Testing Stochastic Dominance with Many Conditioning Variables

Oliver Linton[1], **Myung Hwan Seo**[2], **Yoon-Jae Whang**[2]

[1] University of Cambridge, [2] Seoul National University

1. Introduction

Introduction

- **Stochastic Dominance (SD)**: *A popular ordering rule of various distribution functions.* Useful in
  - Ranking portfolio investment strategies
  - Comparing income distributions or poverty levels
  - Distributional treatment effects
  - Evaluation of forecasting models, etc.
- Attractive in that it does not require restrictive assumptions on the distributions of choice alternatives and preference structure of economic agents or policy makers (Levy (2016), Whang (2019).
- Allowing for conditioning information in decision making is essential in any applied context, which necessitates the concept of *conditional SD*. On the other hand, there is nowadays a plethora of available data for both decision makers and econometricians.
- This paper develops a test of conditional SD with *high-dimensional covariates*, allowing for *dependent observations.*

Hypotheses of Interests

- Let $y_{t\mathbf{1}}$ and $y_{t\mathbf{2}}$ denote the outcomes of interest to be compared and let $\mathcal{F}_{t-1}$ be the information set observed at time $t-1$.
- **First order Stochastic Dominance (FSD)**:

$$\mathcal{H}_0 : \Pr(y_{t\mathbf{1}} \le y|\mathcal{F}_{t-1}) \le \Pr(y_{t\mathbf{2}} \le y|\mathcal{F}_{t-1}) \quad \forall y \text{ a.s.}$$

- **Second order Stochastic Dominance (SSD)**:

$$\mathcal{H}_0 : \int_{-\infty}^{y} \Pr(y_{t\mathbf{1}} \le z|\mathcal{F}_{t-1})\,dz \le \int_{-\infty}^{y} \Pr(y_{t\mathbf{2}} \le z|\mathcal{F}_{t-1})\,dz \ \forall y \text{ a.s.}$$

- In both cases, the alternative hypothesis is the negation of $\mathcal{H}_0$.
- The dimensionality of $\mathcal{F}_{t-1}$ can be large and the conditional distributions of the outcome variables are unknown.

This Paper

- We propose (one-sided) Kolmogorov-Smirnov type tests based on a semi-nonparametric location scale model for the observed outcomes with unknown error distribution.
- We estimate the unknown location and scale functions by the regularized least squares (with thresholding) and the error distribution by the empirical distribution function of the rescaled residuals.
- We establish the weak convergence of the rescaled residual empirical process by developing an exponential inequality and deviation bounds for the regularized estimators with dependent data.
- We propose a smooth stationary bootstrap to compute the p-values and show its asymptotic validity.
- We provide Monte Carlo simulation results and applications to investigate the home bias problem in the stock market.

2. Model

Model

- For $\boldsymbol{j}=\mathbf{1},\mathbf{2}$, our approach builds on the following regression model:

$$y_{t\boldsymbol{j}} = g^{\boldsymbol{j}}(q_t) + \sigma^{\boldsymbol{j}}(q_t)\,\varepsilon_{t\boldsymbol{j}}, \quad t=1,\ldots,n. \tag{1}$$

  – The innovation $\{\varepsilon_{t\boldsymbol{j}}\}$ is an iid sequence with the common distribution $F^{\boldsymbol{j}}$.
  – The covariate $\{q_t\}$ is a k-dimensional vector of controls, whose dimension can be large (as n gets large).
  – We allow $F^{\boldsymbol{j}}(\cdot)$, $g^{\boldsymbol{j}}(\cdot)$, and $\sigma^{\boldsymbol{j}}(\cdot)$ to be nonparametric.
  – We assume that $\{\varepsilon_{t\boldsymbol{j}}\}$ and $\{q_t\}$ are mutually independent, so that

$$F^{\boldsymbol{j}}(y|q) := \Pr\left(y_{t\boldsymbol{j}} \le y|q_t = q\right) = F^{\boldsymbol{j}}\left(\frac{y-g^{\boldsymbol{j}}(q)}{\sigma^{\boldsymbol{j}}(q)}\right).$$

  – Then, the FSD hypothesis can be written as

$$\mathcal{H}_0 : F^{\mathbf{1}}\left(\frac{y-g^{\mathbf{1}}(q)}{\sigma^{\mathbf{1}}(q)}\right) \le F^{\mathbf{2}}\left(\frac{y-g^{\mathbf{2}}(q)}{\sigma^{\mathbf{2}}(q)}\right) \text{ for all } (y,q).$$

Sparsity Assumption

- The key assumption for the function g is that there exists a vector

$$X_t := (X_{t1},\ldots,X_{tp})^{\intercal} := (X_1(q_t),\ldots,X_p(q_t))^{\intercal} = X(q_t)$$

such that

$$g(q_t) = \beta_0^{\intercal} X_t + r_{gt}, \tag{2}$$

where β_0 is sparse and $r_{gt} \to 0$ at a proper order as the dimensionality of X_t expands.

