Some semiparametric models for panels of financial time series with an application to fragmentation of trading

Lena Körber, Oliver Linton and Michael Vogt
London School of Economics, University of Cambridge, and University of Konstanz

CEMMAP, November 2014

Our Contribution

- Methodology for semiparametric panel data models for financial data with large n, T
  - Flexible heterogeneous nonparametric covariate effects
  - Fixed effects in two directions
- Application to assessing the impact on market quality of the introduction of competition between equity trading venues in Europe post MiFID I, 2007
  - Competition has marginally improved quality in the presence of very big downer from the financial crisis

Model

We observe a sample $\{(Y_{it}, X_{it}) : i = 1, \ldots, n,\ t = 1, \ldots, T\}$. The *benchmark* model is a heterogenous nonparametric panel model

$$Y_{it} = \mu_0 + m_i(X_{it}) + \alpha_i + \gamma_t + \varepsilon_{it}$$

with $\mathbb{E}[\varepsilon_{it}|X_{it}] = 0$, where $m_1, \ldots, m_n$ are smooth nonparametric functions, μ_0 is the model constant, α_i is an unknown individual specific effect and γ_t denotes an unknown time specific effect.
For identification we shall assume that $\sum_{i=1}^{n} \alpha_i = \sum_{t=1}^{T} \gamma_t = 0$ and that $\mathbb{E} m_i(X_{it}) = 0$.
The main focus of this paper is on the unknown functions $m_i(.)$ the estimation of which is complicated by the presence of the nuisance parameters $\theta = \{\mu_0, \alpha_i, \gamma_t;\ i = 1, \ldots, n;\ t = 1, \ldots, T\}$. Large n, T.

Our T is not so large so we will look at the restricted model where

$$m_i(\cdot) = \sum_{k=1}^{K} \beta_{ik} \mu_k(\cdot)$$

where

$$\mu_1(\cdot), \ldots, \mu_K(\cdot)$$

are unknown functions, and

$$B = (\beta_{ik})$$

are unknown parameters (loadings), and perhaps K is unknown too.

Parameter of interest is $m_i(1) - m_i(0)$ or some weighted average

$$\sum_{i=1}^{n} w_i \left\{ m_i(1) - m_i(0) \right\} = \sum_{i=1}^{n} w_i \sum_{k=1}^{K} \beta_{ik} \left\{ \mu_k(1) - \mu_k(0) \right\}$$

Alternatively, may want to compare the cross-sectional "distributions" of

$$m_i(1), m_i(0)$$

We will allow "correlated random effects"

$$\mathbb{E}[\alpha_i + \gamma_t | X] \neq 0$$

in general, the unobserved (fixed) effects α_i and γ_t introduce a simultaneity between X and Y. This is important in applications where there are reasons to believe there are common unobserved factors (eg HFT intensity) affecting both X and Y.
For example. For some deterministic functions G_i, H_t and additional random errors η_i, δ_t

$$\begin{aligned} \alpha_i &= G_i(X_{i1}, \ldots, X_{iT}, Y_{i1}, \ldots, Y_{iT}; \eta_i) \\ \gamma_t &= H_t(X_{1t}, \ldots, X_{nt}, Y_{1t}, \ldots, Y_{nt}; \delta_t) \end{aligned}$$

Chamberlain (1982). Restricts the channels through which this endogeneity can work.

Two alternative specifications (non-nested) require modification in our procedures.

1. Fixed effect in X. For some $\{(\delta_i, \alpha_i)\}$ and $\{(\gamma_t, c_t)\}$ arbitrary processes
$$X_{it} = \delta_i + c_t + U_{it}$$

2. Interactive fixed effects
$$X_{it} = \Gamma_i^\intercal f_t + v_{it}$$
$$Y_{it} = \mu_0 + m_i(X_{it}) + \alpha_i^\intercal f_t + \varepsilon_{it}$$
where $f_t \in \mathbb{R}^d$

Outline

1. Brief Literature review
2. Elimination of nuisance parameters
3. Dimensionality reduction or common functions in covariate effects
4. Identification
5. Estimation
6. Distribution Theory
7. Application

Pesaran (2006, Econometrica). Observe $Y_{it}, X_{it}, d_t,$ linear model with multiple factors and loadings varying in cross section

$$\begin{aligned} Y_{it} &= \alpha_i^\intercal d_t + \beta_i^\intercal X_{it} + \gamma_i^\intercal f_t + \varepsilon_{it} \\ X_{it} &= A_i^\intercal d_t + \Gamma_i^\intercal f_t + v_{it} \end{aligned}$$

But common unobserved factor $f_t \in \mathbb{R}^K$ enters $Z = (X, Y)$ in a linear way.
Estimation Idea is that

$$\sum_i w_i Z_{it} \simeq A^\intercal(w) d_t + \Gamma^\intercal(w) f_t$$

for diversified weights. Can obtain f_t upto an affine transformation with enough different weights. Include control function

$$Y_{it} \simeq \alpha_i^{*\intercal} d_t + \beta_i^\intercal X_{it} + \gamma_i^{*\intercal} \overline{Z}_t^w + \varepsilon_{it}$$

More general than our treatment with respect to number of factors effects but less general in functional form and about the way fixed effects are related.

Mammen, Støve, and Tjøstheim (2009, ET). Additive nonparametric regression. Mostly deal with the case with $n \to \infty$ and T fixed and no cross sectional dependence

$$Y_{it} = \mu_0 + \gamma_t + \sum_{j=1}^{J} m_j(X_{jit}) + \varepsilon_{it}.$$

Estimation is by backfitting(and joint estimation of fixed effect). Iterative one dimensional smooths of partial residuals

$$m_j(X_{jit}) \longleftarrow E\left[\overbrace{\left\{Y_{it} - \mu_0 - \gamma_t - \sum_{k\neq j}^{J} m_k(X_{kit})\right\}}^{\text{partial residuals}} |X_{jit}\right]$$

Hastie and Tibshirani (1990). Mammen, Linton, and Nielsen (1999). No concurvity(singularity in joint distribution of $(X_1, \ldots, X_J)$.

