Detection of Units with Pervasive Effects in Large Panel Data Models*

G. Kapetanios
King's College, London

M. H. Pesaran
University of Southern California, and Trinity College, Cambridge

S. Reese
Lund University, Sweden

December 29, 2019

Abstract

The importance of units that influence a large number of other units in a network has become increasingly recognized in the literature. In this paper we propose a new method to detect such pervasive units by basing our analysis on unit-specific residual error variances subject to suitable adjustments due to the multiple testing issues involved. Accordingly, a sequential multiple testing (SMT) procedure is proposed, which allows identification of pervasive units (if any) without *a priori* knowledge of the interconnections amongst cross-section units or availability of a short list of candidate units to search over. The proposed method is applicable even if the cross section dimension exceeds the time series dimension, and most importantly it could end up with none of the units selected as pervasive when this is in fact the case. The SMT procedure exhibits satisfactory small-sample performance in Monte Carlo simulations and compares well relative to existing approaches. We apply the SMT detection method to sectoral indices of U.S. industrial production, U.S. house price changes by states, and the rates of change of real GDP and real equity prices across the world's largest economies.

JEL Classifications: C18, C23, C55
Key Words: Pervasive units, factor models, systemic risk, multiple testing, sequential procedure, cross-sectional dependence.

*We are grateful to two anonymous referees and an Associate Editor for constructive comments and suggestions. We would also like to thank Natalia Bailey, Alexander Chudik, Ron Smith, Vanessa Smith, Donggyu Sul, Takashi Yamagata and Cynthia Yang for helpful comments on a preliminary version of the paper. Simon Reese was a Postdoctoral Scholar Research Associate at USC Dornsife INET while working on this paper.

1 Introduction

Detecting economic agents with influence over a large number of peers has become a relevant issue in several areas of economics. For example, in banking and finance it is of interest to consider formal ways of identifying whether particular financial institutions present systemic risks. At times of economic and financial crises it is often of interest to know if a certain corporation, particularly among financial institutions, is so large and interconnected that its failure could lead to cascade effects with important consequences for the economy as a whole. Such units are often referred to as 'too big to fail' and their existence is debated in the press and in public policy forums, although empirical evidence in support of such claims is often lacking. In cases where information on interconnections across units exist, it is possible to use a network approach to detect the most influential unit in the network and examine its degree of dominance. An important example is input-output data used to analyze the role that individual production units, such as industrial sectors, play in propagating shocks across the economy. A major recent contribution in this area is by Acemoglu, Carvalho, Ozdaglar, and Tahbaz-Salehi (2012) who suggest using the shape parameter of a power law assumed for the degree sequence of a network to measure the extent to which variations in aggregate volatility are affected by shocks to individual units within the network. Further developments are provided by Acemoglu, Akcigit, and Kerr (2016) and Acemoglu, Autor, Dorn, Hanson, and Price (2016). In related work, Pesaran and Yang (2019) propose extremum estimators based on outdegrees of a network to detect and identify influential units in the network and to estimate their degrees of pervasiveness.

In cases where information on network connections is not available, it is still possible to identify individual units with pervasive effects if there is a sufficient number of time series observations (T) on all cross section units (N) under consideration. In this paper we suppose that such time series observations are available and address the problem of jointly determining the number as well as the identity of those units that are *influential* or *pervasive,* in the sense that they influence almost all other units in the panel. From the perspective of economic networks, the central hub in a star network provides a simple example of a pervasive unit. As noted above, the concept of pervasive units is closely related to the notion of 'too big to fail' often used in the context of financial and production networks. However, it is important to bear in mind that the two concepts are not identical. For example, a unit that is too big to fail may become influential mainly in crisis periods, implying a nonlinear behavior that our linear model may not be best equipped to handle.

Our approach shares some features with existing contributions on the same subject (see e.g. Bai and Ng, 2006; Parker and Sul, 2016; Brownlees and Mesters, 2019) but improves on them in a number of respects. First, we allow for the possibility that the data under consideration do not include any pervasive unit in the first place. This is a leading case of interest and, in fact, some of our empirical applications confirm its practical importance. Secondly, we do not require *a priori* information on a potential list of pervasive units or observations on network linkages. This is a key advantage relative to contributions in the production network literature which relies on the availability of input-output tables. Third, our detection procedure can determine pervasive units from a large number of potential candidates, even in the presence of external common factors that could potentially influence all units (including the pervasive units). Finally, our procedure applies even if $N > T$, which is an important consideration in practice where in many applications of interest the number of time series observations is limited either because of unavailability of data or due to structural breaks.

Before proceeding to propose an operational procedure to identify pervasive units, we need to provide a clear mathematical definition of what we mean by pervasive units, using both intuitive and mathematical arguments. As a result, we consider intuitive properties that a pervasive unit should have. As we wish to have a simple structure we choose not to focus on dynamic models that would potentially allow consideration of various concepts of causation. Accordingly, using a standard multi-factor panel data model, we regard a unit as pervasive if it affects a large proportion of other units in the panel. In other words, any shock that impacts a pervasive unit has to impact a large proportion of other units. In contrast, for non-pervasive units there can be shocks that are idiosyncratic and do not affect many other units. Although we do not allow for dynamics, our model can be extended to allow for shocks to be serially correlated.

A major implication of the existence of pervasive units, as defined above, is that the data can be represented by a factor model where variation in the pervasive units is perfectly explained by the true factors. This view on pervasive units reflects the fact that an influential unit can be viewed as a common factor for all other units in the panel. Consequently, factor estimates obtained from the dataset will have close to perfect explanatory power for true pervasive units. Using this result, we consider the residual variance from regressions of individual units on the factor estimates as a metric that quantifies the explanatory ability of the estimated factors. Based on ideas from multiple testing we then construct thresholds that determine whether the residual variance estimated for a given unit is sufficiently small to identify that unit as pervasive. We find that thresholding residual variances across the units provides a powerful approach with a number of desirable characteristics and good small sample performance.

A further defining characteristic of our work is to consider refinements that again make use of multiple testing to allow for the possibility that identified pervasive units may not be fully pervasive - that is they may only affect a subset of cross-sectional units. This further distinguishes our work from existing methods which either do not pay much attention to such weak cross-sectional dependence structures or are unclear about the motivation and nature of these structures. The use of multiple testing focuses on the possibility that some units selected as pervasive might only affect a majority of the units in the panel rather than being fully pervasive with non-zero effects on all units. We feel that local to zero representations of factor loadings, which are sometimes used in the literature, where the magnitude of the loadings depend on the sample size and tend to zero with the sample size, are less persuasive as a model for economic interdependence than the weak dependence formulation that we consider in this paper.

Monte Carlo simulations suggest that our refined thresholding method performs very well in finite-sample, and most importantly, it reliably detects the absence of pervasive units from a dataset with many potential candidates. Furthermore, if influential units are part of the model specification, our detection methodology succeeds in jointly detecting their total number and their identities. The proposed method also works well even if N is much larger than T, and unlike other methods proposed in the literature, its false discovery rate is very low and tends to zero as N and $T \to \infty$.

The proposed detection procedure is applied to sectoral indices of U.S. industrial production (already investigated in the literature), as well as to the rates of change of real GDP and real equity prices across the world's largest economies over the period 1979q2-2016q4. Unlike other detection methods proposed in the literature, we do not find convincing evidence that there are pervasive sectors within the U.S. industrial production, or that there exist pervasive economies or equity markets in the global system, once we adequately allow for the presence of common factors. Finally, we apply the new method to real U.S. house price changes across the 48 mainland states, and find evidence that New York is pervasive, in contrast to the other methods that select states such as New Hampshire, Nevada, North Carolina, Maryland and Virginia (just to mention a few) and not New York as pervasive.

The paper is structured as follows. Section 2 presents a review of the existing literature. Section 3 provides the main setup of our approach and details our theoretical results. Further refinements are discussed in Section 5. Sections 6 and 7 present simulation and empirical evidence on the relative performance of our method compared to existing ones. Formal proofs and additional simulation results are relegated to Appendix A and an online supplement.

Notation: Generic positive finite constants are denoted by C when large, and c when small. They can take different values at different instances. $\rightarrow^p$ denotes convergence in probability as $N, T \rightarrow \infty$. $\lambda_{\max}(\boldsymbol{A})$ and $\lambda_{\min}(\boldsymbol{A})$ denote the maximum and minimum eigenvalues of matrix $\boldsymbol{A}$. $\boldsymbol{A} > 0$ denotes that $\boldsymbol{A}$ is a positive definite matrix. $\|\boldsymbol{A}\|$ and $\|\boldsymbol{A}\|_F$ denote the spectral and Frobenius norm of matrix $\boldsymbol{A}$. If $\{f_n\}_{n=1}^{\infty}$ is any real sequence and $\{g_n\}_{n=1}^{\infty}$ is a sequences of positive real numbers, then $f_n = O(g_n)$, if there exists C such that $|f_n|/g_n \leq C$ for all n. $f_n = o(g_n)$ if $f_n/g_n \rightarrow 0$ as $n \rightarrow \infty$. If $\{f_n\}_{n=1}^{\infty}$ and $\{g_n\}_{n=1}^{\infty}$ are both positive sequences of real numbers, then $f_n = \ominus(g_n)$ if there exists $n_0 \geq 1$ and positive finite constants C_0 and C_1, such that $\inf_{n \geq n_0}(f_n/g_n) \geq C_0$, and $\sup_{n \geq n_0}(f_n/g_n) \leq C_1$.

2 Related literature

Asset pricing models have motivated the earliest approaches aimed at determining whether a given set of observed time series coincides with one of the estimated common factors (principal components) from a large panel dataset. Bai and Ng (2006) regress each observed candidate series onto the estimated factors and propose statistics to test the equality between the model fit from the aforementioned regression and the observed values of a list of (assumed) potential influential variables. The framework considered by these authors is one where economic theory reduces the number of potential influential variables to a small, fixed number of economic indicators that are not part of the large dataset at hand. Consequently, using their framework to identify pervasive units in large datasets without any means of reducing the number of candidates is problematic. The framework considered by these authors is one where economic theory reduces the number of potentially influential variables to a small set of economic indicators that are not part of the large dataset at hand. Consequently, using their framework to identify pervasive units without some *a priori* knowledge about the list of candidate variables will be problematic. This shortcoming was recognized by Parker and Sul (2016) who provide an alternative approach to that suggested by Bai and Ng (2006) and consider the identification of pervasive units by focusing on the idiosyncratic components of the estimated factor model and identify an observed series as pervasive if it can replace at least one of the estimated factors in the factor model without introducing common factors in the idiosyncratic components.[1] [2]

[1] In Parker and Sul (2016) a pervasive unit is referred to as the *dominant leader*.

[2] Further empirical applications of the Parker and Sul method (in a simpler form) are provided by Gaibulloev, Sandler, and Sul (2013) and Greenaway-McGrevy, Mark, Sul, and Wu (2018). Soofi-Siavash (2018) also considers a version of the Parker and Sul method which is applicable to any cross-section unit taken as potentially pervasive, and provides an application to the industrial sectors in the U.S..

In order to address multiple testing concerns, Parker and Sul (2016) propose a rule of thumb to restrict the number of potential pervasive units. However, this only mitigates the problem rather than providing a full solution. A more general solution is provided by Brownlees and Mesters (2019), who formalize the notion of a pervasive unit in a framework very similar to the one used in this study. Their approach consists of using the sample concentration matrix of all the units in the data, and identify as pervasive those units whose concentration matrix column norms are considerably larger than those of the remaining units.[3] Under certain regularity conditions, Brownlees and Mesters show that their procedure consistently partitions the units into pervasive and non-pervasive by ordering the column norms in descending order and then choosing the maximum ratio between two successive, ordered column norms.

The detection method proposed by Brownlees and Mesters (2019) has the advantage that it does not require common external factors in the data to be estimated. But this comes at the expense of requiring the number of time periods to be larger than the number of cross-section units ($T > N$), and the assumption that there exists at least one pervasive unit in the data. These two requirements could be quite restrictive in practice. First, many datasets, notably those involving aggregate economic indicators, have a number of cross section units that is approximately as large as the number of time periods, if not larger. Even if the time dimension of the dataset is sufficiently large, sub-samples of interest (due to structural breaks) might be too short to allow for a separate investigation. Second, it is crucial to allow for the possibility that none of the units in the sample at hand is pervasive. The relevance of this case is apparent by recent contributions that track the effect of sector-specific shocks on aggregate fluctuations. For example, application of a structural model to data on U.S. industrial production leads Foerster, Sarte, and Watson (2011, p.21) to conclude that "[...] linkages alone and uncorrelated sector-specific shocks implies noticeably less co-movement across sectors than in U.S. data." Further evidence is given in Pesaran and Yang (2019) who develop an estimator for the degree of dominance of the most pervasive unit in a network. Their application to U.S. input-output tables reveals that there is "[...] some evidence of sector-specific shock propagation, but [that] such effects do not seem sufficiently strong and long-lasting [...]" in the sense that the aggregate effect of sectoral shocks vanishes as the number of sectors in the economy increases. Finally, while the two studies cited above allow for the absence of pervasive units, they crucially rely on the availability of input-output matrices as a measure of linkages between cross-sections. Comparable information may not always be available, thus making it impossible to use the techniques in these studies. By contrast, the approach proposed in the current paper is applicable to any large dimensional panels without requiring the presence of a minimum number of pervasive units in the panel.

3 Panel data models with pervasive units

Suppose T time series observations are available on N cross section units denoted by x_{it}, for $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$. We are interested in determining the number and identity of pervasive units (if any), in this panel. To define the concept of a pervasive unit, we propose a mathematical formalization of our intuitive idea that pervasive units are those for which any shock that impacts them also has to impact a large proportion of other units. In its most general form, the idea can be formalized in terms of conditional probability distributions

[3] Brownlees and Mesters (2019) employ the term *granular unit* to denote a pervasive unit.

where conditioning is on general σ-fields that represent information sets. Let $(\Omega, \mathcal{F}, P)$ be some probability space that is rich enough for modeling x_{it}, Let $\mathcal{G}_1 \subset \sigma(x_{it}) \subset \mathcal{F}$ and $\mathcal{G}_2 \subset \sigma(x_{it}) \subset \mathcal{F}$, for some i, be some σ-fields and, assuming stationarity for x_{it}, let $F_i(x|.)$ denote the conditional distribution of x_{it}, which, we assume, exists. Then, a unit i is pervasive if for all possible $\mathcal{G}_1 \subset \sigma(x_{it})$ and $\mathcal{G}_2 \subset \sigma(x_{it})$, such that $E_P\left[\sup_x |F_i(x|\mathcal{G}_1) - F_i(x|\mathcal{G}_2)|\right] > 0$, we have that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{j=1, j\neq i}^{N} E_P\left[\sup_x |F_j(x|\mathcal{G}_1) - F_j(x|\mathcal{G}_2)|\right] = c,$$

for some $0 < c < \infty$, where $E_P[.]$ denotes expectation with respect to P. If there exist $\mathcal{G}_1 \subset \sigma(x_{it})$ and $\mathcal{G}_2 \subset \sigma(x_{it})$ such that $E_P\left[\sup_x |F_i(x|\mathcal{G}_1) - F_i(x|\mathcal{G}_2)|\right] > 0$, but

$$\lim_{N\to\infty} \frac{1}{N} \sum_{j=1, j\neq i}^{N} E_P\left[\sup_x |F_j(x|\mathcal{G}_1) - F_j(x|\mathcal{G}_2)|\right] = 0,$$

the unit is not pervasive. This definition becomes clearer if we specialize to the case of conditional expectations, on which we will focus from now, and adopt the following concept of pervasiveness.

Definition 1 *A unit i is pervasive if for all possible $\mathcal{G}_1 \subset \sigma(x_{it})$ and $\mathcal{G}_2 \subset \sigma(x_{it})$, such that $E_P\left[|E(x_{it}|\mathcal{G}_1) - E(x_{it}|\mathcal{G}_2)|\right] > 0$, we have that*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{j=1, j\neq i}^{N} E_P\left[|E(x_{jt}|\mathcal{G}_1) - E(x_{jt}|\mathcal{G}_2)|\right] = c, \tag{1}$$

for some $0 < c < \infty$.

