepsilon x}\right) + \boldsymbol{\Sigma}_{x\varepsilon}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{-1}\boldsymbol{\Sigma}_{\varepsilon x} \quad (30)$$

$$\underset{(N\times T)}{\mathcal{E}} = \begin{bmatrix} \varepsilon_{11} & \cdots & \varepsilon_{1T} \\ \vdots & & \vdots \\ \varepsilon_{N1} & \cdots & \varepsilon_{NT} \end{bmatrix}, \quad \underset{(N\times K_x T)}{\mathbf{X}} = \begin{bmatrix} \mathbf{x}_{11}' & \cdots & \mathbf{x}_{1T}' \\ \vdots & & \vdots \\ \mathbf{x}_{N1}' & \cdots & \mathbf{x}_{NT}' \end{bmatrix}$$

ML/MD estimators

- After concentrating out $\boldsymbol{\theta}_{xx}$, we have

$$\begin{aligned}\log L_{ML}^{con}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_{\varepsilon\varepsilon}, \boldsymbol{\theta}_{x\varepsilon}) &= -\frac{N}{2}\log|\boldsymbol{\Sigma}_{\varepsilon\varepsilon}| - \frac{1}{2}\operatorname{tr}\left[\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{-1}\mathcal{E}'\mathcal{E}\right] \\ &\quad + \frac{N}{2}\log\left|\frac{1}{N}\left(\mathbf{X} - \mathcal{E}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{-1}\boldsymbol{\Sigma}_{\varepsilon x}\right)'\left(\mathbf{X} - \mathcal{E}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{-1}\boldsymbol{\Sigma}_{\varepsilon x}\right)\right|.\end{aligned}$$

ML/MD estimators

MD

- Next, we consider minimum distance(MD) estimators.
- Let us consider the following moment condition:

$$E[\mathbf{s}_i - \boldsymbol{\sigma}_{zz}(\boldsymbol{\theta}_{all})] = \mathbf{0} \tag{31}$$

where $\mathbf{s}_i = \text{vech}[(\mathbf{z}_i - \bar{\mathbf{z}})(\mathbf{z}_i - \bar{\mathbf{z}})']$, $\bar{\mathbf{z}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{z}_i$ and $\boldsymbol{\sigma}_{zz}(\boldsymbol{\theta}_{all}) = \text{vech}\left(\boldsymbol{\Sigma}_{zz}(\boldsymbol{\theta}_{all})\right)$.

- The MD estimator is defined as

$$\widehat{\boldsymbol{\theta}}_{all}^{MD} = \underset{\theta_{all}}{\text{argmin}}[\mathbf{s}_N - \boldsymbol{\sigma}_{zz}(\boldsymbol{\theta}_{all})]'\mathbf{W}_N[\mathbf{s}_N - \boldsymbol{\sigma}_{zz}(\boldsymbol{\theta}_{all})] \tag{32}$$

where $\mathbf{s}_N = \text{vech}(\mathbf{S}_N)$, $\mathbf{W}_N$ is a positive definite weighting matrix.

ML/MD estimators

- We consider three MD estimators depending on the choice of $\mathbf{W}_N$.
- Since the optimal weighting matrix under normality is given by

$$\mathbf{W}_0 = \mathbf{W}(\boldsymbol{\theta}_{all0}) = \frac{1}{2}\mathcal{D}_p'\left(\boldsymbol{\Sigma}(\boldsymbol{\theta}_{all0})^{-1} \otimes \boldsymbol{\Sigma}(\boldsymbol{\theta}_{all0})^{-1}\right)\mathcal{D}_p \tag{33}$$

we consider two weighting matrices which are optimal under normality:

$$\mathbf{W}_{N,1} = \frac{1}{2}\mathcal{D}_p'\left(\mathbf{S}_N^{-1} \otimes \mathbf{S}_N^{-1}\right)\mathcal{D}_p \tag{34}$$

$$\mathbf{W}_{N,2} = \frac{1}{2}\mathcal{D}_p'\left(\boldsymbol{\Sigma}\left(\widetilde{\boldsymbol{\theta}}_{all}\right)^{-1} \otimes \boldsymbol{\Sigma}\left(\widetilde{\boldsymbol{\theta}}_{all}\right)^{-1}\right)\mathcal{D}_p \tag{35}$$

where $\text{vec}(\mathbf{A}) = \mathcal{D}_p\,\text{vech}(\mathbf{A})$ and $\widetilde{\boldsymbol{\theta}}_{all}$ is a preliminary estimator of $\boldsymbol{\theta}_{all}$.

- Since the weighting matrix that is optimal under both normality and non-normality is given by $\boldsymbol{\Omega} = Var(\mathbf{s}_i)$, we consider

$$\mathbf{W}_{N,opt} = \left(\frac{1}{N}\sum_{i=1}^{N}(\mathbf{s}_i - \bar{\mathbf{s}})(\mathbf{s}_i - \bar{\mathbf{s}})'\right)^{-1}, \qquad \bar{\mathbf{s}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{s}_i \tag{36}$$

ML/MD estimators

- Three MD estimators, "MD1", "MD2" and "OMD" are defined as

$$\widehat{\boldsymbol{\theta}}_{all}^{MD1} = \underset{\theta_{all}}{\operatorname{argmin}}[\mathbf{s}_N - \boldsymbol{\sigma}_{zz}(\boldsymbol{\theta}_{all})]'\mathbf{W}_{N,1}[\mathbf{s}_N - \boldsymbol{\sigma}_{zz}(\boldsymbol{\theta}_{all})] \quad (37)$$

$$\widehat{\boldsymbol{\theta}}_{all}^{MD2} = \underset{\theta_{all}}{\operatorname{argmin}}[\mathbf{s}_N - \boldsymbol{\sigma}_{zz}(\boldsymbol{\theta}_{all})]'\mathbf{W}_{N,2}[\mathbf{s}_N - \boldsymbol{\sigma}_{zz}(\boldsymbol{\theta}_{all})] \quad (38)$$

$$\widehat{\boldsymbol{\theta}}_{all}^{OMD} = \underset{\theta_{all}}{\operatorname{argmin}}[\mathbf{s}_N - \boldsymbol{\sigma}_{zz}(\boldsymbol{\theta}_{all})]'\mathbf{W}_{N,opt}[\mathbf{s}_N - \boldsymbol{\sigma}_{zz}(\boldsymbol{\theta}_{all})] \quad (39)$$

- Note that $\widehat{\boldsymbol{\theta}}_{MD1}$ is a one-step estimator whereas $\widehat{\boldsymbol{\theta}}_{MD2}$ is a two-step estimator.
- In practice, we can set $\widetilde{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}_{MD1}$ to obtain $\mathbf{W}_{N,2}$.
- Next, we consider concentration out of MD estimators.

ML/MD estimators

- To derive concentrated out MD estimators, we introduce a new vech operator that changes the order of a vector obtained by standard vech.

vech* operator

Let us define a $p \times p$ symmetric matrix $\boldsymbol{\Sigma}$ as follows:

$$\underset{(p\times p)}{\boldsymbol{\Sigma}} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{21}' \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \quad (40)$$

where $\boldsymbol{\Sigma}$ is $p \times p$, $\boldsymbol{\Sigma}_{11}$ is $p_1 \times p_1$, $\boldsymbol{\Sigma}_{21}$ is $p_2 \times p_1$,$\boldsymbol{\Sigma}_{22}$ is $p_2 \times p_2$.
Define the vech* operator as follows:

$$\text{vech}^*(\boldsymbol{\Sigma}) = \begin{bmatrix} \text{vech}(\boldsymbol{\Sigma}_{11}) \\ \text{vec}(\boldsymbol{\Sigma}_{21}) \\ \text{vech}(\boldsymbol{\Sigma}_{22}) \end{bmatrix} \quad (41)$$

ML/MD estimators

vech* operator

Note that $\text{vech}^*(\boldsymbol{\Sigma})$ and $\text{vech}(\boldsymbol{\Sigma})$ have a relationship as follows:

$$\text{vech}(\boldsymbol{\Sigma}) = \mathcal{R}_{p_1,p_2}\,\text{vech}^*(\boldsymbol{\Sigma}) \tag{42}$$

where

$$\mathcal{R}_{p_1,p_2} = \mathcal{D}_p^+ \begin{bmatrix} \mathcal{K}_{p_1,p} & \mathbf{0} \\ \mathbf{0} & \mathcal{K}_{p_2,p} \end{bmatrix} \begin{bmatrix} \mathcal{D}_{p_1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathcal{K}_{p_2,p_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{p_1 p_2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathcal{D}_{p_2} \end{bmatrix} \tag{43}$$

and $\mathcal{K}_{m,n}$ is a commutation matrix such that $\text{vec}(\mathbf{B}') = \mathcal{K}_{m,n}\,\text{vec}(\mathbf{B})$ for an $m \times n$ matrix $\mathbf{B}$.
Note that $\mathcal{R}_{p_1,p_2}$ is $p(p+1)/2$ dimensional full rank square matrix of zeros or ones.

ML/MD estimators

- Using this new vech* operator, we have

$$\begin{aligned} \boldsymbol{\sigma}_{zz}(\boldsymbol{\theta}) &= \text{vech}(\boldsymbol{\Sigma}_{zz}(\boldsymbol{\theta})) = \mathcal{R}_{T,p_2}\,\text{vech}^*(\boldsymbol{\Sigma}_{zz}(\boldsymbol{\theta})) && (44) \\ &= \mathcal{R}_{T,p_2} \begin{bmatrix} \text{vech}(\boldsymbol{\Sigma}_{yy}) \\ \text{vec}(\boldsymbol{\Sigma}_{xy}) \\ \text{vech}(\boldsymbol{\Sigma}_{xx}) \end{bmatrix} = \mathcal{R}_{T,p_2}\mathbf{A}(\boldsymbol{\theta}_1)\,\boldsymbol{\theta}_2 && (45) \end{aligned}$$

where $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are defined in slide 17

$$\mathbf{A}(\boldsymbol{\theta}_1) = \begin{bmatrix} \mathbf{C}_{\varepsilon\varepsilon}(\mathbf{F}) & -2\mathcal{D}_T^+(\mathbf{I}_T \otimes \mathbf{B}_{12})\,\mathbf{C}_{x\varepsilon}(\mathbf{F}) & \mathcal{D}_T^+(\mathbf{B}_{12} \otimes \mathbf{B}_{12})\,\mathcal{D}_{p_2} \\ \mathbf{0} & \mathbf{C}_{x\varepsilon}(\mathbf{F}) & -(\mathbf{B}_{12} \otimes \mathbf{I}_{p_2})\,\mathcal{D}_{p_2} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix} \tag{46}$$

$$\mathbf{B}_{12} = -(\mathbf{I}_T \otimes \beta') \tag{47}$$

$$\mathbf{C}_{\varepsilon\varepsilon}(\mathbf{F}) = \left[\mathbf{J},\ \mathcal{D}_T^+(\mathbf{F} \otimes \mathbf{F})\mathcal{D}_m\right] \tag{48}$$

$$\mathbf{C}_{x\varepsilon}(\mathbf{F}) = \left[(\mathbf{F} \otimes \mathbf{I}_{p_2}),\ \mathcal{D}_{p_2}^{\dagger}\right] \tag{49}$$

where $\mathbf{J} = [\text{vech}(\mathbf{J}_{11}), \ldots, \text{vech}(\mathbf{J}_{TT})]$ and $\mathbf{J}_{st}$ is a $T \times T$ matrix whose (s,t) element is 1 and 0 otherwise.

ML/MD estimators

- Using this alternative expression, we can solve for $\boldsymbol{\theta}_2$ as follows, which is a function of $\boldsymbol{\theta}_1$:

$$\boldsymbol{\theta}_2 = \mathbf{b}(\boldsymbol{\theta}_1) = \left[\mathbf{A}(\boldsymbol{\theta}_1)'\mathcal{R}'_{T,p_2}\mathbf{W}_N\mathcal{R}_{T,p_2}\mathbf{A}(\boldsymbol{\theta}_1)\right]^{-1}\mathbf{A}(\boldsymbol{\theta}_1)'\mathcal{R}'_{T,p_2}\mathbf{W}_N\mathbf{s}_{zz} \tag{50}$$

- Then, we have the concentrated objective function:

$$Q_{MD}^{con}(\boldsymbol{\theta}_1) = [\mathbf{s}_N - \mathcal{R}_{T,p_2}\mathbf{A}(\boldsymbol{\theta}_1)\mathbf{b}(\boldsymbol{\theta}_1)]' \times \mathbf{W}_N[\mathbf{s}_N - \mathcal{R}_{T,p_2}\mathbf{A}(\boldsymbol{\theta}_1)\mathbf{b}(\boldsymbol{\theta}_1)] \tag{51}$$

where $\mathbf{s}_N = \text{vech}(\mathbf{S}_N)$.

- Compared with ML, $\boldsymbol{\theta}_{\varepsilon\varepsilon}$ and $\boldsymbol{\theta}_{x\varepsilon}$ are further concentrated out.
- Therefore, MD is computationally less demanding than ML.

Properties of ML/MD estimators

- Next, we investigate the theoretical properties of ML and MD estimators.
- If $\mathbf{G}(\boldsymbol{\theta}_{all}) = \frac{\partial\boldsymbol{\sigma}_{zz}(\boldsymbol{\theta}_{all})}{\partial\boldsymbol{\theta}'_{all}}$ is of full rank and other regularity conditions are satisfied, the ML and MD estimators are consistent and asymptotically normal:

$$\widehat{\boldsymbol{\theta}}_{all}^{est} \xrightarrow{p} \boldsymbol{\theta}_{all0}, \qquad (est = ML, MD1, MD2, OMD) \tag{52}$$

$$\sqrt{N}\left(\widehat{\boldsymbol{\theta}}_{all}^{est} - \boldsymbol{\theta}_{all0}\right) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}_\theta), \qquad (est = ML, MD1, MD2) \tag{53}$$

where

$$\boldsymbol{\Sigma}_\theta = \begin{cases} (\mathbf{G}_0'\mathbf{W}_0\mathbf{G}_0)^{-1}\mathbf{G}_0'\mathbf{W}_0\boldsymbol{\Omega}\mathbf{W}_0\mathbf{G}_0(\mathbf{G}_0'\mathbf{W}_0\mathbf{G}_0)^{-1} & \text{when } \mathbf{z}_i \text{ is non-normally distributed} \\ (\mathbf{G}_0'\mathbf{W}_0\mathbf{G}_0)^{-1} & \text{when } \mathbf{z}_i \text{ is normally distributed} \end{cases}$$

The asymptotic distribution of the OMD estimator is given by

$$\sqrt{N}\left(\widehat{\boldsymbol{\theta}}_{all}^{OMD} - \boldsymbol{\theta}_{all0}\right) \xrightarrow{d} N\left(\mathbf{0}, (\mathbf{G}_0'\mathbf{W}_0\mathbf{G}_0)^{-1}\right). \tag{54}$$

Properties of ML/MD estimators

- Unfortunately, when $\mathbf{x}_{it}$ is endogenous, $\mathbf{G}(\boldsymbol{\theta})$ is rank deficient.
- However, by reparametrizating $\boldsymbol{\Sigma}_{x\varepsilon}$, we can address the rank deficiency problem(details are in the paper).
- Note that $\widehat{\boldsymbol{\theta}}_{all}^{ML}$ is asymptotically equivalent to $\widehat{\boldsymbol{\theta}}_{all}^{MD1}$ $\widehat{\boldsymbol{\theta}}_{all}^{MD2}$.
- Under normality, all estimators are efficient.
- Under non-normality, ML, MD1 and MD2 are not asymptotically efficient, but OMD is efficient.
- However, in finite samples, OMD is not necessarily more efficient than ML, MD1 and MD2, since OMD requires estimation of fourth order moments $\mathbf{W}_{N,opt}$(Altonji and Segal, 1996).

Properties of ML/MD estimators

- The novel feature of our approach is that since we use ML/MD, we do not need to use instrumental variables even in the presence of endogeneity.
- Intuition behind this is as follows:
- Using the FOC associated with β given by $\frac{\partial \log L_{ML}(\boldsymbol{\theta})}{\partial \beta} = 0$ and noting $\mathbf{S}_{xx} \approx \boldsymbol{\Sigma}_{xx}$, we have

$$\widehat{\beta}_{ML} \approx \underbrace{\frac{\frac{1}{N}\sum_{i=1}^{N}(\mathbf{x}_i - \bar{\mathbf{x}})' \mathbf{V}^{-1} (\mathbf{y}_i - \bar{\mathbf{y}})}{\frac{1}{N}\sum_{i=1}^{N}(\mathbf{x}_i - \bar{\mathbf{x}})' \mathbf{V}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}})}}_{\substack{\text{Inconsistent GLS estimator} \\ \text{with weight } \mathbf{V}}} - \underbrace{\frac{E\left[\mathbf{x}_i' \mathbf{V}^{-1} \varepsilon_i\right]}{\frac{1}{N}\sum_{i=1}^{N}(\mathbf{x}_i - \bar{\mathbf{x}})' \mathbf{V}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}})}}_{\substack{\text{Bias coming from} \\ Cov(x_{it}, \eta_i) \text{ and } Cov(x_{it}, v_{is})}} \tag{55}$$

where $\mathbf{V} = \boldsymbol{\Sigma}_{\varepsilon\varepsilon} - \boldsymbol{\Sigma}_{\varepsilon x}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{x\varepsilon}$
- This correction is possible since we estimate $Cov(\mathbf{x}_{it}, \boldsymbol{\eta}_i) = \boldsymbol{\Sigma}_{x_t\eta}$ and $Cov(\mathbf{x}_{it}, v_{is}) = \boldsymbol{\sigma}_{x_t v_s}$.

Number of factors

- In practice, we have to determine the number of factors.
- The number of factors, m_0, can be estimated by information criterion such as AIC, BIC and HQIC for ML:

$$
\begin{aligned}
AIC &= -2\log L(\theta) + 2q \\
BIC &= -2\log L(\theta) + \log(N)q \\
HQIC &= -2\log L(\theta) + 2.01\log(\log(N))q
\end{aligned}
$$

  where q denotes the number of parameters.
- GMM version of AIC/BIC/HQIC proposed by Andrews and Lu (2001) can be used for MD estimators.

