Uniform Inference for Conditional Factor Models with Instrumental and Idiosyncratic Betas

Yuan Liao   Xiye Yang
Rutgers University

Dec 2017
Greater NY Econometrics

Introduction

- Consider a factor model

$$Y_{lt} = \boldsymbol{\beta}_{lt}'\mathbf{f}_t + u_{lt}, \quad l \leq p, \quad t \leq T.$$

  Large p, large T.
- $\boldsymbol{\beta}_l$ depends on observed “instruments” $\mathbf{X}$:
  - firm specific characteristics (e.g., Conner et al 15)

    $$\textit{size}, \textit{value}, \textit{momentum}, \textit{volatility}$$
  - common macroeconomic variables (e.g., Ferson and Harvey 99)

    $$\textit{bonds' spread}, \textit{moody' ratings}$$

Introduction

- We propose:

$$\beta_{lt} = \mathbf{g}_{lt}(\mathbf{X}_{lt}) + \gamma_{lt}$$

  - "g" part: captures long-run patterns in beta.
  - "γ" part: captures high-frequency remaining components in beta
- $\mathbf{X}_{lt} = (\mathbf{z}_l, \mathbf{x}_t, \mathbf{w}_{lt})$.

  $\mathbf{z}_l$: firm specific, time-invariant

  $\mathbf{x}_t$: common to stocks, e.g., macroeconomic variables

  $\mathbf{w}_{lt}$: both firm spec. and time-varying.
- This paper:

  Inference about instruments' effects
- Effect of $\mathbf{X}$ on β represents instruments' effects on risks.

Introduction

$$y_{lt} = \beta_{lt}'\mathbf{f}_t + u_{lt}, \quad \boldsymbol{\beta}_{lt} = \mathbf{g}_{lt}(\mathbf{X}_{lt}) + \boldsymbol{\gamma}_{lt}$$

- Main challenge: DO NOT know whether γ_{lt} is close to zero or not.
  Inference about **g** is uniformly valid over a class of DGP for γ
- Two ways of dealing of time-varying models
  1. time-domain kernel smoothing, e.g., Ang et al. (2010)
     $$\text{fix "time" } r, \quad \min \frac{1}{Tp}\sum_{l,t}(y_{lt} - \beta_{lt}'\mathbf{f}_t)^2 K\left(\frac{t-r}{Th}\right)$$
  2. High frequency data (this paper).

Main results

1. The distribution of $\widehat{\mathbf{g}}$ has a discontinuity when $\text{var}(\gamma_{it}) \approx 0$.

$$rate = \frac{1}{\sqrt{pT}} + \sqrt{\frac{\text{var}(\gamma_{it})}{p}}$$

2. Standard "plug-in" for asym. variance is only *pointwise*, NOT uniformly valid over the strength of $\text{var}(\gamma_{it})$.
3. "cross-sectional bootstrap" achieves uniformity.
4. When factors unknown, $\widehat{\mathbf{g}}$ is biased, due to the effect of estimating factors.
5. Forecast using estimated factors: prediction interval is also misleading if time-varying γ_{it} is ignored.

The econometric problem: A Toy example

$$y_{lt} = [\mathbf{x}_l'\boldsymbol{\theta} + \gamma_l]'\mathbf{f}_t + \mathbf{u}_{lt}$$

$$\text{matrix form:} \quad \mathbf{Y} = [\mathbf{G} + \boldsymbol{\Gamma}]\mathbf{F}' + \mathbf{U}, \quad \mathbf{G} = \mathbf{X}\boldsymbol{\theta}$$

Key assumption: $\mathbb{E}(\gamma_l | \mathbf{x}_l) = 0, \quad \mathbb{E}(\mathbf{u}_{lt} | \mathbf{x}_l) = 0.$

Estimations:

$$\widehat{\mathbf{G}} = \mathbf{P}_X \mathbf{Y}\mathbf{F}(\mathbf{F}'\mathbf{F})^{-1} \Rightarrow \widehat{\mathbf{g}}_l$$

- Two understandings:

  (i) first time series reg, then cross-sec reg.