- Analogously, we assume

$$\sigma(q_t) = \gamma_0^{\intercal} X_t + r_{\sigma t}, \tag{3}$$

where γ_0 is sparse and $r_{\sigma t} \to 0$ as the dimension of X_t grows.

Mean Regression

- Estimate β_0 in (2) by the (weighted) LASSO: i.e., the ℓ_1-penalized least squares with penality varying for each element in β:

$$\hat\beta_{lasso} := \arg\min_{\beta} \frac{1}{n}\sum_{t=1}^{n}\left(y_t - X_t^{\intercal}\beta\right)^2 + \lambda\,|D\beta|_1, \tag{4}$$

where D is a diagonal weighting matrix and λ is a tuning parameter.

- We introduce thresholding so that we can control the random variation arising from the imperfect selection of the smallish coefficients:

$$\hat S = \left\{ j : \left|\hat\beta_{lasso,j}\right| > \lambda_{thr} \right\}$$

where the threshold λ_{thr} is strictly larger than λ. (cf. SCAD (Fan and Li, 2001), adaptive LASSO (Zou, 2006)).

- After the selection, we re-estimate β by the OLS on the selected. Then,

$$\hat\beta_{Tasso} := \arg\min_{\beta:\beta_j=0, j\notin\hat S} \sum_{t=1}^{n}\left(y_t - X_t^{\intercal}\beta\right)^2,$$

which is equivalent to the OLS estimator $\hat\beta$ from the linear regression of y_t on $\hat x_t := X_{t,\hat S}$.

- Then, we may set

$$\hat g_t = \hat g(q_t) = X_t^{\intercal}\hat\beta_{Tasso} = \hat x_t^{\intercal}\hat\beta.$$

Skedastic Regression

- *Scale Normalization:* Since σ determines the scale of the error term $e_t = \sigma(q_t)\,\varepsilon_t$, it is natural to impose a certain scale normalization on the distribution of ε_t. We shall assume $E|\varepsilon_t| = 1$, which is convenient for our asymptotic theory.
- To estimate γ, note that the scale normalization is equivalent to the condition that

$$E\left[|e_t|\,|q_t\right] = \sigma(q_t),$$

or

$$|e_t| = \sigma(q_t) + \eta_t \quad \text{and} \quad E\left[\eta_t|q_t\right] = 0.$$

- Since we do not directly observe e_t, we employ the residual $\hat e_t = y_t - \hat x_t^{\intercal}\hat\beta$ and proceed as for the estimation of β.

Feasible Skedastic Regression

- Estimate γ_0 in (3) by the (weighted) LASSO:

$$\hat\gamma_{lasso} := argmin_{\gamma\in\mathbb{R}^p} \frac{1}{n}\sum_{t=1}^{n}\left(|\hat e_t| - X_t^{\intercal}\gamma\right)^2 + \mu\,|Q\gamma|_1,$$

where Q is a diagonal weighting matrix and μ is a tuning parameter.

- We then apply the OLS after thresholding. That is, let

$$\hat S_\gamma = \left\{ j : \left|\hat\gamma_{lasso,j}\right| \ge \mu_{thr} \right\},$$

for some μ_{thr}.

- let $\hat w_t = X_{t,\hat S_\gamma}$ and $\hat\gamma$ denote the OLS estimate of $|\hat e_t|$ on $\hat w_t$. We may also define $\hat\gamma_{Tasso}$ as the thresholded LASSO estimate as for $\hat\beta_{Tasso}$. Then, we may set

$$\hat\sigma_t = \hat\sigma(q_t) = X_t^{\intercal}\hat\gamma_{Tasso} = \hat w_t^{\intercal}\hat\gamma.$$

3. Test Statistics

Test Statistics

When $\boldsymbol{j}=\mathbf{1},\mathbf{2}$, the dataset is given by $\{y_{t\mathbf{1}}, y_{t\mathbf{2}}, q_t\}_{t=1}^n$ and the testing proceeds as follows.