3. Extension

Dynamic panel data model with time-invariant regressor

- We extend the previous model to include dynamics and time-invariant regressors as follows:

$$y_{it} = \alpha y_{it-1} + \beta' \mathbf{x}_{it} + \gamma' \mathbf{w}_i + \mu_{y,t} + \mathbf{f}_t' \boldsymbol{\eta}_i + v_{it}, \qquad (t = 1, ..., T) \quad (56)$$

  where $|\alpha| < 1$, γ and $\mathbf{w}_i$ are $K_w \times 1$.
- For identification of γ, we assume that $\boldsymbol{\iota}_T$ and $\mathbf{F}$ are linearly independent, i.e. $\mathbf{f}_t$ is time-varying.
- For the initial condition y_{i0}, and the time-invariant regressor $\mathbf{w}_i$, we assume

$$y_{i0} = \mu_{y_0} + \xi_{y_0,i}, \quad (57)$$
$$\mathbf{w}_i = \boldsymbol{\mu}_w + \boldsymbol{\xi}_{w,i}, \quad (58)$$

  where $E(y_{i0}) = \mu_{y_0}$, $E(\mathbf{w}_i) = \boldsymbol{\mu}_w$ and $\xi_{y_0,i}$ and $\boldsymbol{\xi}_{w,i}$ are independent over i.

Dynamic panel data model with factor error structure

- For the initial conditions y_{i0}, we assume that

$$Cov(y_{i0}, v_{it}) = 0 \quad (t = 1, ..., T), \qquad Cov(y_{i0}, \boldsymbol{\eta}_i') = \boldsymbol{\sigma}'_{y_0\eta} \quad (59)$$

- There are almost no restrictions in the initial conditions.
  - We do not need to impose a restriction such as mean-stationarity as in the system GMM estimator.
- For the time-invariant regressor $\mathbf{w}_i$, we assume that

$$Cov(\mathbf{w}_i, v_{it}) = \mathbf{0} \quad (t = 1, ..., T), \qquad Cov(\mathbf{w}_i, \boldsymbol{\eta}_i) = \boldsymbol{\Sigma}_{w\eta} \quad (60)$$

- Time-invariant regressor $\mathbf{w}_i$ is allowed to be correlated with unobserved fixed effects $\boldsymbol{\eta}_i$ in an unrestricted way.
- The definitions of other variables are the same as before.

Dynamic panel data model with factor error structure

- The model can be written as

$$\mathbf{B}\mathbf{z}_i = \boldsymbol{\mu} + \mathbf{u}_i$$

where

$$\mathbf{B} = \left[\begin{array}{cccc|cccccc} 1 & 0 & \cdots & 0 & -\alpha & -\boldsymbol{\beta}' & \mathbf{0}' & \cdots & \mathbf{0}' & -\boldsymbol{\gamma}' \\ -\alpha & 1 & & \vdots & 0 & \mathbf{0}' & -\boldsymbol{\beta}' & \ddots & \vdots & \vdots \\ 0 & \ddots & \ddots & \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & \cdots & -\alpha & 1 & 0 & \mathbf{0}' & \cdots & \cdots & -\boldsymbol{\beta}' & -\boldsymbol{\gamma}' \\ \hline 0 & \cdots & \cdots & 0 & & & & & & \\ \vdots & & \vdots & & & \mathbf{I}_{TK_x+K_w+1} & & & & \\ 0 & \cdots & \cdots & 0 & & & & & & \end{array}\right]$$

$$\begin{aligned} \mathbf{z}_i &= \left(y_{i1}, \ldots y_{iT}, y_{i0}, \mathbf{x}'_{i1}, \ldots, \mathbf{x}'_{iT}, \mathbf{w}'_i\right)' = \left(y_{i1}, \ldots y_{iT}, \mathbf{z}'_{2i}\right)', \\ \mathbf{z}_{2i} &= \left(y_{i0}, \mathbf{x}'_{i1}, \ldots, \mathbf{x}'_{iT}, \mathbf{w}'_i\right)' \\ \boldsymbol{\mu} &= \left(\mu_{y_1}, \ldots, \mu_{y_T}, \mu_{y_0}, \boldsymbol{\mu}'_{x_1}, \ldots, \boldsymbol{\mu}'_{x_T}, \boldsymbol{\mu}'_w\right)' \end{aligned}$$

Dynamic panel data model with factor error structure

-

$$\mathbf{u}_i = \begin{bmatrix} \varepsilon_i \\ \xi_{y_0,i} \\ \boldsymbol{\xi}_{x,i} \\ \boldsymbol{\xi}_{w,i} \end{bmatrix} = \begin{bmatrix} \mathbf{F}\boldsymbol{\eta}_i + \mathbf{v}_i \\ \xi_{y_0,i} \\ \boldsymbol{\xi}_{x,i} \\ \boldsymbol{\xi}_{w,i} \end{bmatrix}$$

$$\mathbf{F} = (\mathbf{f}_1, \ldots, \mathbf{f}_T)', \qquad \mathbf{v}_i = (v_{i1}, \ldots v_{iT})', \quad \boldsymbol{\xi}_{x,i} = \left(\boldsymbol{\xi}'_{x_1,i}, \ldots, \boldsymbol{\xi}'_{x_T,i}\right)'$$

- Since $\mathbf{B}$ is invertible, we have

$$\mathbf{z}_i = \mathbf{B}^{-1}\boldsymbol{\mu} + \mathbf{B}^{-1}\mathbf{u}_i$$

- Therefore, noting that $E(\mathbf{z}_i) = \mathbf{B}^{-1}\boldsymbol{\mu}$, we have

$$Var(\mathbf{z}_i) = \boldsymbol{\Sigma}_{zz}(\boldsymbol{\theta}) = \mathbf{B}^{-1}\boldsymbol{\Sigma}_{uu}\left(\mathbf{B}^{-1}\right)'$$

where $\boldsymbol{\Sigma}_{uu}$ is given by

Dynamic panel data model with factor error structure

- $$\boldsymbol{\Sigma}_{uu} = \begin{bmatrix} \boldsymbol{\Sigma}_{\varepsilon\varepsilon} & \\ \boldsymbol{\Sigma}_{z_2\varepsilon} & \boldsymbol{\Sigma}_{z_2z_2} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{vv} + \mathbf{F}\boldsymbol{\Sigma}_{\eta\eta}\mathbf{F}' & \\ \boldsymbol{\Sigma}_{z_2v} + \boldsymbol{\Sigma}_{z_2\eta}\mathbf{F}' & \boldsymbol{\Sigma}_{z_2z_2} \end{bmatrix} \quad (61)$$

  with

  $$\boldsymbol{\Sigma}_{z_2\varepsilon} = E\begin{bmatrix} y_{i0}\varepsilon_i' \\ \xi_{x,i}\varepsilon_i' \\ \xi_{w,i}\varepsilon_i' \end{bmatrix} = \begin{bmatrix} \boldsymbol{\sigma}'_{y_0\eta}\mathbf{F}' \\ \boldsymbol{\Sigma}_{x\eta}\mathbf{F}' + \boldsymbol{\Sigma}_{xv} \\ \boldsymbol{\Sigma}_{w\eta}\mathbf{F}' \end{bmatrix} = \boldsymbol{\Sigma}_{z_2\eta}\mathbf{F}' + \boldsymbol{\Sigma}_{z_2v} \quad (62)$$

  $$\boldsymbol{\Sigma}_{z_2\eta} = \begin{bmatrix} \boldsymbol{\sigma}'_{y_0\eta} \\ \boldsymbol{\Sigma}_{x\eta} \\ \boldsymbol{\Sigma}_{w\eta} \end{bmatrix}, \qquad \boldsymbol{\Sigma}_{z_2v} = \begin{bmatrix} \mathbf{0} \\ \boldsymbol{\Sigma}_{xv} \\ \mathbf{0} \end{bmatrix} \quad (63)$$

- The definitions of $\boldsymbol{\Sigma}_{vv}$, $\boldsymbol{\Sigma}_{\eta\eta}$. $\boldsymbol{\Sigma}_{x\eta}$, $\boldsymbol{\Sigma}_{xv}$ are the same as the static model.
- The estimation procedure is exactly the same as before.
- Note that $\boldsymbol{\theta}_1 = (\alpha, \boldsymbol{\beta}', \boldsymbol{\gamma}', \text{vec}(\mathbf{F}_1)')'$ and $\boldsymbol{\theta}_2$ are remaining parameters such as $\boldsymbol{\sigma}_{y_0\eta}$, vec$(\boldsymbol{\Sigma}_{x\eta})$ etc..

4. Monte Carlo simulation

Static panel data model
Setup

- Consider the following DGP:

$$y_{it} = \mu_{y,t} + \beta x_{it} + \mathbf{f}_t'\boldsymbol{\eta}_i + v_{it}, \qquad (t = 1, ..., T) \tag{64}$$

$$x_{it} = \mu_{x,t} + \rho x_{i,t-1} + \tau_\eta \mathbf{f}_t'\boldsymbol{\eta}_i + \tau_{v0} v_{it} + r_{it} \tag{65}$$

$$x_{i0} = \tilde{\mu}_{x,0} + \tau_\eta \tilde{\mathbf{f}}_0'\boldsymbol{\eta}_i + \varpi_{i0} \tag{66}$$

- For parameter values, we set

$$\beta = 1, \quad \rho = 0.5, \quad \tau_\eta = 0.2, \quad \tau_{v0} = \begin{cases} 0 & \text{strictly exogenous} \\ 0.2 & \text{endogenous} \end{cases}$$

$$Var(v_{it}) = \varsigma_i h_t, \ \varsigma_i \sim \mathcal{U}(0.5, 1.5), \ h_t = 0.5 + \frac{t}{T},$$

$$Var(\varpi_{i0}) = \frac{\tau_{v0}^2 \sigma_{v,i1}^2 + \sigma_r^2}{1 - \rho^2}$$

- $\sigma_r^2 = Var(r_{it})$ is chosen in terms of signal-to-noise ratio(SNR): $SNR = \frac{1}{T}\sum_{t=1}^T \frac{Var(y_{it} - v_{it} | \boldsymbol{\eta}_i, \mathbf{F})}{Var(v_{it})}$ with $SNR = 6$.

Static panel data model
Setup

- The variance of $\boldsymbol{\eta}_i$ is specified as

$$Var(\boldsymbol{\eta}_i) = \begin{cases} 1 & m = 1 \\ \frac{1}{2}\begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} & m = 2 \end{cases} \tag{67}$$

- For the specification of $\mathbf{f}_t$, we let

$$\dot{\mathbf{F}} = \begin{bmatrix} f_{11} & f_{21} \\ \vdots & \vdots \\ f_{1T} & f_{2T} \end{bmatrix} \text{ where } \begin{cases} \dot{f}_{1t} = \sqrt{t} \\ \dot{f}_{2t} = 2|t - (T+1)/2| + 1 \end{cases} \tag{68}$$

- Then,
  - $\mathbf{f}_t = f_{1t}, (m = 1)$ is the first principal component of $\dot{\mathbf{F}}$.
  - $\mathbf{f}_t = (f_{1t}, f_{2t})', (m = 2)$ is the first two principal components of $\dot{\mathbf{F}}$.
  - $\tilde{\mathbf{f}}_0 = \frac{\mathbf{f}_1}{1-\rho}$.
- This ensures that $\frac{1}{T}\sum_{t=1}^T Var(\mathbf{f}_t'\boldsymbol{\eta}_i) = 1$ for any m and T.
- Without loss of generality, we set $\mu_{y,t} = 0$ and $\mu_{x,t} = 0$.

Static panel data model

Setup

- For the sample size, we consider

$$T = \{5\}, \qquad N = \{200, 500, 1000\}$$

- The number of replications is 1,000.
- Significance level is 5%.

Static panel data model

Setup

- We compare six estimators:
  - ML estimator("ML").
  - Three MD estimators("MD1", "MD2", "OMD").
    - "MD1" and "MD2" are asymptotically efficient only when $\mathbf{z}_i$ is normal.
    - "OMD" is efficient regardless the distribution of $\mathbf{z}_i$.
  - One- and two-step GMM estimators proposed by Ahn, Lee and Schmidt (2013) ("GMM1", "GMM2").
    - When x_{it} is strictly exogenous, all periods of x_{it}, i.e. $\mathbf{z}_{it} = (x_{i1}, ..., x_{iT})$ are used as instruments for each period.
    - When x_{it} is endogenous, all lagged x_{it}, i.e. $\mathbf{z}_{it} = (x_{i1}, ..., x_{i,t-1})$ are used as instruments for each period.
    - When x_{it} is endogenous, $m = 2$ and $T = 5$, GMM estimators are not computed since the number of parameters(=5) is larger than that of moment conditions(=3).
- For ML and MD, two standard errors are computed
  - one is assuming normality("Size")
  - the other is robust to non-normality("Size$_{rob}$")

Static panel data model
Setup

- In covariance structure analysis, a multivariate kurtosis is an underlying key parameter which controls a deviation from multivariate normality.
- If data are generated according to this regression form, it is difficult to control the multivariate kurtosis of $\mathbf{z}_i = (\mathbf{y}_i', \mathbf{x}_i')'$.
- Hence, we generate data as follows:

$$\mathbf{z}_i = \mathbf{\Sigma}_{zz,i} (\boldsymbol{\theta}_0)^{1/2} \mathbf{r}_i \tag{69}$$

- For generation of $\mathbf{r}_i$, we follow Yuan and Bentler (1998) and Yanagihara (2007).
- Specifically, $\mathbf{r}_i$ are generated as

$$\mathbf{r}_i \sim \phi_i \mathbf{A} \mathbf{w}_i \tag{70}$$

where $\mathbf{A}$ is a $k \times p$ matrix with $rank(\mathbf{A}) = p$ and $\mathbf{A}'\mathbf{A} = \mathbf{I}_p$. $\mathbf{w}_i = (w_{i1}, ..., w_{ip})'$

Static panel data model
Setup

- We consider two distributions
  - Normal distribution

$$w_{ij} \sim iidN(0,1), \qquad \phi_i = 1 \qquad \mathbf{A} = \mathbf{I}_p, \qquad (\kappa_4 = 0)$$

  - χ^2 distribution

$$\begin{aligned} w_i &\sim (x_i - 4)/\sqrt{8}, \qquad x_i \sim \chi^2_4, \qquad \phi_i = \sqrt{6/\chi^2_8} \\ \mathbf{A} &= \begin{bmatrix} \mathbf{I}_p \\ \boldsymbol{\iota}_p' \end{bmatrix} \left(\mathbf{I}_p + \boldsymbol{\iota}_p \boldsymbol{\iota}_p'\right)^{-1/2}, \quad \left(\kappa_4 = \frac{4.5p^2}{p+1} + \frac{p(p+2)}{2}\right) \end{aligned}$$

where κ_4 denotes the multivariate kurtosis due to Mardia (1970).

Static panel data model

Setup

- Note that $\boldsymbol{\Sigma}_{zz,i}(\boldsymbol{\theta})$ is given by

$$
\begin{aligned}
\boldsymbol{\Sigma}_{zz,i}(\boldsymbol{\theta}) &= \mathbf{B}^{-1}\boldsymbol{\Sigma}_{uu,i}\left(\mathbf{B}^{-1}\right)' \\
\boldsymbol{\Sigma}_{uu,i} &= \begin{bmatrix} \boldsymbol{\Sigma}_{\varepsilon\varepsilon,i} & \boldsymbol{\Sigma}'_{x\varepsilon,i} \\ \boldsymbol{\Sigma}_{x\varepsilon,i} & \boldsymbol{\Sigma}_{xx,i} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{vv,i} + \mathbf{F}\boldsymbol{\Sigma}_{\eta\eta}\mathbf{F}' & \left(\boldsymbol{\Sigma}_{xv,i} + \boldsymbol{\Sigma}_{x\eta}\mathbf{F}'\right)' \\ \boldsymbol{\Sigma}_{xv,i} + \boldsymbol{\Sigma}_{x\eta}\mathbf{F}' & \boldsymbol{\Sigma}_{xx,i} \end{bmatrix}
\end{aligned}
$$

where

$$
\begin{aligned}
\boldsymbol{\Sigma}_{vv,i} &= \operatorname{diag}\left(\sigma^2_{v,i1}, \ldots, \sigma^2_{v,iT}\right), \qquad \boldsymbol{\Sigma}_{\varepsilon\varepsilon,i} = \boldsymbol{\Sigma}_{vv,i} + \mathbf{F}\boldsymbol{\Sigma}_{\eta\eta}\mathbf{F}' \\
\boldsymbol{\Sigma}_{xv,i} &= \boldsymbol{\Xi}^{v}\breve{\boldsymbol{\Sigma}}_{vv,i}, \qquad \breve{\boldsymbol{\Sigma}}_{vv,i}, = \begin{bmatrix} \mathbf{0} \\ \boldsymbol{\Sigma}_{vv,i} \end{bmatrix} \qquad \boldsymbol{\Sigma}_{x\eta} = \tau_{\eta}\boldsymbol{\Xi}^{f}\boldsymbol{\Sigma}_{\eta\eta} \\
\boldsymbol{\Sigma}_{x\varepsilon,i} &= \boldsymbol{\Sigma}_{xv,i} + \boldsymbol{\Sigma}_{x\eta}\mathbf{F}' = \boldsymbol{\Xi}^{v}\breve{\boldsymbol{\Sigma}}_{vv,i} + \tau_{\eta}\boldsymbol{\Xi}^{f}\boldsymbol{\Sigma}_{\eta\eta}\mathbf{F}' \\
\boldsymbol{\Sigma}_{xx,i} &= \tau_{\eta}^{2}\boldsymbol{\Xi}^{f}\boldsymbol{\Sigma}_{\eta\eta}\boldsymbol{\Xi}^{f\prime} + \sigma^2_{\varpi 0,i}\mathbf{b}_{\rho}\mathbf{b}'_{\rho} + \boldsymbol{\Xi}^{v}\boldsymbol{\Sigma}_{\breve{v}\breve{v},i}\boldsymbol{\Xi}^{v\prime} + \sigma^2_{r}\left(\mathbf{C}_{\rho}\mathbf{C}'_{\rho}\right)
\end{aligned}
$$

Static panel data model

Setup

-

$$
\begin{aligned}
\mathbf{b}_{\rho} &= \rho\mathbf{C}_{\rho}\mathbf{e}_1 = \begin{bmatrix} \rho \\ \rho^2 \\ \vdots \\ \rho^{T} \end{bmatrix}, \quad \mathbf{C}_{\rho} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \rho & 1 & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ \rho^{T-1} & \cdots & \rho & 1 \end{bmatrix} \\
\boldsymbol{\Xi}^{f} &= \mathbf{b}_{\rho}\tilde{\mathbf{f}}'_0 + \mathbf{C}_{\rho}\mathbf{F}, \qquad \boldsymbol{\Xi}^{v} = \tau_{v0}\mathbf{C}_{\rho}\mathbf{L}_0
\end{aligned}
$$

Static panel data model

Findings for estimation of β

- All estimators have little bias.
- ML and MD perform very similarly, but computational cost of MD is smaller than ML.
- ML and MD are more efficient than GMM under both normality and non-normality when x is endogenous.
- When x is strictly exogenous, all estimators perform similarly when $N = 500, 1000$ and $m_0 = 1$, but not when $m_0 = 2$.
- In terms of RMSE, ML and MD perform (sometimes substantially) better than GMM.
- Inference based on ML and MD is more accurate than GMM.