This definition basically states that if there are no shocks that affect x_{it} but do not affect a non-zero proportion of the other units then x_{it} is pervasive. Later on we shall see that concepts of weak pervasiveness (or dominance) can be accommodated by, for example, stating that there is $0 < \alpha < 1$, such that a unit is weakly pervasive if, for all possible $\mathcal{G}_1 \subset \sigma(x_{it})$ and $\mathcal{G}_2 \subset \sigma(x_{it})$, such that $E_P\left[|E(x_{it}|\mathcal{G}_1) - E(x_{it}|\mathcal{G}_2)|\right] > 0$, we have that $\lim_{N\to\infty} \frac{1}{N^\alpha} \sum_{j=1, j\neq i}^{N} E_P\left[|E(x_{jt}|\mathcal{G}_1) - E(x_{jt}|\mathcal{G}_2)|\right] = c$, for some $0 < c < \infty$, and $\alpha < 1$.

While this definition is quite general and model free it is not that useful for operationalizing a procedure that detects the number and identity of pervasive units. Therefore, we proceed by specifying that all cross sectional units can be modeled using unobserved common factors. More formally, we consider a data-generating process similar to that used in Brownlees and Mesters (2019), namely

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

$$x_{it} = \boldsymbol{\lambda}_i' \mathbf{g}_t + \sum_{j=1}^{m} b_{ij} x_{jt} + u_{it}, \ \ i = m+1, m+2, ..., N. \tag{3}$$

The non-pervasive units (namely units $i = m+1, m+2, ..., N$) are affected by k common external factors, $\mathbf{g}_t$, and the common innovations of the m pervasive units (namely units $i = 1, 2, ..., m$):

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

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

$$\mathbf{x}_{at} = \boldsymbol{\Lambda}_a \mathbf{g}_t + \mathbf{u}_{at}, \tag{4}$$
$$\mathbf{x}_{bt} = \boldsymbol{\Lambda}_b \mathbf{g}_t + \mathbf{B}\mathbf{x}_{at} + \mathbf{u}_{bt}, \tag{5}$$

for $t = 1, 2, \ldots, T$, where $\mathbf{x}_{at}$ and $\mathbf{x}_{bt}$ are $m \times 1$ and $(N-m) \times 1$ vectors of observations at time t on the pervasive and non-pervasive units, respectively. In what follows we set $n = N - m$. Only the $N \times 1$ vector $\mathbf{x}_t = (\mathbf{x}'_{at};\ \mathbf{x}'_{bt})'$ is observed to the researcher and the true number of pervasive units m as well as their identities are unknown. The partitioning of $\mathbf{x}_t$ into m pervasive units, followed by n non-pervasive units is made exclusively for expositional purposes and in general there is no *a priori* information about the cross-section indexes of potential pervasive cross-sections.

The m pervasive units, $x_{a,jt}$, $j = 1, 2, ..., m$ affect the non-pervasive units, $x_{b,it}$, $i = m+1, m+2, ..., N$ *via* the $n \times m$ matrix of loading coefficients $\mathbf{B} = (b_{ij})$, where $\sup_{ij} |b_{ij}| < K < \infty$. For $x_{a,jt}$ to be a pervasive unit we must have

$$|b_{ij}| > c > 0, \text{ for } i = m+1, m+2, ..., N \tag{6}$$

or, asymptotically equivalently,[4]

$$\sum_{i=m+1}^{N} |b_{ij}| = \ominus(n)\,,\ j = 1, 2, ..., m. \tag{7}$$

In other words, for a unit to be pervasive it must have non-zero effects on almost all other units in the panel or network. The following proposition, established in Section S1 of the online supplement, formalizes the connection between the operational definition of pervasiveness, based on the above factor model and the more primitive Definition 1 in the slightly simplified case where $\boldsymbol{\lambda}_i = \mathbf{0}$, for all i. We impose this condition, which rules out certain pathological cases that will be dealt with more formally by Assumption 2 below, so as to keep this preliminary exposition simple. For simplicity we also assume that u_{it} are independently distributed across i. Limited dependence across i does not affect the results but complicates the exposition.

Proposition 1 *Let the model for x_{it}, $i = 1, ..., N$, $t = 1, ..., T$ be given by (2) and (3), where $\boldsymbol{\lambda}_i = 0$, $i = 1, 2, ..., N$, and u_{it} are independent across i. Let (6) hold for some unit i. Then, unit i is pervasive according to Definition 1.*

Following Chudik, Pesaran, and Tosetti (2011), we can also consider units that are not pervasive but still quite influential. Suppose that there exists an ordering of the non-pervasive units such that unit $x_{a,jt}$ *only* affects the non-pervasive units, $x_{b,it}$, whose index $i \leq \lfloor n^{\alpha_j} \rfloor$, where α_j $(0 < \alpha_j \leq 1)$ is an exponent parameter that measures the degree of pervasiveness of $x_{a,jt}$ in the panel.[5] This requirement can be written as

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

[4] Strictly speaking, we can allow a *finite* number of loadings to be zero.

[5] $\lfloor a \rfloor$ denotes the integer part of a.

or, equivalently,as[6]

$$\sum_{i=m+1}^{N} |b_{ij}| = \ominus\left(n^{\alpha_j}\right), \text{ for } j = 1, 2, ..., m, \tag{8}$$

which is a natural generalization of (7). The unit $x_{a,jt}$ with $\alpha_j < 1$ can be viewed as a weak factor, but as argued in Bailey, Kapetanios, and Pesaran (2016) and Bailey, Kapetanios, and Pesaran (2019), for $x_{a,jt}$ to have pervasive effects on other units we need α_j to be reasonably close to unity. Clearly, the values of $\alpha_j \leq 1/2$ can be ruled out since for such values, $x_{a,jt}$ becomes so weak that it loses many of the standard characteristics, associated with factor variables. In practice, we might need to focus on exponents that fall in the range $2/3 < \alpha_j \leq 1$ before we can be confident that unit $x_{a,jt}$ has non-negligible impacts on other units in the panel dataset. In terms of the general definition (8), for all elements of $\mathbf{x}_{at}$ to be pervasive it is required that $\alpha_j = 1$ for $j = 1, 2, ..., m$, and pervasive units can be regarded as strong factors. While our theory focuses on $\alpha_j = 1$, $j = 1, 2, ..., m$, it can be extended to $\alpha_j \leq 1$, using ideas in the above cited papers. It is also possible to estimate the exponent α_j once the unit $x_{a,jt}$ is selected as pervasive/influential. However, such extensions are beyond the scope of the present paper. The $k \times 1$ vector $\mathbf{g}_t$ contains common "external" factors affecting *both* pervasive and non-pervasive units via the $m \times k$ and the $n \times k$ loading matrices $\mathbf{\Lambda}_a$ and $\mathbf{\Lambda}_b$, respectively. The pervasive units can also be viewed as “internal” factors. Lastly, $\mathbf{u}_{at}$ and $\mathbf{u}_{bt}$ represent the stochastic components of the pervasive and non-pervasive units, respectively. To simplify the exposition we abstract from deterministic effects such as intercepts or linear trends, and without loss of generality assume that x_{it} have zero means and finite variances.

Denoting the common factors (internal and external) by the $p \times 1$ vector $\mathbf{f}_t = (\mathbf{g}_t', \mathbf{u}_{at}',)' = (f_{1t}, f_{2t}, \ldots, f_{pt})'$ where $p = m + k$, then equations (4)-(5) can be written as a restricted static factor model:

$$\begin{aligned} \begin{pmatrix} \mathbf{x}_{at} \\ \mathbf{x}_{bt} \end{pmatrix} &= \begin{pmatrix} \mathbf{A}_a \\ \mathbf{A}_b \end{pmatrix} \mathbf{f}_t + \begin{pmatrix} \mathbf{0} \\ \mathbf{u}_{bt} \end{pmatrix} \\ &= \mathbf{A}\mathbf{f}_t + \mathbf{v}_t, \end{aligned} \tag{9}$$

where $\mathbf{A}_a = (\mathbf{\Lambda}_a, \mathbf{I}_m)$ and $\mathbf{A}_b = (\mathbf{\Lambda}_b + \mathbf{B}\mathbf{\Lambda}_a, \mathbf{B})$. Additionally, denote by $\mathbf{a}_i$ the i-th row of $\mathbf{A} = (\mathbf{A}_a', \mathbf{A}_b')'$. Since a pervasive unit is *de facto* a common factor, then $m \leq p$. It is also shown in Chudik, Pesaran, and Tosetti (2011), that p must be a fixed integer to ensure that $Var(x_{it})$ is bounded in N. Accordingly, we assume that $0 \leq m \leq p < p_{max}$, where p_{max} is an upper bound on p.[7]

We shall also make the following 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 some* $s_f > 0$ *such that*

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

[6] The exponent α_j is defined asymptotically (as $n \to \infty$), and as such can accommodate a finite number of loadings for units $1, 2, ..., n^{\alpha_j}$, to be zero and conversely that a finite number of loadings for units $n^{\alpha_j} + 1, ..., n$, to be non-zero.

[7] The magnitude of m relative to k is immaterial as long as both are fixed.

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

Assumption 2

1. $\mathbf{A}_a$ *and* $\mathbf{A}_b$ *are parameter matrices, satisfying* $Rank\left(\mathbf{A}_a\right) = m$ *and* $Rank\left(\mathbf{A}_b\right) = p$*.*

2. $\inf_i \|\mathbf{a}_i\| > c$*, and* $\sup_i \|\mathbf{a}_i\| < C$*, and for any* $N = n + m$ *(the true number of pervasive units,* m_0*, being finite),*

$$\lambda_{\max}\left(n^{-1}\sum_{i=m+1}^{N}\mathbf{a}_i\mathbf{a}_i'\right) < C < \infty, \tag{10}$$

$$\lambda_{\min}\left(n^{-1}\sum_{i=m+1}^{N}\mathbf{a}_i\mathbf{a}_i'\right) > c > 0. \tag{11}$$

Assumption 3

1. *The* $n \times 1$ *vector* $\mathbf{u}_{bt}$ *is defined by*

$$\mathbf{u}_{bt} = \mathbf{H}\boldsymbol{\varepsilon}_t, \tag{12}$$

where

$$\boldsymbol{\varepsilon}_t = \left(\varepsilon_{m+1,t}, \varepsilon_{m+2,t}, \ldots, \varepsilon_{Nt}\right)' \sim IID\left(\mathbf{0}, \mathbf{I}_n\right), \tag{13}$$

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

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

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

3. $\mathbf{H} = (h_{ij})$ *is an* $n \times n$ *matrix with fixed coefficients, 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 independently distributed for all* i, s*, and* t*.*

Remark 1 *Most of the above assumptions relate closely to those made in the literature on the large dimensional factor models (see Remark 2 below). Restricting the covariance matrix of* $\mathbf{f}_t$ *to be the identity matrix is an innocuous simplification, since the factors are identified only up to a p-dimensional rotation. However, since the methodology proposed in this article goes beyond estimation in a large-dimensional factor model, some of the assumptions made above are slightly stronger than those made in the literature. Covariance stationarity of the common factors is one such restriction but does not rule out conditional heteroskedasticity. Our use of results from the multiple testing literature assumes that the probability distributions of* ε_{it} *and* f_{it} *have exponentially decaying tails. While this assumption is standard in high-dimensional statistics, it implies that all moments of* ε_{it} *and* f_{it} *exist and thus sharpens our assumptions beyond those required for the estimation of unobserved factors. This assumption simplifies the theoretical analysis. It can be relaxed, considerably in the case of* f_{it}*, and replaced with moment* *assumptions, at the cost of more complex proofs. We choose to avoid this complexity as we are mainly focused on suggesting and analyzing a new methodology. Furthermore, to establish consistency of our proposed criterion, we assume ε_{it} to be independently distributed across i and t.*[8] *Still, dependence between the elements of the unit-specific component $\mathbf{u}_{bt}$ is allowed for by Assumption 3 which admits weak cross-section correlation. The rank condition on $\mathbf{A}_a$ and $\mathbf{A}_b$ in Assumption 2 ensures that m is identified, $\mathbf{x}_{at}$ is pervasive, and that there does not exist some linear combination of $\mathbf{f}_t$, of dimension lower than p, that can fully capture the common component of $\mathbf{x}_{bt}$. In particular, it disallows the possibility that $\mathbf{\Lambda}_b+\mathbf{B}\mathbf{\Lambda}_a = \mathbf{0}$, and hence ensures that $\mathbf{g}_t$ acts as external factors for non-pervasive units. Assumption 2 implies strong factors in the sense that the fraction of cross-section units affected is asymptotically non-negligible. This is a standard property of latent factors in the related literature, as is Assumption 4. On this see, for example, Assumptions L and LFE in Bai and Ng (2008).*

Remark 2 *A further consideration concerns how the above assumptions relate to those of the standard factor model literature as set out, for example, in Bai (2003). As noted above our assumptions are stricter, and therefore imply the assumptions made by Bai (2003). In particular, Assumption 1 implies Assumption A of Bai (2003), Assumption 2 implies Assumption B of Bai (2003) and Assumptions 3 and 4 imply Assumptions C, D, E and F1-F2 of Bai (2003) while we note that we have no need for Assumptions F3-F4 of Bai (2003).*

As shown in the next section, it is possible to consistently estimate the parameters of the static factor model (9), even if the variance matrix of the $N \times 1$ vector $\mathbf{v}_t = \left(\mathbf{0}'_{m\times 1}, \mathbf{u}'_{bt}\right)'$, containing the idiosyncratic errors, is singular when $m > 0$. In our theoretical derivations and Monte Carlo simulations we only require that $p_{max} \geq p$ is known and base our analysis on p_{max} principal components of $\{x_{it}$, for $i = 1, 2, \ldots, N,$ and $t = 1, 2, \ldots, T\}$.

4 Identification of pervasive units via thresholding of error variances

The idea behind our detection procedure is simple. It exploits the fact that there is a clear separation between the fit of pervasive and non-pervasive units in terms of the factors, $\mathbf{f}_t$, for sufficiently large sample sizes. In the context of the restricted factor model representation (9), a clear separation between a pervasive and non-pervasive unit could be achieved if the common factors $\mathbf{f}_t$ as well as $\mathbf{x}_t$ are observed. In such a case only pervasive units (with exponent $\alpha = 1$) will be perfectly correlated with $\mathbf{f}_t$. But in practice where only observations on $\mathbf{x}_t$ are available to the econometrician, the fit of cross-section specific observations can only be evaluated in terms of a factor estimate $\hat{\mathbf{f}}_t$. Finite sample errors in the estimation of the true factors entail an imperfect fit, and hence would yield strictly positive residual variances, for all cross-sections (irrespective of whether they are pervasive or not). However, residual variances for non-pervasive units remain bounded away from zero due to their non-degenerate idiosyncratic error, u_{it}, even asymptotically as N and $T \to \infty$. By contrast, corresponding residual variances for pervasive units exclusively contain sampling errors due to the fact that $\mathbf{f}_t$ is replaced by its estimator $\hat{\mathbf{f}}_t$. This latter source of variation vanishes as the sample size increases. In situations

[8] This assumption can be relaxed considerably by requiring ε_{it} to follow a martingale difference process over t, or even to be a strong mixing process with sufficiently small mixing coefficients.

where prior information allows narrowing down the number of potential pervasive units to a small set, it would generally be possible to exploit the behavior of residual variances to develop a statistical test for the absence of idiosyncratic variation in a given cross-section. However, in the context of this study, *any* cross-section unit is considered to be potentially pervasive. This amounts to a high-dimensional statistical problem for which standard testing approaches are ill-suited due to the inherent difficulty of controlling size. For this reason, our detection procedure is based on classification of cross-sections by means of a threshold rule applied to residual variances.