1. For each $\boldsymbol{j}=\mathbf{1},\mathbf{2}$, run the regression of $y_{t\boldsymbol{j}}$ on X_t by the thresholded LASSO to get $\hat S^{\boldsymbol{j}}$.
2. Let $\hat S = \hat S^{\mathbf{1}} \cup \hat S^{\mathbf{2}}$ and define $\hat x_t := X_{t,\hat S}$, i.e. the collection of selected elements of X_t in at least one of the two regressions, and $\hat x(q) = X_{\hat S}(q)$.
3. Let $\hat\beta^{\boldsymbol{j}}$ denote the OLS estimate in the regression of $y_{t\boldsymbol{j}}$ on $\hat x_t$ and define the residual $\hat e_{t\boldsymbol{j}} = y_{t\boldsymbol{j}} - \hat x_t^{\intercal}\hat\beta^{\boldsymbol{j}}$.
4. Likewise, for each $\boldsymbol{j}=\mathbf{1},\mathbf{2}$, run the skedastic regression of $|\hat e_{t\boldsymbol{j}}|$ on X_t and compute $\hat S_\sigma$, $\hat w_t$, $\hat w(q)$, and $\hat\gamma^{\boldsymbol{j}}$ analogously to $\hat S$, $\hat x_t$, $\hat x(q)$, and $\hat\beta^{\boldsymbol{j}}$ in the preceding steps, respectively.
5. For each $\boldsymbol{j}=\mathbf{1},\mathbf{2}$, construct the scaled residual $\hat\varepsilon_{t\boldsymbol{j}} = \left(y_{t\boldsymbol{j}} - \hat x_t^{\intercal}\hat\beta^{\boldsymbol{j}}\right)/\hat\sigma_t^{\boldsymbol{j}}$, its empirical distribution function

$$\hat F^{\boldsymbol{j}}(\tau) = \frac{1}{n}\sum_{t=1}^{n} 1\{\hat\varepsilon_{t\boldsymbol{j}} \le \tau\}$$

and

$$\hat\tau^{\boldsymbol{j}}(y,q) = \left(y - \hat x(q)^{\intercal}\hat\beta^{\boldsymbol{j}}\right)/\hat\sigma^{\boldsymbol{j}}(q).$$

6. Construct the test statistic for the FSD hypothesis

$$T_n = \sqrt{n}\sup_{y,q}\left[\hat F^{\mathbf{1}}\left(\hat\tau^{\mathbf{1}}(y,q)\right) - \hat F^{\mathbf{2}}\left(\hat\tau^{\mathbf{2}}(y,q)\right)\right],$$

and, for the SSD hypothesis,

$$U_n = \sqrt{n}\sup_{y,q}\int_{-\infty}^{y}\left[\hat F^{\mathbf{1}}\left(\hat\tau^{\mathbf{1}}(u,q)\right) - \hat F^{\mathbf{2}}\left(\hat\tau^{\mathbf{2}}(u,q)\right)\right]du.$$

Asymptotic Distributions

- To characterize the asymptotic distributions, recall that $x_t = X_{t,S}$ and $w_t = X_{t,S_\gamma}$ and define

$$\mathcal{D}(y,q) = \mathbb{D}(y,q) + f(\tau(y,q))\,D_1 - \tau(y,q)\,f(\tau(y,q))\,D_2,$$

where $\tau(y,q) = (y-g(q))/\sigma(q)$, and $\mathbb{D}$ and $D = (D_1,D_2)^{\intercal}$ are centered gaussian processes with covariance kernels given by:

$$E\mathbb{D}(y_1,q_1)\,\mathbb{D}(y_2,q_2) = cov\left(1\{\varepsilon_{t\mathbf{1}} \le \tau_1\} - 1\{\varepsilon_{t\mathbf{2}} \le \tau_1\},\ 1\{\varepsilon_{t\mathbf{1}} \le \tau_2\} - 1\{\varepsilon_{t\mathbf{2}} \le \tau_2\}\right)$$

with $\tau_i = (y_i - g(q_i))/\sigma(q_i)$, $i=1,2$, and

$$EDD^{\intercal} = \lim_{n\to\infty} E\begin{bmatrix} \tilde x_t^2(\varepsilon_{t\mathbf{1}}-\varepsilon_{t\mathbf{2}})^2, & \tilde x_t(\varepsilon_{t\mathbf{1}}-\varepsilon_{t\mathbf{2}})\,\tilde w_t(|\varepsilon_{t\mathbf{1}}|-|\varepsilon_{t\mathbf{2}}|) \\ \cdot & \tilde w_t^2(|\varepsilon_{t\mathbf{1}}|-|\varepsilon_{t\mathbf{2}}|)^2 \end{bmatrix}$$

$$E\mathbb{D}(y_1,q_1)\,D = \lim_{n\to\infty} E\left(1\{\varepsilon_{t\mathbf{1}} \le \tau_1\} - 1\{\varepsilon_{t\mathbf{2}} \le \tau_1\}\right) \times \begin{pmatrix} \tilde x_t(\varepsilon_{t\mathbf{1}}-\varepsilon_{t\mathbf{2}}) \\ \tilde w_t(|\varepsilon_{t\mathbf{1}}|-|\varepsilon_{t\mathbf{2}}|) \end{pmatrix},$$

with $\tilde x_t = \mu_x^{\intercal}\left(Ex_tx_t^{\intercal}\right)^{-1}x_t$ and $\tilde w_t = \mu_w^{\intercal}\left(Ew_tw_t^{\intercal}\right)^{-1}w_t\sigma_t$.

