Detection of Units with Pervasive Effects in Large Panel Data Models

G. Kapetanios (a) M. Hashem Pesaran (b) S. Reese (c)

(a) King's College London, University of London

(b) Department of Economics, University of Southern California, USA, and Trinity College, Cambridge, UK

(c) Lund University, Sweden

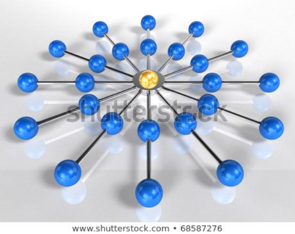
Econometric Society Meeting, San Diego Jan 3-5, 2020

- Detecting economic units whose behavior influences a large number of other units, has become an important policy issue.
- Banks/companies that are deemed to be 'too big to fail' are debated in the press and in public policy forums, although empirical evidence on their existence is often lacking.
- When interconnections are observed (denoted by the adjacency matrix $\mathbf{W}$), such as input-output data in production networks a la Acemoglu et al. 2012, outdegrees of the network (defined by $\mathbf{d}' = (d_1, d_2, ..., d_N) = \mathbf{W}'\mathbf{1}_N$) can be used to detect pervasive/influential units. See Pesaran and Yang (2020, JoE forthcoming).

- The present paper considers the problem of detecting pervasive units when $\mathbf{W}$ is not known, but instead there exists a sufficient number of time series observation (T) on unit-specific characteristics (such as production or prices in multi-sectoral models, or output growths and equity returns in a multi-country global models).
- A new **thresholding multiple testing** method is proposed to detect pervasive units (**if any**) in large panel datasets.
- The detection method is theoretically justified using results from large factor models as well as recent developments on multiple testing.
- The proposed method **(a)** allows for the presence of common (external) factors, **(b)** is capable of identifying networks without a pervasive (or influential unit), and **(c)** is valid for panels with different combinations of N and T (including both cases where $N > T$ and $T > N$).

Pervasive units and their detection

- We characterize a unit in a network as pervasive if it influences almost all other units, analogous to the unit at the hub of a star network.



- In the absence of external factors, the pervasive unit(s) are the common factors, and thus are **perfectly explained by** the **Principle Components** (PCs) used as estimates of the **factors**. This property extends to the networks with external common factors.

- In the general case where the network is also subject to external factors, the number of PCs used must be sufficiently large to allow for the external factors as well.
- Once a sufficient number of PCs are used, we should obtain perfect fit for the regression of pervasive units on the PCs for N and T sufficiently large (but **not** for the non-pervasive units). An extension to dynamic factor models can also be considered but will not be pursued in this paper.
- This motivates using a **thresholding procedure** for estimated **error variances** of the regressions of individual units on the PCs.
- Once a unit passes the threshold test, we then check to see if the selected unit is in fact pervasive using results from the multiple testing literature. See, for example, Bailey, Pesaran and Smith (2019, JoE) on estimation of large covariance matrices.

Related literature

- In the context of linear asset pricing models, Bai and Ng (2006) determine whether a (small) set of observed series coincides with estimated common factors. The observed series could then be regarded as pervasive, although Bai and Ng do not consider such a possibility.

- Parker and Sul (2016) develop an approach in the same factor model context as Bai and Ng and consider identification of pervasive units in a large dataset as a special case.

- Brownlees and Mesters (2019, BM) provide a more general solution - they identify pervasive units from the column sums of the sample concentration matrix (inverse of the sample covariance matrix) of the observations. BM procedure is subject to two limitations: **(a)** It assumes the presence of at least one pervasive unit. **(b)** it requires $T >> N$.

Panel data models with pervasive units and external factors

- Suppose $\mathbf{x}_t = (x_{1t}, x_{2t}, \ldots, x_{Nt})'$ is observed over $t = 1, 2, ..., T$, where x_{it} is the variable of interest (returns, growth rates, rate of inflation) on unit i observed at time t.

- Without loss of generality consider the partition of $\mathbf{x}_t = (\mathbf{x}'_{at}, \mathbf{x}'_{bt})'$, where $\mathbf{x}_{at} = (x_{a,1t}, \ldots, x_{a,mt})'$ is the $m \times 1$ vector of pervasive units and $\mathbf{x}_{bt} = (x_{b,m+1,t}, \ldots, x_{b,Nt})'$ is the $n \times 1$ vector of non-pervasive units with $n = N - m$. A unit is **strongly pervasive** if it affects all other units (there are **degrees** of dominance to be formalized below).

- In addition all units can also be affected by k unobserved **external factors,** $\mathbf{g}_t$.

- The number (m) and the identities of the pervasive units are unknown. In total we could have $p = m + k$ common factors.

- More formally, we consider the DGP (similar to that used by BM):

$$x_{it} = \boldsymbol{\lambda}_i' \mathbf{g}_t + u_{it}, \qquad i = 1, 2, ..., m, \tag{1}$$

$$x_{it} = \boldsymbol{\lambda}_i' \mathbf{g}_t + \sum_{j=1}^{m} b_{ij} x_{jt} + u_{it}, \;\; i = m+1, m+2, ..., N. \tag{2}$$

- The non-pervasive units are affected by the external factors, $\mathbf{g}_t$, and the innovations, $\mathbf{u}_{at} = (u_{1t}, u_{2t}, ..., u_{mt})'$ of the pervasive units that act as **internal factors**:

$$x_{it} = \mathbf{d}_i' \mathbf{g}_t + \sum_{j=1}^{m} b_{ij} u_{jt} + u_{it}, \;\; i = m+1, m+2, ..., N,$$

where $\mathbf{d}_i = \boldsymbol{\lambda}_i + \sum_{j=1}^{m} b_{ij} \boldsymbol{\lambda}_j$ for $i = m+1, m+2, ..., N$.

- The impact of pervasive units, $\mathbf{x}_{at}$, on non-pervasive units, $\mathbf{x}_{bt}$, is governed by the $n \times m$ **loading matrix** $\mathbf{B} = (b_{ij})$.
- For x_{jt}, $j = 1, \ldots, m$ to be **strongly pervasive**, we must have

$$|b_{ij}| > c > 0, \text{ for } i = m+1, m+2, ..., n^{\alpha_j},$$

and

$$b_{ij} = 0, \qquad \text{for } i = n^{\alpha_j} + 1, ..., n,$$

or equivalently

$$\sum_{i=m+1}^{N} |b_{ij}| = \ominus\left(n^{\alpha_j}\right), \text{ for } j = 1, 2, ..., m, \qquad (3)$$

with $\alpha_j = 1$, and $n = N - m$.
- Following Chudik et al. (2011) we could extend our analysis to $\alpha_j \in [1/2, 1)$ for units that are weakly pervasive.

- The model in matrix notation can be written as

$$\mathbf{x}_{at} = \boldsymbol{\Lambda}_a \mathbf{g}_t + \mathbf{u}_{at}, \quad (4)$$

$$\mathbf{x}_{bt} = (\boldsymbol{\Lambda}_b + \mathbf{B}\boldsymbol{\Lambda}_a)\,\mathbf{g}_t + \mathbf{B}\mathbf{u}_{at} + \mathbf{u}_{bt}, \quad (5)$$

or more compactly as

$$\begin{pmatrix} \mathbf{x}_{at} \\ \mathbf{x}_{bt} \end{pmatrix} = \begin{pmatrix} \mathbf{A}_a \\ \mathbf{A}_b \end{pmatrix} \mathbf{f}_t + \begin{pmatrix} \mathbf{0} \\ \mathbf{u}_{bt} \end{pmatrix}, \\ = \mathbf{A}\mathbf{f}_t + \mathbf{v}_t, \quad (6)$$

where $\mathbf{f}_t$ is $p \times 1$ $(p = m + k)$, defined by $\mathbf{f}_t = (\mathbf{g}_t', \mathbf{u}_{at}',)'$, $\mathbf{A}_a = (\boldsymbol{\Lambda}_a, \mathbf{I}_m)$ and $\mathbf{A}_b = (\boldsymbol{\Lambda}_b + \mathbf{B}\boldsymbol{\Lambda}_a, \mathbf{B})$.

Summary of model assumptions

- Factor loadings $\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_N)'$ are treated as fixed parameters. A rank condition on $\mathbf{A}_a$ ensures that m is identified.
- Factors $\mathbf{f}_t$ and errors v_{is} are assumed to be mutually independent and covariance stationary. However, conditional heteroskedasticity is allowed.
- To make use of results from the multiple testing literature we also assume that the distributions of $\mathbf{f}_t$ and v_{is} have exponentially decaying tails. This is standard in high-dimensional statistics.
- Independence of v_{is} across t is imposed for simplicity. However, weak cross-section correlation is allowed for.