  (ii) first cross-sec reg, then time series reg.
- recommend the second understanding (will get to this later)

The "discontinuity issue"

- Fix $l \leq p$,

$$\widehat{\mathbf{g}}_l - \mathbf{g}_l = \underbrace{\frac{1}{pT}\sum_{t=1}^{T}\sum_{m=1}^{p}\mathbf{f}_t h_l(\mathbf{X}_m)u_{mt}}_{A_1} + \underbrace{\frac{1}{p}\sum_{m=1}^{p}\gamma_m h_l(\mathbf{X}_m)}_{A_2}$$

$$A_1 \sim \frac{1}{\sqrt{Tp}}, \quad A_2 \sim \sqrt{\frac{\mathrm{var}(\gamma_m)}{p}}$$

- Case I: If $\mathrm{var}(\gamma_m) \gg 0$,

$$\sqrt{p}(\widehat{\mathbf{g}}_l - \mathbf{g}_l) = \sqrt{p}A_2 + o_P(1)$$

Case II: If $\mathrm{var}(\gamma_m) \approx 0$,

$$\sqrt{pT}(\widehat{\mathbf{g}}_l - \mathbf{g}_l) = \sqrt{pT}A_1 + o_P(1)$$

Uniform Confidence interval

$$\text{asym. var.} = \frac{1}{pT}\text{var}(\sqrt{pT}A_1) + \frac{1}{p}\text{var}(\sqrt{p}A_2)$$

- Usual "plug-in" method: use

$$\frac{1}{pT}\widehat{\text{var}}(\sqrt{pT}A_1) + \frac{1}{p}\widehat{\text{var}}(\sqrt{p}A_2)$$

- But this method is only *pointwise*.
- Goal: CI_τ,

$$\sup_{\mathbb{P}\in\mathcal{P}} |\mathbb{P}(\mathbf{g}_l \in CI_\tau) - (1-\tau)| \to 0.$$

Here $\mathcal{P}$ is a broad class of DGP, containing various strenghs of γ_l.

Poinwise Confidence interval

$$\mathrm{var}(\sqrt{p}A_2) = \frac{1}{p}\sum_{m=1}^{p} h_m(\mathbf{X}_{mt})^2 \mathrm{var}(\gamma_{mt}|\mathbf{X}_{mt})$$

- $\widehat{\mathrm{var}}(\sqrt{p}A_2) - \mathrm{var}(\sqrt{p}A_2)$ equals

$$\underbrace{\frac{1}{p}\sum_{m=1}^{p} h_m^2 \left[\widehat{\gamma}_{mt}\widehat{\gamma}_{mt}' - \gamma_{mt}\gamma_{mt}'\right]}_{\gamma\text{-estimation error}} + \underbrace{\frac{1}{p}\sum_{m=1}^{p} h_m^2 \left[\gamma_{mt}\gamma_{mt}' - \mathbb{E}\gamma_{mt}\gamma_{mt}'\right]}_{\text{LLN error}}$$

- The "γ-est. error" is lower bounded by $O_P(T^{-1})$.
- Hence when $\mathrm{var}(\sqrt{p}A_2) \approx 0$,

$$\widehat{\mathrm{var}}(\sqrt{p}A_2) - \mathrm{var}(\sqrt{p}A_2) \asymp T^{-1} \quad \text{NOT negligible}$$

  leads to over-coverage.
- If ignore $\mathrm{var}(\sqrt{pA_2})$, then when it is "large", leads to under-coverage.

A possible solution

- The optimal rate:

$$\widehat{\text{var}}(\sqrt{p}A_2) - \text{var}(\sqrt{p}A_2) \asymp T^{-1}$$

- To make optimal rate even "faster": regularizations

shrinkage, penalizations, thresholding, trimming...