- Consider the following class of hypotheses regarding the data distribution:

$$\begin{aligned} g^{\mathbf{1}}(q) &= g(q),\ g^{\mathbf{2}}(q) = g(q) + \delta_{1n}(q) \\ \sigma^{\mathbf{1}}(q) &= \sigma(q),\ \sigma^{\mathbf{2}}(q) = \sigma(q) + \delta_{2n}(q) \\ F^{\mathbf{1}}(\tau) &= F(\tau),\ F^{\mathbf{2}}(\tau) = F(\tau) + \delta_{3n}(\tau), \end{aligned}$$

such that $\int \tau dF^{\mathbf{2}}(\tau) = \int \tau dF^{\mathbf{1}}(\tau) = 0$ and $\int |\tau|\,dF^{\mathbf{2}}(\tau) = \int |\tau|\,dF^{\mathbf{1}}(\tau) = 1$.

- The least favorable case (LFC) of the null hypothesis corresponds to $\delta_{in} = 0$, for all $i=1,2,3$.
- We derive the asymptotic distribution of the test statistics under the drifting sequence of models

$$(\delta_{1n}(q), \delta_{2n}(q), \delta_{3n}(\tau)) = \frac{1}{\sqrt{n}}(\delta_1(q), \delta_2(q), \delta_3(\tau)), \tag{5}$$

for all n, where δ_i is continuous and bounded for all i, q, and τ.

- Let

$$\mathcal{B}(y,q) = \frac{\partial F(\tau)}{\partial\tau}\frac{1}{\sigma(q)}(\delta_1(q) + \tau\delta_2(q)) - \delta_3(\tau).$$

and let $\mathcal{P}$ denote the collection of all the joint distributions that satisfy Assumptions A-C.

Theorem 1 *Suppose that Assumptions A, B and C hold. Then, under (5)*

$$T_n \Rightarrow \sup_{y,q}\left[\mathcal{D}(y,q) + \mathcal{B}(y,q)\right], \tag{6}$$

$$U_n \Rightarrow \sup_{y,q}\int_{-\infty}^{y}\left[\mathcal{D}(u,q) + \mathcal{B}(u,q)\right]du. \tag{7}$$

Boosting Power

- We propose to apply a *screening principle*, which is to test certain implications of the null hypothesis with a *higher criticism.* (cf.) Fan et al. (2015).
- One implication of the first order stochastic dominance of $y_t^{\mathbf{1}}$ over $y_t^{\mathbf{2}}$ (conditional on $q_t = q$) is the dominance of the conditional means, i.e.,

$$E\left(Y_t^{\mathbf{1}}|q_t = q\right) \ge E\left(Y_t^{\mathbf{2}}|q_t = q\right).$$

The negation of this implication implies the negation of the null hypothesis.

- Using the conditional mean function $\hat g(q_t) = X_t^{\top}\hat\beta_{Tasso}$, which is estimated to construct our main test statistic T_n, we can screen this implication for a sequence of values of $X_t \in \{x_1,\ldots,x_J\}$ by statistics

$$t_k = 1\left\{\frac{x_k^{\intercal}\left(\hat\beta^{\mathbf{2}} - \hat\beta^{\mathbf{1}}\right)}{\hat\sigma_k} > c^*\right\},\ k=1,\ldots,J,$$

for some scaling $\hat\sigma_k$ and a critical value c^*.

- If $t_k = 1$ for any k, we can stop and conclude that the null is rejected. Otherwise, we resort to our test statistic T_n.

- To justify this initial screening, the value c^* needs to satisfy the high criticism property that

$$\Pr\left\{\max_{k\in\{1,2,\ldots,J\}} \frac{x_k^\intercal\left(\widehat{\beta}^2-\widehat{\beta}^1\right)}{\widehat{\sigma}_k} \le c^*\right\} = 1-o(1)$$

under the null hypothesis.
- As for x_k, we may consider $x_k = X(q_k)$ for a grid of $\{g_k\}$.
- Since

$$\max_k \frac{x_k^\intercal\left(\widehat{\beta}^2-\widehat{\beta}^1\right)}{\widehat{\sigma}_k} \le \max_k \left|\frac{x_k}{\widehat{\sigma}_k}\right|_1 \left|\left(\widehat{\beta}^2-\widehat{\beta}^1\right)\right|_\infty,$$

we may utilize the deviation bounds for $\widehat{\beta}^j$ as in our Lemma 3 and set $c^* = C(\log\log n)(\log s)/\sqrt{n}$
- For scaling, we suggest to set

$$\hat{\sigma}_k^2 = n\sum_{i=1}^{J} x_{ki}^2 \left(\sum_{t=1}^{n} X_{ti}^2\right)^{-2} \sum_{t=1}^{n} X_{ti}^2 \left(x_k^\intercal \hat{\gamma}\right)^2 \pi/2,$$

which corresponds to the case where there is no correlation among the X_{ti}'s. These estimates are uniformly bounded.