Connor, Hagmann, and Linton (2012, Econometrica) consider the model (for returns) with large n and T but again no heterogeneity

$$Y_{it} = \gamma_{0t} + \sum_{j=1}^{J} m_j(X_{ji})\gamma_{jt} + \varepsilon_{it},$$

Allow dependence of weak sort in time and cross section but covariates not time varying and no fixed effect in them.

$$m_j(X_{jit}) \longleftarrow E\left[\overbrace{\frac{1}{\gamma_{jt}}\left\{Y_{it} - \gamma_{0t} - \sum_{k\neq j}^{J} m_k(X_{kit})\gamma_{kt}\right\}}^{\text{partial residuals}} | X_{jit}\right]$$

$$\gamma_t \longleftarrow \left(M_t^{\intercal} M_t\right)^{-1} M_t^{\intercal} Y_t$$

Many other papers especially with time varying parameters.

Kneip, Sickles, and Song (2012, ET).

$$Y_{it} = \mu_0(t) + \alpha_i(t) + \sum_{j=1}^{J} \beta_j X_{jit} + \varepsilon_{it}$$

$$\alpha_i(t) = \sum_{k=1}^{K} \theta_{ik} \mu_k(t),$$

They do not allow individual effects to be related to included covariates, i.e., no endogeneity. Estimation method based on splines.

Elimination of Nuisance Parameters by Fixed Effect Transformation

Denote the time, cross sectional, and global averages by

$$\overline{Y}_i = \frac{1}{T}\sum_{t=1}^{T} Y_{it}, \qquad \overline{Y}_t = \frac{1}{n}\sum_{i=1}^{n} Y_{it}, \qquad \overline{\overline{Y}} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} Y_{it}$$

$$Y_{it}^{fe} = Y_{it} - \overline{Y}_i - \overline{Y}_t + \overline{\overline{Y}}$$

Then note that the fixed effect transformation completely removes the fixed effects $(\alpha_i + \gamma_t)$ with some approximation error depending on m_i and on properties of X_{it}

$$\begin{aligned} Y_{it}^{fe} &= m_i(X_{it}) + \varepsilon_{it} - \frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it} - \frac{1}{n}\sum_{i=1}^{n}\varepsilon_{it} + \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\varepsilon_{it} \\ &\quad - \frac{1}{T}\sum_{t=1}^{T} m_i(X_{it}) - \frac{1}{n}\sum_{i=1}^{n} m_i(X_{it}) + \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} m_i(X_{it}) \\ &= m_i(X_{it}) + \varepsilon_{it} + O_p(T^{-1/2}) + O_p(n^{-1/2}). \end{aligned}$$

Elimination of Nuisance Parameters by Differencing

An alternative approach is to use differencing to eliminate the nuisance parameters. Specifically, the DID transformation

$$\begin{aligned} Y_{ijts}^{did} &= Y_{it} - Y_{is} - (Y_{jt} - Y_{js}) \\ &= m_i(X_{it}) - m_i(X_{is}) - m_j(X_{jt}) + m_j(X_{js}) + u_{ijts}, \end{aligned}$$

where $u_{ijts} = \varepsilon_{it} - \varepsilon_{is} - (\varepsilon_{jt} - \varepsilon_{js})$ is a serially dependent error term. The right hand side is a four term "dyadic" (Fafchamps and Gubert (2007)) additive time series regression model with Y_{ijts}^{did} on $X_{it}, X_{is}, X_{jt}, X_{js}$. Backfitting type estimation. For estimation need $E(\varepsilon_{it}|X_{it}, X_{is}, X_{jt}, X_{js}) = 0$ rather than just $E(\varepsilon_{it}|X_{it}) = 0$ needed for the fixed effect method.

Henderson, Carroll, and Li (2008) propose this method (with just time differencing) in the homogeneous one way model, i.e., $Y_{it} = \mu_0 + \alpha_i + m(X_{it}) + \varepsilon_{it}$.

Common Structure

We allow for heterogeneous functions m_i. In many applications, it is very restrictive to assume that $m_i = m$ for all individuals. However, rather than allowing the functions m_i to vary completely freely across individuals, one may expect some common underlying structure.
For example,

$$m_i(x) = \beta_i m(x),$$

for common function m and parameters β_i. We consider the K-factor specification

$$m_i(x) = \sum_{k=1}^{K} \beta_{ik} \mu_k(x),$$

where $\mu_k(\cdot)$, $k = 1, \ldots, K$ are unknown functions and β_{ik} are unknown constants.
Like CHL (2012) except that same covariate in each component, which makes this a "second moment" estimation problem rather than a first moment one. The joint distribution of $(X, X, \ldots, X)$ is singular.

Identification (ForePlay)

Writing $m(x) = (m_1(x), \ldots, m_n(x))^\intercal$ and $\mu(x) = (\mu_1(x), \ldots, \mu_K(x))^\intercal$, we can represent m as

$$m(x) = B\mu(x)$$

where B is a $n \times K$ matrix with the entries β_{ik}. Identification issue. We assume

The matrix B is orthogonal, i.e. $B^\intercal B = I_K$ and $\int \mu(x)\mu(x)^\intercal dw(x)$ is a diagonal matrix with non-zero diagonal entries.

It follows that once B is known

$$B^\intercal m(x) = B^\intercal B\mu(x) = \mu(x)$$

We have

$$\Omega = \int m(x)m(x)^\intercal w(x)dx = B \int \mu(x)\mu(x)^\intercal w(x)dx \; B^\intercal = BDB^\intercal,$$

where $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_K)$ with $\lambda_k = \int \mu_k^2(x)w(x)dx$.
We obtain B from the eigendecomposition of Ω, thence

$$\mu = B^\intercal m.$$

This equation almost surely identifies the functions μ up to sign.
Practical/Theoretical problem is that Ω is $n \times n$.