- Use

$$\widehat{\text{var}}(\sqrt{p}A_2)1\{\widehat{\text{var}}(\sqrt{p}A_2) \leq k_n\}$$

- Similar idea used for distribution with discontinuity:
  - finding binding moment inequalities
  - inference for parameters on boundaries
  - Random coefficients

Cross-sectional bootstrap

- key assumption: cross-sectional independence
- Resample individuals, keep the entire time series for each sampled indiv.

$$\widehat{\mathbf{g}}_l - \mathbf{g}_l = \underbrace{\frac{1}{pT}\sum_{t=1}^{T}\sum_{m=1}^{p}\mathbf{f}_t h_l(\mathbf{X}_m)u_{mt}}_{A_1} + \underbrace{\frac{1}{p}\sum_{m=1}^{p}\gamma_m h_l(\mathbf{X}_m)}_{A_2}$$

$$\widehat{\mathbf{g}}_l^* - \mathbf{g}_l = \frac{1}{pT}\sum_{t=1}^{T}\sum_{m=1}^{p}\mathbf{f}_t h_l^*(\mathbf{X}_m^*)u_{mt}^* + \frac{1}{p}\sum_{m=1}^{p}\gamma_m^* h_l^*(\mathbf{X}_m^*)$$

- Bootstrap $\text{var}^*(\sqrt{p}A_2) = \frac{1}{p}\sum_m h_m^2\gamma_{mt}\gamma_{mt}'$, so only has

$$\underbrace{\frac{1}{p}\sum_{m=1}^{p} h_m^2\left[\gamma_{mt}\gamma_{mt}' - \mathbb{E}\gamma_{mt}\gamma_{mt}'\right]}_{\text{LLN error}}$$

- avoids γ-estimation error.
- cross-sec. bootstrap leads to uniformly valid C.I.

Why Cross-sectional bootstrap works?

Key differences from Andrews (1999, 2000):

- Different sources of "discontinuity in the asymptotic distribution":

$$\mathbf{s}_{f,t}^{-1}\frac{1}{k_n p\Delta_n}\sum_{i\in I_t^n}\sum_{m=1}^{p}\Delta_i^n\mathbf{F}\Delta_i^n U_m h_{i-1,ml}, \quad \text{and} \quad \underbrace{\frac{1}{p}\sum_{m=1}^{p}\gamma_{m,t}h_{t,ml}}_{(\mathbf{a})},$$

  The "discontinuity" is a consequence of the interplay between the two terms. But the term (**a**) itself is continuous.

- The "γ-estimation error" can be avoided by resampling the cross-sectional units.

- Whether γ_{lt} is near the boundary is unknown, and is uncertain under both the null and the alternative.

Formal treatments: High Frequency Setup

- Sample period $[0, T]$, $T < \infty$.
  - Sampling interval length $\Delta_n \to 0$.
  - Number of "times" on each interval: k_n,
- On each interval:

$$\begin{aligned} \Delta_i^n \mathbf{Y} &= \mathbf{Y}_{i\Delta_n} - \mathbf{Y}_{(i-1)\Delta_n} \\ &= (\mathbf{G}_{i-1} + \mathbf{\Gamma}_{i-1})\Delta_i^n \mathbf{F} + \Delta_i^n \mathbf{U} + \text{neglibigle terms}, \end{aligned} \tag{1.1}$$

- Time variational error: Burkholder-Davis-Gundy Inequality

$$\mathbb{E}\left(\sup_{u \in [0, k_n \Delta_n]} \left| \text{Semimartingale increment over } [t, t+u] \right| \right) = O(\sqrt{k_n \Delta_n})$$

- Known factor case

$$\widehat{\mathbf{G}}_t = \sum_{i \in I_t^n} \mathbf{P}_{i-1} \, \Delta_i^n \mathbf{Y} \, \Delta_i^n \mathbf{F}' \left(\sum_{i \in I_t^n} \Delta_i^n \mathbf{F} \, \Delta_i^n \mathbf{F}' \right)^{-1},$$

$$\widehat{\mathbf{\Gamma}}_t = \sum_{i \in I_t^n} (\mathbf{I}_p - \mathbf{P}_{i-1}) \Delta_i^n \mathbf{Y} \, \Delta_i^n \mathbf{F}' \left(\sum_{i \in I_t^n} \Delta_i^n \mathbf{F} \, \Delta_i^n \mathbf{F}' \right)^{-1}, \tag{1.2}$$