4. Bootstrap

- To compute the critical values, we suggest the *smooth stationary bootstrap*, which combines the methods of Politis and Romano (1994) and Neumeyer (2009) to take care of the complexity of our test statistics due to the temporal dependence and the highly nonlinear nature of the statistics.
- **Stationary Bootstrap**
  1. Let d_t and i_t, $t = 1,\ldots,n$, be random draws from Bernoulli(π_n) and Uniform$\{1,\ldots,n\}$, respectively.
  2. Let $i_1^* = i_1$.
  3. For $t = 2,\ldots,n$, let

$$i_t^* = \left(i_{t-1}^*+1\right)(1-d_t) + i_t d_t$$

  with the convention that $i_{t-1}^*+1 = 1$ if $i_{t-1}^* = n$.
- **Smooth Stationary Bootstrap** of $\mathcal{Z}_n = \{v_1,\ldots,v_n\}$:

$$v_t^* = v_{i_t^*} + a_n\eta_t,$$

where $\eta_t \sim G$ and $a_n \to 0$ as $n\to\infty$

Bootstrap Test Statistic

1. Fix constants a_n and π_n within the interval $(0,1)$ and a smooth distribution function G and generate $\{i_t^*, \eta_t\}$ as described before.
2. For each $\boldsymbol{j} = \mathbf{1},\mathbf{2}$, construct the bootstrap sample $\{x_t^* = \hat{x}_{i_t^*}, w_t^* = \hat{w}_{i_t^*}\}$ and$\{\varepsilon_{t\boldsymbol{j}}^* = \hat{\varepsilon}_{i_t^*,\boldsymbol{j}} + a_n\eta_t\}$, respectively, and then compute

$$y_{t\boldsymbol{j}}^* = x_t^{*\intercal}\hat{\beta}^{\boldsymbol{j}} + w_t^{*\intercal}\hat{\gamma}^{\boldsymbol{j}}\cdot\varepsilon_{t\boldsymbol{j}}^*, \quad t=1,\ldots,n.$$

3. For each $\boldsymbol{j} = \mathbf{1},\mathbf{2}$, obtain the OLS estimates $\hat{\beta}^{\boldsymbol{j}*}$ with the bootstrap sample $\{x_t^*, y_{t\boldsymbol{j}}^*\}$, i.e.,

$$\hat{\beta}^{\boldsymbol{j}*} = \left(\sum_{t=1}^n x_t^* x_t^{*\intercal}\right)^{-1}\sum_{t=1}^n x_t^* y_{t\boldsymbol{j}}^*,$$

and compute the bootstrap OLS residuals $\hat{e}_{t\boldsymbol{j}}^* = y_t^* - x_t^{*\intercal}\hat{\beta}^{\boldsymbol{j}*}$, $t=1,\ldots,n$. Then, compute:

$$\hat{\gamma}^{\boldsymbol{j}*} = \left(\sum_{t=1}^n w_t^* w_t^{*\intercal}\right)^{-1}\sum_{t=1}^n w_t^*\left|\hat{e}_{t\boldsymbol{j}}^*\right|,$$

$$\hat{\varepsilon}_{t\boldsymbol{j}}^* = \hat{e}_{t\boldsymbol{j}}^*\left(w_t^{*\intercal}\hat{\gamma}^{\boldsymbol{j}*}\right)^{-1},$$

$$^{\boldsymbol{j}*}(y,q) = \left(\hat{w}(q)^\intercal\hat{\gamma}^{\boldsymbol{j}*}\right)^{-1}\left(y - \hat{x}(q)^\intercal\hat{\beta}^{\boldsymbol{j}*}\right).$$

4. Define the empirical distribution functions

$$\hat{F}^{\boldsymbol{j}*}(\tau) = \frac{1}{n}\sum_{t=1}^n 1\left\{\hat{\varepsilon}_{t\boldsymbol{j}}^* \le \tau\right\}\ ;\ F^{\boldsymbol{j}*}(\tau) = \frac{1}{n}\sum_{t=1}^n G\left(\frac{\tau-\hat{\varepsilon}_{t\boldsymbol{j}}}{a_n}\right).$$