With this in mind we first extract the first $p_{\max}$ principle components (PC) of the observations x_{it} for $i = 1, 2, \ldots, N;\ t = 1, 2, \ldots, T$. We then compute the residual sum of squares from the regressions of x_{it}, for $i = 1, 2, \ldots, N$ on $\hat{\mathbf{f}}_t$, where $\hat{\mathbf{f}}_t$ is the PC estimator of $\mathbf{f}_t$ with $p = p_{\max}$. Specifically, we compute

$$\hat{\sigma}_{iT}^2 = \frac{\mathbf{x}_i' \mathbf{M}_{\hat{\mathbf{F}}} \mathbf{x}_i}{T}, \text{ for } i = 1, 2, \ldots, N, \tag{14}$$

where $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{iT})'$,

$$\mathbf{M}_{\hat{\mathbf{F}}} = \mathbf{I}_T - \hat{\mathbf{F}} \left(\hat{\mathbf{F}}' \hat{\mathbf{F}} \right)^{-1} \hat{\mathbf{F}}', \tag{15}$$

and $\hat{\mathbf{F}} = (\hat{\mathbf{f}}_1, \hat{\mathbf{f}}_2 \ldots, \hat{\mathbf{f}}_T)'$. 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 as pervasive. Below we proceed by analyzing the asymptotic properties of $\hat{\sigma}_{iT}^2$, and sketching steps that lead to a procedure that consistently, namely with probability tending to one, selects only pervasive units. Formal proofs are given in the appendix. But first we provide an overview of the literature on estimation of $\mathbf{F}$ and $\mathbf{A}$ and derive asymptotic properties for functions of their estimators that are needed to establish our main theoretical results.

4.1 Consistent estimation of A and F by principal components

Let $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$ be the $T \times N$ matrix of observations on x_{it} for $i = 1, 2, \ldots, N; t = 1, 2, \ldots, T$. It is well known that the $T \times T$ matrix $\mathbf{X}\mathbf{X}'$ and the $N \times N$ matrix $\mathbf{X}'\mathbf{X}$ have the same eigenvalues. Denote the first p largest eigenvalues of these two matrices by $(\hat{\rho}_1, \hat{\rho}_2, \ldots, \hat{\rho}_p)$, and the associated orthonormal eigenvectors of $\mathbf{X}\mathbf{X}'$ and $\mathbf{X}'\mathbf{X}$ by the $T \times p$ matrix $\hat{\mathbf{P}}$ and the $N \times p$ matrix $\hat{\mathbf{Q}}$, respectively. Also note that by construction $\hat{\mathbf{P}}'\hat{\mathbf{P}} = \mathbf{I}_p$, and $\hat{\mathbf{Q}}'\hat{\mathbf{Q}} = \mathbf{I}_p$, where $\mathbf{I}_p$ is an $p \times p$ identity matrix. We will express the central results in this section in terms of the PC estimators

$$\hat{\mathbf{A}} = \sqrt{N} \hat{\mathbf{Q}}, \tag{16}$$

$$\hat{\mathbf{F}} = \frac{1}{\sqrt{N}} \mathbf{X} \hat{\mathbf{Q}} = \frac{1}{N} \mathbf{X} \hat{\mathbf{A}}, \tag{17}$$

which satisfy $N^{-1}\hat{\mathbf{A}}'\hat{\mathbf{A}} = \mathbf{I}_p$ and $T^{-1}\hat{\mathbf{F}}'\hat{\mathbf{F}} = \hat{\mathbf{D}}_{NT}$, with $\hat{\mathbf{D}}_{NT} = (NT)^{-1} Diag\left(\hat{\rho}_1, \hat{\rho}_2, \ldots, \hat{\rho}_p\right)$.[9] We shall make use of a number of existing results from the literature on large-dimensional factor models that are expressed in terms of the alternative estimators

$$\tilde{\mathbf{F}} = \sqrt{T} \hat{\mathbf{P}}, \tag{18}$$

$$\tilde{\mathbf{A}} = T^{-1} \mathbf{X}' \tilde{\mathbf{F}}, \tag{19}$$

[9] For example, see Bai and Ng (2008, equation (3.1)).

where (18) satisfies $T^{-1}\tilde{\mathbf{F}}'\tilde{\mathbf{F}} = \mathbf{I}_p$. However, the two pairs of estimators $\left(\hat{\mathbf{A}}, \hat{\mathbf{F}}\right)$ and $\left(\tilde{\mathbf{A}}, \tilde{\mathbf{F}}\right)$ are related via the equalities

$$\hat{\mathbf{F}} = \tilde{\mathbf{F}}\hat{\mathbf{D}}_{NT}^{1/2}, \text{ and } \hat{\mathbf{A}} = \tilde{\mathbf{A}}\hat{\mathbf{D}}_{NT}^{-1/2}, \tag{20}$$

as shown by Bai and Ng (2008, equation (3.2)). Consequently, it follows that

$$\mathbf{M}_{\hat{\mathbf{F}}} = \mathbf{I}_T - \hat{\mathbf{F}}\left(\hat{\mathbf{F}}'\hat{\mathbf{F}}\right)^{-1}\hat{\mathbf{F}} = \mathbf{I}_T - \tilde{\mathbf{F}}\left(\tilde{\mathbf{F}}'\tilde{\mathbf{F}}\right)^{-1}\tilde{\mathbf{F}}' = \mathbf{M}_{\tilde{\mathbf{F}}}, \tag{21}$$

and hence

$$\hat{\sigma}_{iT}^2 = \frac{\mathbf{x}_i'\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{x}_i}{T} = \frac{\mathbf{x}_i'\mathbf{M}_{\tilde{\mathbf{F}}}\mathbf{x}_i}{T}.$$

Furthermore, by Lemma A.3 in Bai (2003), $\hat{\mathbf{D}}_{NT} \xrightarrow{p} \mathbf{D}$, where $\mathbf{D}$ is a diagonal matrix with finite elements (see the proof of Lemma A.1 in Bai (2003)). It follows that

$$\left\|\hat{\mathbf{D}}_{NT}\right\|_F = O_p(1). \tag{22}$$

This latter result, together with the equalities in equation (20), allows us to directly apply existent convergence results on $\tilde{\mathbf{A}}$, $\tilde{\mathbf{F}}$ to our estimators $\hat{\mathbf{A}}$, $\hat{\mathbf{F}}$. For example, Bai and Ng (2002) show in their equation (5) that

$$\frac{1}{T}\left\|\tilde{\mathbf{F}} - \mathbf{F}\mathbf{H}_{NT}\right\|_F^2 = O_p\left(\frac{1}{\delta_{NT}^2}\right), \tag{23}$$

where $\delta_{NT}^2 = \min(N, T)$, and $\mathbf{H}_{NT}$ is a non-singular $p \times p$ matrix that could depend on N and T, so long as its probability limit exists and is non-singular.[10] Using (23) and the fact that $\left\|\mathbf{H}_{NT}^{-1}\right\| = O_p(1)$ holds by the properties of $\mathbf{H}_{NT}$, then

$$\frac{1}{T}\left\|\mathbf{F} - \tilde{\mathbf{F}}\mathbf{Q}_{NT}\right\|_F^2 = O_p\left(\frac{1}{\delta_{NT}^2}\right), \tag{24}$$

where $\mathbf{Q}_{NT} = \mathbf{H}_{NT}^{-1}$, noting that this matrix is non-singular and satisfies $\|\mathbf{Q}_{NT}\|_F = O_p(1)$. Setting $\mathbf{Q}_{NT} = \hat{\mathbf{D}}_{NT}^{1/2}$, and noting the relation between $\tilde{\mathbf{F}}$ and $\hat{\mathbf{F}}$, we also have $T^{-1}\left\|\mathbf{F} - \hat{\mathbf{F}}\right\|_F^2 = O_p\left(\delta_{NT}^{-2}\right)$, and more generally,

$$\frac{1}{T}\left\|\mathbf{F} - \hat{\mathbf{F}}\mathbf{S}_{NT}\right\|_F^2 = O_p\left(\frac{1}{\delta_{NT}^2}\right), \tag{25}$$

for any non-singular $p \times p$ matrix $\mathbf{S}_{NT}$ that satisfies $\|\mathbf{S}_{NT}\|_F = O_p(1)$. Result (23) is well know and plays an important role in our analysis. However, for our analysis we require additional results that, to a large extent, go beyond those existing in the literature. We provide these in the following proposition. To simplify the exposition and without loss of generality, we set $\mathbf{S}_{NT} = \mathbf{I}_p$. Since only the product $\mathbf{F}\mathbf{A}'$ is identified, this restriction is innocuous and implies the normalization $N^{-1}\mathbf{A}'\mathbf{A} = \mathbf{I}_p$.[11]

[10] Matrix $\mathbf{H}_{NT}$ is a $p \times p$ rotation matrix and should not be confused with the $n \times n$ matrix $\mathbf{H}$ defined in Assumption 3.

[11] see also Bai and Ng (2008) and Onatski (2012) who make a similar assumption for $T^{-1}\mathbf{F}'\mathbf{F}$.

Proposition 2 *Under Assumptions 1–4, and setting* $\mathbf{S}_{NT} = \mathbf{I}_p$*, we have*

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

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

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

$$\|\mathbf{V}\mathbf{A}_0\|_F = O_p\left(\sqrt{NT}\right), \tag{D}$$

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

where $\hat{\mathbf{F}}$ *and* $\hat{\mathbf{A}}$ *are defined by (17) and (16), and* $\mathbf{F}_0$ *and* $\mathbf{A}_0$ *denote the true values of* $\mathbf{F}$ *and* $\mathbf{A}$*, and* $\delta_{NT}^2 = \min(N, T)$*.*

The above proposition follows from Lemmas 3-6 set out in Section S1.2 in the online supplement. For general rotation matrices $\mathbf{H}_{NT}, \mathbf{Q}_{NT}$ and $\mathbf{S}_{NT}$, Proposition 2 can be used to obtain

$$\left\|\frac{\mathbf{V}\left(\tilde{\mathbf{A}} - \mathbf{A}_0\mathbf{H}_{NT}^{-1}\right)}{N}\right\|_F = O_p\left(\sqrt{\frac{T}{N}}\frac{1}{\delta_{NT}}\right). \tag{26}$$

The matrix $\mathbf{H}_{NT}$ has been introduced into the expression above in order to ensure compatibility with the results (23) and (25). Again, $\mathbf{V}\left(\tilde{\mathbf{A}} - \mathbf{A}_0\mathbf{H}_{NT}^{-1}\right) = \mathbf{V}\left(\tilde{\mathbf{A}}\mathbf{H}_{NT} - \mathbf{A}_0\right)\mathbf{H}_{NT}^{-1}$, so that

$$\left\|\mathbf{V}\left(\tilde{\mathbf{A}}\mathbf{H}_{NT} - \mathbf{A}_0\right)\right\|_F \leq \|\mathbf{H}_{NT}\|_F \left\|\mathbf{V}\left(\tilde{\mathbf{A}} - \mathbf{A}_0\mathbf{H}_{NT}^{-1}\right)\right\|_F,$$

and letting $\mathbf{H}_{NT} = \hat{\mathbf{D}}_{NT}^{-1/2}\mathbf{S}_{NT}^{-1}$, we have

$$\left\|\mathbf{V}\left(\mathbf{A}_0 - \tilde{\mathbf{A}}\hat{\mathbf{D}}_{NT}^{-1/2}\mathbf{S}_{NT}^{-1}\right)\right\|_F \leq O_p(1)\, O_p\left[\left\|\mathbf{V}\left(\tilde{\mathbf{A}} - \mathbf{A}_0\mathbf{H}_{NT}^{-1}\right)\right\|_F\right].$$

Recall that $\hat{\mathbf{A}} = \tilde{\mathbf{A}}\hat{\mathbf{D}}_{NT}^{-1/2}$ by equation (20). Hence,

$$\begin{aligned}\left\|\mathbf{V}\left(\mathbf{A}_0 - \hat{\mathbf{A}}\mathbf{S}_{NT}^{-1}\right)\right\|_F &= O_p\left[\left\|\mathbf{V}\left(\tilde{\mathbf{A}} - \mathbf{A}_0\hat{\mathbf{D}}_{NT}^{1/2}\mathbf{S}_{NT}\right)\right\|_F\right] \\ &= O_p\left(\frac{\sqrt{NT}}{\delta_{NT}}\right),\end{aligned} \tag{27}$$

by the equality $\mathbf{H}_{NT}^{-1} = \hat{\mathbf{D}}_{NT}^{1/2}\mathbf{S}_{NT}$ and application of result (26). This concludes our discussion of principal components estimators. The preceding results will be extensively used in the next section under the simplifying restrictions $\mathbf{S}_{NT}^{-1} = \mathbf{I}_p$.

4.2 Thresholding of $\hat{\sigma}_{iT}^2$

Equipped with the results of Proposition 2, we consider the asymptotic properties of $\hat{\sigma}_{iT}^2$, defined by (14). Our aim is to develop a threshold C_{NT}^2 such that $N\hat{\sigma}_{iT}^2 \leq C_{NT}^2$ with probability approaching 1 as $N, T \to \infty$ if cross-section i is pervasive, while the reverse is true if cross-section i is non-pervasive. Consider first the case where a given unit i is pervasive. By equation (9), we have $\mathbf{x}_i = \mathbf{F}_0\mathbf{a}_i$, where $\mathbf{F}_0$ is the $T \times m$ matrix of observations on the true factors. Consequently, the sample error variance of unit i, once the effects of estimated factors are filtered out, is given by

$$\begin{aligned} \hat{\sigma}_{iT}^2 &= \frac{\mathbf{a}_i'\mathbf{F}_0'\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{F}_0\mathbf{a}_i}{T}, \\ &= \frac{\mathbf{a}_i'\left(\mathbf{F}_0 - \hat{\mathbf{F}}\right)'\mathbf{M}_{\hat{\mathbf{F}}}\left(\mathbf{F}_0 - \hat{\mathbf{F}}\right)\mathbf{a}_i}{T}. \end{aligned} \tag{28}$$

The last result is obtained noting that $\mathbf{F}_0'\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{F}_0 = \left(\mathbf{F}_0 - \hat{\mathbf{F}}\mathbf{S}_{NT}\right)'\mathbf{M}_{\hat{\mathbf{F}}}\left(\mathbf{F}_0 - \hat{\mathbf{F}}\mathbf{S}_{NT}\right)$, for *any* positive definite matrix, $\mathbf{S}_{NT}$. Now, using (9) and post-multiplying both sides by $\mathbf{A}_0$ we also have

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

where to derive the last step we have made use of the normalization $N^{-1}\mathbf{A}_0'\mathbf{A}_0 = \mathbf{I}_m$. Furthermore, since $\hat{\mathbf{F}} = N^{-1}\mathbf{X}\hat{\mathbf{A}}$ by equation (17), then

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

Using this result in $\mathbf{M}_{\hat{\mathbf{F}}}\left(\mathbf{F}_0 - \hat{\mathbf{F}}\right)$ together with (9) to substitute out $\mathbf{X}$, we obtain

$$\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{F}_0 = \mathbf{M}_{\hat{\mathbf{F}}}\left[\frac{\mathbf{F}_0\mathbf{A}_0'\left(\mathbf{A}_0 - \hat{\mathbf{A}}\right)}{N} + \frac{\mathbf{V}\left(\mathbf{A}_0 - \hat{\mathbf{A}}\right)}{N}\right] - \mathbf{M}_{\hat{\mathbf{F}}}\left(\frac{\mathbf{V}\mathbf{A}_0}{N}\right) \tag{29}$$

The results of Proposition 2 suggest that the leading term in the above expression is $\mathbf{M}_{\hat{\mathbf{F}}}\left(N^{-1}\mathbf{V}\mathbf{A}_0\right)$, whereas the other two terms are of lower order in probability. As detailed in Lemma 1 in the online supplement, this allows us to rewrite expression (28) as

$$N\hat{\sigma}_{iT}^2 = \frac{\mathbf{a}_i'\mathbf{A}_0'\mathbf{V}'\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{V}\mathbf{A}_0\mathbf{a}_i}{NT} + O_p\left(\frac{1}{\delta_{NT}}\right) + O_p\left(\frac{\sqrt{N}}{\delta_{NT}^2}\right), \tag{30}$$

where the orders of probability on the right-hand side of (30) hold uniformly over i. However, if unit i is not pervasive, $N\hat{\sigma}_{iT}^2$ includes the additional two terms:

$$\frac{N\mathbf{v}_i'\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{v}_i}{T} \quad \text{and} \quad \frac{N\mathbf{a}_i'\left(\mathbf{F}_0 - \hat{\mathbf{F}}\right)'\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{v}_i}{T},$$

which are due to the idiosyncratic component $\mathbf{v}_i \neq 0$ in $\mathbf{x}_i$ when i is not pervasive.[12] As shown formally in Appendix A.1, the first term, $N\left(\mathbf{v}_i'\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{v}/T\right)$ diverges at rate N, while the

[12] Note that in the absence of any pervasive units $\mathbf{v}_i = \mathbf{u}_i$. In general, we use $\mathbf{v}_i$ (and $\mathbf{V}$) in line with the general factor model given by (9).

second term, $NT^{-1}\mathbf{a}_i'\left(\mathbf{F}_0-\hat{\mathbf{F}}\right)'\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{v}_i$ is $o_p(N)$. Consequently, a threshold C_{NT}^2 that allows us to distinguish pervasive from non-pervasive units must diverge at a rate considerably slower than N, as $N,T\to\infty$.