Static panel data model

Results under normality ($T = 5$, $m_0 = \{1, 2\}$, Endogenous)

$\mathbf{z}_i \sim$Normal, $T = 5$, $\beta = 1$, Endogenous

	$m_0 = 1$					$m_0 = 2$				
	ML	MD1	OMD	GMM1	GMM2	ML	MD1	OMD	GMM1	GMM2
	$N = 200$									
Mean	1.002	0.997	0.998	1.004	1.004	1.004	1.005	0.969	–	–
StDev	0.046	0.041	0.043	0.079	0.074	0.074	0.075	0.199	–	–
RMSE	0.046	0.041	0.043	0.079	0.074	0.074	0.075	0.202	–	–
Size	7.8	5.1	–	–	–	8.7	7.6	–	–	–
Size_{rob}	8.2	5.6	8.0	2.7	5.1	9.7	9.0	8.7	–	–
	$N = 500$									
Mean	1.001	0.999	0.999	1.002	1.000	1.002	1.002	0.993	–	–
StDev	0.027	0.026	0.027	0.046	0.041	0.043	0.043	0.104	–	–
RMSE	0.027	0.026	0.027	0.046	0.041	0.044	0.043	0.104	–	–
Size	6.4	5.3	–	–	–	7.1	6.4	–	–	–
Size_{rob}	6.3	5.3	6.4	3.5	3.9	7.4	6.8	7.0	–	–
	$N = 1000$									
Mean	0.999	0.998	0.999	1.000	0.999	1.000	1.000	0.998	–	–
StDev	0.018	0.018	0.018	0.031	0.027	0.028	0.028	0.053	–	–
RMSE	0.018	0.018	0.018	0.031	0.027	0.028	0.028	0.053	–	–
Size	4.4	4.2	–	–	–	4.8	4.7	–	–	–
Size_{rob}	4.5	4.0	4.5	3.6	4.2	4.8	4.8	5.1	–	–

Static panel data model

Results under normality ($T = 5$, $m_0 = \{1, 2\}$, Strictly exogenous)

$\mathbf{z}_i \sim$Normal, $T = 5$, $\beta = 1$, Strictly exogenous

	$m_0 = 1$					$m_0 = 2$				
	ML	MD1	OMD	GMM1	GMM2	ML	MD1	OMD	GMM1	GMM2
	$N = 200$					$N = 200$				
Mean	1.001	0.998	0.998	1.005	1.004	1.000	0.999	0.982	1.016	1.015
StDev	0.015	0.015	0.035	0.021	0.018	0.019	0.019	0.134	0.033	0.030
RMSE	0.015	0.015	0.035	0.021	0.019	0.019	0.019	0.136	0.037	0.034
Size	6.5	6.4	–	–	–	5.9	5.5	–	–	–
Size_{rob}	6.6	6.8	10.8	5.6	11.4	6.3	6.4	9.4	12.1	16.8
	$N = 500$					$N = 500$				
Mean	1.000	0.999	1.000	1.002	1.002	1.000	1.000	0.996	1.012	1.012
StDev	0.010	0.010	0.010	0.014	0.011	0.011	0.011	0.064	0.021	0.020
RMSE	0.010	0.010	0.010	0.014	0.011	0.011	0.011	0.064	0.025	0.023
Size	7.0	6.3	–	–	–	6.4	6.0	–	–	–
Size_{rob}	6.8	6.7	10.0	5.3	9.1	6.5	6.3	6.9	14.0	19.2
	$N = 1000$					$N = 1000$				
Mean	1.000	1.000	1.000	1.001	1.001	1.000	1.000	0.997	1.009	1.009
StDev	0.006	0.006	0.007	0.009	0.007	0.008	0.008	0.055	0.016	0.015
RMSE	0.006	0.006	0.007	0.009	0.007	0.008	0.008	0.055	0.019	0.017
Size	5.4	5.4	–	–	–	6.1	6.3	–	–	–
Size_{rob}	5.4	5.3	6.9	5.5	6.6	5.9	6.3	6.7	14.3	21.0

Static panel data model

Results under non-normality ($T = 5$, $m_0 = \{1, 2\}$, Endogenous)

$\mathbf{z}_i \sim \chi^2$, $T = 5$, $\beta = 1$, Endogenous

	$m_0 = 1$					$m_0 = 2$				
	ML	MD1	OMD	GMM1	GMM2	ML	MD1	OMD	GMM1	GMM2
	$N = 200$					$N = 200$				
Mean	1.002	0.995	0.996	0.999	1.003	1.001	1.005	0.925	–	–
StDev	0.058	0.050	0.052	0.090	0.095	0.093	0.097	0.293	–	–
RMSE	0.058	0.050	0.052	0.090	0.095	0.092	0.097	0.302	–	–
Size	16.4	11.5	–	–	–	14.6	12.5	–	–	–
Size_{rob}	10.0	5.2	9.8	1.9	4.2	8.8	7.3	7.0	–	–
	$N = 500$					$N = 500$				
Mean	0.998	0.995	0.995	0.998	0.999	0.998	0.998	0.967	–	–
StDev	0.034	0.032	0.031	0.054	0.052	0.056	0.054	0.181	–	–
RMSE	0.034	0.033	0.031	0.054	0.052	0.056	0.054	0.184	–	–
Size	14.5	12.2	–	–	–	16.0	15.5	–	–	–
Size_{rob}	8.2	6.7	7.8	3.0	4.0	9.3	8.8	8.7	–	–
	$N = 1000$					$N = 1000$				
Mean	1.000	0.998	0.998	1.000	0.999	0.998	0.998	0.983	–	–
StDev	0.024	0.024	0.023	0.041	0.036	0.039	0.038	0.127	–	–
RMSE	0.024	0.024	0.023	0.041	0.036	0.039	0.038	0.128	–	–
Size	14.6	13.7	–	–	–	14.2	14.3	–	–	–
Size_{rob}	7.2	6.2	6.5	3.2	5.4	7.8	7.3	7.3	–	–

Static panel data model

Results under non-normality ($T = 5$, $m_0 = \{1, 2\}$, Strictly exogenous)

$\mathbf{z}_i \sim \chi^2$, $T = 5$, $\beta = 1$, Strictly exogenous

	$m_0 = 1$					$m_0 = 2$				
	ML	MD1	OMD	GMM1	GMM2	ML	MD1	OMD	GMM1	GMM2
	$N = 200$					$N = 200$				
Mean	0.999	0.995	0.993	1.006	1.003	0.999	0.997	0.941	1.015	1.015
StDev	0.021	0.020	0.066	0.030	0.023	0.025	0.024	0.235	0.042	0.038
RMSE	0.021	0.021	0.067	0.030	0.023	0.025	0.025	0.242	0.044	0.041
Size	19.5	16.9	—	—	—	14.4	14.0	—	—	—
Size_{rob}	7.4	6.3	12.2	8.3	14.0	7.4	6.3	10.0	8.0	17.7
	$N = 500$					$N = 500$				
Mean	0.999	0.998	0.999	1.003	1.001	1.000	0.999	0.975	1.016	1.015
StDev	0.013	0.013	0.013	0.019	0.014	0.016	0.016	0.157	0.030	0.027
RMSE	0.013	0.013	0.013	0.019	0.015	0.016	0.016	0.159	0.034	0.030
Size	18.9	19.4	—	—	—	18.4	19.7	—	—	—
Size_{rob}	5.1	5.5	8.4	6.4	10.3	7.3	7.1	7.6	12.2	18.6
	$N = 1000$					$N = 1000$				
Mean	0.999	0.999	0.999	1.001	1.001	1.000	0.999	0.990	1.012	1.012
StDev	0.010	0.010	0.009	0.014	0.010	0.011	0.012	0.100	0.021	0.020
RMSE	0.010	0.010	0.009	0.014	0.010	0.011	0.012	0.101	0.025	0.024
Size	20.4	19.5	—	—	—	18.1	18.2	—	—	—
Size_{rob}	6.3	6.2	6.3	7.0	7.9	7.6	7.0	7.1	12.1	21.0

Static panel data model

Findings for estimating the number of factors

- The performance of model selection based on information criterion is investigated with $T = 7$, $N = \{200, 500, 1000\}$.
- Performance based on AIC, BIC, HQIC is investigated for the candidate $m = \{0, 1, 2\}$ where the true number of factors is $m_0 = \{0, 1, 2\}$ and known fixed effects($f_t = 1$) are included.
- Findings are
  1. AIC based on ML perfoms well under normality, but does not under non-normality.
  2. BIC based on ML/MD2/OMD performs well when $m_0 = 1, 2$, but does not when $m_0 = 2$ and $N = 200$.
  3. HQIC based on ML/MD2 tends to perform best in many cases.

Static panel data model

Results for estimation of the number of factors

Table: Empirical frequency of $\widehat{m}$(%)

$\mathbf{z}_i$ ~Normal, $T = 7$, $m_0 = 0$, Strictly exogenous

	ML			MD2			OMD		
m	AIC	BIC	HQIC	AIC	BIC	HQIC	AIC	BIC	HQIC
	N = 200								
0	**88.9**	**100.0**	**100.0**	**28.3**	**99.4**	**84.6**	**19.1**	**98.9**	**79.2**
1	10.2	0.0	0.0	61.9	0.6	15.3	57.8	1.1	20.1
2	0.9	0.0	0.0	9.8	0.0	0.1	23.1	0.0	0.7
	N = 500								
0	**96.3**	**100.0**	**100.0**	**68.7**	**100.0**	**99.8**	**67.2**	**100.0**	**99.6**
1	3.6	0.0	0.0	28.4	0.0	0.2	28.3	0.0	0.4
2	0.1	0.0	0.0	2.9	0.0	0.0	4.5	0.0	0.0
	N = 1000								
0	**96.9**	**100.0**	**100.0**	**83.3**	**100.0**	**100.0**	**83.7**	**100.0**	**100.0**
1	2.8	0.0	0.0	15.3	0.0	0.0	14.9	0.0	0.0
2	0.3	0.0	0.0	1.4	0.0	0.0	1.4	0.0	0.0

Static panel data model

Results for estimation of the number of factors

Table: Empirical frequency of $\widehat{m}$(%)

$\mathbf{z}_i$ ~Normal, $T = 7$, $m_0 = 1$, Strictly exogenous

	ML			MD2			OMD		
m	AIC	BIC	HQIC	AIC	BIC	HQIC	AIC	BIC	HQIC
	N = 200								
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.9	0.0
1	**91.0**	**100.0**	**100.0**	**41.1**	**99.8**	**92.1**	**38.6**	**93.1**	**93.7**
2	9.0	0.0	0.0	58.9	0.2	7.9	61.4	0.0	6.3
	N = 500								
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	**92.3**	**100.0**	**100.0**	**71.2**	**100.0**	**99.7**	**70.8**	**100.0**	**99.9**
2	7.7	0.0	0.0	28.8	0.0	0.3	29.2	0.0	0.1
	N = 1000								
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	**93.7**	**100.0**	**100.0**	**80.6**	**100.0**	**100.0**	**81.8**	**100.0**	**100.0**
2	6.3	0.0	0.0	19.4	0.0	0.0	18.2	0.0	0.0

Static panel data model

Results for estimation of the number of factors

Table: Empirical frequency of $\widehat{m}(\%)$

$\mathbf{z}_i$ ∼Normal, $T = 7$, $m_0 = 2$, Strictly exogenous

	ML			MD2			OMD		
m	AIC	BIC	HQIC	AIC	BIC	HQIC	AIC	BIC	HQIC
	N = 200								
0	0.0	0.8	0.0	0.0	0.0	0.0	0.0	15.7	0.2
1	0.3	83.8	22.2	0.0	77.7	3.4	0.7	73.3	24.1
2	**99.7**	**15.4**	**77.8**	**100.0**	**22.3**	**96.6**	**99.3**	**11.0**	**75.7**
	N = 500								
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	5.0	0.0	0.0	0.0	0.0	0.0	48.3	0.4
2	**100.0**	**95.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**51.7**	**99.6**
	N = 1000								
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**

Static panel data model

Results for estimation of the number of factors

Table: Empirical frequency of $\widehat{m}(\%)$

$\mathbf{z}_i \sim \chi^2$, $T = 5$, $m_0 = 0$, Strictly exogenous

	ML			MD2-rob			OMD-rob		
m	AIC	BIC	HQIC	AIC	BIC	HQIC	AIC	BIC	HQIC
	N = 200								
0	**36.2**	**99.4**	**89.6**	**50.3**	**92.0**	**83.4**	**24.5**	**99.9**	**90.7**
1	38.0	0.6	9.8	43.7	7.9	16.2	57.5	0.1	9.2
2	25.8	0.0	0.6	6.0	0.1	0.4	18.0	0.0	0.1
	N = 500								
0	**45.2**	**99.3**	**94.2**	**72.0**	**99.3**	**97.1**	**69.2**	**100.0**	**99.9**
1	30.9	0.7	5.3	24.3	0.7	2.9	28.0	0.0	0.1
2	23.9	0.0	0.5	3.7	0.0	0.0	2.8	0.0	0.0
	N = 1000								
0	**59.3**	**99.9**	**97.4**	**81.3**	**99.9**	**99.6**	**82.8**	**100.0**	**100.0**
1	22.2	0.1	2.6	16.7	0.1	0.4	15.5	0.0	0.0
2	18.5	0.0	0.0	2.0	0.0	0.0	1.7	0.0	0.0

Static panel data model

Results for estimation of the number of factors

Table: Empirical frequency of $\widehat{m}(\%)$

$\mathbf{z}_i \sim \chi^2$, $T = 5$, $m_0 = 1$, Strictly exogenous

	ML			MD2-rob			OMD-rob		
m	AIC	BIC	HQIC	AIC	BIC	HQIC	AIC	BIC	HQIC
	N = 200								
0	0.0	0.0	0.0	0.1	0.2	0.1	0.0	59.3	4.3
1	**45.1**	**99.7**	**94.5**	**58.8**	**97.9**	**91.6**	**53.1**	**40.7**	**94.3**
2	54.9	0.3	5.5	41.1	1.9	8.3	46.9	0.0	1.4
	N = 500								
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.5	0.0
1	**43.6**	**99.6**	**95.0**	**73.2**	**99.7**	**98.1**	**74.0**	**93.5**	**99.9**
2	56.4	0.4	5.0	26.8	0.3	1.9	26.0	0.0	0.1
	N = 1000								
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	**48.3**	**99.9**	**96.8**	**78.2**	**100.0**	**99.7**	**83.9**	**100.0**	**100.0**
2	51.7	0.1	3.2	21.8	0.0	0.3	16.1	0.0	0.0

Static panel data model

Results for estimation of the number of factors

Table: Empirical frequency of $\widehat{m}(\%)$

$\mathbf{z}_i \sim \chi^2$, $T = 5$, $m_0 = 2$, Strictly exogenous

	ML			MD2-rob			OMD-rob		
m	AIC	BIC	HQIC	AIC	BIC	HQIC	AIC	BIC	HQIC
	N = 200								
0	0.0	0.4	0.0	0.0	0.5	0.3	0.0	66.6	6.0
1	0.4	64.0	10.1	3.5	84.3	46.2	2.7	32.5	49.2
2	**99.6**	**35.6**	**89.9**	**96.5**	**15.2**	**53.5**	**97.3**	**0.9**	**44.8**
	N = 500								
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	19.6	0.0
1	0.0	2.6	0.1	0.1	25.4	1.0	0.0	69.2	10.3
2	**100.0**	**97.4**	**99.9**	**99.9**	**74.6**	**99.0**	**100.0**	**11.2**	**89.7**
	N = 1000								
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0
1	0.0	0.0	0.0	0.0	0.8	0.1	0.0	18.4	0.2
2	**100.0**	**100.0**	**100.0**	**100.0**	**99.2**	**99.9**	**100.0**	**81.5**	**99.8**

Dynamic panel data model with time-invariant regressor

Setup

- Consider the following DGP:

$$\begin{aligned} y_{it} &= \mu_{y,t} + \alpha y_{i,t-1} + \beta x_{it} + \gamma w_i + \mathbf{f}_t' \boldsymbol{\eta}_i + v_{it}, \qquad (t = 1, ..., T) \\ x_{it} &= \mu_{x,t} + \rho x_{i,t-1} + \tau_\eta \mathbf{f}_t' \boldsymbol{\eta}_i + \tau_{v0} v_{it} + r_{it} \\ x_{i0} &= \tilde{\mu}_{x,0} + \tau_\eta \tilde{\mathbf{f}}_0' \boldsymbol{\eta}_i + \varpi_{i0} \\ w_i &= \mu_w + \kappa_\eta \frac{1}{m} \boldsymbol{\iota}_m' \boldsymbol{\eta}_i + s_i \end{aligned}$$

- For parameter values, we set

$$\begin{aligned} (\alpha, \beta, \gamma) &= (0.2, 0.8, 0.8), (0.8, 0.2, 0.2), \quad \rho = 0.5, \quad \tau_\eta = 0.2 \\ \tau_{v0} &= \begin{cases} 0 & \text{strictly exogenous} \\ 0.2 & \text{endogenous} \end{cases}, \quad \kappa_\eta = 0.2, \end{aligned}$$

- Other variables are generated similarly to the static case with $Var(s_i) = Var(r_{it})$.
- For the sample size, we consider

$$T = 5, \qquad N = \{200, 500, 1000\}$$

Dynamic panel data model with time-invariant regressor

Findings for estimation of α and β

- In terms of bias, ML, MD and OMD perform similarly well, and GMM performs poorly.
- ML is more dispersed than MD, and slightly size distorted when x_{it} is endogenous and $N = 200$.
- MD perfomrs best in almost all cases in terms of RMSE and size.
- OMD performs worse than MD when $N = 200$, but similarly when $N = 500, 1000$.
- Even under non-normality, OMD does not perform best when $N = 200$.
- When $N = 1000$, OMD perfoms better than MD, but the difference is very small.
- GMM is (sometimes severely) biased and has large size distortions.