Jump to assumption details

Identification of pervasive units via thresholding

- The restricted factor model

$$\begin{pmatrix} \mathbf{x}_{at} \\ \mathbf{x}_{bt} \end{pmatrix} = \begin{pmatrix} \mathbf{A}_a \\ \mathbf{A}_b \end{pmatrix} \mathbf{f}_t + \begin{pmatrix} \mathbf{0} \\ \mathbf{u}_{bt} \end{pmatrix}$$

  suggests a simple **detection** procedure **based on the fit of** individual **cross-sections**, x_{it}, in terms of the factors $\mathbf{f}_t$.
- The explanatory power of $\mathbf{f}_t$ should be perfect for pervasive units, but not for non-pervasive units.
- Since $\mathbf{f}_t$ is unobserved, we use a principal components-based estimator $\hat{\mathbf{F}} = \left(\hat{\mathbf{f}}_1, \hat{\mathbf{f}}_2, \ldots, \hat{\mathbf{f}}_T\right)'$ of dimension $T \times p$ (see e.g. Bai and Ng, 2002; Bai, 2003). Recall that $p = m + k$.

- Given $\hat{\mathbf{F}}$, one can compute

$$\hat{\sigma}_{iT}^2 = \frac{\mathbf{x}_i \mathbf{M}_{\hat{\mathbf{F}}} \mathbf{x}_i}{T}, \text{ for } i = 1, \ldots, N$$

where $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{iT})$ and $\mathbf{M}_{\hat{\mathbf{F}}} = \mathbf{I}_T - \hat{\mathbf{F}} \left(\hat{\mathbf{F}}' \hat{\mathbf{F}} \right)^{-1} \hat{\mathbf{F}}'$.

- We then determine a threshold $C_{NT}^2 > 0$ such that if, and only if, $N\hat{\sigma}_{iT}^2 < C_{NT}^2$, then unit i is selected **potentially** as pervasive.

Estimation of the unobserved factors

- $\mathbf{f}_t$ can be consistently estimated up to a rotation matrix, and it is well known that

$$T^{-1}\left\|\mathbf{F}_0-\hat{\mathbf{F}}\mathbf{S}_{NT}\right\|_F^2=O_p\left[\max\left(N^{-1},T^{-1}\right)\right] \qquad (7)$$

where $\mathbf{F}_0$ is the $T\times p$ matrix of true factors, and $\mathbf{S}_{NT}$ is a $p\times p$ rotation matrix.

- This indicates the rate at which $\hat{\sigma}_{iT}^2$ converges to zero, and hence the scaling to be applied to obtain a stochastically bounded expression.

- Since it is only the product $\mathbf{F}_0\mathbf{A}_0$ in factor models that is identified, rather than $\mathbf{F}_0$ and $\mathbf{A}_0$ individually, without loss of generality we set $\mathbf{S}_{NT}=\mathbf{I}_p$.

Deriving a threshold for error variances

- To arrive at a specific threshold, note that, **if** unit i is pervasive, then

$$\hat{\sigma}_i^2 = T^{-1}\mathbf{a}_i'\mathbf{F}_0'\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{F}_0\mathbf{a}_i = T^{-1}\mathbf{a}_i'\left(\mathbf{F}_0 - \hat{\mathbf{F}}\right)'\mathbf{M}_{\hat{\mathbf{F}}}\left(\mathbf{F}_0 - \hat{\mathbf{F}}\right)\mathbf{a}_i$$
$$\leq T^{-1}\left\|\mathbf{F}_0 - \hat{\mathbf{F}}\right\|_F^2 \left\|\mathbf{a}_i\right\|_F^2 \left\|\mathbf{M}_{\hat{\mathbf{F}}}\right\|_F .$$

- Recalling that $T^{-1}\left\|\mathbf{F}_0 - \hat{\mathbf{F}}\right\|_F^2 = O_p\left[\max\left(N^{-1}, T^{-1}\right)\right]$, we consider the following scaled version of $\hat{\sigma}_i^2$:

$$N\hat{\sigma}_i^2 = \frac{\mathbf{a}_i'\mathbf{A}_0'\mathbf{V}'\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{V}\mathbf{A}_0\mathbf{a}_i}{NT} + O_p\left[\max\left(N^{-1/2}, \sqrt{N}T^{-1}\right)\right].$$

Jump to details on this result

- Since $(NT)^{-1}\mathbf{a}_i'\mathbf{A}_0'\mathbf{V}'\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{V}\mathbf{A}_0\mathbf{a}_i \leq (NT)^{-1}\mathbf{a}_i'\mathbf{A}_0'\mathbf{V}'\mathbf{V}\mathbf{A}_0\mathbf{a}_i$, if unit i is pervasive we must have

$$N\hat{\sigma}_i^2 \leq (NT)^{-1}\mathbf{a}_i'\mathbf{A}_0'\mathbf{V}'\mathbf{V}\mathbf{A}_0\mathbf{a}_i + o_p(1),$$

  and it can be shown that (as $N, T \to \infty$ such that $\sqrt{N}/T \to 0$)

$$\Pr\left(N\hat{\sigma}_i^2 > C_{NT}^2\right) \leq \Pr\left(\frac{\mathbf{a}_i'\mathbf{A}_0'\mathbf{V}'\mathbf{V}\mathbf{A}_0\mathbf{a}_i}{NT} > C_{NT}^2\right) + o(1),$$

  where C_{NT}^2 is some positive function of N, T, to be characterized below.

- Multiple testing issues arise since we need to consider N threshold tests.

- Using the assumption of exponentially decaying tails in the distribution of v_{it} we can use Lemma A11 of Chudik, Kapetanios and Pesaran (2018, Econometrica). It holds that

$$\Pr\left(\frac{\mathbf{a}_i'\mathbf{A}_0'\mathbf{V}'\mathbf{V}\mathbf{A}_0\mathbf{a}_i}{NT} > \frac{n}{N}C_{NT}^2\right) \leq T\exp\left[\frac{-\left(1-\pi\right)^2 C_{NT}^2}{2\eta_{in}^2}\left(\frac{n}{N}\right)\right] + o\left(1\right)$$

where $0 < \pi < 1$ and

$$\eta_{in}^2 = n^{-1}\mathbf{a}_i'\mathbf{A}_0'E\left(\mathbf{u}_t\mathbf{u}_t'\right)\mathbf{A}_0\mathbf{a}_i = n^{-1}\mathbf{a}_i'\mathbf{A}_0' \circ_u \mathbf{A}_0\mathbf{a}_i$$

when there are no pervasive units ($m = 0$).

- The expression for η_{in}^2 is unobserved and its estimation will be discussed below.

- Given an upper bound on $\Pr\left(N\hat{\sigma}_i^2 > C_{NT}^2\right)$ in terms of T, η_{in}^2 and C_{NT}^2, it is possible to establish sufficient properties for the positive function C_{NT}^2 such that $\Pr\left(\frac{\mathbf{a}_i'\mathbf{A}_0'\mathbf{V}'\mathbf{V}\mathbf{A}_0\mathbf{a}_i}{NT} > \frac{n}{N}C_{NT}^2\right) \to 0$.
- This condition is met if

$$C_{NT}^2 > \frac{2\log(T)\,\eta_{in}^2}{(1-\pi)^2} \quad \text{or} \quad C_{NT}^2 = 2C\log(T)\,\eta_{in}^2$$

  for some $C > 1$.
- Accordingly, i is selected as a (potential) pervasive unit if

$$\hat{\sigma}_{iT}^2 \leq \frac{2\eta_{in}^2\log(T)}{N}$$

- The threshold $\hat{\sigma}_{iT}^2 \leq \frac{2\eta_{in}^2 \log(T)}{N}$ is small enough to ensure that the sample error variances of non-pervasive units exceed its threshold value with probability approaching 1 as $N, T, \to \infty$.
- For a non-pervasive unit, the expression for $N\hat{\sigma}_{iT}$ is augmented by the two extra terms

$$B_{i7} = \frac{N\mathbf{v}_i'\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{v}_i}{T}, \; B_{i8} = \frac{N\mathbf{a}_i'\left(\mathbf{F}_0 - \hat{\mathbf{F}}\right)\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{v}_i}{T}$$

which are of order $B_{i7} = O_p\left(N\right)$ and $B_{i8} = o_p\left(N\right)$. Thus, $N\hat{\sigma}_{iT}^2 \to \infty$, as N and $T \to \infty$, if i is non-pervasive,

- Implementation of the thresholding procedure requires a consistent estimator of $\eta_{in}^2 = N^{-1}\mathbf{a}_i'\mathbf{A}'\mathbf{\Sigma}_u\mathbf{A}\mathbf{a}_i$.