To establish a minimum requirement on the rate at which C_{NT}^2 must diverge, we return to expression (30). We set $\mathbf{d}_i=\mathbf{A}_0\mathbf{a}_i$ and note that $\mathbf{d}_i'\mathbf{V}'\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{V}\mathbf{d}_i\leq\mathbf{d}_i'\mathbf{V}'\mathbf{V}\mathbf{d}_i$. Suppose that cross-section unit i is pervasive, then it follows that

$$\Pr\left(N\hat{\sigma}_{iT}^2>C_{NT}^2\right)\leq\Pr\left(\frac{\mathbf{d}_i'\mathbf{V}'\mathbf{V}\mathbf{d}_i}{NT}>C_{NT}^2\right)+o\left(1\right), \tag{31}$$

if $\sqrt{N}/\delta_{NT}^2\to0$. Also note that

$$\frac{\sqrt{N}}{\delta_{NT}^2}=\begin{cases}\frac{\sqrt{N}}{T} \text{ if } T\leq N\\ \frac{1}{\sqrt{N}} \text{ otherwise}\end{cases},$$

and so the condition for the remainder term to vanish is $\frac{\sqrt{N}}{T}\to0$, as $N,T\to\infty$.[13] Under this condition, we focus on the first probability term in (31) and note that

$$\mathbf{d}_i'\mathbf{V}'\mathbf{V}\mathbf{d}_i=\sum_{t=1}^{T}\left(\sum_{j=1}^{N}v_{jt}d_{ij}\right)^2=\sum_{t=1}^{T}\left(\mathbf{d}_i'\mathbf{v}_t\right)^2,$$

where as before $\mathbf{d}_i=\mathbf{A}_0\mathbf{a}_i=(d_{i1},d_{i2},\ldots,d_{iN})'$. Note that if the panel contains m pervasive units, $\mathbf{v}_t=(\mathbf{0}_{1\times m},\mathbf{u}_{bt}')'$, where $\mathbf{u}_{bt}=\mathbf{H}\boldsymbol{\varepsilon}_t$. See (12). Partition $\mathbf{d}_i=(\mathbf{d}_{i1}',\mathbf{d}_{i2}')'$, where $\mathbf{d}_{i1}$ and $\mathbf{d}_{i2}$ are the $m\times1$ and $n\times1$ sub-vectors of $\mathbf{d}_i$ (recall that $n=N-m$). Hence

$$\mathbf{d}_i'\mathbf{V}'\mathbf{V}\mathbf{d}_i=\sum_{t=1}^{T}\left(\mathbf{d}_{i2}'\mathbf{u}_{bt}\right)^2=\sum_{t=1}^{T}\left(\mathbf{d}_{i2}'\mathbf{H}\boldsymbol{\varepsilon}_t\right)^2,$$

and by assumption $\mathbf{H}$ is an $n\times n$ matrix with bounded row and column absolute sum norms, and $\boldsymbol{\varepsilon}_t=\left(\varepsilon_{m+1,t},\varepsilon_{m+2,t},\ldots,\varepsilon_{N,t}\right)\sim IID\left(\mathbf{0},\mathbf{I}_n\right)$. Using the above results we can now write

$$\begin{aligned}\frac{\mathbf{d}_i'\mathbf{V}'\mathbf{V}\mathbf{d}_i}{NT}&=\left(\frac{n}{N}\right)\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\mathbf{d}_{i2}'\mathbf{H}\boldsymbol{\varepsilon}_t}{\sqrt{n}}\right)^2\\&=\left(\frac{n}{N}\right)\frac{1}{T}\sum_{t=1}^{T}\left(\varphi_i'\boldsymbol{\varepsilon}_t\right)^2,\end{aligned} \tag{32}$$

where $\varphi_i=n^{-1/2}\mathbf{H}'\mathbf{d}_{i2}=(\varphi_{i1},\varphi_{i2},\ldots,\varphi_{in})'$. Let

$$\eta_{in}^2=\varphi_i'\varphi_i=\frac{1}{n}\mathbf{d}_{i2}'\mathbf{H}\mathbf{H}'\mathbf{d}_{i2}=\frac{1}{n}\mathbf{d}_{i2}'\boldsymbol{\Sigma}_u\mathbf{d}_{i2}, \tag{33}$$

where $\boldsymbol{\Sigma}_u=E\left(\mathbf{u}_{bt}\mathbf{u}_{bt}'\right)=\mathbf{H}\mathbf{H}'$, which is time-invariant, by assumption.[14] We also have

$$\sup_i\eta_{in}^2\leq\sup_i\left(n^{-1}\mathbf{d}_{i2}'\mathbf{d}_{i2}\right)\lambda_{\max}\left(\boldsymbol{\Sigma}_u\right), \text{ and } \inf_i\eta_{in}^2\geq\inf_i\left(n^{-1}\mathbf{d}_{i2}'\mathbf{d}_{i2}\right)\lambda_{\min}\left(\boldsymbol{\Sigma}_u\right),$$

[13] It is possible that this condition can be relaxed if one finds a tighter upper bound for $\left\|\mathbf{A}_0'\left(\mathbf{A}_0-\hat{\mathbf{A}}\right)\right\|$ than result (E) in Proposition 2, and if stationarity is imposed on $E(v_{it}v_{it'})$, where $\mathbf{V}=(v_{it})$. For now, we adhere to Assumptions A-F of Bai (2003), and require that $\frac{\sqrt{N}}{T}\to0$, as $N,T\to\infty$.

[14] However, one can still allow for conditionally time-varying covariances.

where by assumption $0 < c < \lambda_{\min}(\mathbf{\Sigma}_u) \leq \lambda_{\max}(\mathbf{\Sigma}_u) < C < \infty$. Noting that, in view of (10) and (11), we have

$$\sup_i \left(n^{-1}\mathbf{a}_i'\mathbf{A}_b'\mathbf{A}_b\mathbf{a}_i\right) \leq \sup_i \|\mathbf{a}_i\|^2 \lambda_{\max}\left(n^{-1}\sum_{j=m+1}^{N}\mathbf{a}_j\mathbf{a}_j'\right) < C < \infty, \tag{34}$$

and

$$\inf_i \left(n^{-1}\mathbf{a}_i'\mathbf{A}_b'\mathbf{A}_b\mathbf{a}_i\right) \geq \inf_i \|\mathbf{a}_i\|^2 \lambda_{\min}\left(n^{-1}\sum_{j=m+1}^{N}\mathbf{a}_j\mathbf{a}_j'\right) > c > 0, \tag{35}$$

we also have $\sup_i \left(n^{-1}\mathbf{d}_{i2}'\mathbf{d}_{i2}\right) < C$, and $\inf_i \left(n^{-1}\mathbf{d}_{i2}'\mathbf{d}_{i2}\right) > 0$. Hence, it follows that $\sup_i \eta_{in}^2 < C$, and $\inf_i \eta_{in}^2 > 0$, for all n. Now using (32) in (31), we have

$$\begin{aligned}\Pr\left[n\hat{\sigma}_{iT}^2 > \left(\frac{n}{N}\right)C_{NT}^2\right] &\leq \Pr\left[\sum_{t=1}^{T}\left(\varphi_i'\boldsymbol{\varepsilon}_t\right)^2 > T\left(\frac{n}{N}\right)C_{NT}^2\right] + o(1) \\ &\leq \sum_{t=1}^{T}\Pr\left[\left(\varphi_i'\boldsymbol{\varepsilon}_t\right)^2 > \left(\frac{n}{N}\right)C_{NT}^2\right] + o(1).\end{aligned} \tag{36}$$

where the last line applies Lemma A11 in the online supplement to Chudik, Kapetanios, and Pesaran (2018) in order to bound the tail probability of a non-negative sum by the sum of individual tail probabilities. Additionally, letting $\varphi_i'\boldsymbol{\varepsilon}_t = \sum_{j=1}^{n}\varphi_{ij}\varepsilon_{jt}$, we can write

$$\begin{aligned}\Pr\left[\left(\varphi_i'\boldsymbol{\varepsilon}_t\right)^2 > \left(\frac{n}{N}\right)C_{NT}^2\right] &= \Pr\left(\left|\varphi_i'\boldsymbol{\varepsilon}_t\right| > \left(\frac{n}{N}\right)^{1/2}C_{NT}\right) \\ &= \Pr\left(\left|\sum_{j=1}^{n}\varphi_{ij}\varepsilon_{jt}\right| > \left(\frac{n}{N}\right)^{1/2}C_{NT}\right).\end{aligned}$$

In order to proceed from the above expression, we note that under Assumption 3, $Var\left(\sum_{j=1}^{n}\varphi_{ij}\varepsilon_{jt}\right) = \sum_{j=1}^{n}\varphi_{ij}^2 = \eta_{in}^2$, and

$$\Pr\left(|\varepsilon_{jt}| > a\right) \leq C_0\exp\left(-C_1 a^s\right),$$

for all $a > 0$, $s > 0$ and some fixed constants C_0 and C_1. This assumption allows us to employ a concentration inequality in order to bound the tail probability of our expression of interest by an exponential function of its second natural moment. More specifically, by applying Lemma A3 of Chudik, Kapetanios, and Pesaran (2018) we obtain

$$\Pr\left(\left|\sum_{j=1}^{n}\varphi_{ij}\varepsilon_{jt}\right| > C_{NT}\right) \leq \exp\left[\frac{-(1-\pi)^2 C_{NT}^2}{2\eta_{in}^2}\right], \tag{37}$$

for some $0 < \pi < 1$ and $C_{NT} = O\left(n^{\lambda}\right)$, with $0 < \lambda < \frac{s+1}{s+2}$ (note that $n/N = 1 - m/N \approx 1$). Now using (37) in (36) and assuming that unit i is pervasive yields

$$\Pr\left(N\hat{\sigma}_{iT}^2 > \left(\frac{n}{N}\right)C_{NT}^2 \middle|\, i \text{ is pervasive}\right) \leq T\exp\left[\frac{-(1-\pi)^2 C_{NT}^2}{2\eta_{in}^2}\left(\frac{n}{N}\right)\right] + o(1),$$

for some $0 < \pi < 1$, and η_{in}^2 as defined by (33). Hence,

$$\Pr\left[N\hat{\sigma}_{iT}^2 \leq \left(\frac{n}{N}\right) C_{NT}^2 \mid i \text{ is pervasive}\right] \geq 1 - \exp\left[\log(T) - \frac{(1-\pi)^2 C_{NT}^2}{2\eta_{in}^2}\left(\frac{n}{N}\right)\right],$$

and (note that $n/N \to 1$)

$$\Pr\left[N\hat{\sigma}_{iT}^2 > \left(\frac{n}{N}\right) C_{NT}^2 \mid i \text{ is pervasive}\right] \to 0, \text{ as } N, T \to \infty,$$

if

$$\frac{\sqrt{N}}{T} \to 0, \text{ and } \log(T) - \frac{(1-\pi)^2 C_{NT}^2}{2\eta_{in}^2} \to -\infty.$$

This last condition is satisfied if (again setting n/N to unity) if $C_{NT}^2 > \frac{2\log(T)\eta_{in}^2}{(1-\pi)^2}$, or equivalently if $C_{NT}^2 = 2C\eta_{in}^2 \log(T)$, for some $C > 1$. Accordingly, i can be selected as a pervasive unit if $\hat{\sigma}_{iT}^2 \leq \frac{2C\eta_{in}^2 \log(T)}{N}$, for some $C > 1$. Our Monte Carlo results presented in Section 6 show that the simple choice $C = 1$, works well in practice. Additional simulation results which show that the performance of the thresholding rule described here is fairly robust to the exact choice of C can be found in Section S5 of the online supplement accompanying this article.

Furthermore, as shown in Appendix A.2, none of the remainder terms in (30) can exceed the threshold, with probability approaching one if the unit i is pervasive. We also show that $B_{i7} = O_p(N)$ and $B_{i8} = o_p(N)$, and further, using (47), that the residual variance, $\hat{\sigma}_{iT}^2$, will exceed the threshold with probability approaching one if the unit i is *not* pervasive.

An important issue relates to the estimation of η_{in}^2. Since m is unknown and typically small, at the estimation stage we assume $m = 0$, and set $n = N$ and note that in this case $\boldsymbol{\Sigma}_u = E(\mathbf{u}_{bt}\mathbf{u}_{bt}') = E(\mathbf{v}_t\mathbf{v}_t') = \boldsymbol{\Sigma}_v$. Consequently, $\eta_{iN}^2 = N^{-1}\mathbf{a}_i'\mathbf{A}_0'\boldsymbol{\Sigma}_v\mathbf{A}_0\mathbf{a}_i$ for which a consistent estimator can be obtained using the PC estimators of $\mathbf{a}_i$ and $\mathbf{A}_0$, and a suitable threshold estimator of $\boldsymbol{\Sigma}_v$ (or $\boldsymbol{\Sigma}_u$). Also, recall that $\boldsymbol{\Sigma}_u = \mathbf{H}\mathbf{H}'$ where by assumption $\mathbf{H}$ is row and column bounded (see Assumption 3), and hence $\boldsymbol{\Sigma}_u$ is also row-bounded and therefore satisfies the usual sparsity conditions assumed in the literature on estimation of large covariance matrices (see, e.g., El Karoui, 2008 or Bickel and Levina, 2008). Then, η_{iN}^2 can be consistently estimated by

$$\hat{\eta}_{iN}^2 = N^{-1}\hat{\mathbf{a}}_i'\hat{\mathbf{A}}'\tilde{\boldsymbol{\Sigma}}_u\hat{\mathbf{A}}\hat{\mathbf{a}}_i, \tag{38}$$

where $\hat{\mathbf{A}}$ is given by (16), $\hat{\mathbf{a}}_i$ is the OLS estimator of $\mathbf{a}_i$ in the regression of $\mathbf{x}_i$ (the selected pervasive unit) on $\hat{\mathbf{F}}$, where the latter is given by (17), and $\tilde{\boldsymbol{\Sigma}}_u = (\tilde{\sigma}_{ij})$ is a consistent estimator of $\boldsymbol{\Sigma}_u$. Here we use the multiple testing estimator of Bailey, Pesaran, and Smith (2019) given by

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

$$\hat{\sigma}_{ij} = \frac{1}{T}\sum_{t=1}^{T} \hat{u}_{it}\hat{u}_{jt}, \; \hat{\rho}_{ij} = \frac{\hat{\sigma}_{ij}}{\hat{\sigma}_{ii}^{1/2}\hat{\sigma}_{jj}^{1/2}},$$

where $\hat{u}_{it}, t = 1, 2, \ldots, T$ are the OLS residuals from the regression of x_{it} (the selected pervasive unit) on $\hat{\mathbf{F}}$ (including an intercept in all regressions), $\Phi^{-1}(\cdot)$ is the inverse of cumulative distribution function of a standard normal variate, π is the nominal size for the multiple testing procedure, which we set to 1%, and δ is set to 1.5, which allows for possible departures from Gaussian errors, u_{it}. Other estimators can also be used such as the universal thresholding by El Karoui (2008) and Bickel and Levina (2008), and the adaptive thresholding by Cai and Liu (2011).