Dynamic panel data model with time-invariant regressor

Results under normality ($T = 5$, $m_0 = 1$, Endogenous)

$\mathbf{z}_i$ ~Normal, $T = 5$, $(\alpha, \beta, \gamma) = (0.2, 0.8, 0.8)$, $m_0 = 1$, Endogenous

	$\alpha = 0.2$				$\beta = 0.8$				$\gamma = 0.8$			
	ML	MD	OMD	GMM	ML	MD	OMD	GMM	ML	MD	OMD	GMM
	$N = 200$											
Mean	0.202	0.204	0.204	0.194	0.792	0.790	0.790	0.816	0.798	0.796	0.795	0.805
StDev	0.039	0.032	0.034	0.065	0.115	0.090	0.096	0.122	0.043	0.037	0.040	0.070
RMSE	0.039	0.032	0.035	0.065	0.115	0.090	0.097	0.123	0.043	0.037	0.040	0.070
Size	10.7	6.1	—	—	13.8	7.5	—	—	10.0	6.2	—	—
Size_{rob}	10.3	5.3	10.6	7.8	13.0	6.9	12.8	7.2	9.7	5.3	10.5	7.7
	$N = 500$											
Mean	0.199	0.201	0.201	0.196	0.802	0.800	0.799	0.810	0.800	0.799	0.799	0.803
StDev	0.020	0.019	0.019	0.046	0.052	0.051	0.051	0.086	0.023	0.023	0.023	0.048
RMSE	0.020	0.019	0.019	0.046	0.052	0.051	0.051	0.086	0.023	0.023	0.023	0.048
Size	7.7	6.6	—	—	9.1	7.0	—	—	7.3	6.5	—	—
Size_{rob}	7.4	6.3	7.3	5.5	8.7	6.3	8.5	6.0	6.8	6.2	6.5	5.6
	$N = 1000$											
Mean	0.200	0.200	0.200	0.197	0.801	0.800	0.800	0.806	0.800	0.800	0.800	0.802
StDev	0.013	0.013	0.013	0.035	0.036	0.035	0.035	0.065	0.016	0.016	0.016	0.036
RMSE	0.013	0.013	0.013	0.035	0.036	0.035	0.035	0.066	0.016	0.016	0.016	0.036
Size	5.8	6.2	—	—	5.4	4.9	—	—	5.9	5.5	—	—
Size_{rob}	5.5	5.6	6.0	6.1	5.2	4.7	5.6	6.0	5.7	5.8	5.3	5.7

Note: MD=MD1, GMM=GMM2

Dynamic panel data model with time-invariant regressor

Results under normality ($T = 5$, $m_0 = 1$, Endogenous)

$\mathbf{z}_i$ ~Normal, $T = 5$, $(\alpha, \beta, \gamma) = (0.8, 0.2, 0.2)$, $m_0 = 1$, Endogenous

	$\alpha = 0.8$				$\beta = 0.2$				$\gamma = 0.2$			
	ML	MD	OMD	GMM	ML	MD	OMD	GMM	ML	MD	OMD	GMM
	$N = 200$											
Mean	0.799	0.801	0.801	0.789	0.198	0.195	0.197	0.212	0.201	0.200	0.199	0.211
StDev	0.018	0.017	0.018	0.027	0.059	0.049	0.053	0.052	0.024	0.023	0.025	0.034
RMSE	0.018	0.017	0.018	0.029	0.059	0.049	0.053	0.053	0.024	0.023	0.025	0.035
Size	6.1	5.4	—	—	11.6	5.2	—	—	5.9	4.3	—	—
Size_{rob}	6.2	5.1	9.1	11.6	11.8	5.4	9.3	9.3	5.7	4.5	8.4	9.5
	$N = 500$											
Mean	0.799	0.800	0.800	0.796	0.201	0.199	0.199	0.206	0.201	0.200	0.200	0.204
StDev	0.011	0.011	0.011	0.017	0.032	0.030	0.031	0.031	0.015	0.015	0.015	0.022
RMSE	0.011	0.010	0.011	0.018	0.032	0.030	0.031	0.032	0.015	0.015	0.015	0.022
Size	5.7	5.7	—	—	6.8	4.5	—	—	6.6	6.1	—	—
Size_{rob}	5.4	4.9	6.8	8.7	7.0	4.9	5.8	6.9	6.0	5.7	6.2	8.9
	$N = 1000$											
Mean	0.800	0.800	0.800	0.797	0.201	0.200	0.200	0.203	0.200	0.200	0.200	0.202
StDev	0.007	0.007	0.007	0.012	0.023	0.022	0.022	0.023	0.010	0.010	0.010	0.014
RMSE	0.007	0.007	0.007	0.012	0.023	0.022	0.022	0.023	0.010	0.010	0.010	0.015
Size	6.6	6.6	—	—	5.9	5.0	—	—	5.1	5.1	—	—
Size_{rob}	6.3	5.8	6.2	7.1	6.0	5.3	6.3	6.7	5.1	5.0	5.9	5.1

Dynamic panel data model with time-invariant regressor

Results under normality ($T = 5$, $m_0 = 2$, Endogenous)

$\mathbf{z}_i \sim$Normal, $T = 5$, $(\alpha, \beta, \gamma) = (0.2, 0.8, 0.8)$, $m_0 = 2$, Endogenous

	$\alpha = 0.2$				$\beta = 0.8$				$\gamma = 0.8$			
	ML	MD	OMD	GMM	ML	MD	OMD	GMM	ML	MD	OMD	GMM
	$N = 200$											
Mean	0.218	0.218	0.217	0.100	0.759	0.765	0.755	0.927	0.783	0.783	0.775	0.801
StDev	0.070	0.054	0.062	0.250	0.209	0.156	0.185	0.399	0.072	0.056	0.093	1.142
RMSE	0.072	0.057	0.065	0.269	0.213	0.160	0.190	0.419	0.074	0.058	0.096	1.142
Size	7.8	5.2	—	—	10.5	5.4	—	—	7.0	4.8	—	—
Size_{rob}	7.8	5.3	8.6	13.9	10.7	5.9	8.7	8.7	6.7	4.0	6.1	9.1
	$N = 500$											
Mean	0.204	0.207	0.207	0.065	0.792	0.788	0.787	0.985	0.796	0.793	0.793	0.932
StDev	0.034	0.029	0.031	0.236	0.100	0.085	0.089	0.368	0.036	0.033	0.034	0.915
RMSE	0.034	0.030	0.032	0.272	0.100	0.086	0.090	0.412	0.037	0.033	0.035	0.924
Size	5.0	4.3	—	—	7.8	5.6	—	—	5.7	4.9	—	—
Size_{rob}	4.8	4.0	4.5	14.1	7.7	5.4	5.5	7.3	5.5	4.8	5.2	10.5
	$N = 1000$											
Mean	0.201	0.203	0.203	0.069	0.801	0.798	0.798	0.995	0.800	0.798	0.798	0.968
StDev	0.019	0.019	0.019	0.205	0.054	0.052	0.053	0.307	0.022	0.021	0.021	0.979
RMSE	0.019	0.019	0.019	0.243	0.054	0.052	0.053	0.364	0.022	0.021	0.021	0.993
Size	6.0	5.8	—	—	6.7	5.1	—	—	5.2	5.5	—	—
Size_{rob}	5.1	5.1	5.4	13.6	6.4	5.4	5.4	10.0	5.0	4.8	4.8	11.5

Dynamic panel data model with time-invariant regressor

Results under normality ($T = 5$, $m_0 = 2$, Endogenous)

$\mathbf{z}_i \sim$Normal, $T = 5$, $(\alpha, \beta, \gamma) = (0.8, 0.2, 0.2)$, $m_0 = 2$, Endogenous

	$\alpha = 0.8$				$\beta = 0.2$				$\gamma = 0.2$			
	ML	MD	OMD	GMM	ML	MD	OMD	GMM	ML	MD	OMD	GMM
	$N = 200$											
Mean	0.806	0.808	0.806	0.763	0.199	0.199	0.199	0.250	0.195	0.193	0.193	0.179
StDev	0.031	0.028	0.054	0.115	0.116	0.094	0.108	0.201	0.038	0.035	0.038	0.852
RMSE	0.032	0.029	0.054	0.121	0.116	0.094	0.108	0.207	0.038	0.035	0.039	0.852
Size	9.2	9.0	—	—	12.8	6.2	—	—	6.3	6.5	—	—
Size_{rob}	9.3	9.1	11.7	35.4	11.8	6.4	8.4	5.5	6.8	6.9	6.9	17.6
	$N = 500$											
Mean	0.800	0.802	0.802	0.766	0.201	0.199	0.199	0.238	0.200	0.198	0.198	0.265
StDev	0.016	0.015	0.016	0.103	0.054	0.050	0.050	0.179	0.020	0.020	0.020	0.889
RMSE	0.016	0.016	0.016	0.108	0.054	0.050	0.050	0.183	0.020	0.020	0.020	0.890
Size	7.4	7.6	—	—	7.8	6.4	—	—	6.1	5.7	—	—
Size_{rob}	6.5	6.8	7.1	34.5	8.0	6.5	6.2	6.3	5.4	5.7	6.1	23.2
	$N = 1000$											
Mean	0.800	0.801	0.801	0.771	0.200	0.199	0.199	0.222	0.200	0.199	0.199	0.316
StDev	0.011	0.011	0.011	0.094	0.035	0.034	0.035	0.165	0.013	0.013	0.014	1.048
RMSE	0.011	0.011	0.011	0.098	0.035	0.034	0.035	0.166	0.013	0.013	0.014	1.054
Size	5.7	5.8	—	—	6.5	5.6	—	—	5.7	6.1	—	—
Size_{rob}	5.1	5.2	5.9	25.9	6.3	5.1	5.7	6.5	5.7	5.9	6.6	18.9

Dynamic panel data model with time-invariant regressor

Results under normality ($T = 5$, $m_0 = 1$, Strictly exogenous)

$\mathbf{z}_i \sim$Normal, $T = 5$, $(\alpha, \beta, \gamma) = (0.2, 0.8, 0.8)$, $m_0 = 1$, Strictly exogenous

	$\alpha = 0.2$				$\beta = 0.8$				$\gamma = 0.8$			
	ML	MD	OMD	GMM	ML	MD	OMD	GMM	ML	MD	OMD	GMM
	$N = 200$											
Mean	0.199	0.200	0.200	0.208	0.800	0.801	0.801	0.797	0.801	0.798	0.799	0.791
StDev	0.021	0.021	0.023	0.030	0.020	0.019	0.022	0.022	0.028	0.027	0.031	0.035
RMSE	0.021	0.021	0.023	0.031	0.020	0.019	0.022	0.022	0.028	0.027	0.031	0.037
Size	6.4	5.3	—	—	6.0	5.2	—	—	5.6	5.3	—	—
Size_{rob}	6.5	5.1	12.8	16.4	6.4	5.4	12.3	12.6	5.6	4.8	13.2	15.3
	$N = 500$											
Mean	0.200	0.200	0.200	0.204	0.800	0.801	0.801	0.799	0.800	0.799	0.799	0.796
StDev	0.013	0.013	0.013	0.017	0.013	0.013	0.013	0.014	0.017	0.017	0.018	0.021
RMSE	0.013	0.013	0.013	0.018	0.013	0.013	0.013	0.014	0.017	0.017	0.018	0.021
Size	4.8	5.0	—	—	5.7	5.6	—	—	4.2	3.8	—	—
Size_{rob}	4.7	4.7	6.6	8.9	5.9	5.7	8.4	8.6	4.6	4.2	5.8	7.5
	$N = 1000$											
Mean	0.200	0.200	0.200	0.202	0.800	0.800	0.800	0.799	0.800	0.800	0.800	0.798
StDev	0.009	0.009	0.010	0.013	0.008	0.008	0.009	0.009	0.013	0.013	0.013	0.015
RMSE	0.009	0.009	0.010	0.013	0.008	0.008	0.009	0.009	0.013	0.013	0.013	0.015
Size	5.9	5.8	—	—	4.3	3.9	—	—	6.3	5.7	—	—
Size_{rob}	5.7	5.0	6.9	7.9	4.4	4.1	5.5	5.9	5.8	5.3	6.7	6.9

Dynamic panel data model with time-invariant regressor

Results under normality ($T = 5$, $m_0 = 1$, Strictly exogenous)

$\mathbf{z}_i \sim$Normal, $T = 5$, $(\alpha, \beta, \gamma) = (0.8, 0.2, 0.2)$, $m_0 = 1$, Strictly exogenous

	$\alpha = 0.8$				$\beta = 0.2$				$\gamma = 0.2$			
	ML	MD	OMD	GMM	ML	MD	OMD	GMM	ML	MD	OMD	GMM
	$N = 200$											
Mean	0.798	0.799	0.799	0.792	0.200	0.200	0.200	0.201	0.202	0.200	0.200	0.207
StDev	0.018	0.016	0.019	0.027	0.017	0.016	0.018	0.018	0.023	0.022	0.026	0.031
RMSE	0.018	0.016	0.019	0.028	0.017	0.016	0.018	0.018	0.023	0.022	0.026	0.032
Size	7.2	4.8	—	—	5.4	4.8	—	—	5.9	5.3	—	—
Size_{rob}	6.4	4.7	12.4	24.0	6.1	5.0	11.5	11.1	6.6	4.9	12.6	16.1
	$N = 500$											
Mean	0.799	0.799	0.799	0.797	0.201	0.200	0.200	0.201	0.201	0.200	0.201	0.203
StDev	0.011	0.010	0.011	0.017	0.010	0.010	0.011	0.011	0.015	0.014	0.015	0.019
RMSE	0.011	0.010	0.011	0.017	0.010	0.010	0.011	0.011	0.015	0.014	0.015	0.020
Size	6.6	6.0	—	—	5.1	4.7	—	—	6.9	6.5	—	—
Size_{rob}	5.9	5.3	7.5	14.0	5.4	5.0	7.6	8.0	6.7	6.0	8.8	13.2
	$N = 1000$											
Mean	0.799	0.800	0.800	0.799	0.200	0.200	0.200	0.200	0.200	0.200	0.200	0.201
StDev	0.007	0.007	0.008	0.012	0.007	0.007	0.007	0.008	0.010	0.010	0.010	0.013
RMSE	0.007	0.007	0.008	0.012	0.007	0.007	0.007	0.008	0.010	0.010	0.010	0.013
Size	5.9	5.5	—	—	6.5	6.0	—	—	6.0	5.7	—	—
Size_{rob}	5.3	5.2	6.9	11.0	6.4	6.6	6.8	6.8	5.5	5.1	7.3	8.1

Dynamic panel data model with time-invariant regressor

Results under normality ($T = 5$, $m_0 = 2$, Strictly exogenous)

$\mathbf{z}_i \sim$Normal, $T = 5$, $(\alpha, \beta, \gamma) = (0.2, 0.8, 0.8)$, $m_0 = 2$, Strictly exogenous

	$\alpha = 0.2$				$\beta = 0.8$				$\gamma = 0.8$			
	ML	MD	OMD	GMM	ML	MD	OMD	GMM	ML	MD	OMD	GMM
	$N = 200$											
Mean	0.199	0.201	0.200	0.226	0.800	0.799	0.795	0.806	0.801	0.799	0.795	0.764
StDev	0.028	0.027	0.032	0.059	0.025	0.024	0.062	0.036	0.035	0.034	0.067	0.254
RMSE	0.028	0.027	0.032	0.064	0.025	0.024	0.062	0.037	0.035	0.034	0.067	0.256
Size	8.5	6.9	—	—	5.8	5.3	—	—	6.0	5.8	—	—
Size_{rob}	8.5	6.8	11.6	31.2	5.8	5.2	9.4	15.7	6.4	5.5	10.4	25.4
	$N = 500$											
Mean	0.200	0.201	0.201	0.207	0.800	0.800	0.800	0.803	0.800	0.799	0.799	0.793
StDev	0.016	0.016	0.017	0.034	0.015	0.015	0.016	0.020	0.020	0.020	0.021	0.040
RMSE	0.016	0.016	0.017	0.035	0.015	0.015	0.016	0.020	0.020	0.020	0.021	0.041
Size	4.7	4.8	—	—	5.9	5.5	—	—	4.9	4.8	—	—
Size_{rob}	4.6	4.2	6.8	15.4	5.8	5.3	6.4	9.1	5.1	5.3	5.3	11.5
	$N = 1000$											
Mean	0.200	0.200	0.200	0.201	0.800	0.800	0.800	0.801	0.801	0.800	0.800	0.799
StDev	0.012	0.012	0.012	0.020	0.010	0.010	0.010	0.012	0.015	0.015	0.015	0.023
RMSE	0.012	0.012	0.012	0.020	0.010	0.010	0.010	0.012	0.015	0.015	0.015	0.023
Size	6.0	5.6	—	—	4.0	3.8	—	—	5.2	4.5	—	—
Size_{rob}	5.7	5.6	5.9	8.4	3.9	3.4	4.3	5.4	5.0	4.4	5.1	6.5

Dynamic panel data model with time-invariant regressor

Results under normality ($T = 5$, $m_0 = 2$, Strictly exogenous)

$\mathbf{z}_i \sim$Normal, $T = 5$, $(\alpha, \beta, \gamma) = (0.8, 0.2, 0.2)$, $m_0 = 2$, Strictly exogenous

	$\alpha = 0.8$				$\beta = 0.2$				$\gamma = 0.2$			
	ML	MD	OMD	GMM	ML	MD	OMD	GMM	ML	MD	OMD	GMM
	$N = 200$											
Mean	0.798	0.801	0.798	0.824	0.201	0.200	0.200	0.205	0.203	0.199	0.199	0.179
StDev	0.026	0.023	0.051	0.043	0.021	0.020	0.024	0.028	0.033	0.030	0.035	0.081
RMSE	0.026	0.023	0.051	0.049	0.021	0.020	0.024	0.029	0.033	0.030	0.035	0.084
Size	11.0	9.1	—	—	6.2	5.3	—	—	7.2	6.5	—	—
Size_{rob}	10.7	8.6	14.5	40.5	6.8	5.2	9.3	13.5	7.4	6.7	9.5	21.8
	$N = 500$											
Mean	0.799	0.800	0.800	0.831	0.201	0.200	0.200	0.203	0.202	0.200	0.200	0.168
StDev	0.015	0.014	0.015	0.038	0.013	0.013	0.013	0.018	0.020	0.018	0.019	0.043
RMSE	0.015	0.014	0.015	0.049	0.013	0.013	0.013	0.018	0.020	0.018	0.019	0.053
Size	8.5	6.7	—	—	6.1	5.9	—	—	7.0	6.6	—	—
Size_{rob}	8.5	6.5	8.0	43.0	6.7	6.0	7.4	10.6	6.7	6.0	7.7	33.9
	$N = 1000$											
Mean	0.799	0.800	0.800	0.828	0.200	0.200	0.200	0.202	0.201	0.200	0.200	0.171
StDev	0.011	0.010	0.011	0.034	0.009	0.009	0.009	0.012	0.013	0.013	0.013	0.037
RMSE	0.011	0.010	0.011	0.044	0.009	0.009	0.009	0.012	0.013	0.013	0.013	0.047
Size	7.3	6.7	—	—	6.1	6.0	—	—	5.9	6.1	—	—
Size_{rob}	6.7	6.0	7.3	46.5	6.3	6.0	6.4	9.7	5.6	5.6	6.3	40.5