Then construct the bootstrap statistics

$$T_n^* = \sqrt{n}\sup_{y,q}\left[\hat{F}^{\mathbf{1}*}\left(\hat{\tau}^{\mathbf{1}*}(y,q)\right) - \hat{F}^{\mathbf{2}*}\left(\hat{\tau}^{\mathbf{2}*}(y,q)\right) - \left(F^{\mathbf{1}*}\left(\hat{\tau}^{\mathbf{1}}(y,q)\right) - F^{\mathbf{2}*}\left(\hat{\tau}^{\mathbf{2}}(y,q)\right)\right)\right],$$

$$U_n^* = \sqrt{n}\sup_{y,q}\int_{-\infty}^{y}\left[\hat{F}^{\mathbf{1}*}\left(\hat{\tau}^{\mathbf{1}*}(y,q)\right) - \hat{F}^{\mathbf{2}*}\left(\hat{\tau}^{\mathbf{2}*}(y,q)\right) - \left(F^{\mathbf{1}*}\left(\hat{\tau}^{\mathbf{1}}(y,q)\right) - F^{\mathbf{2}*}\left(\hat{\tau}^{\mathbf{2}}(y,q)\right)\right)\right]du.$$

Asymptotic Properties of the Test Statistics

Theorem 2 *Suppose that Assumptions A-D hold. Let c_α^* denote the bootstrap critical value of level α for T_n. Then, under $\mathcal{H}_0$, we have*

$$\limsup_n \Pr\{T_n > c_\alpha^*\} \le \alpha$$

for any $0<\alpha<1$, while under $\mathcal{H}_1$,

$$\Pr\{T_n > c_\alpha^*\} \to 1$$

for any $0<\alpha<1$. The same holds true for U_n.

5. Monte Carlo Simulations

Simulation Designs

- For $\boldsymbol{j} = \mathbf{1}$ or $\boldsymbol{j} = \mathbf{2}$, the true DGP is

$$y_t^{\boldsymbol{j}} = \beta^{\boldsymbol{j}} q_{1,t} + c_v\cdot\left(|q_{1,t}|+1\right)\cdot\varepsilon_t$$

where $c_v = 0.3, \varepsilon_t$ is an i.i.d. normal with mean 0 and $\mathbb{E}[|\varepsilon_t|] = 1$. The explanatory variables $q_{i,t}$ are generated by:

$$q_{i,t} = a + bq_{i,t-1} + e_{i,t}$$

where $i = 1,2$, $a = 0$, $b = 0.5$, and $t = 1,2,\cdots n$.
- We estimate the model based on $X_t = X(q_{1,t}, q_{2,t})$, which are transformations (powers and interaction terms up to polynomial order of 10) of $q_{1,t}, q_{2,t}$ and are common for $\boldsymbol{j} = \mathbf{1},\mathbf{2}$ and some additional variables (described below).
- The parameters for LASSO is $\lambda = c_{v'}\cdot\sqrt{\log p/n}$. The threshold parameter is $\lambda_{thr} = 2\lambda$.

Size

- **Additional Variables:** We increase p by adding more terms to X_t with three different ways.
  1. Grow polynomial order of $q_{1,t}, q_{2,t}$ constructing X_t.
  2. Generate more $q_{3,t}, q_{4,t}, \ldots$ pairs and add them to X_t.
  3. Add lagged $q_{1,t}, q_{2,t}$ terms and its powers (up to 10).
- **Tables 1-3:** Rejection rates at the significance level of $\alpha = 0.05$ with the true parameter value of $\beta^1 = \beta^2 = 1$ out of 1000 simulation iterations.
- **Table 4:** Rejection rates with with different values of $b = 0.3, 0.4, \ldots, 0.9$ to examine the effect of higher serial correlation in q_t.

Table 1: *Rejection probability with higher polynomial orders*

order	10	15	20	25	30	35	40
$n\setminus p$	65	135	230	350	495	665	860
100	0.077	0.062	0.065	0.073	0.068	0.082	0.084
200	0.075	0.048	0.060	0.055	0.048	0.060	0.047
300	0.055	0.043	0.048	0.053	0.058	0.045	0.051
400	0.052	0.056	0.055	0.041	0.047	0.051	0.044
500	0.048	0.051	0.047	0.045	0.051	0.039	0.046

Table 2: *Rejection probability with additional q pairs*

New Pairs	1	3	5	7	10	13	15
$n\setminus p$	130	260	390	520	715	910	1040
100	0.071	0.070	0.073	0.085	0.058	0.063	0.068
200	0.061	0.062	0.064	0.057	0.062	0.056	0.061
300	0.058	0.070	0.062	0.044	0.072	0.054	0.055
400	0.048	0.067	0.062	0.057	0.069	0.075	0.052
500	0.048	0.054	0.058	0.062	0.067	0.065	0.054