We transform it into a system of dimension K. Let $W = (\omega_{ki})$ be a $K \times n$ matrix such that $S = WB$ is a full-rank $K \times K$ matrix. Then $g = Wm$ with $g = (g_1, \dots, g_K)^\intercal$ and

$$g_k(x) = \sum_{i=1}^{n} \omega_{ki} m_i(x) = \sum_{i=1}^{n} \omega_{ki} \sum_{j=1}^{K} \beta_{ij} \mu_j(x),$$

whence

$$g(x) = S\mu(x),$$

where S has the entries $S_{kj} = \sum_{i=1}^{n} \omega_{ki} \beta_{ij}$. We now impose that
The matrix S is orthonormal, i.e. $S^\intercal S = I_K$ and $\int \mu(x)\mu(x)^\intercal dw(x)$ is a diagonal matrix (with non-zero diagonal entries).
Once S is known we have

$$S^\intercal g(x) = S^\intercal S\mu(x)g(x) = \mu(x)$$

We have that

$$\Sigma = \int g(x)g(x)^\intercal w(x)dx = S\int \mu(x)\mu(x)^\intercal w(x)dx\; S^\intercal = SDS^\intercal,$$

where $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_K)$ with $\lambda_k = \int \mu_k^2(x)w(x)dx$.
Therefore,

$$\mu(\cdot) = S^\intercal g(\cdot).$$

It follows that μ is identified up to sign.

Nonparametric Estimation of m_i

Let h be a scalar bandwidth and $K(\cdot)$ a kernel satisfying $\int K(u)du = 1$ and $K_h(\cdot) = h^{-1}K(h^{-1}\cdot)$. Then define local linear kernel smoother

$$Q_T(\theta_0, \theta_1; x) = \sum_{t=1}^{T} K_h(X_{it} - x)\{Y_{it}^{fe} - \theta_0 - \theta_1(X_{it} - x)\}^2,$$

Then let $\widehat{\theta}_0, \widehat{\theta}_1$ minimize $Q_T(\theta_0, \theta_1; x)$ with respect to θ_0, θ_1 and let $\widehat{m}_i(x) = \widehat{\theta}_0$,
Consistent under some conditions e.g. iid; but not under fixed effect in X.
Good at boundary points.

Estimation Algorithm

1. First construct estimates of the functions $g_1, \ldots, g_K$ according to

$$\widehat{g}_k(x) = \sum_{i=1}^{n} \omega_{ki} \widehat{m}_i(x).$$

2. Then estimate the matrix Σ by

$$\widehat{\Sigma} = \int \widehat{g}(x)\widehat{g}(x)^\intercal w(x) dx.$$

3. Then estimate the eigenvalues and eigenvectors by

$$\widehat{\Sigma} = \widehat{S}\widehat{D}\widehat{S}^\intercal,$$

   i.e. by performing an eigenvalue decomposition of $\widehat{\Sigma}$.

4. Let

$$\widehat{\mu}(\cdot) = \widehat{S}^\intercal \widehat{g}(\cdot).$$

[5]. We then estimate the loadings by the least squares estimator

$$\widehat{\beta}_i = \left[\frac{1}{T} \sum_{t=1}^{T} w(X_{it}) \widehat{\mu}(X_{it}) \widehat{\mu}(X_{it})^{\intercal} \right]^{-1} \frac{1}{T} \sum_{t=1}^{T} w(X_{it}) \widehat{\mu}(X_{it}) Y_{it}^{fe}.$$

Assume that K is known but in practice choose between them using model selection tools. Theory allows some zero eigenvalues so only need an upper bound on K.

[6] Let

$$\widehat{m}_i^e(x) = \widehat{\beta}_i^{\intercal} \widehat{\mu}(x)$$

[7] Continue, update fixed effects estimators

Method is SIMPLE
Iterated version may converge to minimum of (Gaussian LIKELIHOOD)

$$\sum_{i=1}^{n} \sum_{t=1}^{T} \left[Y_{it} - \mu_0 - m_i(X_{it}) - \alpha_i - \gamma_t \right]^2$$

Distribution Theory

Theory established under some alternative specifications. Need CLT for objects like

$$\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} K_h(X_{it}-x)\varepsilon_{it}$$

1. $\{X_{it}, \varepsilon_{it}\}$ are i.i.d across i and t;
2. $\{X_{it}, \varepsilon_{it}\}$ are stationary and weakly dependent across t and iid across i
3. $\{X_{it}, \varepsilon_{it}\}$ are stationary and weakly dependent across t and i (this is based on an ordering of i that we do not know, Connor and Koraczyck, 1993). eg $\mathrm{cov}(\varepsilon_{it}, \varepsilon_{js}) = \gamma(||(i-j, t-s)||)$, where $\gamma(u) \to 0$ as $u \to \infty$
4. $\{X_{it}, \varepsilon_{it}\}$ are locally stationary (Dahlhaus, 1997) and weakly dependent across t and i
5. Fixed effect in X? See below

This allows general class of volatility processes

Locally stationary nonparametric regression (Vogt, 2011 Mannheim PhD, Forthcoming in Annals of Statistics)

X_t as in Dahlhaus definition has (local) stationary density $f_u(x)$ for $u \sim t/T$. Excludes $X_t = t/T$.

Results are similar to the stationary case in every regard except limiting constants are averages of local densities and variances etc.

Can generalize this in the cross-section dimension for some ordering of i that we do not observe.

Our theory involves:

1. Obtain expansion for $\widehat{m}_i(x)$ (Large T) and hence $\widehat{g}_k(x)$ (Large n)
2. Obtain asymptotics for $\widehat{\Sigma}$
3. Derive rate of convergence of $\widehat{S}$ using theory of estimation of eigenvalues and eigenvectors of sample matrices
4. Obtain asymptotics for $\widehat{\mu}(x)$ and $\widehat{\beta}_i$
5. Obtain asymptotics for $\widehat{m}_i(x) = \widehat{\beta}_i^{\intercal}\widehat{\mu}(x)$

THEOREM. Under some regularity conditions

$$\sup_{x\in[0,1]} \|\widehat{\mu}(x) - \mu(x)\| = O_p\left(\sqrt{\frac{\log nT}{nTh}}\right)$$

Moreover, for any fixed point $x \in (0,1)$,

$$\sqrt{Th}\left(\widehat{m}_i(x) - m_i(x)\right) \xrightarrow{d} N\left(0, ||K||_2^2 \frac{\sigma_i^2(x)}{f_i(x)}\right)$$

$$\sqrt{nTh}(\widehat{\mu}(x) - \mu(x)) \xrightarrow{d} N\left(0, ||K||_2^2 S^\intercal VS\right).$$