Dynamic panel data model with time-invariant regressor

Results under non-normality ($T = 5$, $m_0 = 1$, Endogenous)

$\mathbf{z}_i \sim \chi^2$, $T = 5$, $(\alpha, \beta, \gamma) = (0.2, 0.8, 0.8)$, $m_0 = 1$, Endogenous

	$\alpha = 0.2$				$\beta = 0.8$				$\gamma = 0.8$			
	ML	MD	OMD	GMM	ML	MD	OMD	GMM	ML	MD	OMD	GMM
	$N = 200$											
Mean	0.206	0.206	0.207	0.198	0.784	0.788	0.780	0.811	0.792	0.792	0.792	0.799
StDev	0.062	0.040	0.046	0.077	0.173	0.107	0.124	0.138	0.073	0.054	0.057	0.087
RMSE	0.062	0.041	0.046	0.077	0.174	0.108	0.126	0.139	0.073	0.054	0.058	0.086
Size	22.6	15.6	–	–	29.8	17.5	–	–	25.2	18.9	–	–
Size_{rob}	14.9	6.9	13.6	15.3	21.1	9.1	17.8	12.7	11.8	7.0	12.8	13.7
	$N = 500$											
Mean	0.201	0.202	0.201	0.198	0.797	0.795	0.795	0.806	0.799	0.798	0.799	0.801
StDev	0.027	0.026	0.025	0.058	0.072	0.067	0.065	0.105	0.035	0.034	0.033	0.064
RMSE	0.027	0.026	0.025	0.058	0.072	0.067	0.065	0.105	0.035	0.034	0.033	0.064
Size	17.4	15.7	–	–	17.3	14.3	–	–	22.5	20.3	–	–
Size_{rob}	6.9	5.7	7.4	10.3	9.0	7.5	8.6	10.2	7.1	6.0	7.6	9.7
	$N = 1000$											
Mean	0.199	0.200	0.200	0.198	0.800	0.799	0.798	0.805	0.800	0.799	0.800	0.802
StDev	0.018	0.018	0.017	0.045	0.048	0.047	0.044	0.081	0.024	0.024	0.023	0.048
RMSE	0.018	0.018	0.017	0.045	0.048	0.047	0.044	0.082	0.024	0.024	0.023	0.048
Size	16.4	15.2	–	–	14.7	13.5	–	–	21.6	20.9	–	–
Size_{rob}	5.6	5.8	7.1	9.2	7.2	5.8	6.7	8.5	7.1	6.1	7.1	7.2

Dynamic panel data model with time-invariant regressor

Results under non-normality ($T = 5$, $m_0 = 1$, Endogenous)

$\mathbf{z}_i \sim \chi^2$, $T = 5$, $(\alpha, \beta, \gamma) = (0.8, 0.2, 0.2)$, $m_0 = 1$, Endogenous

	$\alpha = 0.8$				$\beta = 0.2$				$\gamma = 0.2$			
	ML	MD	OMD	GMM	ML	MD	OMD	GMM	ML	MD	OMD	GMM
	$N = 200$											
Mean	0.797	0.799	0.800	0.783	0.203	0.199	0.196	0.218	0.202	0.199	0.197	0.215
StDev	0.028	0.024	0.035	0.039	0.078	0.056	0.062	0.059	0.037	0.035	0.037	0.048
RMSE	0.029	0.024	0.035	0.042	0.078	0.056	0.062	0.061	0.037	0.035	0.037	0.050
Size	19.1	16.2	–	–	22.1	10.3	–	–	18.8	16.2	–	–
Size_{rob}	11.2	8.0	12.5	22.7	15.0	5.3	11.4	12.1	9.4	8.3	12.2	18.6
	$N = 500$											
Mean	0.800	0.800	0.801	0.795	0.199	0.197	0.197	0.206	0.201	0.200	0.199	0.205
StDev	0.014	0.013	0.013	0.021	0.040	0.036	0.036	0.037	0.021	0.021	0.020	0.027
RMSE	0.014	0.013	0.013	0.022	0.040	0.036	0.036	0.038	0.021	0.021	0.020	0.028
Size	13.6	13.3	–	–	14.0	10.3	–	–	18.5	17.8	–	–
Size_{rob}	5.2	5.3	6.1	9.9	8.0	5.2	7.1	8.6	7.4	7.9	9.1	10.8
	$N = 1000$											
Mean	0.799	0.800	0.800	0.798	0.199	0.198	0.197	0.202	0.201	0.200	0.199	0.202
StDev	0.010	0.010	0.010	0.015	0.029	0.027	0.026	0.027	0.015	0.014	0.014	0.019
RMSE	0.010	0.010	0.010	0.015	0.029	0.027	0.026	0.027	0.015	0.014	0.014	0.020
Size	14.8	14.6	–	–	13.6	11.6	–	–	17.2	17.0	–	–
Size_{rob}	5.3	4.0	5.8	8.0	6.8	5.0	6.0	6.1	5.8	5.1	7.0	6.7

Dynamic panel data model with time-invariant regressor

Results under non-normality ($T = 5,\ m_0 = 2,$ Endogenous)

$\mathbf{z}_i \sim \chi^2,\quad T = 5,\quad (\alpha, \beta, \gamma) = (0.2, 0.8, 0.8),\quad m_0 = 2,$ Endogenous

	$\alpha = 0.2$				$\beta = 0.8$				$\gamma = 0.8$			
	ML	MD	OMD	GMM	ML	MD	OMD	GMM	ML	MD	OMD	GMM
	$N = 200$											
Mean	0.228	0.226	0.228	0.100	0.745	0.757	0.729	0.916	0.772	0.774	0.757	0.883
StDev	0.088	0.068	0.078	0.262	0.254	0.186	0.225	0.394	0.116	0.101	0.189	1.105
RMSE	0.092	0.073	0.083	0.281	0.260	0.191	0.236	0.410	0.119	0.104	0.193	1.107
Size	19.3	15.0	—	—	22.2	11.9	—	—	18.6	16.2	—	—
Size_{rob}	11.5	7.6	12.8	14.0	15.6	7.1	11.0	10.2	9.8	6.5	10.6	9.5
	$N = 500$											
Mean	0.205	0.209	0.209	0.080	0.789	0.784	0.783	0.960	0.794	0.791	0.789	0.877
StDev	0.046	0.039	0.040	0.238	0.124	0.102	0.108	0.365	0.055	0.048	0.059	1.064
RMSE	0.046	0.040	0.041	0.266	0.125	0.103	0.110	0.398	0.055	0.048	0.060	1.066
Size	14.9	13.2	—	—	16.2	11.4	—	—	18.0	16.6	—	—
Size_{rob}	6.9	5.2	6.6	16.4	9.6	5.5	6.8	9.2	7.3	5.9	6.8	10.5
	$N = 1000$											
Mean	0.202	0.204	0.206	0.062	0.796	0.792	0.789	0.984	0.798	0.796	0.794	0.997
StDev	0.028	0.026	0.027	0.212	0.078	0.071	0.072	0.326	0.034	0.032	0.033	0.984
RMSE	0.029	0.027	0.028	0.253	0.078	0.071	0.073	0.374	0.034	0.032	0.033	1.003
Size	16.4	15.2	—	—	17.5	14.1	—	—	20.1	19.0	—	—
Size_{rob}	5.2	4.4	5.4	15.5	8.0	6.1	6.3	9.1	4.6	4.1	5.4	12.3

Dynamic panel data model with time-invariant regressor

Results under non-normality ($T = 5,\ m_0 = 2,$ Endogenous)

$\mathbf{z}_i \sim \chi^2,\quad T = 5,\quad (\alpha, \beta, \gamma) = (0.8, 0.2, 0.2),\quad m_0 = 2,$ Endogenous

	$\alpha = 0.8$				$\beta = 0.2$				$\gamma = 0.2$			
	ML	MD	OMD	GMM	ML	MD	OMD	GMM	ML	MD	OMD	GMM
	$N = 200$											
Mean	0.810	0.811	0.806	0.753	0.206	0.203	0.199	0.247	0.187	0.186	0.181	0.256
StDev	0.044	0.038	0.090	0.137	0.145	0.112	0.127	0.250	0.060	0.054	0.058	1.066
RMSE	0.045	0.040	0.090	0.145	0.145	0.112	0.127	0.254	0.062	0.056	0.061	1.067
Size	20.8	17.9	—	—	20.4	10.8	—	—	15.3	14.1	—	—
Size_{rob}	14.4	11.3	16.1	33.6	13.6	6.8	10.5	9.4	9.0	8.7	10.1	15.5
	$N = 500$											
Mean	0.801	0.803	0.804	0.759	0.204	0.202	0.202	0.255	0.199	0.197	0.196	0.245
StDev	0.023	0.021	0.021	0.111	0.071	0.061	0.063	0.181	0.028	0.027	0.027	0.910
RMSE	0.023	0.021	0.022	0.118	0.071	0.061	0.063	0.189	0.028	0.027	0.028	0.911
Size	17.9	16.0	—	—	15.4	10.5	—	—	16.5	16.0	—	—
Size_{rob}	8.2	7.5	8.9	39.5	7.9	5.3	5.9	8.7	6.5	6.4	6.6	23.6
	$N = 1000$											
Mean	0.800	0.801	0.802	0.761	0.201	0.200	0.198	0.228	0.200	0.199	0.198	0.287
StDev	0.015	0.015	0.014	0.100	0.049	0.045	0.043	0.171	0.019	0.019	0.018	0.711
RMSE	0.015	0.015	0.014	0.107	0.049	0.045	0.043	0.173	0.019	0.019	0.018	0.716
Size	17.5	16.8	—	—	16.6	13.8	—	—	16.6	15.5	—	—
Size_{rob}	6.8	6.7	7.5	32.8	7.9	5.5	6.0	6.9	5.7	5.0	6.0	23.4

Dynamic panel data model with time-invariant regressor

Results under non-normality ($T = 5$, $m_0 = 1$, Strictly exogenous)

$\mathbf{z}_i \sim \chi^2$, $T = 5$, $(\alpha, \beta, \gamma) = (0.2, 0.8, 0.8)$, $m_0 = 1$, Strictly exogenous

	$\alpha = 0.2$				$\beta = 0.8$				$\gamma = 0.8$			
	ML	MD	OMD	GMM	ML	MD	OMD	GMM	ML	MD	OMD	GMM
	$N = 200$											
Mean	0.200	0.202	0.201	0.213	0.799	0.800	0.796	0.793	0.799	0.795	0.797	0.784
StDev	0.027	0.026	0.028	0.035	0.028	0.027	0.029	0.029	0.040	0.039	0.041	0.043
RMSE	0.027	0.026	0.028	0.038	0.028	0.027	0.029	0.029	0.040	0.039	0.041	0.046
Size	14.9	13.0	–	–	17.3	15.1	–	–	18.9	15.7	–	–
Size_{rob}	7.0	3.9	13.8	25.4	6.9	4.9	15.8	21.8	7.0	5.5	13.6	22.7
	$N = 500$											
Mean	0.200	0.201	0.200	0.207	0.799	0.800	0.799	0.797	0.800	0.799	0.800	0.793
StDev	0.018	0.017	0.017	0.023	0.018	0.017	0.017	0.018	0.027	0.026	0.025	0.029
RMSE	0.018	0.017	0.017	0.024	0.018	0.017	0.017	0.018	0.027	0.026	0.025	0.030
Size	16.8	15.7	–	–	18.7	17.2	–	–	20.9	19.2	–	–
Size_{rob}	6.4	4.9	7.6	15.2	6.0	4.9	8.7	12.9	6.8	5.6	9.6	16.4
	$N = 1000$											
Mean	0.199	0.200	0.200	0.203	0.800	0.800	0.799	0.798	0.800	0.800	0.800	0.797
StDev	0.012	0.012	0.012	0.016	0.012	0.012	0.012	0.012	0.019	0.019	0.018	0.020
RMSE	0.012	0.012	0.012	0.016	0.012	0.012	0.012	0.013	0.019	0.019	0.018	0.020
Size	15.0	14.3	–	–	18.3	16.9	–	–	22.1	20.9	–	–
Size_{rob}	5.1	4.3	5.4	9.5	5.2	4.7	6.4	8.9	5.5	5.1	8.1	9.0

Dynamic panel data model with time-invariant regressor

Results under non-normality ($T = 5$, $m_0 = 1$, Strictly exogenous)

$\mathbf{z}_i \sim \chi^2$, $T = 5$, $(\alpha, \beta, \gamma) = (0.8, 0.2, 0.2)$, $m_0 = 1$, Strictly exogenous

	$\alpha = 0.8$				$\beta = 0.2$				$\gamma = 0.2$			
	ML	MD	OMD	GMM	ML	MD	OMD	GMM	ML	MD	OMD	GMM
	$N = 200$											
Mean	0.796	0.798	0.799	0.789	0.201	0.199	0.198	0.200	0.203	0.198	0.198	0.208
StDev	0.027	0.022	0.026	0.031	0.021	0.020	0.022	0.021	0.036	0.033	0.039	0.040
RMSE	0.028	0.022	0.026	0.033	0.021	0.020	0.022	0.021	0.036	0.033	0.039	0.041
Size	20.1	15.7	–	–	12.2	10.6	–	–	19.7	16.0	–	–
Size_{rob}	11.7	7.9	18.9	34.1	7.5	5.7	14.0	17.9	9.0	7.3	19.0	25.9
	$N = 500$											
Mean	0.799	0.800	0.800	0.797	0.200	0.199	0.199	0.199	0.201	0.200	0.199	0.203
StDev	0.014	0.013	0.013	0.021	0.013	0.013	0.013	0.013	0.021	0.020	0.020	0.025
RMSE	0.014	0.013	0.013	0.021	0.013	0.013	0.013	0.013	0.021	0.020	0.020	0.025
Size	15.7	13.0	–	–	13.9	12.7	–	–	19.2	18.0	–	–
Size_{rob}	7.1	5.8	8.7	19.3	6.3	5.7	7.9	10.3	7.1	7.1	9.8	15.7
	$N = 1000$											
Mean	0.799	0.800	0.800	0.799	0.200	0.199	0.199	0.199	0.201	0.200	0.199	0.201
StDev	0.010	0.009	0.009	0.015	0.009	0.009	0.009	0.009	0.014	0.014	0.014	0.018
RMSE	0.010	0.009	0.009	0.015	0.009	0.009	0.009	0.009	0.014	0.014	0.014	0.018
Size	14.1	13.2	–	–	12.2	12.7	–	–	17.8	16.9	–	–
Size_{rob}	4.7	4.0	6.0	12.8	5.7	5.3	7.2	8.5	5.6	4.7	6.8	10.7

Dynamic panel data model with time-invariant regressor

Results under non-normality ($T = 5$, $m_0 = 2$, Strictly exogenous)

$\mathbf{z}_i \sim \chi^2$, $T = 5$, $(\alpha, \beta, \gamma) = (0.2, 0.8, 0.8)$, $m_0 = 2$, Strictly exogenous

	$\alpha = 0.2$				$\beta = 0.8$				$\gamma = 0.8$			
	ML	MD	OMD	GMM	ML	MD	OMD	GMM	ML	MD	OMD	GMM
	$N = 200$											
Mean	0.203	0.205	0.202	0.237	0.798	0.795	0.782	0.801	0.796	0.794	0.781	0.760
StDev	0.037	0.035	0.046	0.064	0.033	0.031	0.105	0.044	0.050	0.047	0.112	0.088
RMSE	0.038	0.035	0.046	0.074	0.033	0.031	0.107	0.044	0.050	0.047	0.114	0.096
Size	18.9	15.9	—	—	16.0	13.5	—	—	19.6	17.4	—	—
Size_{rob}	10.3	7.1	15.3	41.4	7.9	4.9	12.3	19.9	8.8	6.0	12.7	31.0
	$N = 500$											
Mean	0.201	0.202	0.201	0.216	0.799	0.798	0.798	0.801	0.799	0.798	0.799	0.785
StDev	0.022	0.022	0.022	0.045	0.021	0.020	0.020	0.027	0.032	0.031	0.031	0.056
RMSE	0.022	0.022	0.022	0.048	0.021	0.020	0.020	0.027	0.032	0.031	0.031	0.058
Size	16.3	14.9	—	—	17.6	15.5	—	—	21.3	19.8	—	—
Size_{rob}	7.3	5.9	7.8	24.9	6.5	5.4	7.2	12.8	7.6	6.9	8.8	20.6
	$N = 1000$											
Mean	0.200	0.200	0.200	0.202	0.800	0.800	0.799	0.801	0.800	0.799	0.800	0.800
StDev	0.015	0.015	0.015	0.028	0.015	0.015	0.014	0.018	0.022	0.021	0.021	0.035
RMSE	0.015	0.015	0.015	0.028	0.015	0.015	0.014	0.018	0.022	0.021	0.021	0.035
Size	14.9	14.5	—	—	16.2	15.4	—	—	20.8	20.3	—	—
Size_{rob}	5.3	4.8	5.9	12.0	5.6	5.4	6.0	9.5	5.0	4.3	6.1	10.8

Dynamic panel data model with time-invariant regressor

Results under non-normality ($T = 5$, $m_0 = 2$, Strictly exogenous)

$\mathbf{z}_i \sim \chi^2$, $T = 5$, $(\alpha, \beta, \gamma) = (0.8, 0.2, 0.2)$, $m_0 = 2$, Strictly exogenous

	$\alpha = 0.8$				$\beta = 0.2$				$\gamma = 0.2$			
	ML	MD	OMD	GMM	ML	MD	OMD	GMM	ML	MD	OMD	GMM
	$N = 200$											
Mean	0.799	0.802	0.794	0.828	0.201	0.199	0.196	0.204	0.200	0.196	0.193	0.172
StDev	0.033	0.028	0.093	0.044	0.026	0.024	0.034	0.032	0.045	0.039	0.047	0.059
RMSE	0.033	0.028	0.093	0.052	0.026	0.024	0.034	0.032	0.045	0.039	0.048	0.065
Size	21.3	15.4	—	—	13.6	9.8	—	—	18.2	15.4	—	—
Size_{rob}	13.7	8.2	19.0	45.4	7.6	4.6	11.2	16.9	10.7	7.5	14.0	26.8
	$N = 500$											
Mean	0.799	0.800	0.801	0.828	0.201	0.200	0.200	0.204	0.202	0.199	0.198	0.173
StDev	0.022	0.019	0.020	0.040	0.016	0.016	0.016	0.021	0.027	0.025	0.026	0.047
RMSE	0.022	0.019	0.020	0.049	0.016	0.016	0.016	0.022	0.027	0.025	0.026	0.054
Size	21.5	18.0	—	—	12.6	10.7	—	—	19.3	16.5	—	—
Size_{rob}	10.2	6.3	12.0	45.1	6.9	6.0	8.0	14.6	7.6	6.4	10.8	30.5
	$N = 1000$											
Mean	0.798	0.799	0.800	0.831	0.200	0.200	0.200	0.202	0.202	0.201	0.199	0.170
StDev	0.015	0.014	0.013	0.036	0.011	0.011	0.011	0.015	0.019	0.018	0.017	0.039
RMSE	0.015	0.014	0.013	0.047	0.011	0.011	0.011	0.015	0.019	0.018	0.017	0.049
Size	19.3	16.2	—	—	12.0	11.6	—	—	17.5	17.2	—	—
Size_{rob}	7.5	6.3	8.2	46.9	4.5	4.4	6.7	12.2	5.8	5.1	6.9	36.9

5. Conclusion

Conclusion

- In this paper, we proposed a covariance structure analysis approach to estimation of panel data models with endogenous variables and factor error structure.
- We showed that rank deficiency problem arises when x_{it} is endogenous and proposed a method to address it.
- Monte Carlo simulation results showed that the ML and MD estimators perform better than GMM in most cases.