- Weak cross-section correlation in $\mathbf{u}_t$ implies sparsity of the $n \times n$ cross-section covariance matrix $\mathbf{\Sigma}_u$. Hence, a suitable thresholding estimator $\tilde{\mathbf{\Sigma}}_u = (\tilde{\sigma}_{ij})$ can be used for consistent estimation.

- We use the multiple testing estimator of Bailey et al. (2019, JoE), given by

$$\tilde{\sigma}_{ij} = \hat{\sigma}_{ij} I\left(|\hat{\rho}_{ij}| > \frac{c_\pi(N)}{\sqrt{T}}\right), \; c_\pi(N) = \Phi^{-1}\left(1 - \frac{\pi}{2N^\delta}\right)$$

$$\hat{\sigma}_{ij} = T^{-1}\sum_{t=1}^{T} \hat{u}_{it}\hat{u}_{jt}, \; \hat{\rho}_{ij} = \frac{\hat{\sigma}_{ij}}{\sqrt{\hat{\sigma}_{ii}\hat{\sigma}_{jj}}}$$

- Factor loadings $\mathbf{a}_i$ and $\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_N)'$ are estimated by simple least squares regression of $\mathbf{x}_i$ on $\hat{\mathbf{F}}$.

Defining the main thresholding method

Algorithm 1 (σ^2 thresholding)

Let $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$ *be the* $T \times N$ *matrix of observations on all the* N *units in the panel. Suppose that* $p \leq p_{\max}$*, where* $p_{\max}$ *is selected a priori to be sufficiently large. Compute* $\hat{\mathbf{F}} = \mathbf{N}^{-1/2}\mathbf{X}\hat{\mathbf{Q}}$*, where* $\hat{\mathbf{Q}}$ *is the* $N \times p_{\max}$ *matrix whose columns are the orthonormalized eigenvectors of* $\mathbf{X}'\mathbf{X}$*, such that* $N^{-1}\hat{\mathbf{Q}}'\hat{\mathbf{Q}} = \mathbf{I}_{p_{\max}}$*.*
Compute $\hat{\mathbf{a}}_i = \left(\hat{\mathbf{F}}'\hat{\mathbf{F}}\right)^{-1}\hat{\mathbf{F}}'\mathbf{x}_i$ *and* $\hat{\sigma}_i^2 = T^{-1}\mathbf{x}_i'\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{x}_i$*.*
Consider the p_{max} *smallest values* $\hat{\sigma}^2_{(1)}, \hat{\sigma}^2_{(2)}, \ldots, \hat{\sigma}^2_{(p_{max})}$*. Then, select unit* (i) *to be pervasive if*

$$\hat{\sigma}^2_{(i)} \leq \frac{2\hat{\eta}^2_{(i)N}\log(T)}{N},$$

where $\hat{\eta}^2_{iN} = N^{-1}\hat{\mathbf{a}}_i'\hat{\mathbf{A}}'\tilde{\mathbf{\Sigma}}_u\hat{\mathbf{A}}\hat{\mathbf{a}}_i$*,* $\hat{\mathbf{A}} = (\hat{\mathbf{a}}_1, \hat{\mathbf{a}}_2, \ldots, \hat{\mathbf{a}}_N)'$*, and* $\tilde{\mathbf{\Sigma}}_u$ *is the multiple testing threshold estimator of* $E(\mathbf{u}_t\mathbf{u}_t')$*.*

Consistency of the main thresholding method

Theorem 1

Suppose that observations on x_{it}, for $i = 1, 2, \ldots, N$, and $t = 1, 2, \ldots, T$ are generated according to the general linear factor model set out above, with m pervasive units.
Let I_D be the set of indices of the pervasive units, and I_{ND} its complement, with I_D allowed to be an empty set.
Let $\hat{I}_D$ and $\hat{I}_{ND}$ be their estimates based on Algorithm 1.
Let Assumptions 1-4 hold and $\frac{\sqrt{N}}{T} \to 0$.
Then as N and $T \to \infty$, jointly, we have

$$\lim_{N,T\to\infty} \Pr\left(\left\{\hat{I}_D = I_D\right\} \cap \left\{\hat{I}_{ND} = I_{ND}\right\}\right) = 1$$

A sequential thresholding method

- σ^2 thresholding performs well, but can be improved upon.
- A simple adjustment is a sequential procedure to detect pervasive units one at a time. This procedure includes pervasive units detected at earlier steps as observed factors in the subsequent analysis.
- Formally, we replace the static factor model with an augmented factor model

$$x_{it} = \mathbf{f}_t^* \mathbf{a}_i^* + \mathbf{x}_{at}^{*\prime} \mathbf{b}_{ai}^* + v_{it}, \quad t = 1, 2, \ldots, T; \; i = 1, 2, \ldots, N_1. \tag{8}$$

- where $\mathbf{x}_{at}^*$ is a $r \times 1$ vector of identified pervasive units. $\mathbf{f}_t^*$ is a $p_{\max} - r$ vector of unobserved common factors.

- starting with $r = 0$ identified pervasive units, $N_1 = N - r$ and some $p_{\max} > m + 1$ the sequential procedure consists of the following two steps:

Algorithm 2 (S$-\sigma^2$ thresholding)

1. *Conduct σ^2 thresholding using model (8) with $m^* = p_{\max} - r$ estimated factors. Let $\tilde{m}$ be the estimated number of pervasive units estimated using Algorithm 1.*
   If $\tilde{m} = 0$, stop and conclude that there are r pervasive units.
2. *If $\tilde{m} > 0$, obtain $i^* = \arg\min_i \hat{\sigma}_i^2$. Append $\mathbf{x}_{i^*}$ to $\mathbf{X}_a^*$ and drop $\mathbf{x}_{i^*}$ from $\mathbf{X}$. Update r to $r + 1$ and N_1 to $N_1 - 1$.*

Consistency of sequential thresholding

Corollary 2

Suppose that observations on x_{it}, for $i = 1, 2, \ldots, N$, and $t = 1, 2, \ldots, T$ are generated according to the general linear factor model given above with m pervasive units.
Let I_D be the set of indices of the pervasive units, and I_{ND} its complement, with I_D allowed to be an empty set.
Let $\hat{I}_D$ and $\hat{I}_{ND}$ be their estimates based on $S{-}\sigma^2$ thresholding.
Let Assumptions 1-4 hold and $\frac{\sqrt{N}}{T} \to 0$.
Then as N and $T \to \infty$, jointly, we have

$$\lim_{N,T\to\infty} \Pr\left(\left\{\hat{I}_D = I_D\right\} \cap \left\{\hat{I}_{ND} = I_{ND}\right\}\right) = 1$$

Sequential thresholding with a multiple testing hurdle

- The risk of falsely detecting a pervasive unit can be further reduced with an additional multiple testing (MT) hurdle, applied to a newly identified pervasive unit detected by the above sequential procedure.
- The MT hurdle constitutes a diagnostic check on whether the identified pervasive unit is a sufficiently *strong* factor: It evaluates whether its corresponding slope coefficient is non-zero for most non-pervasive units.

Algorithm 3 (MT hurdle)

1. *Given an identified pervasive unit* i^*, r *previously identified pervasive units* $\mathbf{x}_{at}^*$ *and* $p_{\max} - r - 1$ *unobserved factors* $\mathbf{f}_t^*$, *estimate*

$$x_{jt} = x_{i^*t}\gamma_j^* + \mathbf{f}_t^{*\prime}\mathbf{a}_j^* + \mathbf{x}_{at}^{*\prime}\mathbf{b}_{aj}^* + v_{jt}, \ \ t = 1, 2, \ldots, T$$

for each $j = 1, \ldots, i^* - 1, i^* + 1, \ldots, N$.

2. *Carry out* $N_1 - 1$ *individual t-type tests to check the significance of the slope parameters* $\hat{\gamma}_j^*$ *for all* $j \neq i$ *using the MT critical value* $\Phi^{-1}\left[1 - \frac{0.01}{2(N_1 - 2)}\right]$. *These tests have the form* $t_j^* = \hat{\gamma}_j^* \sqrt{\sum_{t=1}^T x_{i^*t}^2 \left(T^{-1}\sum_{t=1}^T \hat{v}_{jt}^2\right)^{-1}}$.

3. *Let* M *denoted the number of rejections among these tests. If* $\log(M)/\log(N) > 1/2$, *conclude that unit* i^* *has passed the MT hurdle. If not, conclude that the hurdle has failed, and the unit in question is declared as non-dominant.*

- Starting with $r = 0$ identified pervasive units, $N_1 = N - r$ and some $p_{\max} > m + 1$, the sequential procedure with MT hurdle consists of the following three steps:

Algorithm 4 (SMT$-\sigma^2$ thresholding)

1. *Conduct σ^2 thresholding using model (8) with $m^* = p_{\max} - r$ estimated factors. Let $\tilde{m}$ be the estimated number of pervasive units estimated using Algorithm 1.*
   If $\tilde{m} = 0$, stop and conclude that there are r pervasive units.
2. *If $\tilde{m} > 0$, obtain $i^* = \arg\min_i \hat{\sigma}_i^2$.*
   Apply the MT hurdle to this cross-section unit. If the hurdle is failed, stop and conclude that there are r pervasive units.
3. *If the hurdle is passed, append $\mathbf{x}_{i^*}$ to $\mathbf{X}_a^*$ and drop $\mathbf{x}_{i^*}$ from $\mathbf{X}$.*
   Update r to $r + 1$ and N_1 to $N_1 - 1$.