Our threshold detection algorithm, referred to as σ^2 thresholding, can be summarized as follows:

Algorithm 1 *Let $\mathbf{x}_i$ be the $T \times 1$ vector of observations on the i-th unit in the panel, and $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$ 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. Compute $\hat{\mathbf{F}} = \frac{1}{\sqrt{N}}\mathbf{X}\hat{\mathbf{Q}}$, where $\hat{\mathbf{Q}}$ is the $N \times p_{\max}$ matrix whose columns are the orthonormal eigenvectors of $\mathbf{X}'\mathbf{X}$, such that $N^{-1}\hat{\mathbf{Q}}'\hat{\mathbf{Q}} = \mathbf{I}_{p_{\max}}$. Compute $\hat{\mathbf{a}}_i$, $\hat{v}_{it}$ and $\hat{\sigma}^2_{iT}$ to be the OLS estimator, residual and residual variance of the regression of $\mathbf{x}_i$ on $\hat{\mathbf{F}}$, namely*

$$\begin{aligned}
\hat{\mathbf{a}}_i &= \left(\hat{\mathbf{F}}'\hat{\mathbf{F}}\right)^{-1}\hat{\mathbf{F}}'\mathbf{x}_i, \\
\hat{\mathbf{u}}_i &= (\hat{u}_{i1}, \hat{u}_{i2}, \ldots, \hat{u}_{iT})' = \mathbf{M}_{\hat{\mathbf{F}}}\mathbf{x}_i = \left[\mathbf{I}_T - \hat{\mathbf{F}}\left(\hat{\mathbf{F}}'\hat{\mathbf{F}}\right)^{-1}\hat{\mathbf{F}}'\right]\mathbf{x}_i, \\
\hat{\sigma}^2_{iT} &= T^{-1}\mathbf{x}_i'\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{x}_i.
\end{aligned}$$

Sort $\hat{\sigma}^2_{iT}$ in ascending order and denote the sorted series by $\hat{\sigma}^2_{(1)T}$, $\hat{\sigma}^2_{(2)T}, \ldots, \hat{\sigma}^2_{(N)T}$ with $\hat{\sigma}^2_{(i)T}$ being the i^{th} smallest value. Consider the cross-section indexes $i_1, i_2, \ldots, i_{p_{\max}}$ corresponding to $\hat{\sigma}^2_{(1)T}$, $\hat{\sigma}^2_{(2)T}, \ldots, \hat{\sigma}^2_{(p_{\max})T}$. Then, select unit $j \in \{i_1, i_2, \ldots i_{,p_{\max}}\}$ to be pervasive if

$$\hat{\sigma}^2_{jT} \leq \frac{2\hat{\eta}^2_{jN}\log T}{N}, \tag{39}$$

where $\hat{\eta}^2_{jN}$ is given by (38).

The following theorem provides a formal summary statement of the preceding analysis.

Theorem 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 by (4) and (5) 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. Denote by $\hat{I}_D$ and $\hat{I}_{DN}$ their estimates based on the threshold criteria (39). Let Assumptions 1-4 hold and $\frac{\sqrt{N}}{T} \to 0$. Then as N and $T \to \infty$, jointly, we have*

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

This theorem establishes that the proposed error variance threshold criterion is consistent, in the sense that it correctly selects the pervasive (if any) and the non-pervasive units asymptotically.

Remark 3 *Note that both the theoretical exposition above and the formal arguments of the appendix apply both to the case of no external factors as well as the case where all units are affected by a finite number of external factors, represented by $\mathbf{g}_t$ in (4) and (5).*

5 A sequential, multiple testing version of the σ^2 thresholding

The σ^2 thresholding procedure, has good but not exceptional small sample properties as we illustrate in Section S4 of the online supplement to this article. However, it provides a basis for further development. The first point to note is that while the method is good at detecting the presence of pervasive units, in general it tends to pick too many units as pervasive. Finite sample adjustments are needed to achieve a more conservative detection outcome. A simple and effective refinement of the main method is a sequential algorithm that detects pervasive units one at a time. Considering a sequential algorithm suggests the use of pervasive units that have been identified at earlier steps of the procedure as observed factors. This reduces the number of unobserved factors to be estimated for any given maximum number of considered factors, p_{max}. Therefore, the static factor model (9) employed to conduct σ^2 thresholding is replaced by the augmented factor model

$$x_{it} = \mathbf{f}_t^{*\prime}\mathbf{a}_i^* + \mathbf{x}_{at}^{*\prime}\mathbf{b}_{ai}^* + v_{it},\ t = 1, 2, \ldots, T; i = 1, \ldots N_1, \tag{40}$$

where $\mathbf{x}_{at}^*$ is a $r \times 1$ vector of identified pervasive units (the row t of the $T \times r$ matrix $\mathbf{X}_a^*$), $\mathbf{f}_t^*$ is a $p_{max} - r$ vector of unobserved common factors and v_{it} constitutes the idiosyncratic variation of unit i at time point t. With regards to the procedure of Section 4.2, the role of σ^2 thresholding is not to determine the number and the identities of the pervasive units directly. Instead, σ^2 thresholding is used to determined whether or not there is evidence of remaining pervasive units in the data, conditional on the pervasive units that have been identified already. Being initiated with $r = 0$ (i.e. no identified pervasive units), $N_1 = N - r = N$ and some chosen value of p_{max} subject to the condition $p_{max} \geq m + 1$, the sequential algorithm, referred to as $S - \sigma^2$ thresholding is an iteration of the following two-step procedure:

Algorithm 2 *1. Conduct σ^2 thresholding using model (40), with $m^* = p_{max} - r$ estimated factors. Let $\widetilde{m}$ be the estimated number of pervasive units estimated using Algorithm 1. If $\widetilde{m} = 0$, stop and conclude that there are r pervasive units.*

2. If $\hat{m} > 0$, 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 $N_1 - 1$.

The above two steps are repeated until either $\widetilde{m} = 0$ in the first step or $r = p_{max}$ at the end of step 2. The number of pervasive units is then $\hat{m} = r$ and their identities correspond to the indices of the columns in the initial $T \times N$ vector $\mathbf{X}$ that are found in the $T \times r$ matrix $\mathbf{X}_a^* = (\mathbf{x}_{a1}^*; \ldots; \mathbf{x}_{aT}^*)'$.

Effectively the method constructs residuals of the remaining units on the selected units and repeats the selection on these residuals. The use of residuals in the algorithm's steps requires further theoretical refinements. These are discussed in Appendix A.4 where it is shown that our proposed threshold is valid only if $N < T$, which is a more restrictive condition than that of Theorem 2. In particular, we prove the following result:

Corollary 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 given by (4) and (5) 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. Denote by $\hat{I}_D$ and $\hat{I}_{DN}$ their estimates based on $S-\sigma^2$ thresholding. Let Assumptions 1-4 hold and $\frac{N}{T} \to 0$. Then as N and $T \to \infty$, jointly, we have

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

If $N \geq T$, then an alternative threshold could be considered. This is given by

$$\hat{\sigma}^2_{iT} \leq \frac{2\hat{\sigma}^2_{iu}\log T}{T},$$

where $\hat{\sigma}^2_{iu} = T^{-1}\sum_{t=1}^{T}\left(x_{it} - \mathbf{X}^{*\prime}_{at}\hat{\boldsymbol{\gamma}}^*_i\right)^2$, with $\hat{\gamma}^*_i$ being the estimated vector of slope coefficients from a regression of $\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{x}_i$ on $\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{X}^*_a$. This is justified in Appendix A.4. As its small sample properties are inferior to those of our main procedure we do not pursue this further in the main paper but only in Section S4 of the online supplement. However, it is important to note that it provides a theoretical justification for our general methodology when $N > T$.

Finally, the sequential algorithm can be supplemented with an additional multiple testing hurdle in order to reduce the risk of falsely detecting a pervasive unit in small samples. Analogous to the basic sequential algorithm discussed above, the extended algorithm is initiated with $r = 0$, $N_1 = N$ and some chosen value of p_{max} subject to the condition $p_{max} \geq m + 1$. It consists of the following five steps which are repeated until the estimated number of pervasive units $\widetilde{m}$ in the first step is equal to zero:

Algorithm 3

1. *Conduct σ^2 thresholding using model (40) and $m^* = p_{max} - r$ estimated factors. Let $\widetilde{m}$ be the estimated number of pervasive units estimated using Algorithm 1. If $\widetilde{m} = 0$, stop and conclude that there are r pervasive units.*

2. *If $\widetilde{m} > 0$, obtain $i^* = \arg\min_i \hat{\sigma}^2_i$. For each $j = 1, \ldots i^* - 1, i^* + 1, \ldots, N_1$ estimate the model*
$$x_{jt} = \mu^*_j + x_{i^*t}\gamma^*_j + \mathbf{f}^{*\prime}_t\mathbf{a}^*_j + \mathbf{x}^{*\prime}_{at}\mathbf{b}^*_{aj} + v_{jt},\ t = 1, 2, \ldots, T, \tag{41}$$
where the unobserved factors $\mathbf{f}^*_t$ are estimated by the eigenvectors associated to the $p_{max} - r$ largest eigenvalues of $\mathbf{X}_{-i^*}\mathbf{X}'_{-i^*}$ with $\mathbf{X}_{-i^*} = (\mathbf{x}_1; \ldots; \mathbf{x}_{i^*-1}; \mathbf{x}_{i^*+1}; \ldots; \mathbf{x}_{N_1})$.

3. *Carry out $N_1 - 1$ individual t-type tests to check the statistical significance of the slope parameters $\hat{\gamma}^*_j$ for all $j \neq i$ in (41). These tests are based on statistics of the form*
$$t^*_j = \sqrt{T}\hat{\gamma}^*_j\left(\frac{T^{-1}\sum_{t=1}^{T}x^2_{i^*t}}{T^{-1}\sum_{t=1}^{T}\hat{v}^2_{jt}}\right)^{1/2}, \tag{42}$$
where $v_{jt} = x_{jt} - \hat{\mu}^*_j - x_{i^*t}\hat{\gamma}^*_j - \hat{\mathbf{f}}^{*\prime}_t\hat{\mathbf{a}}^*_j - \mathbf{x}^{*\prime}_{at}\hat{\mathbf{b}}^*_{aj}$. Let the critical value for each of these tests be given by $\Phi^{-1}\left[1 - \frac{\pi}{2(N_1-2)}\right]$, where the nominal size of the individual tests, π, is chosen by the investigator. In our analysis we set $\pi = 0.01$. The null hypothesis $\hat{\gamma}^*_j \neq 0$, is reject, if $\left|t^*_j\right| > \Phi^{-1}\left[1 - \frac{\pi}{2(N_1-2)}\right]$.

4. *Let M denote the number of rejections among these $N_1 - 1$ tests. If $\log(M)/\log(N) \leq 1/2$, stop and conclude that there are r pervasive units.*

5. *If* $\log(M)/\log(N) > 1/2$*, append* $\mathbf{x}_{i^*}$ *to* $\mathbf{X}_a^*$ *and eliminate* $\mathbf{x}_{i^*}$ *from* $\mathbf{X}$*. Update* r *to* $r+1$ *and* N_1 *to* $N_1 - 1$

We refer to this algorithm as Sequential-MT σ^2 thresholding or $SMT - \sigma^2$ thresholding for short. Two remarks concerning algorithm 3 are in order. First, deviating from a standard t-statistic when conducting $N_1 - 1$ significance tests is a necessary adjustment to account for the nonstandard properties of the auxiliary regression (41). If i^* denotes the index of a true pervasive unit, then the set of regressors $\left(x_{i^*t};\ \hat{\mathbf{f}}_t^{*\prime}\right)$ is asymptotically multicollinear since $\hat{\mathbf{f}}_t^{*\prime}$ is consistent for the space spanned by all common factors driving x_{jt}, including x_{i^*t}. As shown in Appendix A.5, this characteristic of the model affects the properties of a test statistic for the statistical significance of γ_j^* and is resolved by replacing the standard estimator of $Var\left(\hat{\gamma}_j^*\right)$ with a different standardization for $\hat{\gamma}_j^*$. A further important point is that we need to have an estimator of the *full* common factor space, such as $\hat{\mathbf{f}}_t^*$. Otherwise, even non-pervasive units will appear significant in (41) since the impact of external factors and pervasive units will turn them into a proxy for unaccounted sources of common variation.

Second, the rule $\log(M)/\log(N) \leq 1/2$, or $M \leq N^{1/2}$ is motivated by the fact that if a factor enters only M units, where $M = o(N^{1/2})$, then, it is considered to be a very weak factor, and under certain conditions, it is not detectable using principal components - see, e.g., Bailey, Kapetanios, and Pesaran (2016). Again, after stopping the sequential algorithm, the number of pervasive units is $\hat{m} = r$ and their identities correspond to the indices of the columns in the initial $T \times N$ vector $\mathbf{X}$ that are found in the $T \times r$ matrix $\mathbf{X}_a^* = (\mathbf{x}_{a1}^*; \ldots; \mathbf{x}_{aT}^*)'$.

The additional multiple testing step ends up being very effective in small samples and is therefore our preferred approach. We conclude this section by presenting the Theorem below, which states that both Algorithms 2 and 3 share the consistency properties of 1.

Theorem 3 *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 by (4) and (5) 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. Denote by* $\hat{I}_D$ *and* $\hat{I}_{DN}$ *their estimates based on either Algorithm 2 or 3. Let Assumptions 1-4 hold and* $\frac{\sqrt{N}}{T} \to 0$*. Then as* N *and* $T \to \infty$*, jointly, we have that*

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

6 A comparative analysis of detection procedures by Monte Carlo simulations

Using Monte Carlo simulations we now investigate the small sample performance of our new method relative to the methods proposed by Parker and Sul (2016, henceforth PS) and Brownlees and Mesters (2019, BM in the following).[15] The PS method yields identical outcomes irrespective of whether the observations are standardized to have in sample zero means and unit standard deviations or not. Our proposed method, being based on residuals, is also not affected by demeaning of the observations and the scaling is done through the determination of the unit-specific thresholds, and hence standardization will not be an issue. In contrast,

[15] The detection methods of Parker and Sul and Brownless and Mesters are described in some detail in Section S3 of the online supplement.

BM's detection method can be quite sensitive to standardization in finite samples, although asymptotically it should not matter whether the individual series in the panel are standardized. The BM method is also applied either including all the N units, or only the $N/2$ most connected units when selecting the pervasive units.[16] Accordingly, we consider four variants of the BM method: modified and unmodified with and without standardization. We shall refer to these variants as BM and BM (standardized) when only the $N/2$ most connected units are considered at the selection stage, and unmodified BM and unmodified BM (standardized) when all the N units are included when selecting the pervasive units. In the paper we focus on the modified version of BM, and give the results for their unmodified version in Section S7 of the online supplement. It is clear that fewer units will be detected when the modified version is used, but the effect of standardization is less clear cut. Amongst the various σ^2 thresholding procedures discussed, we focus on $SMT - \sigma^2$ thresholding as described by Algorithm 3.[17]

In accordance with the formal presentation in Section 3, we simulate the pervasive unit model as

$$\mathbf{x}_{ta} = \boldsymbol{\mu}_a + \boldsymbol{\Lambda}_a \mathbf{g}_t + \mathbf{u}_{at}, \tag{43}$$

$$\mathbf{x}_{tb} = \boldsymbol{\mu}_b + \mathbf{B}\mathbf{x}_{ta} + \boldsymbol{\Lambda}_b \mathbf{g}_t + \mathbf{u}_{bt}, \tag{44}$$

for $t = 1, 2, \ldots, T$. The $N \times 1$ vector of fixed effects, $\boldsymbol{\mu} = (\boldsymbol{\mu}_a; \boldsymbol{\mu}_b)'$, are drawn from $IIDU(0, 1)$. The $k_0 \times 1$ vectors $\mathbf{g}_t$, for $t = 1, 2, \ldots, T$, representing the unobserved common factors, are generated as $\mathbf{g}_t = \mathbf{R}_g^{1/2}(\mathbf{g}_{*,t} - 2\boldsymbol{\tau}_k)/2$, where $\boldsymbol{\tau}_k = (1, 1, \ldots, 1)'$, $\mathbf{g}_{*,t}$ is a $k \times 1$ vector generated as $IID\chi^2(2)$, $\mathbf{R}_g^{1/2}$ is the square root of the $k_0 \times k_0$ matrix $\mathbf{R}_g$ defined by

$$\mathbf{R}_g = (1 - \rho_g)\mathbf{I}_k + \rho_g \boldsymbol{\tau}_k \boldsymbol{\tau}_k',$$

where ρ_g represents the pair-wise correlation coefficients of the distinct (i, j) elements of $\mathbf{g}_t$, assumed to be the same across all i and $j = 1, 2, .., k$. Specifically, $Cov(\mathbf{g}_t) = \mathbf{R}_g$. Similarly, the $m_0 \times 1$ vector $\mathbf{u}_{at}$ is generated analogously as

$$\mathbf{u}_{at} = \mathbf{R}_a^{1/2}(\mathbf{u}_{at}^* - 2\tau_m)/2, \ \mathbf{R}_a = (1 - \rho_a)\mathbf{I}_m + \rho_a \boldsymbol{\tau}_m \boldsymbol{\tau}_m',$$

where $\mathbf{u}_{at}^* \sim IID\chi^2(2)$. It follows that $Cov(\mathbf{u}_{at}) = \mathbf{R}_a$, and ρ_a represents the pair-wise correlations of the elements of $\mathbf{u}_{at}$, assumed to be the same across all pairs. The $m_0 \times k_0$ matrix $\boldsymbol{\Lambda}_a$ and the $n \times k_0$ matrix $\boldsymbol{\Lambda}_b$ are obtained as $IIDU(0, 1)$. The correlation coefficients ρ_g and ρ_a are drawn from $U(0.2, 0.8)$, and are allowed to vary across replications.