Future works

- Investigate the relationship between our ML/MD and the moment conditions proposed in the literature.
- In this paper, all variables are assumed to be continuous. However, in practice, it is important to extend to include discrete variables such as dummy variables.

Reference

Abowd, J. M. and D. Card (1989) "On the Covariance Structure of Earnings and Hours Changes," *Econometrica*, 57, 411-445.

Ahn, S. C., H.Y. Lee, and P. Schmidt (2001) "GMM estimation of Linear Panel Data Models with Time-Varying Individual Effects," *Journal of Econometrics*, 101, 219-255.

Ahn, S. C., Y. H. Lee, and P. Schmidt (2013) "Panel Data Models with Multiple Time-Varying Individual Effects," *Journal of Econometrics*, 174, 1-14.

Altonji, J. G. and L. M. Segal (1996) "Small-sample Bias in GMM Estimation of Covariance Structures," *Journal of Business & Economic Statistics*, 14, 353–366.

Andrews, D. W. K. and B. Lu (2001) "Consistent Model and Moment Selection Procedures for GMM Estimation with Application to Dynamic Panel Data Models," *Journal of Econometrics*, 101, 123-164.

Bai, J. (2009) "Panel Data Models with Interactive Fixed Effects," *Econometrica*, 77, 1229-1279.

Bai, J. (2013a) "Fixed-Effects Dynamic Panel Models, a Factor Analytical Method," *Econometrica*, 81, 285–314.

Reference

Bai, J. (2013b) "Likelihood Approach to Dynamic Panel Models with Interactive Effects." mimeo.

Bollen, K.A. and J.E. Brand (2010) "A general panel model with random and fixed effects: A structural equations approach," *Social Forces*, 89, 1–34.

Browne, M. W. (1974) "Generalized Least Squares Estimators in the Analysis of Covariances," *South African Statistical Journal*, 8, p. 24.

Chudik, A. and M.H. Pesaran (2015) "Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors," *Journal of Econometrics*, 188, 393-420.

Hayakawa, K. (2012) "GMM Estimation of a Short Dynamic Panel Data Model with Interactive Fixed Effects," *Journal of the Japan Statistical Society*, 42, 109–123.

Hayakawa, K., M.H. Pesaran, and L.V. Smith (2018) "Short T Dynamic Panel Data Models with Individual and Interactive Time Effects."

Lillard, L. A. and R. J. Willis (1978) "Dynamic Aspects of Earning Mobility," *Econometrica*, 46, 985-1012.

Moon, H.R. and M. Weidner (2015) "Linear regression for panel with unknown number of factors as interactive fixed effects," *Econometrica*, 83, 1543–1579.

Reference

Moral-Benito, E. (2013) "Likelihood-based estimation of dynamic panels with predetermined regressors," *Journal of Business & Economic Statistics*, 31, 451–472.

Moral-Benito, E., P. Allison, and R. Williams (2019) "Dynamic panel data modelling using maximum likelihood: an alternative to Arellano-Bond." forthcoming in *Applied Economics*.

Pesaran, M. H. (2006) "Estimation and Inference in Large Heterogeneous Panels with Cross Section Dependence," *Econometrica*, 74, 967-1012.

Robertson, D. and V. Sarafidis (2015) "IV Estimation of Panels with Factor Residuals," *Journal of Econometrics*, 185, 526-541.

Multivariate Stochastic Volatility with Realized Volatility and Pairwise Realized Correlation

Yasuhiro Omori

Faculty of Economics, University of Tokyo

with Yuta Yamauchi

January 31, 2019

University of Tokyo

Outline

1. Introduction

To model the multivariate asset returns, we consider

- Latent stochastic processes for time-varying variance and covariances.
  → How to guarantee the positive definiteness for the covariance matrix?
- 'Asymmetry" or "Cross leverage effects" in stock market (the negative correlations between today's return and tomorrow's volatility).
  → How to reduce too many parameters and latent variables?

- Using observations: daily returns and 'realized covariances' (using high frequency data) to estimate many parameters efficiently for multivariate models.
  → More accurate parameter estimates. Posterior distributions are not sensitive to the specifications of priors.
- Forecasting time-varying covariances to optimize the portfolio of financial assets.
  → Model comparison based on the portfolio performances

Univariate SV model with leverage

y_t : observed log-return of stock prices.

$$
\begin{aligned}
y_t &= \epsilon_t \exp(h_t/2), \quad t = 1, \ldots, n, \\
h_{t+1} &= \mu + \phi(h_t - \mu) + \eta_t, \quad t = 1, \ldots, n-1, \\
(\epsilon_t, \eta_t)', &\sim N_2(\mathbf{0}, \Sigma), \\
\Sigma &= \begin{pmatrix} 1 & \rho\tau \\ \rho\tau & \tau^2 \end{pmatrix}, \\
&\quad |\phi| < 1, \quad h_1 \sim N(\mu, \tau^2/(1-\phi^2)).
\end{aligned}
$$

$\rho < 0 \rightarrow$ leverage effect

e.g. Omori, Chib, Shephard and Nakajima (2007) JOE, Omori and Watanabe (2008) CSDA

Realized Volatility

Assuming the log price $p(s)$ follows the diffusion process, $dp(s) = \mu(s)ds + \sigma(s)dW(s)$, the variance for day t is defined as the integral of $\sigma^2(s)$ over the interval $(t-1, t)$

$$
IV_t = \int_{t-1}^{t} \sigma^2(s)ds,
$$

If we have n intraday returns during a day t, $\{r_{t,i}\}_{i=0}^{n-1}$, the realized volatility is defined as their squared sum for the day t,

$$
RV_t = \sum_{i=0}^{n-1} r_{t,i}^2,
$$

Then, $RV_t \rightarrow IV_t$ in probability.

Properties of Realized measures

- RV is the consistent estimator of Integrated variance in ideal markets
- In real markets, there are some problems in realized variances and covariances:
  - Non-trading hours (overnight)
    may lead to the underestimation of the integrated variance.
  - Market microstructure noise
    Bid-ask bounces, discreteness of price changes, differences in trade sizes ...
  - Nonsynchronous trading (for multiple returns)
    may lead to the underestimation of the correlations (Epps effect).

Due to these noises, realized measures can be biased estimators.

Introduction
Multivariate SV model
MCMC algorithm
Empirical Studies
Conclusion
Univariate SV model
Realized Volatility
Realized SV model

Realized SV model

Using additional information ($x_t = \log RV_t$), we set RSV model

$$
\begin{aligned}
y_t &= \exp(h_t/2)\epsilon_t \\
x_t &= \xi + h_t + u_t, \quad u_t \perp\!\!\!\perp \epsilon_t, \eta_t \\
h_{t+1} &= \mu + \phi(h_t - \mu) + \eta_t.
\end{aligned}
$$

- ξ and u_t correct the biases and noises of RV caused by the microstructure noise and non-trading hours automatically within models.
- Posterior standard deviations of parameters become smaller by using additional information.$\rightarrow$ Efficient Estimation.
  Takahashi, Omori, and Watanabe (2009). CSDA

- $x_t = \xi + \psi h_t + u_t$ may be used. But, in forecasting performances, the model $\psi \equiv 1$ performs better.
- Similar approach in GARCH model: Realized GARCH Hansen, Huang and Shek (2012) JAE

The correlations of multivariate returns are also important for

- portfolio optimization
- risk management

and thus, in the multivariate SV models, we will further consider

- time varying volatility, correlations and (cross) leverage effects.

2. Multivariate SV model

1. Basic MSV model.
2. Dynamic Correlations.
3. Realized volatilities and correlations.
4. Leverage effects.

$\mathbf{y}_t$:$p \times 1$ return vector at t.
$\mathbf{h}_t$: log volatility vector at t.

(1) Basic MSV model

$$\begin{aligned}\mathbf{y}_t &= \mathbf{m}_t + \mathbf{V}_t^{1/2}\varepsilon_t, \quad t = 1, \ldots, n, \\ \mathbf{h}_{t+1} &= \boldsymbol{\mu} + \boldsymbol{\phi} \odot (\mathbf{h}_t - \boldsymbol{\mu}) + \boldsymbol{\eta}_t, \quad t = 1, \ldots, n-1, \\ \mathbf{m}_{t+1} &= \mathbf{m}_t + \mathbf{w}_t, \quad \mathbf{w}_t \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma}_m), \boldsymbol{\Sigma}_m : \text{diagonal},\end{aligned}$$

where

$$\begin{aligned}\mathbf{V}_t &= \text{diag}\left(\exp(h_{1t}), \ldots, \exp(h_{pt})\right), \\ \mathbf{h}_t &= (h_{1t}, \ldots, h_{pt})' \quad \boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)', \quad |\phi_i| < 1, \\ \begin{pmatrix} \varepsilon_t \\ \boldsymbol{\eta}_t \end{pmatrix} &\sim \mathcal{N}_{2p}(\mathbf{0}, \boldsymbol{\Sigma}), \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{\varepsilon\varepsilon} & \boldsymbol{\Sigma}_{\varepsilon\eta} \\ \boldsymbol{\Sigma}_{\eta\varepsilon} & \boldsymbol{\Sigma}_{\eta\eta} \end{pmatrix},\end{aligned}$$

with $\boldsymbol{\Sigma}_{\varepsilon\eta} \neq \mathbf{O}$, and correlations are assumed to be constant.
Ishihara and Omori (2012). CSDA without $\mathbf{m}_t$.

(2) Dynamic Correlations

We consider dynamic correlations

$$\varepsilon_t \sim \mathcal{N}_p(\mathbf{0}, \mathbf{R}_t),$$

where we assume

$$\begin{aligned}\rho_{ijt} &= \frac{\exp(g_{ijt}) - 1}{\exp(g_{ijt}) + 1}, \\ g_{ij,t+1} &= g_{ijt} + \zeta_{ijt}, \quad \zeta_{ijt} \sim \mathcal{N}(0, \sigma_\zeta^2),\end{aligned}$$

- Many parameters and latent variables
- Correlation matrices should be positive definite.
- Alternative MSV models: Cholesky SV (Shirota et al. (2017)), Matrix exponential SV (Ishihara, Omori and Asai (2017)), Dynamic equicorrelation SV (Kurose and Omori (2016,2017)).
- DCC-GARCH (Engle (2002)).

(3) Realized Volatilities and Correlations

Let RV_{jt} and $RCorr_{ijt}$ denote the realized volatility of the j-th asset return and the realized correlation between the i-th and jth asset returns at time t. Define

$$x_{it} = \log RV_{jt}, \quad w_{ijt} = \log\{(1 + RCorr_{ijt})/(1 - RCorr_{ijt})\},$$

and consider additional measurement equations:

$$\begin{aligned} x_{it} &= \xi_{it} + h_{it} + u_{it}, \quad u_{it} \sim \mathcal{N}(0, \sigma^2_{uj}), \\ w_{ijt} &= \delta_{ij} + g_{ijt} + v_{ijt}, \quad v_{ijt} \sim \mathcal{N}(0, \sigma^2_{v,ij}), \end{aligned}$$

Coefficients for h_{it} and g_{ijt} are set to ones for simplicity.

Positive definite correlation matrix

We sample ρ_{ijt} (g_{ijt}) to guarantee the positive definiteness of the correlation matrix $\mathbf{R}_t$ given other correlations and parameters. Let $\boldsymbol{\rho}_{jt} = (\rho_{j1,t}, \ldots, \rho_{j,j-1,t}, \rho_{j,j+1,t}, \ldots, \rho_{jpt})'$ and $\mathbf{R}_{-j,t}$ denote the matrix which excludes j-th row and j-th column of $\mathbf{R}_t$. Noting that

$$|\mathbf{R}_t| = |\mathbf{R}_{-j,t}| \times |1 - \boldsymbol{\rho}'_{jt}\mathbf{R}^{-1}_{-j,t}\boldsymbol{\rho}_{jt}|$$

we sample ρ_{ijt} under the constraint $\boldsymbol{\rho}'_{jt}\mathbf{R}^{-1}_{-j,t}\boldsymbol{\rho}_{jt} < 1$. Initial values for ρ_{ijt}'s are zero to satisfy the positive definiteness.

Proposition 1

The condition for $\rho_{ij,t}$ to guarantee that $\mathbf{R}_t$ is positive definite is $\rho_{ij,t} \in (L_{ijt}, U_{ijt})$ where bounds L_{ijt} and U_{ijt} are given by

$$\frac{-\mathbf{b}_j' \boldsymbol{\rho}_{i,-j,t} \pm \sqrt{(\mathbf{b}_j' \boldsymbol{\rho}_{i,-j,t})^2 - a_j(\boldsymbol{\rho}_{i,-j,t}' \mathbf{C}_j \boldsymbol{\rho}_{i,-j,t} - 1)}}{a_j},$$

and $\boldsymbol{\rho}_{i,-j,t}$ is the vector excluding the j-th element of $\boldsymbol{\rho}_{it}$, a_j is the (j,j)-th element of $\mathbf{R}_{it}^{-1}$, $\mathbf{b}_j$ is the vector excluding a_j from the j-th column of $\mathbf{R}_{it}^{-1}$, and $\mathbf{C}_j$ is the matrix excluding the j-th row and j-th column from $\mathbf{R}_{it}^{-1}$.

Introduction
Multivariate SV model
MCMC algorithm
Empirical Studies
Conclusion
Basic MSV model
Dynamic Correlations
Leverage effect

Remark:

1. Dynamic volatilities and correlations. $\rightarrow$ General dynamics.
2. Additional information through realized volatilities and pairwise realized correlations (when some of realized variances and correlations are missing, we can use only available realized measures). $\rightarrow$ Stable and accurate estimates.
3. Proposed model is independent of the ordering of assets.
4. Covariance matrices are always positive definite.
5. Leverages between $\mathbf{y}_t$ and $\mathbf{h}_{t+1}$?

(4) Extension to the model with leverage effects

Consider

$$\begin{pmatrix} \mathbf{y}_t \\ \mathbf{h}_{t+1} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mathbf{m}_t \\ \boldsymbol{\mu} + \boldsymbol{\Phi}(\mathbf{h}_t - \boldsymbol{\mu}) \end{pmatrix}, \begin{pmatrix} \mathbf{V}_t^{1/2}\mathbf{R}_t\mathbf{V}_t^{1/2} & \mathbf{V}_t^{1/2}\mathbf{R}_t^{1/2}\boldsymbol{\Lambda}' \\ \boldsymbol{\Lambda}\mathbf{R}_t^{1/2\prime}\mathbf{V}_t^{1/2} & \boldsymbol{\Psi} + \boldsymbol{\Lambda}\boldsymbol{\Lambda}' \end{pmatrix}\right).$$

so that

$$\begin{aligned} E(\mathbf{h}_{t+1}|\mathbf{y}_t) &= \boldsymbol{\mu} + \boldsymbol{\Phi}(\mathbf{h}_t - \boldsymbol{\mu}) + \boldsymbol{\Lambda}\mathbf{R}_t^{-1/2}\mathbf{V}_t^{-1/2}(\mathbf{y}_t - \mathbf{m}_t), \\ Var(\mathbf{h}_{t+1}|\mathbf{y}_t) &= \boldsymbol{\Psi}. \end{aligned}$$

Introduction
Multivariate SV model
MCMC algorithm
Empirical Studies
Conclusion
Basic MSV model
Dynamic Correlations
Leverage effect

Remark:

1. Alternative specification:

$$\begin{pmatrix} \mathbf{y}_t \\ \mathbf{h}_{t+1} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \mathbf{m}_t \\ \boldsymbol{\mu} + \boldsymbol{\Phi}(\mathbf{h}_t - \boldsymbol{\mu}) \end{pmatrix}, \begin{pmatrix} \mathbf{V}_t^{1/2}\mathbf{R}_t\mathbf{V}_t^{1/2} & \mathbf{V}_t^{1/2}\boldsymbol{\Lambda}' \\ \boldsymbol{\Lambda}\mathbf{V}_t^{1/2} & \boldsymbol{\Psi} + \boldsymbol{\Lambda}\mathbf{R}_t^{-1}\boldsymbol{\Lambda}' \end{pmatrix}\right),$$

so that

$$\begin{aligned} E(\mathbf{h}_{t+1}|\mathbf{y}_t) &= \boldsymbol{\mu} + \boldsymbol{\Phi}(\mathbf{h}_t - \boldsymbol{\mu}) + \boldsymbol{\Lambda}\mathbf{R}_t^{-1}\mathbf{V}_t^{-1/2}(\mathbf{y}_t - \mathbf{m}_t), \\ &= \boldsymbol{\mu} + \boldsymbol{\Phi}(\mathbf{h}_t - \boldsymbol{\mu}) + \boldsymbol{\Lambda}\mathbf{R}_t^{-1/2\prime}\mathbf{R}_t^{-1/2}\mathbf{V}_t^{-1/2}(\mathbf{y}_t - \mathbf{m}_t), \\ Var(\mathbf{h}_{t+1}|\mathbf{y}_t) &= \boldsymbol{\Psi}. \end{aligned}$$

However, we do not use this specification since it is difficult to interpret $\boldsymbol{\Lambda}$ and to justify the time-varying unconditional covariance matrix for $\mathbf{h}_{t+1}$'s where $\mathbf{h}_t$ may not be stationary and its initial distribution is unknown.