A roadmap for the remainder of this talk

1. Monte Carlo study results concerning the finite-sample properties of $SMT - \sigma^2$ thresholding and the procedures proposed by BM (2018), and PS (2016).
2. Three empirical applications
   2.1 Sectorial industrial production in the U.S.
   2.2 Economic growth and equity markets worldwide
   2.3 Housing prices in the U.S.

A Monte Carlo Study

- We now investigate how well the unknown number of pervasive units (m_0), as well as their identities are estimated in finite samples.
- Amongst the thresholding procedures discussed today, results for SMT$-\sigma^2$ thresholding (Algorithm 4) will be presented.
- We compare its performance relative to that of PS and BM.
- Two performance criteria are being considered:
  1. The frequency of correctly identifying only the true pervasive units
  2. The average number of non-pervasive units falsely detected

- We consider four different cases, distinguished by the presence or absence of pervasive units and external factors:

	Pervasive units	
Factors	none	1 or 2
none	**Design set I**	**Design set III**
1 or 2	**Design set II**	**Design set IV**

- In all cases errors are generated with spatial effects, which allow for weak error cross-sectional dependence.
- Designs **II** and **IV** allow for common exposure to external shocks.
- Designs **I** and **II** do not contain a pervasive unit (ie $m_0 = 0$).
- Additional MCs are carried with weakly pervasive units.

Formal model setup for simulations

- Formally, we simulate the model

$$\mathbf{x}_{ta} = \boldsymbol{\mu}_a + \boldsymbol{\Lambda}_a \mathbf{g}_t + \mathbf{u}_{at}, \tag{9}$$

$$\mathbf{x}_{tb} = \boldsymbol{\mu}_b + \mathbf{B}\mathbf{x}_{ta} + \boldsymbol{\Lambda}_b \mathbf{g}_t + \mathbf{u}_{bt}, \tag{10}$$

- The elements of $\boldsymbol{\mu} = (\boldsymbol{\mu}_a; \boldsymbol{\mu}_b)$ are $IIDU(0, 1)$.
- $\mathbf{g}_t = \mathbf{R}_g^{1/2} \left(\mathbf{g}_{*,t} - 2\boldsymbol{\tau}_k \right)$ is $k_0 \times 1$, the elements of $\mathbf{g}_{*,t}$ being $IID\chi^2 \left(2\right)$ and $\mathbf{R}_g = (1 - \rho_g) \mathbf{I}_k + \rho_g \boldsymbol{\tau}_k \boldsymbol{\tau}_k'$. Additionally, $\rho_g \sim U \left(0.2, 0.8\right)$. The $m_0 \times 1$ vector $\mathbf{h}_t$ is generated analogously.
- The elements of the $m_0 \times k_0$ matrix $\boldsymbol{\Lambda}_a$ and the $n \times k_0$ matrix $\boldsymbol{\Lambda}_b$ are $IIDU \left(0, 1\right)$.

- The $(N - m_0) \times m_0$ matrix $\mathbf{B} = (b_{ij})$ has elements

$$b_{ij} \begin{cases} \sim IIDU\,(0,1) \text{ if } i \leq \left\lfloor (N - m_0)^{\alpha} \right\rfloor \\ = 0 \text{ otherwise} \end{cases}$$

where setting $\alpha = 1$ results in pervasive units, and $\alpha = 0.8$ entails weakly pervasive units.

- Model errors are generated as heterogeneous AR(1) processes with weakly cross-sectionally correlated innovations. Namely,

$$u_{it} = \rho_i u_{it-1} + \left(1 - \rho_i^2\right)^{1/2} \varepsilon_{it}$$

where $\rho_i \sim IIDU\,(0.2, 0.5)$.

- Here, $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, \ldots, \varepsilon_{nt})' = \boldsymbol{\Sigma}^{1/2}\mathbf{R}_u^{1/2}\boldsymbol{\zeta}_t$ with $\boldsymbol{\Sigma} = diag\left(\sigma_{11}, \sigma_{22}, \ldots, \sigma_{nn}\right)$ and

$$\mathbf{R}_u = \begin{pmatrix} 1 & \rho_u & \rho_u^2 & \cdots & \rho_u^{n-1} \\ \rho_u & 1 & \rho_u & \cdots & \rho_u^{n-2} \\ \rho_u^2 & \rho_u & 1 & \cdots & \rho_u^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_u^{n-1} & \rho_u^{n-2} & \rho_u^{n-3} & \cdots & 1 \end{pmatrix}.$$

- We set $\rho_u = 0.5$, $\sigma_{ii} = \sigma_{*,ii}/4 + 0.5$ and $\sigma_{*,ii} \sim IID\chi^2\left(2\right)$. Lastly, the $n \times 1$ vector of innovations $\boldsymbol{\zeta}_t$ has elements $\zeta_{it} = \left(\zeta_{*,it} - 2\right)/2$, where $\zeta_{*,it} \sim IID\chi^2\left(2\right)$.

- Experiments are carried out for all combinations of $N \in \{50, 100, 200, 500\}$, $T \in \{60, 110, 210, 250\}$.

- All combinations of $m_0 \leq 2$ and $k_0 \leq 2$ are considered. Results are obtained for $\alpha \in \{0.8, 1\}$.

Design sets I&II: no pervasive units

Table 1: Empirical frequency of correctly identifying the absence of a pervasive unit ($m_0 = 0$)

	SMT$-\sigma^2$				PS				
	$k_0 = 0$				$k_0 = 0$				
N\T	60	110	210	250	N\T	60	110	210	250
50	100	100	100	100	50	99.4	99.2	99.6	99.8
100	100	100	100	100	100	100	100	100	100
200	100	100	100	100	200	100	100	100	100
500	100	100	100	100	500	100	100	100	100
	$k_0 = 1$				$k_0 = 1$				
N\T	60	110	210	250	N\T	60	110	210	250
50	88.4	86.4	82.7	80.3	50	53.2	92.0	97.3	97.7
100	94.1	92.3	90.7	88.9	100	75.5	98.5	100	100
200	99.8	99.2	99.4	99.2	200	90.6	100	100	100
500	100	100	100	100	500	92.9	100	100	100
	$k_0 = 2$				$k_0 = 2$				
N\T	60	110	210	250	N\T	60	110	210	250
50	61.6	55.9	47.7	44.3	50	81.0	80.1	69.5	69.5
100	84.0	74.5	64.2	60.9	100	86.6	85.7	63.1	57.4
200	98.6	97.7	94.2	94.1	200	82.5	66.1	46.3	39.7
500	100	100	100	99.9	500	99.4	46.8	22.6	17.6

Notes: SMT$-\sigma^2$ thresholding is implemented with $p_{max} = m_0 + k_0 + 1$. PS refers to the method of Parker and Sul (2016). BM is left out since all corresponding values are 0.

Table 2: Average number of non-pervasive units falsely selected as pervasive ($m_0 = 0$)

	$k_0 = 0$				$k_0 = 1$				$k_0 = 2$			
	SMT$-\sigma^2$				SMT$-\sigma^2$				SMT$-\sigma^2$			
N\T	60	110	210	250	60	110	210	250	60	110	210	250
50	0	0	0	0	0.1	0.2	0.2	0.2	0.4	0.5	0.6	0.7
100	0	0	0	0	0.1	0.1	0.1	0.1	0.2	0.3	0.4	0.4
200	0	0	0	0	0	0	0	0	0	0	0.1	0.1
500	0	0	0	0	0	0	0	0	0	0	0	0
	PS				PS				PS			
N\T	60	110	210	250	60	110	210	250	60	110	210	250
50	0	0	0	0	0.9	0.2	0.2	0.1	0.7	1.2	1.8	1.8
100	0	0	0	0	0.3	0	0	0	1.0	1.4	3.7	4.2
200	0	0	0	0	0.1	0	0	0	3.2	6.7	10.7	12.0
500	0	0	0	0	0.1	0	0	0	0	26.2	38.4	41.0
	BM				BM				BM			
N\T	60	110	210	250	60	110	210	250	60	110	210	250
50	4.1	3.7	4.7	4.9	3.9	4.0	4.5	4.9	3.9	3.8	4.5	4.7
100	n/a	3.6	3.6	4.1	n/a	3.5	3.7	4.2	n/a	3.7	3.6	4.0
200	n/a	n/a	3.2	3.1	n/a	n/a	3.2	3.0	n/a	n/a	3.1	3.0
500	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

Notes: See the notes to Table 1. BM refers to the modified detection method used in Section 6 of Brownlees and Mesters (2018).