- $\mathbf{P}_{i-1}$: projection matrix using sieve basis of $\mathbf{X}_{i-1}$.
-
$$\widehat{\mathbf{g}}_{lt} - \mathbf{g}_{lt} = (A_1 + A_2) + small$$

- Unknown factor case: Replace $\Delta_i^n \mathbf{F}$ by $\widehat{\Delta_i^n \mathbf{F}}$.
- How do we estimate $\widehat{\Delta_i^n \mathbf{F}}$?
  - Usual methods: PCA on
    $$(\Delta^n \mathbf{Y})'(\Delta^n \mathbf{Y}), \quad T \times T$$
  - We use: PCA on
    $$(\mathbf{P}\Delta^n \mathbf{Y})'(\mathbf{P}\Delta^n \mathbf{Y}), \quad T \times T$$
- $\mathbf{P}\Delta^n \mathbf{Y}$ is "noise-free" :

$$\mathbf{P}\Delta^n \mathbf{Y} = \mathbf{P}\mathbf{G}\Delta^n \mathbf{F}' + \underbrace{\mathbf{P}\boldsymbol{\Gamma}\Delta^n \mathbf{F}' + \mathbf{P}\Delta^n \mathbf{U} + \text{drift}}_{\approx 0}$$

Then

$$(\mathbf{P}\Delta^n \mathbf{Y})'(\mathbf{P}\Delta^n \mathbf{Y}) = \Delta^n \mathbf{F}'\mathbf{G}'\mathbf{P}\mathbf{G}\Delta^n \mathbf{F}' + \textit{errors}$$

- The errors vanish as $p \to \infty$.
- $\widehat{\Delta_i^n \mathbf{F}}$ is also a GMM-type estimator.

Unknown factor case

$$\widehat{\mathbf{g}}_{lt} - \mathbf{H}_n \mathbf{g}_{lt} - bias = \mathbf{H}_n (A_1 + A_2) + small$$

- bias: effect of estimating factors.
- Case I: If $\mathrm{var}(\gamma_{lt}) \gg 0$,

$$\sqrt{p}(\widehat{\mathbf{g}}_{lt} - \mathbf{H}_n \mathbf{g}_{lt}) = \sqrt{p}\mathbf{H}_n A_2 + o_P(1)$$

  Case II: If $\mathrm{var}(\gamma_{lt}) \approx 0$,

$$\sqrt{pk_n}(\widehat{\mathbf{g}}_{lt} - \mathbf{H}_n \mathbf{g}_{lt} - bias) = \sqrt{pk_n}\mathbf{H}_n A_1 + o_P(1)$$

-
$$bias = bias(\mathrm{var}(\Delta_i^n \mathbf{U}))$$

  where $\mathrm{var}(\Delta_i^n \mathbf{U})$ is $p \times p$.
- So bias-correction requires estimating a high-dim. cov. matrix.

Bootstrap confidence interval

Fixed $l \leq p$

1. Estimate $\widehat{\mathbf{g}}_{lt}$, (and $\{\widehat{\Delta_i^n F}\}_{i \in I_t^n}$ if factors unknown.)
2. Take cross-sectional bootstrap $\{\Delta_i^n Y_1^*, ..., \Delta_i^n Y_p^*\}_{i \in I_t^n}$

   $$\text{and} \quad \{X_{i1}^*, ..., X_{ip}^*\}_{i \in I_t^n}$$

   Do NOT bootstrap factors.
3. Estimate $\widehat{\mathbf{g}}_{lt}^*$.
4. Repeat 2-3 B times, obtain $\widehat{\mathbf{g}}_{lt}^{*1}, ..., \widehat{\mathbf{g}}_{lt}^{*B}$.