Remark:

2. Need to select the decomposition $\mathbf{R}_t = \mathbf{R}_t^{1/2}\mathbf{R}_t^{1/2\prime}$.
   (1) Spectral decomposition (independent of the odering of asset returns) (2) Cholesky decomposition
   $\rightarrow$ For example, we may select the decomposition with best portfolio performance for a practical purpose.
3. Parsimonious parameterization.

$$\mathbf{\Lambda} = \{\boldsymbol{\lambda}_1, \cdots, \boldsymbol{\lambda}_q, \mathbf{0}, \ldots, \mathbf{0}\},$$

   where we set $q = 1$ in our empirical studies. This would imply the leverage effect between the market factor and individual latent volatilities.

3. MCMC algorithm

1. Initialize.
2. Generate $\{\mathbf{h}_t\}_{t=1}^n|\cdot$.
3. Generate $\{\mathbf{g}_t\}_{t=1}^n|\cdot$.
4. Generate $\{\mathbf{m}_t\}_{t=1}^n|\cdot$.
5. Generate $\boldsymbol{\mu}|\cdot, \xi|\cdot, \boldsymbol{\delta}|\cdot, \mathbf{\Lambda}|\cdot$.
6. Generate $\phi|\cdot$.
7. Generate $\boldsymbol{\Psi}|\cdot$.
8. Generate $\{\sigma_{u,i}^2\}_{i=1}^p|\cdot$, $\{\sigma_{v,ij}^2\}_{i,j=1}^p|\cdot$, $\{\sigma_{\zeta,ij}^2\}_{i,j=1}^p|\cdot$, $\{\sigma_{mi}^2\}_{i=1}^p|\cdot$.
9. Go to 2.

- Generation of $\{\mathbf{h}_t\}_{t=1}^n$. $\rightarrow$ Single move sampler using MH algorithm. Sample $\mathbf{h}_t$ given $\mathbf{h}_s$ $(s \neq t)$. The multi-move sampler may not be necessary since the sampling efficiency is high enough by using the additional measurement equations based on high frequency data.
- Generation of $\{\mathbf{g}_t\}_{t=1}^n$. Single move sampler using Gibbs sampler subject to the constraint for the positive definiteness of $\mathbf{R}_t$.
- Generation of $\{\mathbf{m}_t\}_{t=1}^n$. Consider the linear and Gaussian state space model using the auxiliary variables $\tilde{\mathbf{y}}_t$

$$\begin{aligned} \tilde{\mathbf{y}}_t &= \mathbf{m}_t + \tilde{\boldsymbol{\epsilon}}_t, \quad \tilde{\boldsymbol{\epsilon}}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}_t), \\ \mathbf{m}_{t+1} &= \mathbf{m}_t + \boldsymbol{\nu}_t, \quad \boldsymbol{\nu}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_m), \end{aligned}$$

Generate $\mathbf{m}$ using a simulation smoother.

- Generation of $\boldsymbol{\mu}, \boldsymbol{\xi}, \boldsymbol{\delta}, \boldsymbol{\Lambda}$. Assuming normal priors, the conditional posterior distributions are conditionally independent normal distributions.
- Generation of ϕ. Generate a candidate for ϕ from a normal distribution truncated on the region $|\phi_{ii}| < 1$ and conduct MH algorithm.
- Generation of $\boldsymbol{\Psi}$. Assuming inverse Wishart priors, we conduct MH algorithm using the the inverse Wishart distribution as the proposal.
- Generation of $\{\sigma_{u,i}^2\}_{i=1}^p$, $\{\sigma_{v,ij}^2\}_{i,j=1}^p$, $\{\sigma_{\zeta,ij}^2\}_{i,j=1}^p$, $\{\sigma_{mi}^2\}_{i=1}^p$. Assuming inverse gamma priors, their conditional posterior distributions are conditionally independent inverse gamma distributions.

Proposition 2

Suppose that the prior distribution of $\boldsymbol{\Lambda}$ given $\boldsymbol{\Psi}$ is $\mathrm{N}_{p,p}(\mathbf{M}_0, \boldsymbol{\Psi} \otimes \boldsymbol{\Gamma}_0)$. Then the conditional posterior distribution of $\boldsymbol{\Lambda}$ given other parameters and latent variables is $\mathrm{N}_{p,p}(\mathbf{M}_1, \boldsymbol{\Psi} \otimes \boldsymbol{\Gamma}_1)$ where

$$\begin{aligned} \mathbf{M}_1 &= \left(\mathbf{A} + \boldsymbol{\Gamma}_0^{-1}\right)^{-1}\left(\mathbf{B} + \boldsymbol{\Gamma}_0^{-1}\mathbf{M}_0\right), \quad \boldsymbol{\Gamma}_1 = \left(\mathbf{A} + \boldsymbol{\Gamma}_0^{-1}\right)^{-1}, \\ \mathbf{A} &= \sum_{t=1}^{T-1} \mathbf{z}_t\mathbf{z}_t', \quad \mathbf{B} = \sum_{t=1}^{T-1} \mathbf{z}_t\boldsymbol{\eta}_t', \end{aligned}$$

and $\mathbf{z}_t = \mathbf{R}_t^{-1/2}\mathbf{V}_t^{-1/2}(\mathbf{y}_t - \mathbf{m}_t)$ and $\boldsymbol{\eta}_t = \mathbf{h}_{t+1} - \boldsymbol{\mu} - \boldsymbol{\Phi}(\mathbf{h}_t - \boldsymbol{\mu})$.

Proposition 3

Let $\boldsymbol{\Lambda} = [\boldsymbol{\lambda}_1, \cdots, \boldsymbol{\lambda}_q, \mathbf{0}, \ldots, \mathbf{0}]$ and $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1', \ldots, \boldsymbol{\lambda}_q')'$. If the prior distribution of $\boldsymbol{\lambda}$ is assumed to be normal, $\boldsymbol{\lambda} \sim \mathrm{N}(\mathbf{m}_0, \boldsymbol{\Gamma}_0)$, then the conditional posterior distribution of $\boldsymbol{\lambda}$ is $\boldsymbol{\lambda}|\cdot \sim \mathrm{N}(\mathbf{m}_1, \boldsymbol{\Gamma}_1)$ where where

$$\begin{aligned} \mathbf{m}_1 &= \boldsymbol{\Gamma}_1\left\{\boldsymbol{\Gamma}_0^{-1}\mathbf{m}_0 + (\mathbf{I}_q \otimes \boldsymbol{\Psi}^{-1}\mathbf{B}')\mathrm{vec}\left(\{\mathbf{e}_1, \ldots, \mathbf{e}_q\}\right)\right\}, \\ \boldsymbol{\Gamma}_1 &= \left(\boldsymbol{\Gamma}_0^{-1} + \mathbf{A}_{1:q,1:q} \otimes \boldsymbol{\Psi}^{-1}\right)^{-1}, \end{aligned}$$

$\mathbf{A}, \mathbf{B}$ are defined in (1), $\mathbf{A}_{1:q,1:q}^{-1}$ denotes the first q rows and the q columns of $\mathbf{A}$, $\mathrm{vec}(\mathbf{X}) \equiv (\mathbf{x}_1', \ldots, \mathbf{x}_m')'$ denotes a vectorization of the matrix $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$, and $\otimes$ denotes Kronecker product.

Predictive mean and covariance for portfolio optimization

Let $\mathbf{m}_{t+1|t} \equiv E[\mathbf{y}_{t+1}|\mathcal{F}_t]$ and $\boldsymbol{\Sigma}_{t+1|t} \equiv Var[\mathbf{y}_{t+1}|\mathcal{F}_t]$ given the current information set $\mathcal{F}_t$. Then, add the following several steps to each MCMC iteration:

1. Generate $\mathbf{h}_{n+1}^{(i)}, \mathbf{g}_{n+1}^{(i)}, \mathbf{m}_{n+1}^{(i)} | \{\mathbf{z}_t\}_{t=1}^n, \boldsymbol{\theta}^{(i)}, \{\mathbf{h}_t^{(i)}\}_{t=1}^n, \{\mathbf{g}_t^{(i)}\}_{t=1}^n, \{\mathbf{m}_t^{(i)}\}_{t=1}^n$.
2. Store $\mathbf{m}_{n+1|n}^{(i)}$, $\mathbf{V}_{n+1|n}^{(i)}$, $\mathbf{R}_{n+1|n}^{(i)}$ and $\boldsymbol{\Sigma}_m^{(i)}$.
3. Compute

$$\hat{\mathbf{m}}_{n+1|n} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{m}_{n+1|n}^{(i)},$$

$$\hat{\boldsymbol{\Sigma}}_{n+1|n} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\Sigma}_{n+1|n}^{(i)} = \frac{1}{N}\sum_{i=1}^{N} \left(\mathbf{V}_{n+1|n}^{1/2(i)} \mathbf{R}_{n+1|n}^{(i)} \mathbf{V}_{n+1|n}^{1/2(i)} + \boldsymbol{\Sigma}_m^{(i)} \right).$$

Portfolio strategies

We consider the minimum-variance strategy for the portfolio optimization.

- $\mu_{p,t+1}$, $\sigma_{p,t+1}^2$: the conditional mean and variance of the portfolio return, $r_{p,t+1}$.
- r_f: the risk free asset return (the federal funds rate).
- ω_t: the vector of portfolio weights for stock returns.

$$\mu_{p,t+1} = \boldsymbol{\omega}_t' \mathbf{m}_{t+1|t} + (1 - \boldsymbol{\omega}_t' \mathbf{1}) r_f, \quad \sigma_{p,t+1}^2 = \boldsymbol{\omega}_t' \boldsymbol{\Sigma}_{t+1|t} \boldsymbol{\omega}_t.$$

Portfolio strategies

Minimum-variance strategy

$$\min_{\boldsymbol{\omega}_t} \sigma^2_{p,t+1} \quad \text{s.t.} \quad \mu_{p,t+1} = \mu^*_p,$$

where μ^*_p is the target expected return. The solution is

$$\hat{\boldsymbol{\omega}}_t = \boldsymbol{\Sigma}^{-1}_{t+1|t}(\mathbf{m}_{t+1|t} - r_f \mathbf{1})\frac{\mu^*_p - r_f}{\kappa_t},$$

where $\kappa_t = (\mathbf{m}_{t+1|t} - r_f\mathbf{1})'\boldsymbol{\Sigma}^{-1}_{t+1|t}(\mathbf{m}_{t+1|t} - r_f\mathbf{1})$.

4. Empirical Studies – Data

- 9 US stocks. JP Morgan (JPM), International Business Machine (IBM), Microsoft (MSFT), Exxon Mobil (XOM), Alcoa (AA), American Express (AXP), Du Pont (DD), General Electric (GE), and Coca Cola (KO) in NYSE.
- Daily returns (close-to-close) and realized covariances (open-to-close) for $p = 9$ stocks from Oxford Man Institute Realized Library (e.g. Noureldin, Shephard and Sheppard (2012) JAE).
- The dataset also includes Bank of America (BA). However, since it has the extremely high volatility period after the financial crisis, it is excluded from our empirical studies.

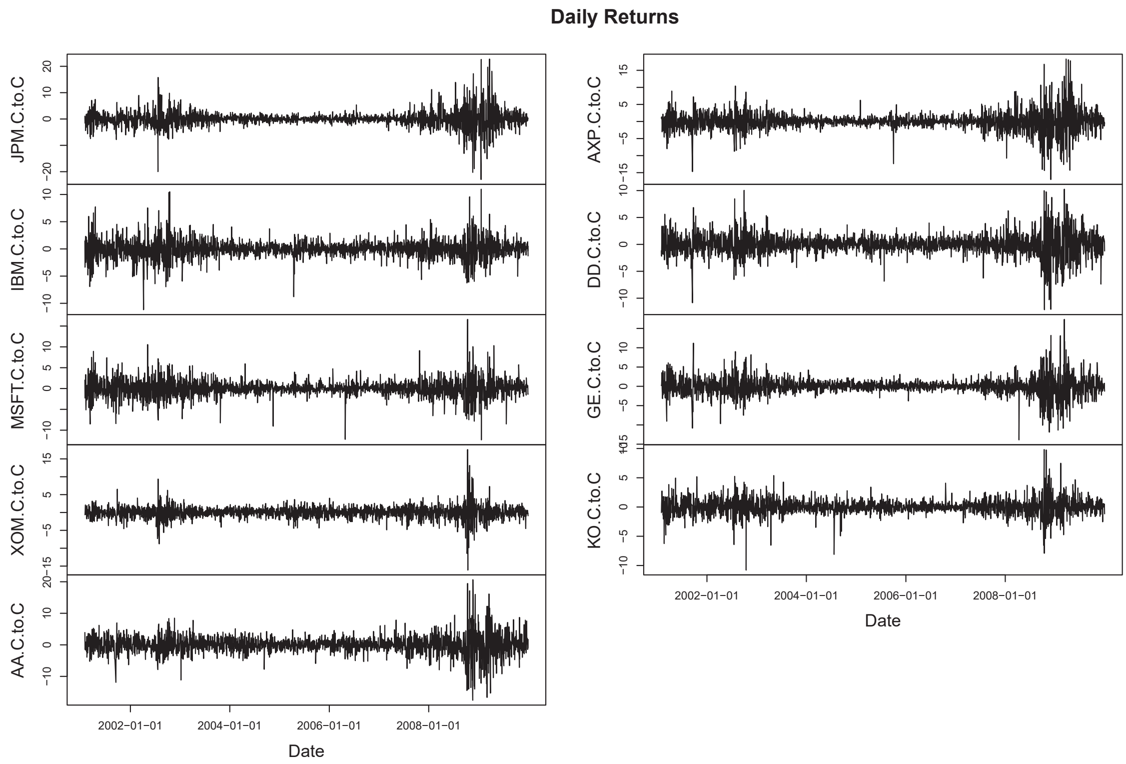
4. Empirical Studies – Data

- The realized covariance is calculated via 5 minutes intraday returns with subsampling.
- The number of obs: is $n = 2242$ (Feb 1, 2001– Dec 31, 2009) and we estimate the proposed model with parsimonious leverage specification $q = 1$).
- Realized measures may have biases due to the microstructure noise, non-trading hours, nonsynchronous trading and so forth.
- Flat priors or typical priors are used in empirical studies.

4. Empirical Studies – Data

- For the portfolio optimization, we conduct the rolling window estimation using past 1742 observations for 500 forecasting periods (Jan 9, 2008– Dec31, 2009). The results are obtained for the basic MSV model and the proposed model with and without leverage.

Application to U.S. stock returns



Introduction
Multivariate SV model
MCMC algorithm
Empirical Studies
Conclusion

Data
Estimation results
Portfolio performances

Application to U.S. stock returns

Parameter	Mean	95% interval	IF
ξ_1	-0.520	[-0.582,-0.47]	101.7
ξ_2	-0.554	[-0.61,-0.495]	100.5
ξ_3	-0.549	[-0.594,-0.501]	89.6
ξ_4	-0.442	[-0.487,-0.394]	88.1
ξ_5	-0.537	[-0.582,-0.494]	77.6
ξ_6	-0.586	[-0.651,-0.533]	108.8
ξ_7	-0.428	[-0.48,-0.376]	101.8
ξ_8	-0.535	[-0.589,-0.474]	100.1
ξ_9	-0.322	[-0.376,-0.263]	93.2

*Negative biases in realized variances (due to non-trading hours).

Application to U.S. stock returns

Parameter	Mean	95% interval	IF
μ_1	1.221	[1.02,1.43]	7.99
μ_2	0.644	[0.495,0.794]	16.13
μ_3	0.914	[0.759,1.07]	8.50
μ_4	0.684	[0.541,0.827]	11.09
μ_5	1.582	[1.43,1.73]	7.73
μ_6	1.136	[0.918,1.36]	9.05
μ_7	0.873	[0.726,1.02]	12.46
μ_8	0.861	[0.67,1.05]	10.38
μ_9	0.203	[0.0553,0.352]	15.08

Application to U.S. stock returns

Parameter	Mean	95% interval	IF
ϕ_1	0.914	[0.904,0.924]	19.2
ϕ_2	0.888	[0.874,0.901]	22.0
ϕ_3	0.900	[0.887,0.913]	19.1
ϕ_4	0.890	[0.876,0.904]	20.9
ϕ_5	0.907	[0.895,0.92]	19.7
ϕ_6	0.926	[0.916,0.935]	24.9
ϕ_7	0.899	[0.886,0.911]	21.5
ϕ_8	0.908	[0.897,0.92]	21.3
ϕ_9	0.903	[0.889,0.916]	21.4

*High persistence in log volatilities.