Here, $V = (V_{k,k'})_{k,k'=1,\ldots,K}$ with

$$V_{k,k'} = \lim_{n\to\infty} n \sum_{i=1}^n \omega_{ki}\omega_{k'i} \frac{\sigma_i^2(x)}{f_i(x)} \quad ; \quad \sigma_i^2(x) = \mathbb{E}[\varepsilon_{it}^2 | X_{it} = x].$$

THEOREM. Under some regularity conditions for any fixed i,

$$\sqrt{T}(\widehat{\beta}_i - \beta_i) \xrightarrow{d} N(0, \Gamma_i^{-1}\Psi_i(\Gamma_i^{-1})^\intercal),$$

where:

$$\Gamma_i = \mathbb{E}[w(X_{i0})\mu(X_{i0})\mu(X_{i0})^\intercal]$$

$$\Psi_i = \sum_{l=-\infty}^{\infty} \mathrm{cov}(\chi_{i0}, \chi_{il})$$

$$\chi_{it} = \{w(X_{it})\mu(X_{it}) - \mathbb{E}[w(X_{it})\mu(X_{it})]\}\varepsilon_{it} - \mathbb{E}[w(X_{it})\mu(X_{it})]m_i(X_{it})$$

Pooled Estimator of m_i

Let $\Delta_{n,T} = \max\{T^{-1/2}, (nT)^{-2/5}\}$. We have

$$\widehat{m}_i^e(x) - m_i(x) = (\widehat{\beta}_i - \beta_i)^\intercal \mu(x) + \beta_i^\intercal(\widehat{\mu}(x) - \mu(x)) + o_p(\Delta_{n,T}).$$

The first term on the right hand side is of order $T^{-1/2}$, while the second term is of order $(nT)^{-2/5}$ under our conditions. The leading term is determined by the ratio n^4/T : if $n^4/T \to 0$, then the larger term is the second one of order $(nT)^{-2/5}$, while if $n^4/T \to \infty$, the larger term is the first one of order $T^{-1/2}$. The knife edge case where n^4/T stays bounded away from zero and infinity allows both terms to contribute to the limiting behaviour.

It follows that $\widehat{m}_i^e(x)$ is asymptotically normal, and at a faster rate than $\widehat{m}_i(x)$, which converges at rate $T^{2/5}$.

In our application below, we are interested in the parameter

$$c_i = m_i(1) - m_i(0),$$

which measures the difference between monopoly and competition for stock i. Letting $\widehat{c}_i = \widehat{m}_i^e(1) - \widehat{m}_i^e(0)$ we have

$$\begin{aligned} \widehat{c}_i - c_i &= (\widehat{\beta}_i - \beta_i)^\intercal(\mu(1) - \mu(0)) + \beta_i^\intercal(\widehat{\mu}(1) - \mu(1)) \\ &\quad -\beta_i^\intercal(\widehat{\mu}(0) - \mu(0)) + o_p(\Delta_{n,T}), \end{aligned}$$

Under the null hypothesis that $c_i = 0$, we should observe that

$$\sqrt{nTh}\widehat{c}_i \xrightarrow{d} N\left(0, \tau_i\right), \quad \tau_i = \beta_i^\intercal(S^*)^\intercal[V(1) + V(0)]S^*\beta_i,$$

which could form the basis of a test. Specifically, we can estimate the asymptotic variance τ_i consistently by $\widehat{\tau}_i$, then

$$t_i = \frac{\widehat{c}_i}{\sqrt{\widehat{\tau}_i / nTh}} \Longrightarrow N(0,1).$$

Alternatively, compare the marginal cross-sectional distribution of $m_i(1)$ with that of $m_i(0)$. First order, second order, et al dominance

- Let X_1 and X_2 be two random variables (incomes, returns/prospects) at either two different points in time, or for different regions or countries, or with or without a program (treatment).
- X_1 First Order Stochastic Dominates X_2,

$$X_1 \succeq_{FSD} X_2 \text{ if } F_1(x) \leq F_2(x), \ \forall x$$

- X_1 Second Order Stochastic Dominates X_2

$$X_1 \succeq_{SSD} X_2 \text{ if } \int_{-\infty}^{x} F_1(t)dt \leq \int_{-\infty}^{x} F_2(t)dt, \ \forall x.$$

Robustness of the estimation method

So far, we have worked under the simplifying assumption that the number K of common component functions $\mu_1, \ldots, \mu_K$ is known. We now drop this assumption and take into account that K is usually not observed in applications. We only suppose that there is some known upper bound $\overline{K}$ of the number of component functions and we do our procedure using the upper bound.

THEOREM. For all $k = 1, \ldots, K$

$$\sup_{x \in I_h} \left| \widetilde{\mu}_k(x) - \mu_k(x) \right| = O_p\Big(\sqrt{\frac{\log nT}{nTh}} \Big)$$

For $k = K+1, \ldots, \overline{K}$,

$$\int \widetilde{\mu}_k^2(x) w(x) dx = o_p\Big(\frac{1}{\sqrt{nTh}} \Big)$$

Taken together, these results show that our procedure is robust to overestimating the number of component functions K.

Our estimator of K is defined as

$$\widehat{K} = \min\left\{ k \in \{1, \dots, \overline{K}\} \mid \frac{\widetilde{\lambda}_1 + \dots + \widetilde{\lambda}_k}{\widetilde{\lambda}_1 + \dots + \widetilde{\lambda}_{\overline{K}}} \geq 1 - \delta_{n,T} \right\}.$$

The intuition behind this estimator is simple: Under our assumptions, the matrix $\overline{\Sigma}$ has K non-zero eigenvalues, i.e. the first K entries of $\overline{\lambda}$ are non-zero. The first K entries of the estimator $\widetilde{\lambda}$ thus converge to some positive values, whereas the other ones approach zero as the sample size increases. Hence, the ratio

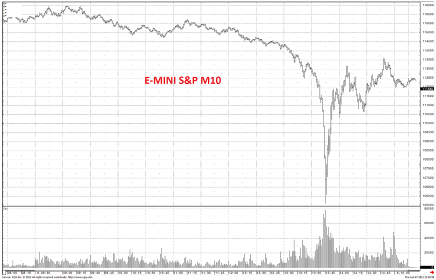
$$\frac{\widetilde{\lambda}_1 + \dots + \widetilde{\lambda}_k}{\widetilde{\lambda}_1 + \dots + \widetilde{\lambda}_{\overline{K}}}$$

should converge to a number strictly smaller than 1 for $k < K$ and to 1 for $k \geq K$. This suggests that $\widehat{K}$ consistently estimates the true number of components K.