   Let q_τ be the quantile of $|\widehat{\mathbf{g}}_{lt}^{*b} - \widehat{\mathbf{g}}_{lt}|$
5. $$CI_\tau = [\widehat{\mathbf{g}}_{lt} \pm q_\tau] \quad \text{known factors}$$
   $$CI_\tau = [\widehat{\mathbf{g}}_{lt} - \widehat{bias} \pm q_\tau] \quad \text{unknown factors}$$

Discussion of $E(\gamma|\mathbf{X}) = 0$

Suppose

$$\boldsymbol{\beta}_l = \mathbf{g}(\mathbf{x}_l) + \boldsymbol{\gamma}_l$$

and $\boldsymbol{\beta}_l$ were "known".

- If $E(\boldsymbol{\gamma}_l|\mathbf{X}_l) \neq 0$, then need IV $E(\boldsymbol{\gamma}_l|\mathbf{w}_l) = 0$. Define

$$\mathcal{T} : \mathbf{g} \to E(\mathbf{g}(\mathbf{x}_l)|\mathbf{w}_l)$$

  Then

$$\mathcal{T}\mathbf{g} = E(\boldsymbol{\beta}_l|\mathbf{w}_l)$$

- $\mathbf{g}$ is identified iff $\mathbf{x}_l|\mathbf{w}_l$ is *complete* (Newey and Powell 03).
- Even if $\mathbf{g}$ is ident., $\mathcal{T}^{-1}$ is discont. $\Rightarrow$ ill-posed problem. (Hall and Horowitz 06, Chen and Pouzo 12).

Simulations

- set $\mathbf{\Gamma} = \sqrt{w_\gamma} N(0,1)$, with various w_γ.
- Construct CI using three methods:

(i) the bootstrap confidence interval: $\widehat{g} \pm q_{\tau, bootstrap}$

(ii) the "over-coverage" confidence interval: $\widehat{g} \pm 1.96\sqrt{\widehat{\mathrm{var}}(A_1) + \widehat{\mathrm{var}}(A_2)}$
use plug-in estimated variance for γ.

(iii) the "under-coverage" confidence interval: $\widehat{g} \pm 1.96\sqrt{\widehat{\mathrm{var}}(A_1)}$
ignore γ.

Simulations

Table: Coverage probabilities, nominal probability= 95%

			w_γ = strength of γ			
k_n	p		0.001	0.1	1	3
10	300	bootstrap	0.946	0.947	0.950	0.948
		over	0.993	0.990	0.962	0.953
		under	0.931	0.923	0.437	0.294
200	300	bootstrap	0.948	0.946	0.949	0.953
		over	0.991	0.973	0.948	0.953
		under	0.929	0.756	0.111	0.069

1. Bootstrap is good uniformly over $\boldsymbol{\Gamma}$.
2. The “over” is conservative when γ is weak, and becomes better as γ is stronger.
3. The “under” has under coverages as γ is stronger.
4. Bootstrap is good even if k_n is small.

Out of sample forecast

- We forecast

$$y_{t+h} = \mathbf{b}' y_t + \rho' \mathbf{F}_t + v_t, \quad t = 1, ..., D$$

e.g.,

$$y_t = \text{int. volatility } = \int_{(t-1)T}^{tT} \sigma_s^2 ds$$

Each interval $[(t-1)T, tT]$ represents "one day".

- Goal: prediction interval for

$$y_{D+h|D} = \mathbf{b}' y_D + \rho' \mathbf{F}_D$$

using estimated factors.