Table 3: *Rejection probability with lagged q terms*

Max lag	5	10	15	20	25	30	35	40
$n\setminus p$	165	265	365	465	565	665	765	865
100	0.072	0.072	0.082	0.095	0.088	0.070	0.079	0.078
200	0.064	0.075	0.071	0.070	0.067	0.045	0.067	0.070
300	0.048	0.074	0.082	0.091	0.077	0.060	0.084	0.070
400	0.070	0.073	0.068	0.068	0.071	0.072	0.062	0.070
500	0.063	0.071	0.078	0.070	0.086	0.087	0.075	0.067

Table 4: *Rejection probability with different AR coefficients b*

$n\setminus b$	0.3	0.4	0.5	0.6	0.7	0.8	0.9
100	0.082	0.089	0.087	0.093	0.089	0.088	0.132
200	0.068	0.072	0.080	0.073	0.090	0.088	0.089
300	0.076	0.077	0.082	0.068	0.078	0.070	0.092
400	0.082	0.093	0.057	0.086	0.079	0.081	0.083
500	0.088	0.085	0.072	0.066	0.079	0.060	0.094

Power

- We fix Max Lag = 30 so that $p = 665$ and evaluate the power performance of our test in three ways.

1. **Table 5:** Change $\beta^2 = 1.0, 1.1, \ldots, 2.0$.
2. **Table 6:** Shift y^2 by adding $\alpha = 0.1, \ldots, 1.0$.
3. **Table 7:** Change the error distribution by letting ε_t^2 follow $(Z^2-1)/0.9680$, i.e. chi-square with one degrees of freedom normalized to mean 0 and the first absolute moment 1 and compare it with normal distribution with mean 0 and the first absolute moment 1.

Table 5: *Rejection probability with β^2 being* $1.0, 1.1, \cdots, 1.5$

$n\setminus\beta^2$	1.0	1.1	1.2	1.3	1.4	1.5
100	0.090	0.124	0.284	0.433	0.636	0.801
200	0.065	0.135	0.357	0.625	0.816	0.935
300	0.079	0.181	0.465	0.764	0.936	0.976
400	0.091	0.192	0.557	0.866	0.969	0.981
500	0.082	0.238	0.686	0.933	0.977	0.987

Table 6: *Rejection probability after shifting y^2 by α*

$n\setminus\alpha$	0.0	0.1	0.2	0.3	0.4	0.5
100	0.070	0.251	0.488	0.741	0.918	0.988
200	0.086	0.324	0.764	0.970	0.987	0.994
300	0.079	0.452	0.860	0.983	0.994	0.986
400	0.082	0.530	0.933	0.981	0.989	0.994
500	0.081	0.581	0.968	0.983	0.992	0.998

Table 7: *Rejection probability of normal vs. chi-square error distribution*

n	Rejection Prob.
100	0.043
200	0.094
300	0.182
400	0.327
500	0.417

6. Application

Home Bias Puzzle

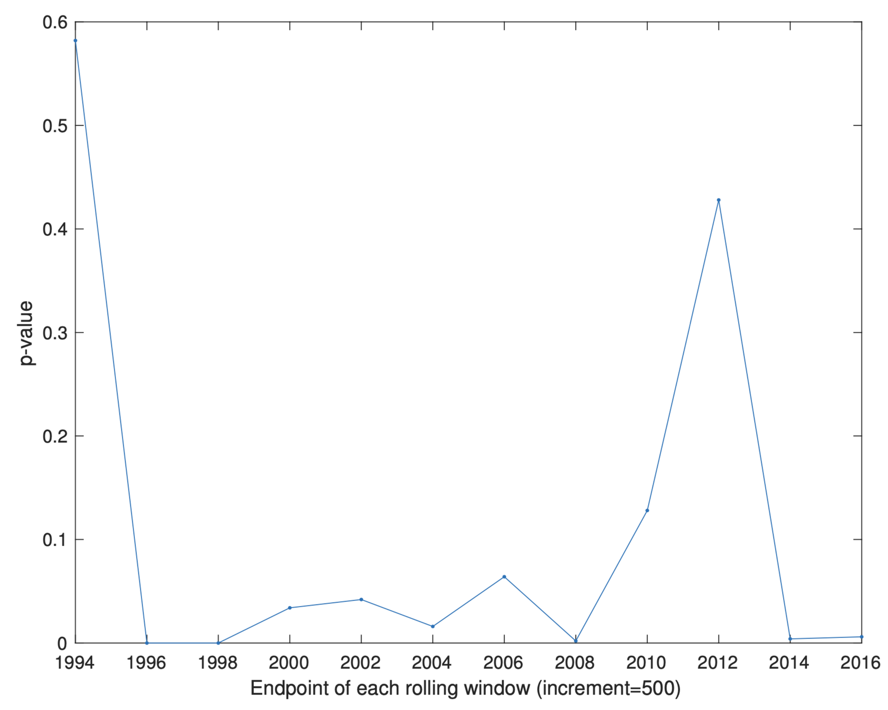
- We apply our method to the comparison of US and Global equity returns (Home Bias Puzzle).
- The home bias puzzle as been investigated by many authors including Chan, Covrig & Ng (2005), French & Poterba (1991), Lewis (1999)). Levy & Levy (2014) argue that *despite a significant reduction in implicit and explicit transaction costs around the world, the US home bias in stock and bond returns has not disappeared.*
- The dataset comes from the Fama-French US and Global risk premium daily series from 7 August 1992 to 30 June 2016 obtained from Kenneth French's Data Library
- We test the dominance of the US series over the global series with 674 conditioning variables detailed below in Table 8.