Application

Application follows from my part in UK Government Office for Science Foresight project on *Future of Computer Based Trading in Financial Markets*.
http://www.bis.gov.uk/foresight/our-work/projects/current-projects/computer-trading.
International project with hundreds of academics and practitioners involved: O'Hara, Leland, Hendershott, Foucault, Menkveld, Farmer etc.

Flash Crash May 6th, 2010



SEC/CFT report 2010: Automated sell order, High frequency trading, fragmentation of trading, market maker obligations, stub quotes, NMS trading rules etc.
Can we predict flash crash now?
Easley, D., Lopez de Prado, M., O'Hara, M., 2011a. The microstructure of the "flash crash": flow toxicity, liquidity crashes, and the probability of informed trading. Journal of Portfolio Management 37 (2), 118–128.
Anderson and Bondarenko VPIN and the Flash Crash (2011→2013) SSRN-id2292605.

Paul Krugman represents the contrarian view of high frequency trading:

> *"It's hard to imagine a better illustration [of social uselessness] than high frequency trading. The stock market is supposed to allocate capital to its most productive uses, for example by helping companies with good ideas raise money. But it's hard to see how traders who place their orders one-thirtieth of a second faster than anyone else do anything to improve that social function ... we've become a society in which the big bucks go to bad actors, a society that lavishly rewards those that make us poorer" .*

System latency following Moores law or even more

System	Implementation Date	Latency (Microseconds)
SETS	<2000	600000
SETS1	Nov 2001	250000
SETS2	Jan 2003	100000
SETS3	Oct 2005	55000
TradElect	June 18, 2007	15000
TradElect 2	October 31, 2007	11000
TradElect 3	September 1, 2008	6000
TradElect 4	May 2, 2009	5000
TradElect 4.1	July 20, 2009	3700
TradElect 5	March 20, 2010	3000
Millenium	February 14, 2011	113

Transactions and quote updating related to this and speeded up likewise. Capacity also increased.

Costs of speed$\downarrow$ Benefits? Limit orders give options to trade to other traders. Black and Scholes (1973, JPE) call option price

$$f(S, X, \tau, r_f, \sigma) = S \cdot \Phi(d_+) - X \cdot e^{-r_f \cdot \tau} \cdot \Phi(d_-)$$

where

$$d_\pm = \frac{\log \frac{S}{X} + \left(r_f \pm \frac{\sigma^2}{2}\right) \cdot \tau}{\sigma \cdot \sqrt{\tau}}$$

and Φ is the standard normal cdf. At the money, $S = X$, as $\tau \to 0$

$$f(S, X, \tau, r_f, \sigma) = \frac{1}{\sqrt{2\pi}} S \cdot \sigma \sqrt{\tau} + O(\tau).$$

There is a positive albeit small value in an order that only sits for small time.

Market Fragmentation

We investigate the effects of equity trading market fragmentation (aka competition) on market quality in the UK.
In 2007, the implementation of the “Markets in Financial Instruments Directive (MiFID)" ended the monopoly of primary exchanges across Europe. Market participants can now execute their trades on traditional primary exchanges or on new exchanges known as Multilateral Trading Facilities (MTF) or Systematic Internalizers (SI).
Since the implementation of MiFID, fragmentation of trading flows increased significantly. In February 2012, the volume of FTSE100 stocks traded via the London Stock Exchange had declined to 50%, and the volume of DAX stocks traded via the German Stock Exchange had decreased to 63%.
Computer-based High Frequency Trading enabled by and almost required in Fragmented environment.

- Competition between venues for market share is different from price competition between traders. However, higher competition between trading venues can improve market quality: technological innovation, improves efficiency and reduces the fees that have to be paid by investors. Foucault and Menkveld (2008).
- On the other hand, might think that security exchanges are natural monopolies. Consolidated exchanges enjoy economies of scale because establishing a new exchange requires the payment of a high fixed cost. Every additional trade lowers the average cost of the exchange. In addition, a single, consolidated exchange market creates network externalities. The larger the market, the more trading opportunities exist that attract even more traders.
- On the other other hand, smart order routers may create a virtual consolidated market place. [although in Europe no reg NMS so trade throughs and crossed markets allowed and observed].

It is an empirical question. Some evidence

- O'Hara and Ye (2009) US markets. Matching methodology. Negative effects of fragmentation on illiquidity and transaction costs and positive effects on efficiency.
- Gresse (2011) uses data on stocks listed on the LSE and Euronext exchanges in Amsterdam, Paris and Brussels to examine the effect of market fragmentation on liquidity. In her study, fragmentation is measured by the reciprocal of the Herfindahl index. Gresse measures liquidity by the sum of quantities market participants are willing to trade at the best bid and ask prices. Using a linear panel model, she finds that market fragmentation improves liquidity both locally and globally.
- Local liquidity is available to investors that trade only on primary exchanges, while global liquidity can be accessed by investors that are connected to different exchanges via, for example, Smart Order Routing Technologies.

Degryse et al. (2011) investigate the same question using a data set of Dutch stocks and their analysis is methodologically similar to Gresse (2011). In contrast to Gresse (2011), however, the authors distinguish between fragmentation in visibleand darktrading platforms. This distinction is based on the pre-trade transparency requirements. They find that

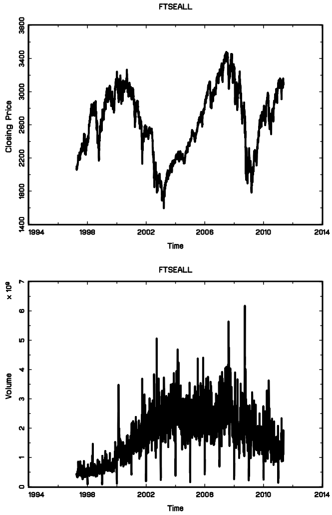
- visible fragmentation improves global liquidity, but has a negative effect on local liquidity.
- Dark fragmentation has a negative effect on both local and global liquidity.