- Forecast is low frequency discrete time;

Estimating factors: high-frequency data

$$\Delta_i^n \mathbf{Y} = \text{drift} + [\mathbf{G}_{i-1} + \mathbf{\Gamma}_{i-1}]\Delta_i^n \mathbf{F} + \Delta_i^n \mathbf{U}$$

$$\mathbf{F}_t = \int_{(t-1)T}^{tT} d\mathbf{F}_s$$

- Step 1 : Estimate $\Delta_i^n \mathbf{F}$

$$\widetilde{\mathbf{F}}_t = \sum_{i \in \text{Day t}} \widehat{\Delta_i^n \mathbf{F}}$$

De-mean (to remove drifts)

$$\widehat{\mathbf{F}}_t = \widetilde{\mathbf{F}}_t - \frac{1}{D}\sum_{t=1}^{D} \widetilde{\mathbf{F}}_t$$

- Step 2: Run OLS on

$$y_{t+h} = y_t + \rho' \widehat{\mathbf{F}}_t + error$$

- Key assumptions: $\mathbf{G}$ is time-invariant. But $\mathbf{\Gamma}$ is still time varying.

Estimating factors

$$y_{t+h} = y_t + \rho' \mathbf{F}_t + v_t, \quad t = 1, ..., D$$
$$\widehat{y}_{D+h|D} - y_{D+h|D} = A_1 + ... + A_4 + small$$

- $A_1 = \frac{1}{D} \sum_{t=1}^{D} w_t v_t$
- $A_2 = \mathbf{M}' \frac{1}{D} \sum_{t=1}^{D} \mathbf{F}_t$
- $A_3 = \frac{1}{p} \sum_{i=1}^{p} U_{iD} H_i$
- $$A_4 = \frac{1}{p} \left[\sum_{t=(D-1)T}^{DT} \Delta_t^n \mathbf{F}' \mathbf{\Gamma}_t' + \frac{1}{D-h} \sum_{t=1}^{D-h} w_d \sum_{s=(t-1)T}^{tD} \Delta_s^n \mathbf{F}' \mathbf{\Gamma}_s' \right] \mathbf{H}$$

  $A_4 = 0$ if $\mathbf{\Gamma}_t$ is time-invariant, but not so in general.
- So forecast interval is misleading if either (i) ignore $\mathbf{\Gamma}_t$, or (ii) treat $\mathbf{\Gamma}_t$ time-invariant.

Forecast intervals

- Using this expansion,

$$\mathrm{var}(\widehat{y}_{D+h|D} - y_{D+h|D}) = \frac{1}{D}\mathrm{var}(\sqrt{D}\mathbf{A}_1) + \frac{1}{D}\mathrm{var}(\sqrt{D}\mathbf{A}_2)$$
$$+\frac{1}{p}\mathrm{var}(\sqrt{p}\mathbf{A}_3) + \frac{1}{p}\mathrm{var}(\sqrt{p}\mathbf{A}_4)$$

- “Plug-in” estimated var works in this case.

$$\sup_{\mathbb{P}\in\mathcal{P}} \left|\mathbb{P}\left(y_{D+h|D} \in \widehat{y}_{D+h|D} \pm z_{\tau/2}\widehat{s}_n\right) - (1-\tau)\right| \to 0.$$

Empirical Study

Data

- Time period: 2006 to 2013.
- Sampling frequency: 5-min.
- 380 S&P500 component stocks.
- Firm characteristics: size, book to market ratio, momentum and volatility.
- Known factors: Market, HML, SMB and RMW.

Cross-sectional means of estimated **G**

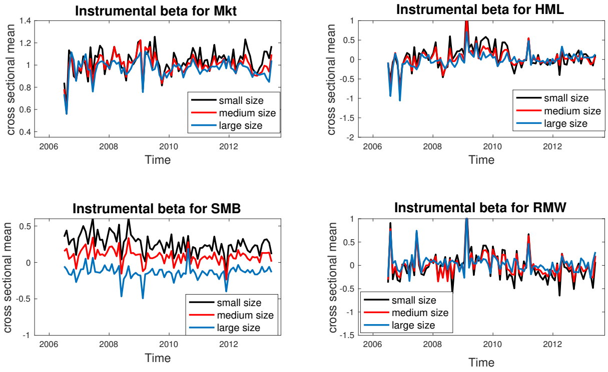

Figure: Grouped by size .

Size has negative effect on the betas for SMB risk factor.