Discussing MC results without pervasive units

- SMT$-\sigma^2$ thresholding performs well, even in the presence of external common factors, so long as N is sufficiently large.
- Its average number of false discoveries is at most 0.7.
- BM always *incorrectly* selects at least one pervasive unit since it assumes $m_0 > 0$. The average number of false discoveries of BM procedure is 3 to 4.
- PS outperforms SMT$-\sigma^2$ thresholding somewhat if N is small and if there are external factors. But PS seems to break down for $k_0 = 2$ as N is increased.

Design sets III&IV: Correct specification rates

Table 3: Empirical frequency of correctly identifying only the true strongly pervasive units ($m_0 = 1$, and $\alpha = 1$)

$k_0 = 0$

N\T	SMT-σ^2 60	SMT-σ^2 110	SMT-σ^2 210	SMT-σ^2 250	PS 60	PS 110	PS 210	PS 250	BM 60	BM 110	BM 210	BM 250
50	97.7	99.9	100	100	61.0	92.6	95.1	95.5	56.4	98.9	100	100
100	100	100	100	100	80.8	99.4	100	100	n/a	78.9	100	100
200	100	100	100	100	91.2	99.9	100	100	n/a	n/a	92.1	100
500	100	100	100	100	94.0	100	100	100	n/a	n/a	n/a	n/a

$k_0 = 1$

N\T	SMT-σ^2 60	SMT-σ^2 110	SMT-σ^2 210	SMT-σ^2 250	PS 60	PS 110	PS 210	PS 250	BM 60	BM 110	BM 210	BM 250
50	58.9	80.6	82.9	82.3	0.5	0	0	0	50.3	97.6	100	100
100	68.1	88.4	93.3	93.0	0.1	0	0	0	n/a	72.9	100	100
200	79.1	97.8	99.6	99.5	0	0	0	0	n/a	n/a	88.6	100
500	82.1	99.9	100	100	0	0	0	0	n/a	n/a	n/a	n/a

$k_0 = 2$

N\T	SMT-σ^2 60	SMT-σ^2 110	SMT-σ^2 210	SMT-σ^2 250	PS 60	PS 110	PS 210	PS 250	BM 60	BM 110	BM 210	BM 250
50	52.5	61.7	61.1	55.5	0	0	0	0	44.3	97.0	99.9	100
100	65.3	75.9	74.7	74.2	0	0	0	0	n/a	69.2	100	100
200	72.7	95.6	97.1	96.0	0	0	0	0	n/a	n/a	87.4	100
500	77.1	99.4	100	100	0	0	0	0	n/a	n/a	n/a	n/a

Notes: See the notes to Table 2.

Table 4: Empirical frequency of correctly identifying only the true strongly pervasive units ($m_0 = 2$, and $\alpha = 1$)

	SMT-σ^2				PS				BM			
	$k_0 = 0$				$k_0 = 0$				$k_0 = 0$			
N\T	60	110	210	250	60	110	210	250	60	110	210	250
50	55.8	85.0	96.0	97.7	0.2	0.1	0.2	0.1	27.3	92.2	99.5	99.7
100	58.9	87.3	98.2	98.6	0	0	0	0	n/a	48.2	100	100
200	59.0	88.8	98.4	98.9	0.1	0	0	0	n/a	n/a	67.7	100
500	60.9	94.8	100	100	0	0	0	0	n/a	n/a	n/a	n/a
	$k_0 = 1$				$k_0 = 1$				$k_0 = 1$			
N\T	60	110	210	250	60	110	210	250	60	110	210	250
50	36.2	67.3	79.1	79.5	0	0	0	0.1	21.0	86.6	98.1	98.5
100	41.7	78.5	91.5	92.4	0	0	0	0	n/a	39.9	99.9	100
200	43.5	87.6	98.3	99.3	0	0	0	0	n/a	n/a	57.2	99.9
500	46.0	96.2	100	100	0	0	0	0	n/a	n/a	n/a	n/a
	$k_0 = 2$				$k_0 = 2$				$k_0 = 2$			
N\T	60	110	210	250	60	110	210	250	60	110	210	250
50	38.9	61.3	63.0	60.5	0	0	0	0	19.2	80.8	96.6	97.7
100	48.4	73.3	79.6	79.6	0	0	0	0	n/a	36.4	100	99.9
200	47.5	86.9	96.8	97.1	0	0	0	0	n/a	n/a	52.3	99.7
500	41.0	94.6	99.9	100	0	0	0	0	n/a	n/a	n/a	n/a

Notes: See the notes to Table 2.

Design sets III&IV: Average number of false detections

Table 5: Average number of non-pervasive units falsely selected as pervasive units ($m_0 = 1$, and $\alpha = 1$)

	SMT-σ^2					PS					BM			
$k_0 = 0$														
N\T	60	110	210	250	N\T	60	110	210	250	N\T	60	110	210	250
50	0	0	0	0	50	0.7	0.2	0.2	0.2	50	1.3	0	0	0
100	0	0	0	0	100	0.2	0	0	0	100	n/a	0.5	0	0
200	0	0	0	0	200	0.1	0	0	0	200	n/a	n/a	0.2	0
500	0	0	0	0	500	0.1	0	0	0	500	n/a	n/a	n/a	n/a
$k_0 = 1$														
N\T	60	110	210	250	N\T	60	110	210	250	N\T	60	110	210	250
50	0.2	0.1	0.2	0.2	50	1.1	1.8	2.5	2.7	50	1.6	0	0	0
100	0.1	0.1	0.1	0.1	100	2.1	3.1	5.4	5.9	100	n/a	0.8	0	0
200	0	0	0	0	200	6.0	10.4	14.1	15.8	200	n/a	n/a	0.3	0
500	0	0	0	0	500	0	37.2	46.8	47.5	500	n/a	n/a	n/a	n/a
$k_0 = 2$														
N\T	60	110	210	250	N\T	60	110	210	250	N\T	60	110	210	250
50	0.4	0.3	0.4	0.5	50	1.8	3.0	3.7	3.8	50	1.8	0	0	0
100	0.1	0.2	0.3	0.3	100	2.9	4.3	7.0	7.5	100	n/a	0.9	0	0
200	0	0	0	0	200	7.8	13.4	16.5	17.0	200	n/a	n/a	0.3	0
500	0	0	0	0	500	0	41.2	46.7	46.9	500	n/a	n/a	n/a	n/a

Notes: See the notes to Table 2.

Table 6: Average number of non-pervasive units falsely selected as pervasive units ($m_0 = 2$, and $\alpha = 1$)

SMT-σ^2

$k_0 = 0$

N\T	60	110	210	250
50	0	0	0	0
100	0	0	0	0
200	0	0	0	0
500	0	0	0	0

$k_0 = 1$

N\T	60	110	210	250
50	0.2	0.1	0.1	0.1
100	0	0	0	0
200	0	0	0	0
500	0	0	0	0

$k_0 = 2$

N\T	60	110	210	250
50	0.3	0.3	0.3	0.4
100	0.1	0.1	0.2	0.2
200	0	0	0	0
500	0	0	0	0

PS

$k_0 = 0$

N\T	60	110	210	250
50	0.7	1.2	1.5	1.5
100	1.0	1.6	3.1	3.3
200	3.5	6.7	9.5	10.7
500	0	25.3	35.7	38.0

$k_0 = 1$

N\T	60	110	210	250
50	2.2	3.2	3.8	3.7
100	3.9	5.3	7.1	7.2
200	9.8	14.6	16.0	16.0
500	0	43.0	44.6	44.6

$k_0 = 2$

N\T	60	110	210	250
50	2.7	3.6	3.7	3.8
100	5.1	6.9	8.1	8.0
200	10.4	16.0	16.5	16.2
500	0	43.6	44.1	44.1

BM

$k_0 = 0$

N\T	60	110	210	250
50	0.8	0	0	0
100	n/a	0.2	0	0
200	n/a	n/a	0.1	0
500	n/a	n/a	n/a	n/a

$k_0 = 1$

N\T	60	110	210	250
50	1.0	0	0	0
100	n/a	0.4	0	0
200	n/a	n/a	0.1	0
500	n/a	n/a	n/a	n/a

$k_0 = 2$

N\T	60	110	210	250
50	1.0	0	0	0
100	n/a	0.4	0	0
200	n/a	n/a	0.1	0
500	n/a	n/a	n/a	n/a

Notes: See the notes to Table 2.