Application to U.S. stock returns

Parameter	Mean	95% interval	IF
$\sigma_{u,1}$	0.285	[0.27,0.3]	39.8
$\sigma_{u,2}$	0.286	[0.271,0.303]	57.4
$\sigma_{u,3}$	0.291	[0.278,0.304]	13.9
$\sigma_{u,4}$	0.274	[0.261,0.287]	20.3
$\sigma_{u,5}$	0.318	[0.304,0.332]	16.5
$\sigma_{u,6}$	0.307	[0.293,0.321]	20.3
$\sigma_{u,7}$	0.291	[0.278,0.304]	19.6
$\sigma_{u,8}$	0.303	[0.289,0.317]	17.9
$\sigma_{u,9}$	0.294	[0.281,0.308]	17.6

Application to U.S. stock returns

Parameter	Mean	95% interval	IF
$\sigma_{m,1}$	0.0873	[0.0673,0.108]	118
$\sigma_{m,2}$	0.0702	[0.0533,0.089]	126
$\sigma_{m,3}$	0.0756	[0.0532,0.106]	129
$\sigma_{m,4}$	0.0871	[0.0669,0.106]	117
$\sigma_{m,5}$	0.0976	[0.0734,0.135]	129
$\sigma_{m,6}$	0.0882	[0.0709,0.118]	119
$\sigma_{m,7}$	0.0874	[0.0627,0.111]	123
$\sigma_{m,8}$	0.0776	[0.0602,0.108]	125
$\sigma_{m,9}$	0.0666	[0.0481,0.0879]	124

*Small variances for the mean process

Application to U.S. stock returns (spectral decomp)

Parameter	Mean	95% interval	IF
λ_1	-0.0626	[-0.0852,-0.0403]	6.82
λ_2	-0.0541	[-0.0757,-0.033]	6.64
λ_3	-0.0430	[-0.0638,-0.0216]	7.30
λ_4	-0.0518	[-0.0722,-0.0311]	5.78
λ_5	-0.0424	[-0.0625,-0.0219]	7.24
λ_6	-0.0518	[-0.0735,-0.0303]	6.32
λ_7	-0.0536	[-0.0736,-0.0331]	9.07
λ_8	-0.0538	[-0.0767,-0.0308]	7.92
λ_9	-0.0436	[-0.0637,-0.0235]	9.72

*Leverage effects exist

Application to U.S. stock returns (Cholesky decomp)

Parameter	Mean	95% interval	IF
λ_1	-0.0552	[-0.0769,-0.0337]	7.64
λ_2	-0.0375	[-0.0576,-0.0173]	4.03
λ_3	-0.0280	[-0.0483,-0.00741]	7.33
λ_4	-0.0351	[-0.0548,-0.0151]	5.64
λ_5	-0.0329	[-0.0525,-0.0133]	7.10
λ_6	-0.0383	[-0.0596,-0.0172]	6.58
λ_7	-0.0385	[-0.058,-0.0191]	8.84
λ_8	-0.0467	[-0.0685,-0.0243]	9.38
λ_9	-0.0365	[-0.0562,-0.0171]	7.92

*Leverage effects exist

Application to U.S. stock returns (spectral decomp)

Parameter	Mean	95% interval	IF
ρ_1^*	-0.199	[-0.306,-0.113]	8
ρ_2^*	-0.187	[-0.302,-0.100]	8
ρ_3^*	-0.153	[-0.261,-0.067]	7
ρ_4^*	-0.199	[-0.327,-0.101]	7
ρ_5^*	-0.179	[-0.322,-0.076]	8
ρ_6^*	-0.177	[-0.288,-0.089]	8
ρ_7^*	-0.224	[-0.374,-0.114]	10
ρ_8^*	-0.171	[-0.278,-0.087]	8
ρ_9^*	-0.187	[-0.334,-0.083]	11

where $\rho_i^* = Corr(z_{1t}, h_{i,t+1})$, $\mathbf{z}_t = \mathbf{R}_t^{-1/2}\mathbf{V}_t^{-1/2}(\mathbf{y}_t - \mathbf{m}_t)$ with $q = 1$. Leverage effects exist.

Application to U.S. stock returns (Cholesky decomp)

Parameter	Mean	95% interval	IF
ρ_1^*	-0.170	[-0.266,-0.092]	7
ρ_2^*	-0.118	[-0.199,-0.049]	4
ρ_3^*	-0.091	[-0.174,-0.021]	8
ρ_4^*	-0.118	[-0.206,-0.045]	7
ρ_5^*	-0.123	[-0.239,-0.044]	7
ρ_6^*	-0.119	[-0.205,-0.048]	7
ρ_7^*	-0.141	[-0.244,-0.061]	9
ρ_8^*	-0.144	[-0.237,-0.067]	10
ρ_9^*	-0.148	[-0.271,-0.058]	8

where $\rho_i^* = Corr(z_{1t}, h_{i,t+1})$, $\mathbf{z}_t = \mathbf{R}_t^{-1/2}\mathbf{V}_t^{-1/2}(\mathbf{y}_t - \mathbf{m}_t)$ with $q = 1$. Leverage effects exist, but with small absolute values.

Application to U.S. stock returns (spectral decomp)

Par.	Mean	95% interval	IF	Par.	Mean	95% interval	IF
ρ^*_{11}	-0.195	[-0.305,-0.107]	6	ρ^*_{21}	-0.009	[-0.103,0.071]	35
ρ^*_{12}	-0.183	[-0.305,-0.091]	10	ρ^*_{22}	0.016	[-0.084,0.096]	41
ρ^*_{13}	-0.147	[-0.254,-0.063]	8	ρ^*_{23}	-0.001	[-0.141,0.106]	64
ρ^*_{14}	-0.199	[-0.334,-0.097]	6	ρ^*_{24}	0.037	[-0.054,0.116]	41
ρ^*_{15}	-0.187	[-0.343,-0.080]	11	ρ^*_{25}	0.025	[-0.096,0.114]	51
ρ^*_{16}	-0.166	[-0.276,-0.079]	7	ρ^*_{26}	-0.013	[-0.109,0.070]	31
ρ^*_{17}	-0.225	[-0.400,-0.111]	8	ρ^*_{27}	-0.012	[-0.125,0.081]	45
ρ^*_{18}	-0.173	[-0.284,-0.084]	8	ρ^*_{28}	0.021	[-0.079,0.099]	39
ρ^*_{19}	-0.183	[-0.337,-0.075]	13	ρ^*_{29}	0.054	[-0.066,0.143]	55

where $\rho^*_{1i} = Corr(z_{1t}, h_{i,t+1})$ and $\rho^*_{2i} = Corr(z_{2t}, h_{i,t+1})$ with $q = 2$. Leverage effects exist only for the first component.

Introduction
Multivariate SV model
MCMC algorithm
Empirical Studies
Conclusion
Data
Estimation results
Portfolio performances

Application to U.S. stock returns (Cholesky decomp)

Par.	Mean	95% interval	IF	Par.	Mean	95% interval	IF
ρ^*_{11}	-0.168	[-0.271,-0.087]	7	ρ^*_{21}	-0.029	[-0.089, 0.024]	5
ρ^*_{12}	-0.132	[-0.230,-0.056]	6	ρ^*_{22}	-0.113	[-0.200,-0.043]	7
ρ^*_{13}	-0.100	[-0.189,-0.029]	7	ρ^*_{23}	-0.073	[-0.147,-0.009]	6
ρ^*_{14}	-0.120	[-0.212,-0.046]	4	ρ^*_{24}	-0.019	[-0.080, 0.036]	7
ρ^*_{15}	-0.128	[-0.240,-0.045]	9	ρ^*_{25}	-0.044	[-0.117, 0.018]	8
ρ^*_{16}	-0.125	[-0.215,-0.051]	8	ρ^*_{26}	-0.028	[-0.093, 0.029]	5
ρ^*_{17}	-0.149	[-0.262,-0.064]	7	ρ^*_{27}	-0.081	[-0.160,-0.016]	8
ρ^*_{18}	-0.145	[-0.239,-0.066]	6	ρ^*_{28}	-0.011	[-0.071, 0.042]	5
ρ^*_{19}	-0.151	[-0.269,-0.058]	10	ρ^*_{29}	-0.003	[-0.069, 0.052]	7

where $\rho^*_{1i} = Corr(z_{1t}, h_{i,t+1})$ and $\rho^*_{2i} = Corr(z_{2t}, h_{i,t+1})$ with $q = 2$. Leverage effects exist for two components.

Application to U.S. stock returns

Table: Posterior means (standard deviation)

δ	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$	$i=6$	$i=7$	$i=8$
$j=2$	-0.313 (0.0285)							
$j=3$	-0.265 (0.0305)	-0.333 (0.0529)						
$j=4$	-0.301 (0.0482)	-0.18 (0.0349)	-0.149 (0.065)					
$j=5$	-0.41 (0.0651)	-0.227 (0.0273)	-0.246 (0.038)	-0.287 (0.0425)				
$j=6$	-0.629 (0.04)	-0.265 (0.032)	-0.295 (0.0476)	-0.353 (0.0449)	-0.391 (0.0359)			
$j=7$	-0.472 (0.0473)	-0.276 (0.0293)	-0.219 (0.0299)	-0.312 (0.0252)	-0.532 (0.0351)	-0.477 (0.032)		
$j=8$	-0.526 (0.0534)	-0.37 (0.0405)	-0.319 (0.0647)	-0.34 (0.0367)	-0.44 (0.0288)	-0.563 (0.0478)	-0.478 (0.0331)	
$j=9$	-0.194 (0.0428)	-0.0807 (0.033)	-0.0983 (0.0258)	-0.249 (0.0501)	-0.15 (0.0293)	-0.246 (0.0301)	-0.158 (0.026)	-0.172 (0.0573)

*Negative biases in realized correlations (due to non-synchronous trading). – Epps effect

Application to U.S. stock returns

Table: Posterior means (standard deviation)

σ_v	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$	$i=6$	$i=7$	$i=8$
$j=2$	0.329 (0.00554)							
$j=3$	0.325 (0.0053)	0.325 (0.00531)						
$j=4$	0.329 (0.00538)	0.329 (0.0057)	0.31 (0.00524)					
$j=5$	0.329 (0.00544)	0.322 (0.00527)	0.314 (0.00509)	0.328 (0.00546)				
$j=6$	0.35 (0.00586)	0.331 (0.00555)	0.315 (0.00532)	0.319 (0.00565)	0.332 (0.00554)			
$j=7$	0.338 (0.0056)	0.338 (0.00575)	0.319 (0.00547)	0.334 (0.00562)	0.344 (0.00571)	0.338 (0.00579)		
$j=8$	0.334 (0.00565)	0.32 (0.00553)	0.304 (0.00523)	0.322 (0.00553)	0.317 (0.00518)	0.332 (0.00576)	0.326 (0.00548)	
$j=9$	0.314 (0.00531)	0.329 (0.00544)	0.307 (0.00516)	0.317 (0.00562)	0.32 (0.00518)	0.321 (0.00532)	0.337 (0.0058)	0.328 (0.00555)

Application to U.S. stock returns

Table: Posterior means (standard deviation)

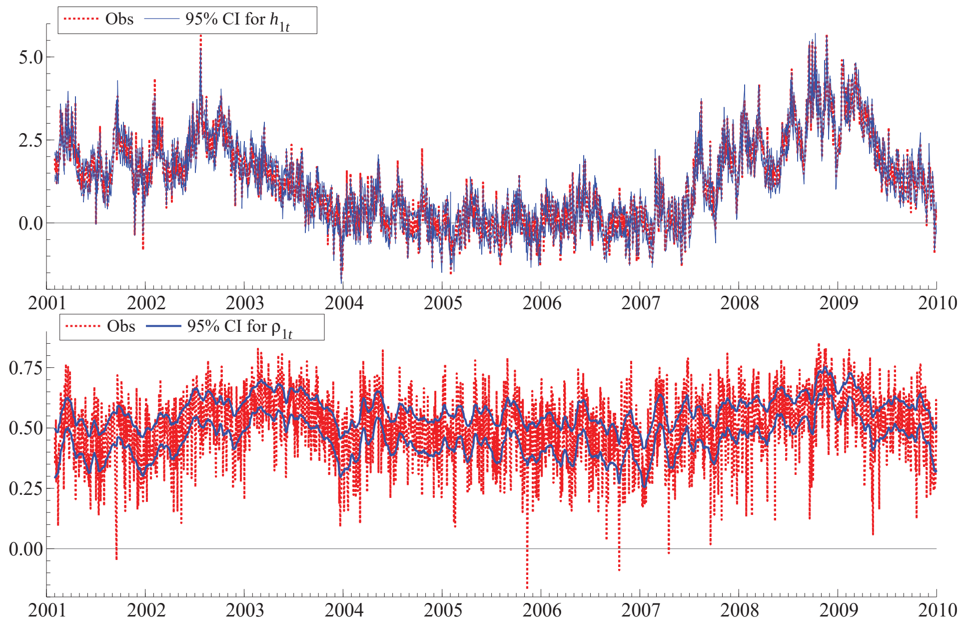
σ_ζ	$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 5$	$i = 6$	$i = 7$	$i = 8$
$j = 2$	0.047							
	(0.00497)							
$j = 3$	0.0387	0.0431						
	(0.00398)	(0.00426)						
$j = 4$	0.0606	0.0615	0.0595					
	(0.00448)	(0.00566)	(0.0046)					
$j = 5$	0.0468	0.0437	0.042	0.0465				
	(0.00414)	(0.00376)	(0.00397)	(0.00446)				
$j = 6$	0.0578	0.0455	0.0442	0.0686	0.0471			
	(0.0048)	(0.00506)	(0.00478)	(0.0055)	(0.00446)			
$j = 7$	0.0411	0.0436	0.041	0.0611	0.0531	0.0492		
	(0.00421)	(0.00483)	(0.00537)	(0.00486)	(0.00509)	(0.00486)		
$j = 8$	0.0484	0.0521	0.0549	0.0707	0.0485	0.0535	0.049	
	(0.00485)	(0.00494)	(0.00527)	(0.00512)	(0.00353)	(0.0055)	(0.00477)	
$j = 9$	0.0485	0.0422	0.0402	0.0641	0.0432	0.0452	0.0441	0.0476
	(0.0049)	(0.00465)	(0.00404)	(0.0055)	(0.00418)	(0.00456)	(0.00512)	(0.0043)

Application to U.S. stock returns

Table: Posterior means (standard deviation)

Ψ	$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 5$	$i = 6$	$i = 7$	$i = 8$	$i = 9$
$j = 1$	0.165								
	(0.0105)								
$j = 2$	0.129	0.142							
	(0.00851)	(0.00944)							
$j = 3$	0.119	0.114	0.127						
	(0.00793)	(0.00743)	(0.00866)						
$j = 4$	0.111	0.105	0.0966	0.124					
	(0.00774)	(0.00708)	(0.00671)	(0.00839)					
$j = 5$	0.104	0.0882	0.082	0.0891	0.103				
	(0.00729)	(0.00648)	(0.00614)	(0.00624)	(0.00773)				
$j = 6$	0.137	0.112	0.105	0.102	0.0935	0.142			
	(0.00882)	(0.00788)	(0.00751)	(0.00736)	(0.00697)	(0.0101)			
$j = 7$	0.116	0.102	0.0967	0.101	0.0891	0.108	0.115		
	(0.00797)	(0.00719)	(0.00674)	(0.00681)	(0.00631)	(0.00737)	(0.00805)		
$j = 8$	0.141	0.124	0.116	0.109	0.1	0.128	0.113	0.157	
	(0.00917)	(0.00831)	(0.00778)	(0.00758)	(0.00712)	(0.00852)	(0.00775)	(0.0105)	
$j = 9$	0.101	0.0954	0.0866	0.0861	0.069	0.091	0.0866	0.0992	0.103
	(0.00715)	(0.00657)	(0.00612)	(0.00606)	(0.00545)	(0.00648)	(0.00616)	(0.00695)	(0.00727)

95 % credible intervals (blue, dashed) of h_{1t} and ρ_{1t}



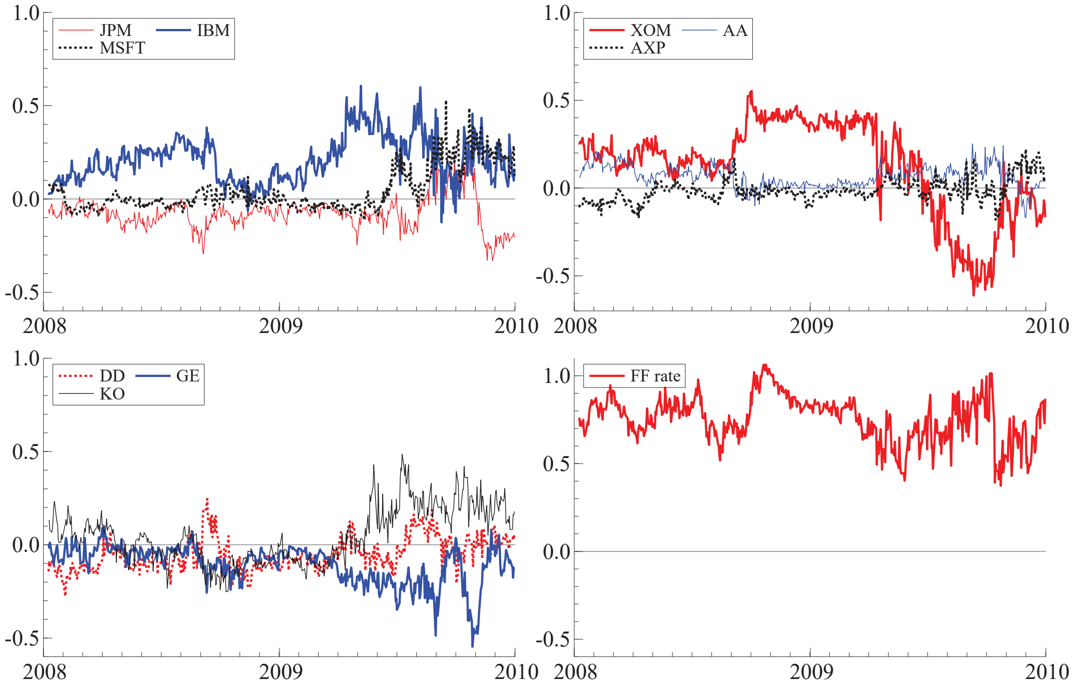
Top:$x_{1t} - \xi_1$ (red, dotted). Bottom:
$\{\exp(w_{21,t} - \delta_{21}) - 1\}/\{\exp(w_{21,t} - \delta_{21}) + 1\}$ (red, dotted).

Portfolio performances for competing models

Cumulative values of realized objective functions.

Minimum-Variance			
	$\mu_p^* = 0.004$	$\mu_p^* = 0.01$	$\mu_p^* = 0.1$
MSV	1.172	6.536	1184
CRSV	0.748	4.448	730
MRSV	0.526	2.943	510
MRSV-L1-C	0.272	1.601	262
MRSV-L2-C	0.264	1.552	255
MRSV-L1-S	**0.249**	**1.430**	**232**
MRSV-L1-S (constant mean)	0.568	3.032	543
DCC-GARCH	2.662	11.962	2537
Equal weight	1425	1425	1425

Portfolio weights in MRSV-L1-S ($\mu_p^* = 0.01$)



Conclusion

We propose

- Multivariate SV model with leverage.
- Dynamic correlations.
- Efficient parameter estimation method using realized covariances.
- The model is independent of order of asset returns.

In empirical studies,

- The model seems to capture the dynamics of log conditional variances and correlations.
- High persistences in h_{it} (log conditional volatilities) are found.
- Better portfolio performances than other volatility models.

- Negative biases are found for realized volatilities and correlations (negative x_i's and δ_{ij}'s)
- Variances of measurement errors u_{it}'s and $v_{ij,t}$'s are larger than those of latent h_{it}'s and $g_{ij,t}$'s.
- The mean processes of the returns are almost constant.