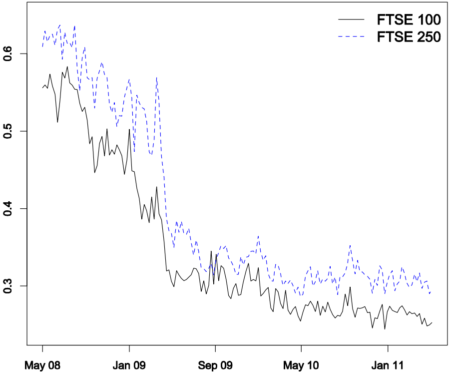
All use linear (in parameter) panel specifications and DID method

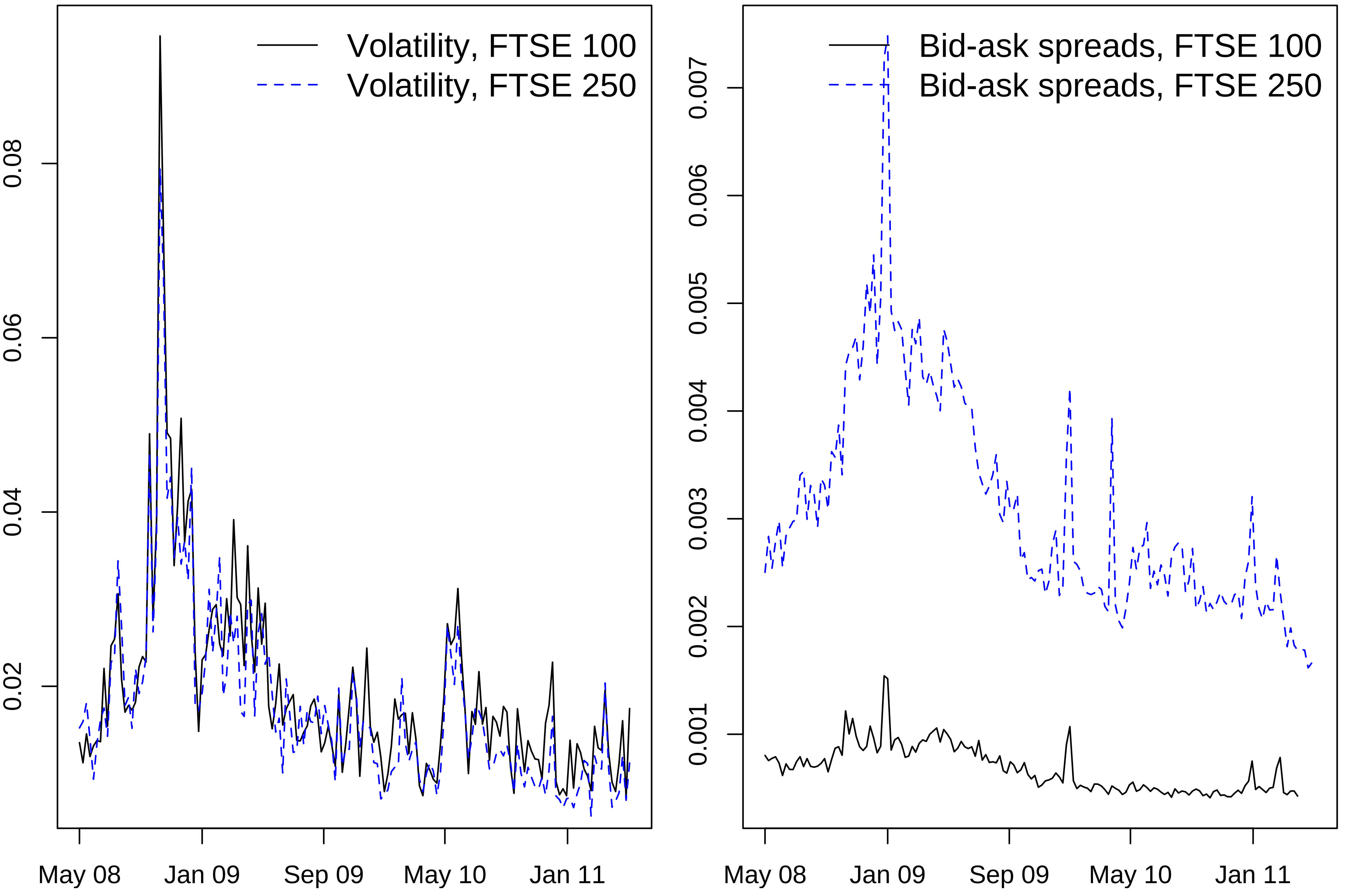
Data

Weekly UK data (Fidessa): volume traded for the all the stocks in the FTSE-100 index and the FTSE-250 index, as well as where that volume was traded over the period 2008-2011. The data distinguish between lit (public exchanges with visible order book), dark pools (invisible order book), otc (over the counter), and si (systematic internalizer) venues. Bloomberg daily data for prices and volumes.

- Results presented for individual stocks of FTSE350 May 2008- June 2011 ($n \sim 350$ and $T \sim 150$)
- Measure fragmentation X_{it} using Fidessa weekly data by
  - Percentage volume traded off the LSE lit venue
  - Herfindahl index of the trading volumes $\sum w_i^2$; equals to one if all concentrated in one venue (measures competition between venues).
  - Same for the lit versus dark
- Market Quality Outcome variables Y_{it} using Bloomberg daily data:
  - log of volatility (high-low and realized vol over week)
  - log of liquidity (1/Amihud return per unit vol averaged over week and log of bid-ask spreads)
  - Market efficiency ACF(1)
- Time period coincides with financial crisis so many reasons why market quality should drop absolutely. Our panel model tries to control for this using cross section and time series variation in the fragmentation.