The importance of the pervasive units for the non-pervasive units is represented by the $(N - m_0) \times m_0$ loading matrix $\mathbf{B} = (b_{ij})$. To allow for the possibility of both strong and weak pervasive units, b_{ij} are generated as

$$b_{ij} \begin{cases} \sim IIDU(0, 1) & \text{if } i \leq \lfloor (N - m_0)^\alpha \rfloor \\ = 0 & \text{otherwise.} \end{cases}, \text{ for } i = 1, 2, \ldots, N - m_0; \ j = 1, 2, \ldots, m_0, \tag{45}$$

where as introduced in (8), α is the exponent that measures the degree of dominance of $\mathbf{x}_{ta}$ in the panel. For the sake of simplicity, all pervasive units are assumed to have the same degree

[16] In their simulation analysis BM seem to be using the unmodified version of their method without standardization, whilst in their empirical applications they apply the modified version after standardization. See Section 6 of Brownlees and Mesters (2019).

[17] Simulation results for other two variants of σ^2 thresholding, described by Algorithms 1 and 2, are provided in Section S4 of the online supplement.

of dominance so that a subscript on α is redundant . When $\alpha = 1$ the units are pervasive, in the sense that they have non-zero effects on all the $N - m_0$ non-pervasive units. This is the standard case in the common factor literature and ensures that $\lim_{N\to\infty}(N - m_0)^{-1}\mathbf{B}'\mathbf{B}$ is a positive definite $m_0 \times m_0$ matrix. This condition clearly breaks down when $\alpha < 1$. As we noted before, $\mathbf{x}_{ta}$ are then referred to as weakly pervasive units.

The errors $\mathbf{u}_{bt} = (u_{b,it})$ are generated as heterogeneous first order autoregressive processes

$$u_{b,it} = \rho_i u_{b,it-1} + (1 - \rho_i^2)^{1/2}\varepsilon_{it}, \text{ for } t = -49, \dots, 0, 1, 2, \dots, T; i = 1, 2, \dots, N - m_0,$$

where $\rho_i \sim IIDU(0.2, 0.5)$. The errors ε_{it} are allowed to be cross-sectionally weakly correlated. To achieve this we set $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, \dots, \varepsilon_{nt})' = \boldsymbol{\Sigma}^{1/2}\mathbf{R}_b^{1/2}\boldsymbol{\zeta}_t$, $n = N - m$, with $\boldsymbol{\Sigma} = diag(\sigma_{11}, \sigma_{22}, \dots, \sigma_{nn})$, and

$$\mathbf{R}_b = \begin{pmatrix} 1 & \rho_b & \rho_b^2 & \dots & \rho_b^{n-1} \\ \rho_b & 1 & \rho_b & \dots & \rho_b^{n-2} \\ \rho_b^2 & \rho_b & 1 & \dots & \rho_b^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_b^{n-1} & \rho_b^{n-2} & \rho_b^{n-3} & \dots & 1 \end{pmatrix}.$$

We set $\rho_b = 0.5$, $\sigma_{ii} = \sigma_{*,ii}/4 + 0.5$, and $\sigma_{*,ii} \sim IID\chi^2(2)$, thus ensuring that $E(\sigma_{ii}) = 1$. Lastly, $\zeta_{it} = (\zeta_{*,it} - 2)/2$, where $\zeta_{*,it} \sim IID\chi^2(2)$. In order to avoid dependence of $\mathbf{u}_{bt}$ on its starting values we discard the first 50 observations. All random variables are redrawn at the start of each replication of the simulation experiments.

We carry out all the different experiments for the following N and T combinations:

$$N \in \{50, 100, 200, 500\} \text{ and } T \in \{60, 110, 210, 250\}.$$

These N and T values allow for both cases where $T > N$, which is required for the BM procedure to be applicable, as well as when $T < N$, which often arises in empirical applications, and can be considered using our proposed method and the PS procedure.

The above setup allows us to control the number of pervasive units, m_0, the number of external factors, k_0, as well as the degree of dominance of the pervasive unit, α. We consider all $m_0 \le 2$ and $k_0 \le 2$ combinations, namely

$$\{m_0, k_0\} = \{0,0\}, \{0,1\}, \{0,2\}, \{1,0\}, \{1,1\}, \{1,2\}, \{2,0\}, \{2,1\}, \{2,2\}.$$

In cases where $m_0 > 0$, we experiment with two values of $\alpha = 1$ and $\alpha = 0.8$. Our theoretical derivations relate to the case of strongly pervasive units, namely when $\alpha = 1$. However, in practice it is more likely that the pervasive units are not strong, but still quite influential, which we represent by the choice of $\alpha = 0.8$. In the production network literature where the degree of the dominance can be computed from input-output tables, α is estimated to lie in the region of $0.7 - 0.8$. See Acemoglu, Carvalho, Ozdaglar, and Tahbaz-Salehi (2012) and Pesaran and Yang (2019, Definition 1).

Finally, all simulations in this section are conducted using $2,000$ replications.

6.1 MC results

The first scenario to consider is one without any pervasive units ($m_0 = 0$). The results for $SMT{-}\sigma^2$ and the PS procedures are summarized in Table 1, which gives the empirical frequency of correctly estimating m_0 to be 0. This table does not include the detection procedure proposed by BM, since the BM method pre-assumes that $m_0 > 0$, and therefore always *incorrectly* selects at least one pervasive unit. As can be seen from this table, the $SMT - \sigma^2$ thresholding performs very well, even in the presence of external common factor (namely when $k_0 > 0$), so long as N is sufficiently large. It is only outperformed by the PS procedure when N is small ($N = 50$) and there are external factors ($k_0 > 0$). Table 2 reports the average number of non-pervasive units (across replications) that are falsely selected as pervasive by $SMT - \sigma^2$, PS and BM. In this regard, $SMT - \sigma^2$ and PS perform perfectly when there are no external factors ($k_0 = 0$), and register a small number of incidence of false discovery when $k_0 = 1$, and N relatively small. However, the PS procedure seems to break down when the number of external factors is increased to $k_0 = 2$, and its average number of false discoveries reaches 41 with $N = 500$ and $T = 250$. However, the $SMT - \sigma^2$ thresholding continues to perform well even for $k_0 = 2$. As can be seen from Table 2, the average number of false discoveries of $SMT - \sigma^2$ thresholding is at most 0.7 over all values of N and T, and tends to zero as N is increased. By contrast, the BM procedure will always falsely selects non-pervasive units as pervasive even for panels with N and T large (subject to $T > N$). The average number of false discoveries for the BM procedure lies in range of 3 to 4, and is unaffected by standardization. However, modification of the BM procedure seems to play a crucial role in controlling the number of false discoveries. If we use the unmodified version of BM the average number of false discoveries rise dramatically and can reach around 100 for $N = 200$ and $T = 250$, with standardization only helping marginally. See Section S7 of the online supplement for details.

Table 1: Empirical frequency of correctly identifying the absence of a pervasive unit

	$SMT - \sigma^2$				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 using Algorithm 3, with $p_{max} = m_0 + k_0 + 1$, where m_0 is the true number of pervasive units (if any) and k_0 is the number of external factors. Threshold in the σ^2 thresholding step is given by $\hat{\sigma}_{iT}^2 \leq 2\hat{\eta}_{iN}^2 N^{-1} \log(T)$. PS refers to the Parker and Sul (2016) method by setting the number of potential pervasive units to $N/10$ per estimated factor, with the number of factors selecting using IC_{p2} criterion of Bai and Ng (2002). See also online supplement.

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

$SMT - \sigma^2$

	$k_0 = 0$				$k_0 = 1$				$k_0 = 2$			
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

	$k_0 = 0$				$k_0 = 1$				$k_0 = 2$			
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

	$k_0 = 0$				$k_0 = 1$				$k_0 = 2$			
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

BM (standardized)

	$k_0 = 0$				$k_0 = 1$				$k_0 = 2$			
N \T	60	110	210	250	60	110	210	250	60	110	210	250
50	4.1	3.4	3.3	3.2	3.9	3.3	3.6	3.9	3.6	3.4	3.5	3.6
100	n/a	4.1	3.0	2.9	n/a	3.4	3.4	3.4	n/a	3.2	3.3	3.4
200	n/a	n/a	3.4	2.8	n/a	n/a	3.0	2.8	n/a	n/a	3.0	2.7
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: The $SMT - \sigma^2$ and PS methods are as described in the notes to Table 1. BM refers to the modified detection method used in Section 6 of Brownlees and Mesters (2019). BM (standardized) stands for application of BM to data that have been recentered and rescaled so that each cross-section specific time series has an average of zero and a variance of one. BM methods are not applicable (n/a) if $T < N$.

Consider now cases where the DGP contains one or two pervasive units. Table 3 reports the empirical frequency of correctly estimating the number and the identity of the pervasive units by all the three detection procedures. The top panel of the table gives the results when there is one pervasive unit ($m_0 = 1$), with and without external factors, namely for $k_0 = 0, 1$ and 2. The lower part of the table gives the empirical frequencies when $m_0 = 2$, and $k_0 = 0, 1$ and 2. For the BM procedure we are only able to provide results when $T > N$. The relative performance of the three detection procedures very much depends on whether the observations are affected by an external factor, and the relative sizes of N and T. For example, the PS method works very well only if $m_0 = 1$ and $k_0 = 0$, and breaks down completely if there are external factors or if there is more than one pervasive unit. The BM method performs well when it is known that $m_0 \geq 1$ and $T > N$. By contrast, our proposed method works reasonably well for all values of m_0 and k_0, and continues to be applicable even if $T < N$. Amongst the three methods considered only the $SMT - \sigma^2$ thresholding method is able to select the true pervasive units with probability approaching unity as both N and T become large. Not surprisingly, the small sample performance of $SMT - \sigma^2$ thresholding deteriorates as the number of common factors, be it pervasive units or external factors, is increased. In Table 4 we again consider the average number of false discoveries. The results are similar to the ones obtained earlier, with $SMT-\sigma^2$ procedure performing best overall. It is also interesting to note that standardization of observations affect the BM procedure adversely. This is particularly pronounced when $m_0 = 2$. Again, the modification of the BM procedure is critical for its performance. When the BM procedure is applied without modification we again obtain a large number of false discoveries, as can be seen from the results in Section S7 of the online supplement.

The above findings continue to hold when the DGP contains weakly pervasive units instead of pervasive units. Table 5 reports the results for models with weakly pervasive rather than strong pervasive units, where the exponent of cross-sectional dependence of the pervasive unit(s), is set to $\alpha = 0.8$ instead of $\alpha = 1$. (see (45) for a definition of α). Not surprisingly, the empirical frequency of correctly identifying the true weakly pervasive units is generally lower as compared to the case where the pervasive units are strong. Nevertheless, $SMT-\sigma^2$ thresholding and BM procedure perform reasonably well even in this case. Of course, the BM method is applicable only in the case of panels with $T > N$ and if it is known that $m_0 > 0$. In cases where both BM and $SMT-\sigma^2$ thresholding are applicable, the proposed method seems to perform somewhat better, particularly when $T-N$ is not that large. Finally, considering the average number of non-pervasive units, falsely selected, in Table 6 we again note very similar patterns to those present in Table 4, with $SMT-\sigma^2$ again performing best.

7 Empirical Applications

In this section we present empirical applications that showcase our proposed detection methodology. We consider three different applications, and report the pervasive units (if any) selected by $SMT-\sigma^2$ thresholding, as well as the methods of Parker and Sul (2016) and Brownlees and Mesters (2019). As in the MC section, we focus on the modified version of the BM procedure (where selection is based only on the $N/2$ most connected units), but report results with and without standardization of the individual time series.[18]

7.1 U.S. industrial production

We begin with a panel of monthly observations on production of $N = 138$ industrial sectors of the U.S. economy over the period 1972m1-2007m12. This data set has been compiled by Foerster, Sarte, and Watson (2011), and used by Brownlees and Mesters (2019) to study the presence of pervasive production sectors in the U.S.[19] As noted previously, by construction BM method will end up finding at least one pervasive sector. In fact, Brownlees and Mesters (2019) find between 2 and 5 pervasive sectors, predominantly related to the production of light motor vehicles and aluminum products. They arrive at these results by applying their modified detection procedure to sectoral growth rates after standardization. In addition to determining which sectors are pervasive, the authors rank different sectors according to their level of dominance by ordering the column norms of the inverse sample covariance matrix. A comparison of this ranking with one based on the explanatory power of estimated common

[18] Estimation results for unmodified BM without restrictions on the maximum number of pervasive units can be found in Section S8 of the online supplement.

[19] In their study, Foerster, Sarte, and Watson (2011) make use of a quarterly version of this data set, and BM choose monthly frequency to ensure $T > N$, which their detection procedure requires.