Discussing MC results with pervasive units

- Performance of SMT$-\sigma^2$ thresholding improves steadily as both N and T increase.
- The existence of common factors leads to a deterioration of small sample performance. However, the average number of falsely selected units as pervasive is close to zero.
- By construction, BM can only be applied if $T > N$. If this condition is satisfied, BM procedure performs well.
- PS works well only if $m_0 = 1$ and $k_0 = 0$. It breaks down completely if there are external factors or more than one pervasive units.
- These observations are unchanged if pervasive units are weakly pervasive ($\alpha = 0.8$) rather than strongly pervasive ($\alpha = 1$).

jump to corresponding MC results

- SMT$-\sigma^2$ thresholding has best overall performance.

Empirical application 1: U.S. industrial production

- We apply SMT$-\sigma^2$ thresholding to growth rates in monthly industrial production in $N = 138$ U.S. industrial sectors.
- The full sample length 1972m1-2007m12 as well as the subsamples 1972m1-1983m12 and 1984m1-2007m12 are investigated.
- To cover a wide range of possible factors, we consider $p_{\max} = \{2, 3, 4, 5, 6\}$.
- The competitor methods of Parker and Sul (2016) as well as Brownlees and Mesters (2018) are applied as a benchmark.

Table 7: Pervasive units in sector-wise industrial production in the U.S.

	Full sample (1972m1 - 2007m12)		
Approach:	SMT$-\sigma^2$	PS	BM
p_{max}	2, 3, 4, 5, 6	1†	
Number of pervasive units:	0	0	1
Identities:			Fluid Milk
	Sub-sample A (1972m1 - 1983m12)		
Approach:	SMT$-\sigma^2$	PS	BM
p_{max}	2, 3, 4, 5, 6	1†	
Number of pervasive units:	0	1	2
Identities:		Plastics Products	Commercial and Service Industry Machinery; Bakeries and Tortilla

†: This value minimizes the IC_{p2} criterion of Bai and Ng (2002) for selecting the number of common factors. Maximum number of factors is set to 10.

Notes: Data taken from Foerster et al. (2011)

Sub-sample B (1984m1 - 2007m12)			
Approach:	SMT$-\sigma^2$	PS	BM
p_{max}	2, 3, 4, 5, 6	$2^{\dagger}$	
Number of pervasive units:	0	19	12
Identities:		*	**

†: This value minimizes the IC_{p2} criterion of Bai and Ng (2002) for selecting the number of common factors. Maximum number of factors is set to 10.

*: Cheese; Breweries; Carpet and Rug Mills; Sawmills and Wood Preservation; Reconstituted Wood Products; Artificial and Synthetic Fibers and Filaments; Plastics Products; Tires; Rubber Products Ex Tires; Lime and Gypsum Products; Foundries; Fabricated Metals: Forging and Stamping; Boiler, Tank, and Shipping Containers; Machine Shops; Turned Products; and Screws, Nuts, and Bolts; Coating, Engraving, Heat Treating, and Allied Activities; Metal Valves Except Ball and Roller Bearings; Metalworking Machinery; Other Electrical Equipment; Travel Trailers and Campers.

**: Fluid Milk; Commercial and Service Industry Mach; Plastics Products;Other Miscellaneous Manufacturing; Metal Valves Except Ball and Roller Bearings; Bakeries and Tortilla; Medical Equipment and Supplies; Newspaper Publishers; Navigational/Measuring/Electromedical/Control Instruments; Architectural and Structural Metal Products; Metalworking Machinery; Printing and Related Support Activities.

Notes: Data taken from Foerster et al. (2011)

- Using SMT-σ^2 thresholding, we do not detect any strongly pervasive sector, in the case of all three sample periods considered.
- Results from the applications of BM and PS procedures are mixed, and vary across sample periods. PS finds 0 units over the full sample, 1 unit over sub-sample A and 19 units as dominant over sub-sample B. BM selects 1 unit as pervasive over the full sample, 2 units over the sub-sample A and 12 units over the sub-sample B.
- BM provide their own analysis of this data using standardized data and find a few pervasive sectors related to automobiles and trucks. Our results are computed without standardization. Recall that the MC results suggest that BM procedure performs much better when data is not standardized.

Empirical application 2: Pervasive economies and equity markets

- Our second application uses quarterly observations on *real GDP* and *real equity prices* in 33 and 26 large economies, respectively.
- Data is transformed into rates of changes.
- Data is taken from the latest vintage of the GVAR database, covering the period 1979Q2-2016Q4.

Table 8: Pervasive unit detection methods applied to cross country rates of change of real GDP (33 countries) and real equity prices (26 markets) over the period 1979Q2-2016Q4 (151 time periods)

	Rate of change of real GDP				
Approach:	SMT−σ^2			PS	BM
p_{max}	2	$\{3,4,5\}$	6	$1^{\dagger}$	
Number of pervasive units:	0	1	0	0	2
Identities:		France			France Spain

	Rate of change of real equity prices		
Approach:	SMT−σ^2	PS	BM
p_{max}	2, 3, 4, 5, 6	$2^{\dagger}$	
Number of pervasive units:	0	6	6
Identities:		France Germany Malaysia Netherlands Singapore Thailand	USA Netherlands UK Canada Switzerland Germany

† This value minimizes the IC_{p2} criterion of Bai and Ng (2002) for selecting the number of common factors. Maximum number of factors is set to 10.

Empirical application 3: U.S. house prices changes

- In a third application, we investigate the existence of U.S. states whose house prices exert a persistent impact on house prices in all other states.
- We consider Freddie Mac House Price Indexes for 48 U.S. states, leaving out Alaska, Hawaii and the District of Columbia.
- Data is observed quarterly over the period 1975Q1-2014Q4, deflated by consumer prices indexes and transformed into rates of change.

Table 9: Estimated U.S. states with pervasive housing market

Approach:	SMT$-\sigma^2$			PS	BM
p_{max}	2	3	4, 5, 6	$5^{\dagger}$	
Number of pervasive units:	1	2	0	2	4
Identities:	New York	Kentucky New York		New Hampshire Nevada	North Carolina Maryland Virginia Connecticut

†: This value minimizes the IC_{p2} criterion of Bai and Ng (2002) for selecting the number of common factors. Maximum number of factors is set to 10.

Notes: Data taken from Freddie Mac House Price Indexes and Yang (2018).

Conclusion

- We suggest a rigorously developed approach for the detection of pervasive units in large panel datasets.
- Our method relies on the observation that pervasive units are common factors for all other units.
- This suggests a thresholding procedure to identify as pervasive units the set of cross-sections that is almost perfectly explained by estimated common factors from the data.

Conclusion

- Using a factor model framework enables us to detect pervasive units without further information on linkages between units.
- Drawing on results from the multiple testing literature, we allow the number of potential pervasive units to be very large.
- Our method performs very well in a wide array of scenarios, including both external factors as well as the absence of pervasive units.

Background slides

Detailed statement of model assumptions.

Assumption 1

1. $\mathbf{f}_t$ is a covariance-stationary stochastic process with $E\left(\mathbf{f}_t\mathbf{f}_t'\right) = \mathbf{I}_p$.
2. There exist sufficiently large positive constants C_0 and C_1 and $s_f > 0$ such that

$$\sup_t \Pr\left(\left|f_{jt}\right| > a\right) \leq C_0 \exp\left(-C_1 a^{s_f}\right) \text{ for each } j = 1, 2, \ldots, p.$$

3. $T^{-1}\sum_{t=1}^{T}\mathbf{f}_t\mathbf{f}_t' \xrightarrow{p} \mathbf{I}_p$ and $T^{-1}\sum_{t=1}^{T}\left[\left\|\mathbf{f}_t\right\|^j - E\left(\left\|\mathbf{f}_t\right\|^j\right)\right] \xrightarrow{p} \mathbf{0},\ j = 3, 4.$

Assumption 2

1. $\mathbf{A}_a$ and $\mathbf{A}_b$ are parameter matrices, the former satisfying $Rank\left(\mathbf{A}_a\right) = m \geq 0$.
2. $\inf_i \|\mathbf{a}_i\| > c$, and $\sup_i \|\mathbf{a}_i\| < C$, and for any $N = n + m$ (m being a finite integer)

$$\lambda_{\max}\left(n^{-1}\sum_{i=m+1}^{N}\mathbf{a}_i\mathbf{a}_i'\right) < C < \infty,$$

$$\lambda_{\min}\left(n^{-1}\sum_{i=m+1}^{N}\mathbf{a}_i\mathbf{a}_i'\right) > c > 0.$$

Assumption 3

1. The $n \times 1$ vector $\mathbf{u}_t$ is defined by

$$\mathbf{u}_t = \mathbf{H}\boldsymbol{\varepsilon}_t$$

where

$$\boldsymbol{\varepsilon}_t = \left(\varepsilon_{m+1,t}, \varepsilon_{m+2,t}, \ldots, \varepsilon_{N,t}\right)' \sim IID\left(0, \mathbf{I}_n\right),$$

and $\sup_i T^{-1} \sum_{t=1}^{T} \sum_{t'=1}^{T} \left| Cov\left(\varepsilon_{it}, \varepsilon_{it'}\right)\right| < C < \infty$.