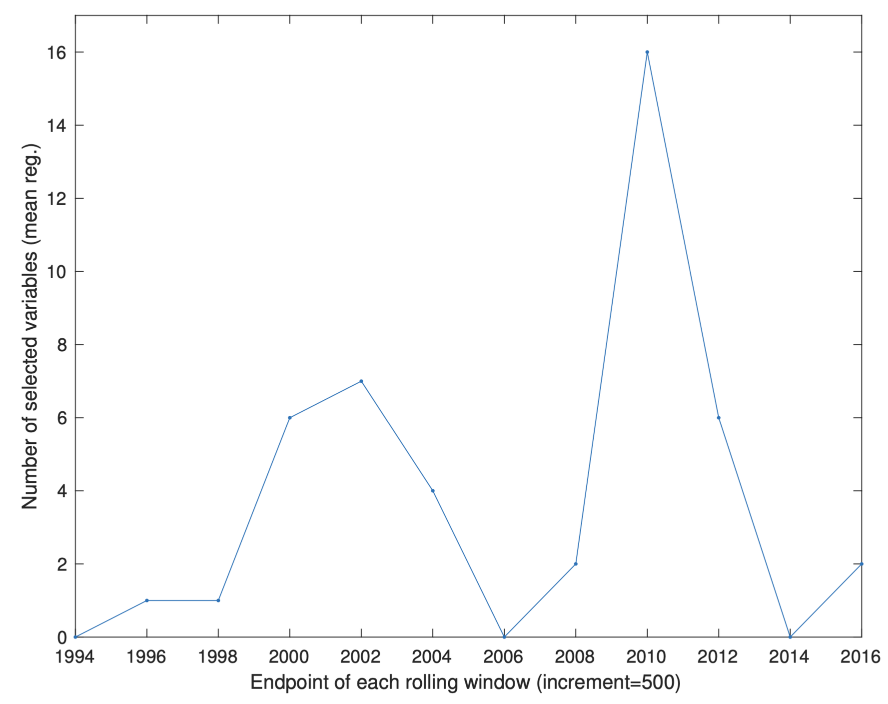
Table 8: *Description of the conditioning variables*

Index	Description
# 1-40	Lagged returns (max. lag = 20)
# 41-200	Powers of lags
# 201-600	Interactions
# 601-638	Momentum measures
# 639-657	Changes in trading volume
# 658-665	Relative strength Indices
# 666-669	Moving average oscillators
# 670-674	Day of the week dummy variables

Results

- We conduct a non-overlapping rolling window analysis of size 500, roughly two years. We plot the series of p-values reported from 12 windows.
- Throughout, λ is set to be $\sqrt{\log(np)/n} = 0.1595$ and the LASSO threshold constant is set to be 2.
- The result reveals that the null hypothesis suggesting the conditional dominance of the US series over the global series is rejected at the 5% level of significance, except for the periods of 1992-1994, 2004-2006, and 2008-2012 where we do not have sufficient statistical evidence to conclude so. It appears that those years have been somewhat different relative to the rest of the sample.





Selection Process

- For the period from 07/08/2000 to 06/08/2002, we calculate the sample correlations between the conditioning variables and the US series, and rank them in descending order of the absolute value of the correlations.
- Table 9 reports the correlations of 7 "selected variables" from the mean regression (cf. Figure 3); the result suggests that the selected variables tends to be those with high correlations in general, with 6 out of 7 variables listed on top 9 out of 674.

Table 9: *Correlations of the selected variables, an example*

Variable Index	Rank	Correlation (abs.)	Sign of the correlation
# 256	1	0.169272187	+
# 234	2	0.151388803	+
# 424	3	0.145195343	–
# 253	4	0.143038571	–
# 351	6	0.137275434	+
# 650	9	0.129063574	+
# 1	291	0.042979460	+