Table 3: Empirical frequency of correctly identifying only the true strongly pervasive units ($m_0 > 0$, and $\alpha = 1$)

Part A: $m_0 = 1$

	$SMT-\sigma^2$				PS				BM				BM (standardized)			
$k_0=0$																
N \T	60	110	210	250	60	110	210	250	60	110	210	250	60	110	210	250
50	97.7	99.9	100	100	61.0	92.6	95.1	95.5	56.4	98.9	100	100	51.9	98.2	99.7	99.9
100	100	100	100	100	80.8	99.4	100	100	n/a	78.9	100	100	n/a	73.5	100	100
200	100	100	100	100	91.2	99.9	100	100	n/a	n/a	92.1	100	n/a	n/a	89.3	100
500	100	100	100	100	94.0	100	100	100	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
$k_0=1$																
N \T	60	110	210	250	60	110	210	250	60	110	210	250	60	110	210	250
50	58.9	80.6	82.9	82.3	0.5	0	0	0	50.3	97.6	100	100	31.0	76.4	91.0	93.3
100	68.1	88.4	93.3	93.0	0.1	0	0	0	n/a	72.9	100	100	n/a	52.1	99.0	99.8
200	79.1	97.8	99.6	99.5	0	0	0	0	n/a	n/a	88.6	100	n/a	n/a	72.8	99.9
500	82.1	99.9	100	100	0	0	0	0	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
$k_0=2$																
N \T	60	110	210	250	60	110	210	250	60	110	210	250	60	110	210	250
50	52.5	61.7	61.1	55.5	0	0	0	0	44.3	97.0	99.9	100	17.0	50.4	65.3	67.8
100	65.3	75.9	74.7	74.2	0	0	0	0	n/a	69.2	100	100	n/a	35.0	94.7	96.2
200	72.7	95.6	97.1	96.0	0	0	0	0	n/a	n/a	87.4	100	n/a	n/a	58.6	98.5
500	77.1	99.4	100	100	0	0	0	0	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

Part B: $m_0 = 2$

	$SMT-\sigma^2$				PS				BM				BM (standardized)			
$k_0=0$																
N \T	60	110	210	250	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	6.9	42.3	69.0	72.8
100	58.9	87.3	98.2	98.6	0	0	0	0	n/a	48.2	100	100	n/a	19.0	94.9	98.4
200	59.0	88.8	98.4	98.9	0.1	0	0	0	n/a	n/a	67.7	100	n/a	n/a	39.9	97.8
500	60.9	94.8	100	100	0	0	0	0	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
$k_0=1$																
N \T	60	110	210	250	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	1.8	8.9	18.2	19.4
100	41.7	78.5	91.5	92.4	0	0	0	0	n/a	39.9	99.9	100	n/a	6.8	59.0	63.5
200	43.5	87.6	98.3	99.3	0	0	0	0	n/a	n/a	57.2	99.9	n/a	n/a	16.9	75.7
500	46.0	96.2	100	100	0	0	0	0	n/a	n/a	n/a	n/a				
$k_0=2$																
N \T	60	110	210	250	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	0.4	1.3	3.2	2.4
100	48.4	73.3	79.6	79.6	0	0	0	0	n/a	36.4	100	99.9	n/a	2.2	25.0	29.9
200	47.5	86.9	96.8	97.1	0	0	0	0	n/a	n/a	52.3	99.7	n/a	n/a	6.9	46.8
500	41.0	94.6	99.9	100	0	0	0	0	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

Notes: See the notes to Table 1, where the $SMT-\sigma^2$ thresholding and the PS detection methods are described. BM refers to the detection methods proposed by Brownlees and Mesters (2019) and are implemented as set out in the online supplement. BM methods are not applicable (n/a) if $T < N$. m_0 is the true number of pervasive units, and α measures their degree of pervasiveness, as defined by (45).

Table 4: Average number of non-pervasive units falsely selected as pervasive units ($m_0 > 0$, and $\alpha = 1$)

Part A: $m_0 = 1$

	$SMT - \sigma^2$				PS				BM				BM (standardized)			
$k_0 = 0$																
N \T	60	110	210	250	60	110	210	250	60	110	210	250	60	110	210	250
50	0	0	0	0	0.7	0.2	0.2	0.2	1.3	0	0	0	1.5	0	0	0
100	0	0	0	0	0.2	0	0	0	n/a	0.5	0	0	n/a	0.6	0	0
200	0	0	0	0	0.1	0	0	0	n/a	n/a	0.2	0	n/a	n/a	0.3	0
500	0	0	0	0	0.1	0	0	0	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
$k_0 = 1$																
N \T	60	110	210	250	60	110	210	250	60	110	210	250	60	110	210	250
50	0.2	0.1	0.2	0.2	1.1	1.8	2.5	2.7	1.6	0	0	0	2.3	0.5	0.2	0.1
100	0.1	0.1	0.1	0.1	2.1	3.1	5.4	5.9	n/a	0.8	0	0	n/a	1.3	0	0
200	0	0	0	0	6.0	10.4	14.1	15.8	n/a	n/a	0.3	0	n/a	n/a	0.7	0
500	0	0	0	0	0	37.2	46.8	47.5	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
$k_0 = 2$																
N \T	60	110	210	250	60	110	210	250	60	110	210	250	60	110	210	250
50	0.4	0.3	0.4	0.5	1.8	3.0	3.7	3.8	1.8	0	0	0	2.9	1.3	0.9	0.8
100	0.1	0.2	0.3	0.3	2.9	4.3	7.0	7.5	n/a	0.9	0	0	n/a	1.9	0.1	0.1
200	0	0	0	0	7.8	13.4	16.5	17.0	n/a	n/a	0.3	0	n/a	n/a	1.1	0
500	0	0	0	0	0	41.2	46.7	46.9	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

Part B: $m_0 = 2$

	$SMT - \sigma^2$				PS				BM				BM (standardized)			
$k_0 = 0$																
N \T	60	110	210	250	60	110	210	250	60	110	210	250	60	110	210	250
50	0	0	0	0	0.7	1.2	1.5	1.5	0.8	0	0	0	1.8	0.3	0.2	0.2
100	0	0	0	0	1.0	1.6	3.1	3.3	n/a	0.2	0	0	n/a	0.8	0	0
200	0	0	0	0	3.5	6.7	9.5	10.7	n/a	n/a	0.1	0	n/a	n/a	0.3	0
500	0	0	0	0	0	25.3	35.7	38.0	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
$k_0 = 1$																
N \T	60	110	210	250	60	110	210	250	60	110	210	250	60	110	210	250
50	0.2	0.1	0.1	0.1	2.2	3.2	3.8	3.7	1.0	0	0	0	2.7	1.3	0.9	0.9
100	0	0	0	0	3.9	5.3	7.1	7.2	n/a	0.4	0	0	n/a	1.6	0.1	0.1
200	0	0	0	0	9.8	14.6	16.0	16.0	n/a	n/a	0.1	0	n/a	n/a	0.8	0
500	0	0	0	0	0	43.0	44.6	44.6	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
$k_0 = 2$																
N \T	60	110	210	250	60	110	210	250	60	110	210	250	60	110	210	250
50	0.3	0.3	0.3	0.4	2.7	3.6	3.7	3.8	1.0	0	0	0	3.0	2.4	2.4	2.3
100	0.1	0.1	0.2	0.2	5.1	6.9	8.1	8.0	n/a	0.4	0	0	n/a	2.3	0.4	0.3
200	0	0	0	0	10.4	16.0	16.5	16.2	n/a	n/a	0.1	0	n/a	n/a	1.3	0.1
500	0	0	0	0	0	43.6	44.1	44.1	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

Notes: See the notes to Tables 1, 2 and 3.

Table 5: Empirical frequency of correctly identifying only the true weakly pervasive units ($m_0 > 0$ with $\alpha = 0.8$)

Part A: $m_0 = 1$

	$SMT-\sigma^2$					PS					BM					BM (standardized)			
	$k_0 = 0$					$k_0 = 0$					$k_0 = 0$					$k_0 = 0$			
N \T	60	110	210	250	N \T	60	110	210	250	N \T	60	110	210	250	N \T	60	110	210	250
50	51.2	80.6	95.4	97.5	50	0	1.3	12.4	16.5	50	21.9	72.0	93.3	95.2	50	25.9	71.1	85.4	88.0
100	87.2	98.9	100	100	100	0	0.1	11.0	21.7	100	n/a	37.5	99.3	100	100	n/a	42.7	98.4	99.4
200	97.5	100	100	100	200	0	0.1	5.2	19.4	200	n/a	n/a	52.5	99.6	200	n/a	n/a	58.4	99.0
500	97.7	100	100	100	500	0	0	9.7	35.9	500	n/a	n/a	n/a	n/a	500	n/a	n/a	n/a	n/a
	$k_0 = 1$					$k_0 = 1$					$k_0 = 1$					$k_0 = 1$			
N \T	60	110	210	250	N \T	60	110	210	250	N \T	60	110	210	250	N \T	60	110	210	250
50	37.4	65.2	79.1	78.9	50	0.05	0	0	0	50	19.0	69.7	92.9	93.9	50	13.1	32.9	46.1	45.8
100	65.1	90.5	93.7	93.5	100	0	0	0	0	100	n/a	34.0	99.0	99.9	100	n/a	25.1	82.1	85.0
200	84.6	99.4	99.6	99.5	200	0	0	0	0	200	n/a	n/a	51.8	99.2	200	n/a	n/a	41.9	90.3
500	82.7	99.9	100	100	500	0	0	0	0	500	n/a	n/a	n/a	n/a	500	n/a	n/a	n/a	n/a
	$k_0 = 2$					$k_0 = 2$					$k_0 = 2$					$k_0 = 2$			
N \T	60	110	210	250	N \T	60	110	210	250	N \T	60	110	210	250	N \T	60	110	210	250
50	37.7	53.3	58.2	54.5	50	0	0	0	0	50	19.4	68.8	90.9	94.1	50	8.2	14.9	17.1	18.5
100	64.2	79.7	75.3	74.4	100	0	0	0	0	100	n/a	33.5	99.4	99.8	100	n/a	15.9	53.8	60.5
200	82.7	98.2	97.1	96.0	200	0	0	0	0	200	n/a	n/a	51.7	99.1	200	n/a	n/a	28.7	76.2
500	80.9	100	100	100	500	0	0	0	0	500	n/a	n/a	n/a	n/a	500	n/a	n/a	n/a	n/a

Part B: $m_0 = 2$

	$SMT-\sigma^2$					PS					BM					BM (standardized)			
	$k_0 = 0$					$k_0 = 0$					$k_0 = 0$					$k_0 = 0$			
N \T	60	110	210	250	N \T	60	110	210	250	N \T	60	110	210	250	N \T	60	110	210	250
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	50	1.8	5.5	9.3	11.5
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	100	n/a	3.8	44.0	50.9
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	200	n/a	n/a	11.1	64.5
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	500	n/a	n/a	n/a	n/a
	$k_0 = 1$					$k_0 = 1$					$k_0 = 1$					$k_0 = 1$			
N \T	60	110	210	250	N \T	60	110	210	250	N \T	60	110	210	250	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	50	0.5	0.7	0.9	0.7
100	16.5	61.3	88.8	91.5	100	0	0	0	0	100	n/a	12.1	94.5	96.4	100	n/a	1.8	12.1	11.9
200	11.4	61.8	94.7	97.1	200	0	0	0	0	200	n/a	n/a	20.3	92.5	200	n/a	n/a	3.9	29.9
500	1.8	32.3	84.6	91.7	500	0	0	0	0	500	n/a	n/a	n/a	n/a	500	n/a	n/a	n/a	n/a
	$k_0 = 2$					$k_0 = 2$					$k_0 = 2$					$k_0 = 2$			
N \T	60	110	210	250	N \T	60	110	210	250	N \T	60	110	210	250	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	50	0	0.2	0	0
100	26.6	65.3	79.3	80.2	100	0	0	0	0	100	n/a	11.8	92.5	96.0	100	n/a	0.7	1.6	1.5
200	17.6	75.7	96.1	96.6	200	0	0	0	0	200	n/a	n/a	19.5	91.4	200	n/a	n/a	1.9	10.0
500	2.4	36.5	89.2	93.8	500	0	0	0	0	500	n/a	n/a	n/a	n/a	500	n/a	n/a	n/a	n/a

Notes: The DGP and the detection methods are the same as those in Table 3, except that the pervasive units are generated with $\alpha = 0.8$, defined by (45).

Table 6: Average number of non-pervasive units falsely selected as pervasive units ($m_0 > 0$, and $\alpha = 0.8$)

Part A: $m_0 = 1$

$k_0 = 0$

	$SMT-\sigma^2$				PS				BM				BM (standardized)			
N \T	60	110	210	250	60	110	210	250	60	110	210	250	60	110	210	250
50	0	0	0	0	3.7	3.4	2.2	2.0	2.8	0.7	0.2	0.1	2.5	0.5	0.2	0.2
100	0	0	0	0	8.2	7.6	2.6	1.8	n/a	2.0	0	0	n/a	1.8	0	0
200	0	0	0	0	16.8	10.6	3.5	1.9	n/a	n/a	1.3	0	n/a	n/a	1.1	0
500	0	0	0	0	41.6	22.3	2.8	1.2	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

$k_0 = 1$

	$SMT-\sigma^2$				PS				BM				BM (standardized)			
N \T	60	110	210	250	60	110	210	250	60	110	210	250	60	110	210	250
50	0.2	0.2	0.2	0.2	2.5	3.4	4.2	4.3	2.9	0.8	0.2	0.1	2.9	1.7	1.3	1.3
100	0.1	0.1	0.1	0.1	3.6	4.9	6.8	7.1	n/a	2.3	0	0	n/a	2.2	0.3	0.2
200	0	0	0	0	6.4	10.9	14.2	15.0	n/a	n/a	1.4	0	n/a	n/a	1.6	0.1
500	0	0	0	0	0.1	29.0	39.4	40.3	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

$k_0 = 2$

	$SMT-\sigma^2$				PS				BM				BM (standardized)			
N \T	60	110	210	250	60	110	210	250	60	110	210	250	60	110	210	250
50	0.5	0.4	0.5	0.5	2.7	4.0	4.6	4.6	3.0	0.8	0.2	0.1	3.1	2.7	2.8	2.7
100	0.2	0.2	0.3	0.3	3.8	5.3	7.6	8.0	n/a	2.2	0	0	n/a	2.5	1.1	0.9
200	0	0	0	0	7.8	12.8	15.3	15.7	n/a	n/a	1.3	0	n/a	n/a	2.0	0.5
500	0	0	0	0	0	35.6	40.5	40.6	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

Part B: $m_0 = 2$

$k_0 = 0$

	$SMT-\sigma^2$				PS				BM				BM (standardized)			
N \T	60	110	210	250	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	2.5	1.9	2.2	2.1
100	0	0	0	0	7.0	4.5	1.4	1.1	n/a	1.3	0	0	n/a	1.8	0.4	0.4
200	0	0	0	0	12.6	5.7	2.1	1.3	n/a	n/a	0.6	0	n/a	n/a	1.0	0.1
500	0	0	0	0	32.9	11.0	2.1	1.2	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

$k_0 = 1$

	$SMT-\sigma^2$				PS				BM				BM (standardized)			
N \T	60	110	210	250	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	2.9	3.4	4.0	4.2
100	0.2	0.1	0.1	0.1	6.2	7.0	7.3	7.2	n/a	1.4	0	0	n/a	2.3	1.5	1.4
200	0	0	0	0	12.1	14.4	14.9	14.9	n/a	n/a	0.8	0	n/a	n/a	1.7	0.5
500	0	0	0	0	0.1	36.8	37.6	37.5	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

$k_0 = 2$

	$SMT-\sigma^2$				PS				BM				BM (standardized)			
N \T	60	110	210	250	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	3.3	4.1	5.0	5.0
100	0.3	0.2	0.2	0.2	6.1	7.0	8.0	8.2	n/a	1.4	0	0	n/a	3.0	3.1	3.0
200	0.1	0	0	0	11.4	14.8	15.2	15.3	n/a	n/a	0.8	0	n/a	n/a	2.1	1.1
500	0	0	0	0	0	37.0	38.0	38.1	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

Notes: See the notes to Table 5.

factors on sector-specific series is provided, revealing substantial differences in the suggested list of highly influential sectors.

We apply all the three detection methods to the full dataset as well as to the two sub-samples, 1972m1–1983m12 and 1984m1–2007m12, investigated in Foerster, Sarte, and Watson (2011). For application of the PS method we selected the number of factors using the IC_{p2} criterion of Bai and Ng (2002). We set the maximum number of factors to 10 and obtain 1 common factor for the full sample and the first sub-sample, and 2 common factors for the second sub-sample. In application of the $SMT - \sigma^2$ we do not need to estimate the number of factors, but set a maximum value for $p = m + k$. To this end and to cover a wide range of possible factors, and to check the robustness of the $SMT - \sigma^2$ thresholding to the choice of p_{max}, we tried all the values of p_{max} in the range $\{2, 3, 4, 5, 6\}$.

The results are summarized in Table 7. The top panel of the table gives the results for the full sample, followed by the two sub-sample results. Starting with $SMT - \sigma^2$ thresholding, we find that no sector is identified as pervasive, with the result being robust to the choice of p_{max} and the sample period. This conclusion is in line with the estimates obtained by Pesaran and Yang (2019) who make use of input-output tables for the whole U.S. economy. The PS procedure arrives at the same outcome and does not detect any pervasive sector when the full sample is used, but identifies Plastic Products as pervasive in the first sub-sample, and as many as 19 sectors as pervasive in the second sub-sample. The list of these 19 sectors is given at the bottom of Table 7, and includes a diverse array of sectors such as Cheese, Breweries, Plastic Products, Shipping Containers, and more.