2. There exist sufficiently large positive constants C_0 and C_1 and $s_\varepsilon > 0$ such that

$$\sup_{i,t} \Pr\left(\left|\varepsilon_{it}\right| > a\right) \leq C_0 \exp\left(-C_1 a^{s_\varepsilon}\right).$$

Assumption 3 (cont.)

3. $\mathbf{H} = (h_{ij})$ is an $n \times n$ matrix with fixed coefficient, with bounded row and column sum norms, formally $\|\mathbf{H}_1\| = \sup_j \sum_{i=1}^{n} |h_{ij}| < C$, and $\|\mathbf{H}_\infty\| = \sup_i \sum_{j=1}^{n} |h_{ij}| < C$. Furthermore, $\lambda_{\min}(\mathbf{H}\mathbf{H}') > c > 0$.

Assumption 4 $\mathbf{f}_t$ and ε_{is} are independent for all i, s, t.

back to assumption summary

Deriving the leading term in residual variances

Assume that unit i is pervasive. We want to show that

$$N\hat{\sigma}_i^2 \leq \frac{\mathbf{a}_i'\mathbf{A}_0'\mathbf{V}'\mathbf{V}\mathbf{A}_0\mathbf{a}_i}{NT} + O_p\left[\max\left(N^{-1/2}, \sqrt{N}T^{-1}\right)\right].$$

1. Note that $\mathbf{M}_{\hat{\mathbf{F}}}\hat{\mathbf{F}} = \mathbf{0}$, implying

$$\begin{aligned}\hat{\sigma}_i^2 &= \frac{\mathbf{a}_i'\mathbf{F}_0'\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{F}_0\mathbf{a}_i}{T} \\ &= \frac{\mathbf{a}_i'\left(\mathbf{F}_0' - \hat{\mathbf{F}}\right)\mathbf{M}_{\hat{\mathbf{F}}}\left(\mathbf{F}_0 - \hat{\mathbf{F}}\right)\mathbf{a}_i}{T}\end{aligned}$$

2. The estimator of $\mathbf{F}_0$ is $\hat{\mathbf{F}} = N^{-1}\mathbf{X}\hat{\mathbf{A}}$, where $\hat{\mathbf{A}} = \sqrt{N}\hat{\mathbf{Q}}$ and $\hat{\mathbf{Q}}$ is the orthonormal eigenvectors associated to the p largest eigenvalues of $\mathbf{X}'\mathbf{X}$. The latter allows us to assume (w.l.o.g.) that $N^{-1}\mathbf{A}_0'\mathbf{A}_0 = \mathbf{I}_p$.

3. Since $\mathbf{X} = \mathbf{F}_0\mathbf{A}_0' + \mathbf{V}$, we have

$$\frac{\mathbf{X}\mathbf{A}_0}{N} = \mathbf{F}_0 + \frac{\mathbf{V}\mathbf{A}_0}{N},$$

and after solving for $\mathbf{F}_0$,

$$\mathbf{F}_0 - \hat{\mathbf{F}} = \frac{\mathbf{X}\left(\mathbf{A}_0 - \hat{\mathbf{A}}\right)}{N} - \frac{\mathbf{V}\mathbf{A}_0}{N}.$$

4. Substituting out $\mathbf{X}$ above plus some algebra allow us to write

$$\mathbf{M}_{\hat{\mathbf{F}}}\left(\mathbf{F}_0 - \hat{\mathbf{F}}\right) = \frac{\mathbf{M}_{\hat{\mathbf{F}}}\left(\mathbf{F}_0 - \hat{\mathbf{F}}\right)\mathbf{A}_0'\left(\mathbf{A}_0 - \hat{\mathbf{A}}\right)}{N} + \frac{\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{V}\left(\mathbf{A}_0 - \hat{\mathbf{A}}\right)}{N} - \frac{\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{V}\mathbf{A}_0}{N}.$$

5. Hence, estimated error variances can be decomposed into

$$N\hat{\sigma}_i^2 = B_{i1} + B_{i2} + \ldots + B_{i6},$$

where

$$B_{i1} = \frac{\mathbf{a}_i'\mathbf{A}_0'\mathbf{V}'\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{V}\mathbf{A}_0\mathbf{a}_i}{NT},$$

$$B_{i2} = 2\frac{\mathbf{a}_i'\mathbf{A}_0'\mathbf{V}'\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{V}\left(\mathbf{A}_0 - \hat{\mathbf{A}}\right)\mathbf{a}_i}{NT},$$

$$B_{i3} = 2\frac{\mathbf{a}_i'\mathbf{A}_0'\mathbf{V}'\mathbf{M}_{\hat{\mathbf{F}}}\left(\mathbf{F}_0 - \hat{\mathbf{F}}\right)\mathbf{A}_0'\left(\mathbf{A}_0 - \hat{\mathbf{A}}\right)\mathbf{a}_i}{NT},$$

and

$$B_{i4} = \frac{\mathbf{a}_i'\left(\mathbf{A_0} - \hat{\mathbf{A}}\right)'\mathbf{V}'\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{V}\left(\mathbf{A}_0 - \hat{\mathbf{A}}\right)\mathbf{a}_i}{NT},$$

$$B_{i5} = 2\frac{\mathbf{a}_i'\left(\mathbf{A_0} - \hat{\mathbf{A}}\right)'\mathbf{V}'\mathbf{M}_{\hat{\mathbf{F}}}\left(\mathbf{F}_0 - \hat{\mathbf{F}}\right)\mathbf{A}_0'\left(\mathbf{A}_0 - \hat{\mathbf{A}}\right)\mathbf{a}_i}{NT},$$

$$B_{i6} = \frac{\mathbf{a}_i'\left(\mathbf{A_0} - \hat{\mathbf{A}}\right)'\mathbf{A}_0\left(\mathbf{F}_0 - \hat{\mathbf{F}}\right)'\mathbf{M}_{\hat{\mathbf{F}}}\left(\mathbf{F}_0 - \hat{\mathbf{F}}\right)\mathbf{A}_0'\left(\mathbf{A}_0 - \hat{\mathbf{A}}\right)\mathbf{a}_i}{NT}.$$

The five terms $B_{i2}, B_{i3}, \ldots, B_{i6}$ require upper bounds on the difference between true model parameters and their estimators, i.e. $\left(\mathbf{F}_0 - \hat{\mathbf{F}}\right)$ and $\left(\mathbf{A_0} - \hat{\mathbf{A}}\right)$.

6. Let $\delta_{NT} = \min\left(\sqrt{N}, \sqrt{T}\right)$ Analogous to $T^{-1}\left\|\mathbf{F}_0 - \hat{\mathbf{F}}\right\|_F^2 = O_p\left(\delta_{NT}^{-2}\right)$, it holds that

$$\left\|\mathbf{F}_0 - \hat{\mathbf{F}}\right\|_F = O_p\left(\frac{\sqrt{T}}{\delta_{NT}}\right); \quad \left\|\mathbf{A}_0 - \hat{\mathbf{A}}\right\|_F = O_p\left(\frac{\sqrt{N}}{\delta_{NT}}\right)$$

$$\left\|\mathbf{V}\left(\mathbf{A}_0 - \hat{\mathbf{A}}\right)\right\|_F = O_p\left(\frac{\sqrt{NT}}{\delta_{NT}}\right); \quad \left\|\mathbf{V}\mathbf{A}_0\right\|_F = O_p\left(\sqrt{NT}\right)$$

$$\left\|\mathbf{A}_0'\left(\mathbf{A}_0 - \hat{\mathbf{A}}\right)\right\|_F = O_p\left(\frac{N}{\delta_{NT}}\right)$$

which is shown using the theoretical framework of Bai and Ng (2002) and Bai (2003).