Cross-sectional means of estimated **G**

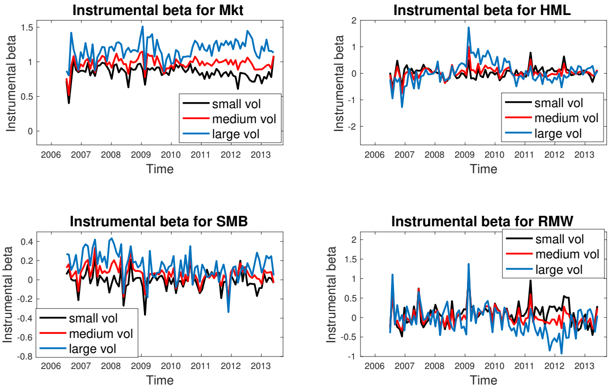

Figure: Grouped by volatility .

Volatility has positive effect on the betas for market factor.

Table: cross-sectional and time-series standard deviations of **G** and **Γ**

	Mkt		HML		SMB		RMW	
size	**G**	**Γ**	**G**	**Γ**	**G**	**Γ**	**G**	**Γ**
	Averaged (over time) cross-sectional std							
small	0.221	0.499	0.409	1.140	0.144	0.613	0.336	1.405
medium	0.179	0.446	0.352	1.063	0.112	0.524	0.281	1.290
large	0.165	0.411	0.333	0.994	0.142	0.466	0.260	1.245
	Averaged (over firms) times-series std							
small	0.195	0.471	0.407	1.095	0.175	0.590	0.431	1.332
medium	0.152	0.416	0.324	0.993	0.140	0.511	0.353	1.214
large	0.148	0.389	0.315	0.908	0.133	0.460	0.336	1.124

G has much smaller standard deviations than **Γ** in both cross-section and time series

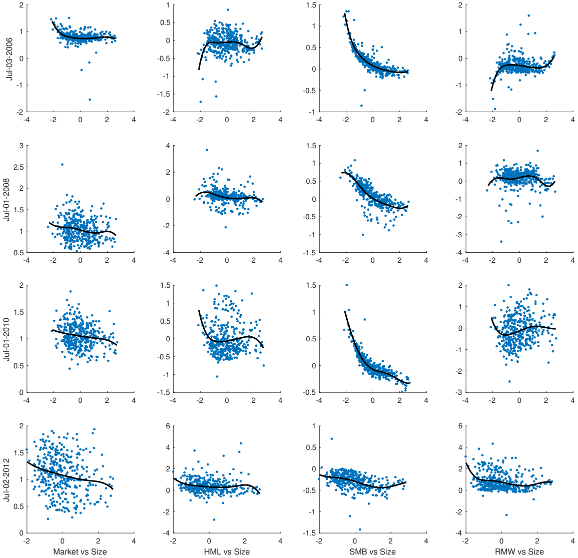
Jul-03-2006
Jul-01-2008
Jul-01-2010
Jul-02-2012
Market vs Size
HML vs Size
SMB vs Size
RMW vs Size

Table: Cross-sectional Proportion of significant **G** , 2006

	positive significance				negative significance			
size	Mkt	HML	SMB	RMW	Mkt	HML	SMB	RMW
small	1	0.261	0.870	0.134	0	0.177	0.004	0.184
large	1	0.133	0.047	0.154	0	0.229	0.544	0.062

Table: Cross-sectional Proportion of signs for β, **Γ** , 2006

	positive significance				negative significance			
size	Mkt	HML	SMB	RMW	Mkt	HML	SMB	RMW
				β				
small			0.71				0.28	
large			0.37				0.62	
				Γ				
small			0.51				0.52	
large			0.48				0.52	

Other directions

- test relevance of instruments: $g = x^T\theta + \gamma$. $H_0 : \theta_i = 0$.
  Confidence interval should be uniform in γ.
- Improve Fama-Macbeth regression using g in place of β.
- Inference about long-run instrumental effects: $\int_0^T g_s ds$
- Test $\gamma = 0$