FTSEALL
Closing Price
Time
FTSEALL
Intraday Volatility
Time
FTSEALL
Volume
Time
FTSEALL
Standardized Daily Return
Time


FTSE 100
FTSE 250
0.6
0.5
0.4
0.3
May 08
Jan 09
Sep 09
May 10
Jan 11

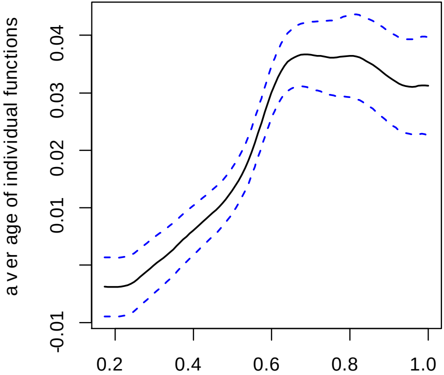

Volatility, FTSE 100
Volatility, FTSE 250
0.02
0.04
0.06
0.08
May 08
Jan 09
Sep 09
May 10
Jan 11
Bid-ask spreads, FTSE 100
Bid-ask spreads, FTSE 250
0.001
0.002
0.003
0.004
0.005
0.006
0.007
May 08
Jan 09
Sep 09
May 10
Jan 11

Pesaran Model

$$Y_{it} = \mu_0 + \beta_i X_{it} + \delta_i X_{it}^2 + \rho_i \overline{X}_t + \vartheta_i VIX_t + \varepsilon_{it}$$

	Volatility	Liquidity	Bid-ask spread
Intercept	-4.80***	-10.72***	-7.87***
	(0.07)	(0.17)	(0.12)
Herf.	0.88***	-2.06***	1.85***
	(0.21)	(0.34)	(0.31)
Herf. squared	-0.85***	1.14***	-2.20***
	(0.22)	(0.42)	(0.35)
VIX	0.02***	0.01***	0.02***
	(0.00)	(0.00)	(0.00)
Cross-sect. mean of Herf.	1.05***	1.23***	1.75***
	(0.11)	(0.19)	(0.18)
Num. obs.	34220	34253	34363

Nonparametric pooled Volatility against Herf (Average of m_i)



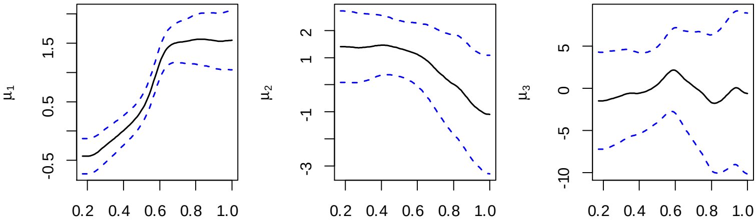
This seems to be saying that as fragmentation increases, first volatility rises and then it falls. It may reach a lower bound where further improvement is not evident.

However, note that there is in our sample relatively few observations in the region where quality falls, since we are nearly a year after Chi-X launch.

Perfect competition case better than monopoly case, but the road there is not easy.

There are a lot of different shapes in individual curves; average effect disguises that.

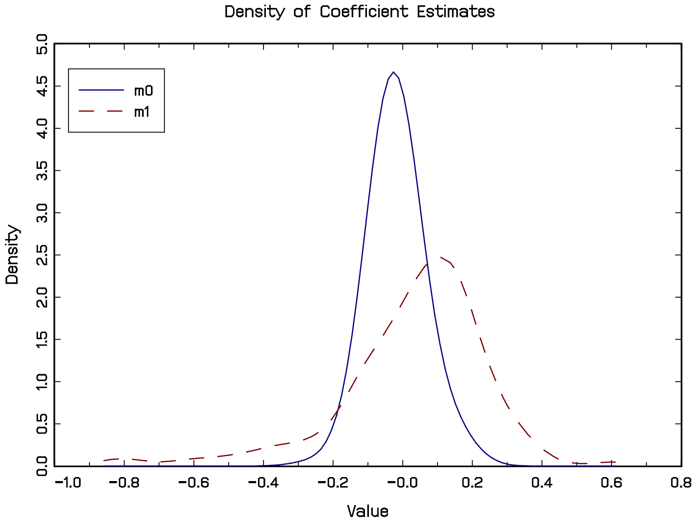
Our model. Volatility against Herf common functions $K = 3$

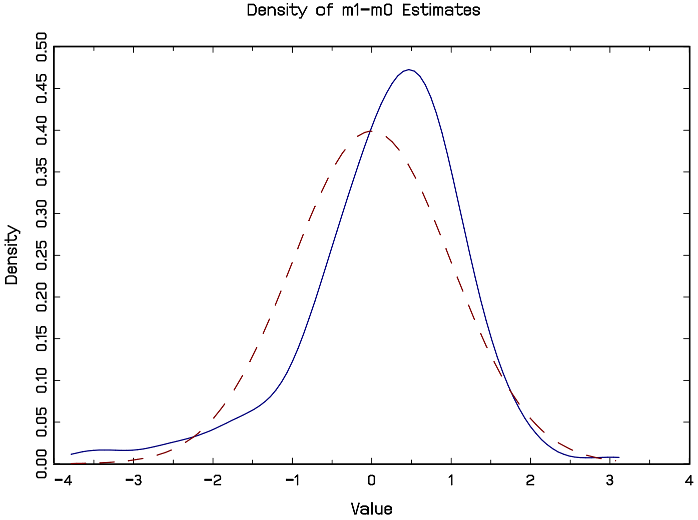


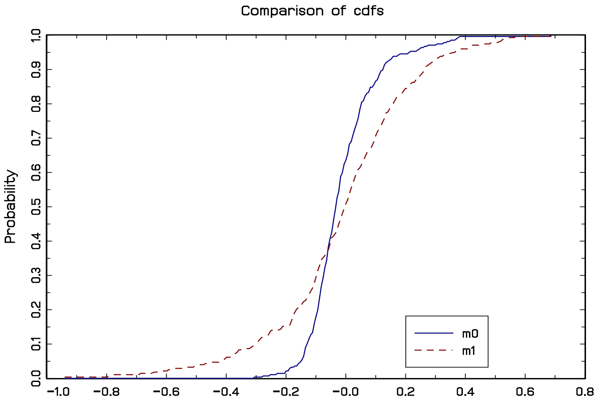
Also tried $K = 5$ and $K = 10$ but curves not significant and doesn't change the main curve

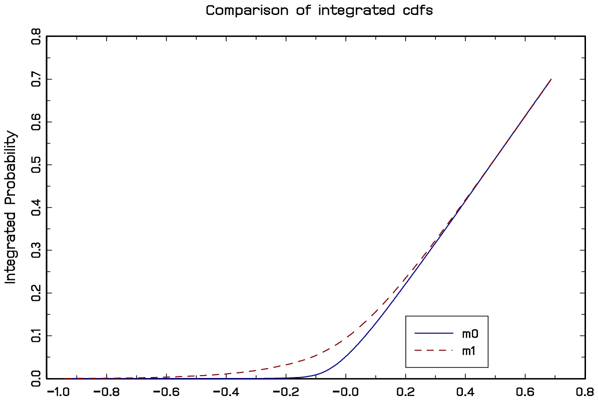
Main curve similar to average effect; second curve shows downward slope but less significance; third curve pretty flat relatively

Individual stocks have some variation in response due to different loadings on μ_1, μ_2, and μ_3.