7. We can hence arrive at

$$B_{i2} + B_{i3} + \ldots + B_{i6} = O_p\left(\frac{1}{\delta_{NT}}\right) + O_p\left(\frac{\sqrt{N}}{\delta_{NT}^2}\right).$$

It immediately follows that

$$N\hat{\sigma}_i^2 = \frac{\mathbf{a}_i'\mathbf{A}_0'\mathbf{V}'\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{V}\mathbf{A}_0\mathbf{a}_i}{NT} + O_p\left[\max\left(N^{-1/2}, \sqrt{N}T^{-1}\right)\right].$$

Lastly, given that the first term above is non-negative and that $\mathbf{M}_{\hat{\mathbf{F}}} = \mathbf{I}_T - \mathbf{P}_{\hat{\mathbf{F}}}$, we can conclude

$$N\hat{\sigma}_i^2 \leq \frac{\mathbf{a}_i'\mathbf{A}_0'\mathbf{V}'\mathbf{V}\mathbf{A}_0\mathbf{a}_i}{NT} + O_p\left[\max\left(N^{-1/2}, \sqrt{N}T^{-1}\right)\right].$$

■

Back to main slides

MC results with weakly pervasive units

Table 10: Empirical frequency of correctly identifying only the true weakly pervasive units ($m_0 = 1$, and $\alpha = 0.8$)

	SMT-σ^2				PS				BM			
	$k_0 = 0$				$k_0 = 0$				$k_0 = 0$			
N\T	60	110	210	250	60	110	210	250	60	110	210	250
50	51.2	80.6	95.4	97.5	0	1.3	12.4	16.5	21.9	72.0	93.3	95.2
100	87.2	98.9	100	100	0	0.1	11.0	21.7	n/a	37.5	99.3	100
200	97.5	100	100	100	0	0.1	5.2	19.4	n/a	n/a	52.5	99.6
500	97.7	100	100	100	0	0	9.7	35.9	n/a	n/a	n/a	n/a
	$k_0 = 1$				$k_0 = 1$				$k_0 = 1$			
N\T	60	110	210	250	60	110	210	250	60	110	210	250
50	37.4	65.2	79.1	78.9	0.05	0	0	0	19.0	69.7	92.9	93.9
100	65.1	90.5	93.7	93.5	0	0	0	0	n/a	34.0	99.0	99.9
200	84.6	99.4	99.6	99.5	0	0	0	0	n/a	n/a	51.8	99.2
500	82.7	99.9	100	100	0	0	0	0	n/a	n/a	n/a	n/a
	$k_0 = 2$				$k_0 = 2$				$k_0 = 2$			
N\T	60	110	210	250	60	110	210	250	60	110	210	250
50	37.7	53.3	58.2	54.5	0	0	0	0	19.4	68.8	90.9	94.1
100	64.2	79.7	75.3	74.4	0	0	0	0	n/a	33.5	99.4	99.8
200	82.7	98.2	97.1	96.0	0	0	0	0	n/a	n/a	51.7	99.1
500	80.9	100	100	100	0	0	0	0	n/a	n/a	n/a	n/a

Notes: See the notes to Table 2.

Table 11: Empirical frequency of correctly identifying only the true weakly pervasive units ($m_0 = 2$, and $\alpha = 0.8$)

	SMT-σ^2				PS				BM				
	$k_0 = 0$					$k_0 = 0$					$k_0 = 0$			

N\T	60	110	210	250	N\T	60	110	210	250	N\T	60	110	210	250
	$k_0 = 0$					$k_0 = 0$					$k_0 = 0$			
50	6.6	31.6	63.7	67.3	50	0.1	0.9	1.4	1.5	50	6.2	48.7	78.7	82.1
100	13.7	57.4	89.2	92.6	100	0	0.1	2.4	3.4	100	n/a	13.0	94.5	98.1
200	7.7	48.2	88.1	92.0	200	0	0.1	1.7	2.7	200	n/a	n/a	23.7	94.3
500	0.9	23.0	71.0	79.3	500	0	0	1.6	2.9	500	n/a	n/a	n/a	n/a
	$k_0 = 1$					$k_0 = 1$					$k_0 = 1$			
N\T	60	110	210	250	N\T	60	110	210	250	N\T	60	110	210	250
50	8.9	32.0	60.6	63.7	50	0	0	0	0	50	5.7	47.8	74.4	80.2
100	16.5	61.3	88.8	91.5	100	0	0	0	0	100	n/a	12.1	94.5	96.4
200	11.4	61.8	94.7	97.1	200	0	0	0	0	200	n/a	n/a	20.3	92.5
500	1.8	32.3	84.6	91.7	500	0	0	0	0	500	n/a	n/a	n/a	n/a
	$k_0 = 2$					$k_0 = 2$					$k_0 = 2$			
N\T	60	110	210	250	N\T	60	110	210	250	N\T	60	110	210	250
50	13.3	33.5	50.1	50.4	50	0	0	0	0	50	5.1	44.8	72.6	77.2
100	26.6	65.3	79.3	80.2	100	0	0	0	0	100	n/a	11.8	92.5	96.0
200	17.6	75.7	96.1	96.6	200	0	0	0	0	200	n/a	n/a	19.5	91.4
500	2.4	36.5	89.2	93.8	500	0	0	0	0	500	n/a	n/a	n/a	n/a

Notes: See the notes to Table 2.

Table 12: Average number of non-pervasive units falsely selected as pervasive units ($m_0 = 1$, and $\alpha = 0.8$)

SMT-σ^2

$k_0 = 0$

N\T	60	110	210	250
50	0	0	0	0
100	0	0	0	0
200	0	0	0	0
500	0	0	0	0

$k_0 = 1$

N\T	60	110	210	250
50	0.2	0.2	0.2	0.2
100	0.1	0.1	0.1	0.1
200	0	0	0	0
500	0	0	0	0

$k_0 = 2$

N\T	60	110	210	250
50	0.5	0.4	0.5	0.5
100	0.2	0.2	0.3	0.3
200	0	0	0	0
500	0	0	0	0

PS

$k_0 = 0$

N\T	60	110	210	250
50	3.7	3.4	2.2	2.0
100	8.2	7.6	2.6	1.8
200	16.8	10.6	3.5	1.9
500	41.6	22.3	2.8	1.2

$k_0 = 1$

N\T	60	110	210	250
50	2.5	3.4	4.2	4.3
100	3.6	4.9	6.8	7.1
200	6.4	10.9	14.2	15.0
500	0.1	29.0	39.4	40.3

$k_0 = 2$

N\T	60	110	210	250
50	2.7	4.0	4.6	4.6
100	3.8	5.3	7.6	8.0
200	7.8	12.8	15.3	15.7
500	0	35.6	40.5	40.6

BM

$k_0 = 0$

N\T	60	110	210	250
50	2.8	0.7	0.2	0.1
100	n/a	2.0	0	0
200	n/a	n/a	1.3	0
500	n/a	n/a	n/a	n/a

$k_0 = 1$

N\T	60	110	210	250
50	2.9	0.8	0.2	0.1
100	n/a	2.3	0	0
200	n/a	n/a	1.4	0
500	n/a	n/a	n/a	n/a

$k_0 = 2$

N\T	60	110	210	250
50	3.0	0.8	0.2	0.1
100	n/a	2.2	0	0
200	n/a	n/a	1.3	0
500	n/a	n/a	n/a	n/a

Notes: See the notes to Table 2.

Table 13: Average number of non-pervasive units falsely selected as pervasive units ($m_0 = 2$, and $\alpha = 0.8$)

	SMT-σ^2				PS				BM			
	$k_0 = 0$				$k_0 = 0$				$k_0 = 0$			
N\T	60	110	210	250	60	110	210	250	60	110	210	250
50	0.1	0.1	0	0	3.7	2.8	2.8	2.9	2.1	0.3	0	0
100	0	0	0	0	7.0	4.5	1.4	1.1	n/a	1.3	0	0
200	0	0	0	0	12.6	5.7	2.1	1.3	n/a	n/a	0.6	0
500	0	0	0	0	32.9	11.0	2.1	1.2	n/a	n/a	n/a	n/a
	$k_0 = 1$				$k_0 = 1$				$k_0 = 1$			
N\T	60	110	210	250	60	110	210	250	60	110	210	250
50	0.2	0.2	0.2	0.2	4.0	4.6	4.8	4.7	2.2	0.4	0.1	0
100	0.2	0.1	0.1	0.1	6.2	7.0	7.3	7.2	n/a	1.4	0	0
200	0	0	0	0	12.1	14.4	14.9	14.9	n/a	n/a	0.8	0
500	0	0	0	0	0.1	36.8	37.6	37.5	n/a	n/a	n/a	n/a
	$k_0 = 2$				$k_0 = 2$				$k_0 = 2$			
N\T	60	110	210	250	60	110	210	250	60	110	210	250
50	0.5	0.4	0.4	0.5	3.9	4.2	4.2	4.2	2.1	0.3	0.1	0
100	0.3	0.2	0.2	0.2	6.1	7.0	8.0	8.2	n/a	1.4	0	0
200	0.1	0	0	0	11.4	14.8	15.2	15.3	n/a	n/a	0.8	0
500	0	0	0	0	0	37.0	38.0	38.1	n/a	n/a	n/a	n/a

Notes: See the notes to Table 2.

back to main MC results