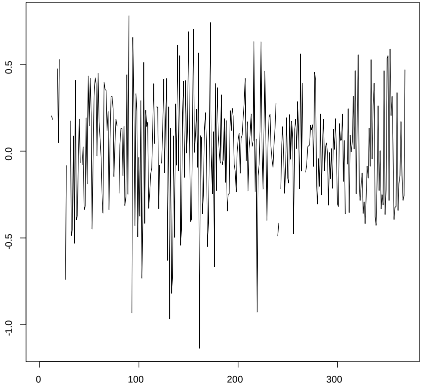

Density of Coefficient Estimates
m0
m1
Density
Value
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
−1.0
−0.8
−0.6
−0.4
−0.2
−0.0
0.2
0.4
0.6
0.8


Density of m1-m0 Estimates
Density
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0.45
0.50
-4
-3
-2
-1
0
1
2
3
4
Value

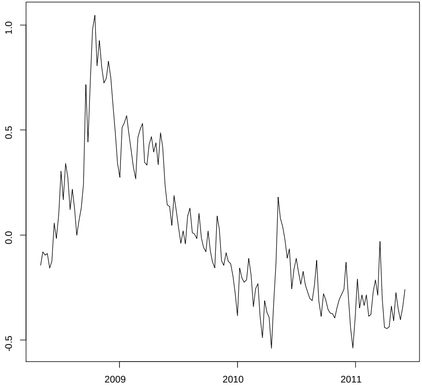

Comparison of cdfs
Probability
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
−1.0
−0.8
−0.6
−0.4
−0.2
−0.0
0.2
0.4
0.6
0.8
m0
m1


Comparison of integrated cdfs
Integrated Probability
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
−1.0
−0.8
−0.6
−0.4
−0.2
−0.0
0.2
0.4
0.6
0.8
m0
m1

Individual effects α_i against market capitalization rank. There is not much relationship (at the LSE, tick size is related to market cap, price level, and segment)



Common trend γ_t against time



Similar to market wide volatility

Conclusions and extensions

- Empirical results suggest that overall competition between trading venues has improved the performance of equity markets according to volatility and liquidity but not in a simple way.
- Both heterogeneity and nonlinearity important here
- Körber, Linton, and Vogt (2013). Extensive parametric results based on quadratic Pesaran model with multiple common factors. Quantile estimation. Variance regression estimation (variability of market quality). Control for market cap. Split into permanent and temporary volatility. Split into overnight and within day volatility. Split into common and idiosyncratic volatility and jump versus continuous component. Split into visible frag and dark. Instrumental variables. Differences in differences. Fixed effect transformation. Similar results, mostly.

Fixed effect in covariate

Suppose that the regressors have the structure

$$X_{it} = \delta_i + c_t + U_{it},$$

where δ_i and c_t are independent of the process $\{(\varepsilon_{it}, U_{it}),\ i = 1, \ldots, n,\ t = 1, \ldots, T\}$ which obeys either of 1-4 above. This causes inconsistency (but asymptotic unbiasedness) of the estimator $\widehat{m}_i(x)$. This is because

$$\begin{aligned} Y_{it}^{fe} &= m_i(X_{it}) + \varepsilon_{it} - E\left[m_i(X_{it})\,|\delta_i\right] - E\left[m_i(X_{it})\,|c_t\right] \\ &\quad + O_p(n^{-1/2}) + O_p(T^{-1/2}) \\ &\simeq m_i(X_{it}) - \overline{\overline{m_i}}(\delta_i) - \underline{\underline{m}}(c_t) + \varepsilon_{it} \end{aligned}$$

However, note that

$$\overline{X}_i - \overline{\overline{X}} = \delta_i + O_p(T^{-1/2})$$
$$\overline{X}_t - \overline{\overline{X}} = c_t + O_p(n^{-1/2})$$

so that these quantities are approximately observable. Therefore, we have essentially

$$Y_{it}^{fe} = m_i(X_{it}) - \overline{\overline{m_i}}(\overline{X}_i) - \underline{\underline{m}}(\overline{X}_t) + \varepsilon_{it} + Rem_{it}^*,$$

where Rem_{it}^* is a small error term. This is an additive panel regression but with covariates $X_{it}, \overline{X}_i, \overline{X}_t$.

Körber, Linton, and Vogt (2013, in progress) show that backfitting (pooled, then cross-section, then time series) works in the homogeneous case ($m_i = m$)

Interactive Fixed effects

Suppose model is

$$\begin{aligned} Y_{it} &= Q_i(f_t) + m_i(X_{it}) + \varepsilon_{it} \\ X_{it} &= R_i(f_t) + v_{it} \end{aligned}$$

Then

$$\frac{1}{n}\sum_{i=1}^{n} X_{it} \simeq \overline{R}(f_t)$$

for some $\overline{R}$. If $\overline{R}$ is monotonic, then can fit the additive model

$$Y_{it} = \overline{Q}_i(\overline{X}_t) + m_i(X_{it}) + \varepsilon_{it}$$

Restrictions on the loadings

There are $K \times n$ loading parameters to estimate and describe. In practice, too many.
Random coefficient assumptions

$$\beta_i \sim f_\theta$$

Estimate the parameters θ by sp qmle
Shrinkage (assumes that many β_{ij} are small or zero)

$$LS(\beta) + \lambda \sum_{i,j} \left| \beta_{ij} \right|$$