The results from the application of the BM procedure are mixed and depend on whether the observations are standardized, and the sample period considered.[20] Still, due to the underlying assumptions of this approach, at least one pervasive sector is found in all cases. As can be seen from the last two columns of Table 7, for the full sample BM selects Fluid Milk as the pervasive sector if observations are not standardized, and selects Automobiles and Light Duty Motor Vehicles, and Motor Vehicle Parts, as pervasive when observations are standardized. For the two sub-samples the results are much more dispersed, and only Motor Vehicle Parts is included in the list of the pervasive sectors for all sub-samples when the observations are standardized.

In addition to splitting the sample at the end of 1983, we also applied our detection method to rolling samples with window sizes of 10, 12, 15 and 20 years, in order to obtain further evidence on how the number and identity of pervasive units could be subject to change. As previously, the maximum admissible number of common factors and pervasive units is set to $p_{max} \in \{2, 3, 4, 5, 6\}$. For the sake of brevity, only $SMT - \sigma^2$ thresholding is considered. The results unanimously confirm our previous finding that there is no pervasive sector in the U.S. industrial production.

7.2 Are there pervasive economies or equity markets in the global economy?

In a second application, we use quarterly observations on *real GDP* and *real equity prices* over a number of countries and equity markets spanning the period 1979q2-2016q4, providing $T = 151$

[20] The detection outcomes also very much depend on whether one uses the modification of the BM procedure or not. The results for unmodified BM is in Section S8 of the online supplement.

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

	Full sample (1972m1 - 2007m12)			
Approach:	$SMT-\sigma^2$	PS	BM	BM (standardized)
p_{max}	2, 3, 4, 5, 6	$1^{\dagger}$		
Number of pervasive units:	0	0	1	2
Identities:			Fluid Milk	Automobiles and Light Duty Motor Vehicles; Motor Vehicle Parts
	Sub-sample A (1972m1 - 1983m12)			
Approach:	$SMT-\sigma^2$	PS	BM	BM (standardized)
p_{max}	2, 3, 4, 5, 6	$1^{\dagger}$		
Number of pervasive units:	0	1	2	1
Identities:		Plastics Products	Commercial and Service Industry Machinery; Bakeries and Tortilla	Motor Vehicle Parts
	Sub-sample B (1984m1 - 2007m12)			
Approach:	$SMT-\sigma^2$	PS	BM	BM (standardized)
p_{max}	2, 3, 4, 5, 6	$2^{\dagger}$		
Number of pervasive units:	0	19	12	5
Identities:		*	**	Motor Vehicle Parts; Automobiles and Light Duty Motor Vehicles; Aluminum Extruded Products; Miscellaneous Aluminum Materials; Motor Vehicle Bodies

†: This value minimizes the IC_{p2} criterion of Bai and Ng (2002) for selecting the number of common factors. The 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, Sarte, and Watson (2011).

observations for each country.[21] Data on real GDP is available for 33 countries and account for over 90 percent of global output. The equity price observations are available for 26 countries, and include all major equity markets.

7.2.1 Cross country output growth

A recent investigation of cross country correlation of real GDP growth rates is given in Cesa-Bianchi, Pesaran, and Rebucci (2019), and shows that accounting for one common factor is enough to reduce average pairwise cross country correlations to almost zero. Despite this suggestive evidence for the presence of only one factor in GDP, we consider a wider set of choices concerning the number of latent factors, and experiment with $p_{max} \in \{2, 3, 4, 5, 6\}$ when applying σ^2 thresholding. As in the previous application, the results from the application of $SMT - \sigma^2$ thresholding are compared to the other two detection procedures (BM and BM standardized as well as PS). The results are summarized in Table 8. In this application $SMT - \sigma^2$ thresholding selects 1 country (France) as pervasive in terms of GDP growth when $p_{max} = 3, 4$ or 5, and selects no pervasive country if $p_{max} = 2$ or 6. Given the cross country growth evidence provided by Cesa-Bianchi, Pesaran, and Rebucci (2019) it is more reasonable to rely on the detection evidence when $p_{max} = 2$, which is compatible with assuming one common global technology factor (i.e. $k_0 = 1$) with one possible pervasive country, say U.S., (with $m_0 = 1$) which gives $p_{max} = 2$. Also if we use the IC_{p2} criterion of Bai and Ng (2002) to select the number of factors across country growth rates we also end up with one factor. (see footnote 1 of Table 8). So we conclude that there is no compelling evidence for the presence of a pervasive country in terms of output growth, and the detection of France as a pervasive economy when $p_{max} = 3, 4$ and 5, is most likely a false discovery. This conclusion is also supported when we consider the result obtained from the application of the PS procedure to the GDP growth series. In contrast, BM procedure selects France and Spain as pervasive economies when the growth series are not standardized, and selects an additional 9 economies (a total of 11 economies out of 33) as pervasive, if observations are standardized. This outcome is difficult to interpret and most likely reflects the tendency of the BM procedure to over-select as documented in the MC section.

7.2.2 Cross market rate of change of real equity prices

The results for the rate of change of real equity prices are summarized in the lower panel of Table 8. In this application $SMT - \sigma^2$ thresholding is the only method not identifying any of the equity markets as pervasive. Both PS and BM procedures select 6 markets as pervasive, and agree only on Germany and Netherlands as the pervasive equity markets. Interestingly enough, BM only selects Netherlands as pervasive when observations are standardized. Once again we find the BM detection method to be highly sensitive to standardization of observations.

Finally, it is important to bear in mind, that not finding a pervasive unit does not mean that the global economy is not subject to global shocks. Our results suggest that once we allow for the possibility of global shocks, it is difficult to find convincing evidence that any country can be singled out as pervasive. This result is also compatible with the presence of influential economies such as U.S., China, Japan and Germany as having important global and regional impacts in the world economy.

[21] Cross country data is taken from the latest vintage of the GVAR data set as described in Mohaddes and Raissi (2018).

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:	$SMT-\sigma^2$			PS	BM	BM (standardized)
p_{max}	2	$\{3,4,5\}$	6	1†		
Number of pervasive units:	0	1	0	0	2	11
Identities:		France			France Spain	Italy, UK Spain, Malaysia France, Belgium USA, Finland Germany, South Africa Canada

	Rate of change of real equity prices			
Approach:	$SMT-\sigma^2$	PS	BM	BM (standardized)
p_{max}	2, 3, 4, 5, 6	2†		
Number of pervasive units:	0	6	6	1
Identities:		France Germany Malaysia Netherlands Singapore Thailand	USA Netherlands UK Canada Switzerland Germany	Netherlands

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

Note: Data taken from the GVAR database (Mohaddes and Raissi, 2018).

7.3 U.S. house price changes

It is well established that house price changes in the U.S. are governed by common national and regional factors (see e.g. Holly, Pesaran, and Yamagata, 2010; Bailey, Holly, and Pesaran, 2016), and it is of interest to investigate if any of these common factors are due to the dominance of particular states amongst the 48 mainland states of the U.S.. To this end we consider state-level quarterly data on real house prices over the 1975q1–2014q4 period ($T = 160$).[22] In our analysis we use the rate of change of real house prices, after seasonal adjustment, with nominal house prices deflated by the state-level consumer price indices.

To investigate whether house price changes in any of 48 mainland U.S. states could be regarded as pervasive for the rest of the states, as in the previous applications, we implement $SMT-\sigma^2$ thresholding with $p_{max} = \{2,3,4,5,6\}$. The PS and BM methods are applied as before. The results are summarized in Table 9. As can be seen there are significant differences in the outcomes depending on the method used. In the case of $SMT-\sigma^2$ thresholding New York is identified as pervasive when the maximum number of common factors is set to 2 and 3. No pervasive unit is found for $p_{max} \in \{4,5,6\}$, and Kentucky is also selected as pervasive when $p_{max} = 3$, which could be false discovery. The BM procedure identifies many more states as pervasive with no clear geographical patterns. Without standardization, BM selects North

[22] House price data is taken from Freddie Mac House Price Indexes (`http://www.freddiemac.com/research/indices/house-price-index.html`). State-level consumer price indexes were taken from Yang (2018) who updated a previously constructed dataset of Bailey, Holly, and Pesaran (2016).

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

Approach:	$SMT - \sigma^2$			PS	BM	BM (standardized)
p_{max}	2	3	4, 5, 6	5[†]		
Number of pervasive units:	1	2	0	2	4	6
Identities:	New York	Kentucky New York		New Hampshire Nevada	North Carolina Maryland Virginia Connecticut	Connecticut New Hampshire Massachusetts Maryland Virginia Rhode Island

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

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

Carolina, Maryland, Virginia and Connecticut as pervasive, whilst with standardization three additional states are selected as pervasive, namely New Hampshire, Massachusetts and Rhode Island. Connecticut is not selected when we use BM (standardized). We take these results as weak evidence for the influential role of the north-eastern part of the United States with New York being the most plausible candidate. By contrast, PS detects two pervasive units in two opposite corners of the U.S., namely New Hampshire and Nevada, thus providing a less coherent picture compared to the other two approaches.

8 Concluding remarks

Recent developments in network and panel literature have emphasized the importance of some key units for interdependencies among economic agents. For example, financial networks can be resilient with no units playing an unduly important (i.e. 'systemic') role while others may have pervasive units that need close monitoring. There is a small literature on how to detect such units but all existing methods are either not rigorously analyzed or have drawbacks such as assuming, rather that ascertaining, the presence of at least one pervasive unit, or considering panel datasets where the time dimension is larger than the number of cross-section units.

We contribute to this literature by proposing a new thresholding method which is rigorously developed using theory on large factor models as well as recent developments on multiple testing. It has good small sample properties and allows for the presence of no pervasive units while being able to detect weakly influential cross-section entities. The ideas developed in this paper can also be applied to other aspects of the dependence across units. Tail dependence is particularly relevant in risk analysis and it would be interesting to develop detection methods that identify units with pervasive effects by quantiles.

Appendix

This appendix provides the proofs of the main results. The auxiliary lemmas used are stated and proven in Section S1 of the online supplement.

A Proof of main results

A.1 Analyzing residual variances for non-pervasive units

Recalling that $\mathbf{x}_i = \mathbf{F}_0\mathbf{a}_i + \mathbf{v}_i$ and

$$\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{F}_0 = \frac{\mathbf{M}_{\hat{\mathbf{F}}}\left(\mathbf{F}_0 - \hat{\mathbf{F}}\right)\mathbf{A}_0'\left(\mathbf{A}_0 - \hat{\mathbf{A}}\right)}{N} + \frac{\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{V}\left(\mathbf{A}_0 - \hat{\mathbf{A}}\right)}{N} - \mathbf{M}_{\hat{\mathbf{F}}}\left(\frac{\mathbf{V}\mathbf{A}_0}{N}\right), \tag{46}$$

we can expand residual variances, $\hat{\sigma}_i^2 = T^{-1}\mathbf{x}_i'\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{x}_i$, into the sum of eight terms.

$$N\hat{\sigma}_i^2 = B_{i1} + B_{i2} + \ldots B_{i6} + B_{i7} + B_{i8}.$$

The final two terms, B_{i7} and B_{i8}, are zero if i is the pervasive unit. The first 6 terms are defined and discussed in the proof of Lemma 1 in the online supplement. In this section, we are concerned with the last two terms which only appear in the case of non-pervasive units. These terms are given by

$$B_{i7} = \frac{N\mathbf{v}_i'\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{v}_i}{T}, \text{ and } B_{i8} = 2\frac{N\mathbf{v}_i'\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{F}_0\mathbf{a}_i}{T}.$$

First, note that

$$B_{i7} = N\left(\frac{\mathbf{v}_i'\mathbf{M}_F\mathbf{v}_i}{T}\right) + \frac{N\mathbf{v}_i'\left(\mathbf{M}_{\hat{\mathbf{F}}} - \mathbf{M}_F\right)\mathbf{v}_i}{T}.$$

Since $T^{-1}\mathbf{v}_i'\mathbf{M}_F\mathbf{v}_i = O_p(1)$, it holds that $B_{i7} = O_p(N)$. Consider now B_{i8}, and note that, using (46), we obtain

$$\begin{aligned}\mathbf{v}_i'\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{F}_0\mathbf{a}_i &= \frac{\mathbf{v}_i'\mathbf{M}_{\hat{\mathbf{F}}}\left(\mathbf{F}_0 - \hat{\mathbf{F}}\right)\mathbf{A}_0'\left(\mathbf{A}_0 - \hat{\mathbf{A}}\right)\mathbf{a}_i}{N} + \frac{\mathbf{v}_i'\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{V}\left(\mathbf{A}_0 - \hat{\mathbf{A}}\right)\mathbf{a}_i}{N} - \frac{\mathbf{v}_i'\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{V}\mathbf{A}_0\mathbf{a}_i}{N}\\ &= \frac{B_{i81}}{N} + \frac{B_{i82}}{N} - \frac{B_{i83}}{N}.\end{aligned}$$

We examine the terms B_{i81}, B_{i82} and B_{i83}, in turn. For B_{i81} we have

$$\left|\mathbf{v}_i'\mathbf{M}_{\hat{\mathbf{F}}}\left(\mathbf{F}_0 - \hat{\mathbf{F}}\right)\mathbf{A}_0'\left(\mathbf{A}_0 - \hat{\mathbf{A}}\right)\mathbf{a}_i\right| \leq \|\mathbf{v}_i\|\,\|\mathbf{M}_{\hat{\mathbf{F}}}\|\left\|\mathbf{F}_0 - \hat{\mathbf{F}}\right\|\left\|\mathbf{A}_0'\left(\mathbf{A}_0 - \hat{\mathbf{A}}\right)\right\|\|\mathbf{a}_i\|.$$

Recall that by Assumption 2, $\|\mathbf{a}_i\| = O_p(1)$, whereas results (A) and (E) yield

$$\left\|\mathbf{F}_0 - \hat{\mathbf{F}}\right\|_F = O_p\left(\frac{\sqrt{T}}{\sqrt{\min(N,T)}}\right),$$

and

$$\left\|\mathbf{A}_0'\left(\mathbf{A}_0 - \hat{\mathbf{A}}\right)\right\|_F = O_p\left(\frac{N}{\sqrt{\min(N,T)}}\right).$$

Furthermore, since $\mathbf{M}_{\hat{\mathbf{F}}}$ is an idempotent matrix we also have $\|\mathbf{M}_{\hat{\mathbf{F}}}\| = O_p(1)$. Lastly, note that $\|\mathbf{v}_i\| = O_p\left(\sqrt{T}\right)$ holds by Assumption 3 and the fact that $\mathbf{v}_i = \mathbf{u}_i$ for any non pervasive unit, i. Consequently,

$$|B_{i81}| = \left|\mathbf{v}_i'\mathbf{M}_{\hat{\mathbf{F}}}\left(\mathbf{F}_0 - \hat{\mathbf{F}}\right)\mathbf{A}_0'\left(\mathbf{A}_0 - \hat{\mathbf{A}}\right)\mathbf{a}_i\right| = O_p\left(\frac{NT}{\min(N,T)}\right).$$

Next,

$$|B_{i82}| = \left|\mathbf{v}_i'\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{V}\left(\mathbf{A}_0 - \hat{\mathbf{A}}\right)\mathbf{a}_i\right| \leq \|\mathbf{v}_i\| \|\mathbf{M}_{\hat{\mathbf{F}}}\| \left\|\mathbf{V}\left(\mathbf{A}_0 - \hat{\mathbf{A}}\right)\right\| \|\mathbf{a}_i\|.$$

Again, recall that by result (C),

$$\left\|\mathbf{V}\left(\mathbf{A}_0 - \hat{\mathbf{A}}\right)\right\|_F = O_p\left(\frac{\sqrt{NT}}{\sqrt{\min(N,T)}}\right).$$

So

$$|B_{i82}| = \left|\mathbf{v}_i'\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{V}\left(\mathbf{A}_0 - \hat{\mathbf{A}}\right)\mathbf{a}_i\right| = O_p\left(\frac{T\sqrt{N}}{\sqrt{\min(N,T)}}\right).$$

Next, $|B_{i83}| = |\mathbf{v}_i'\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{V}\mathbf{A}_0\mathbf{a}_i| \leq \|\mathbf{v}_i\| \|\mathbf{M}_{\hat{\mathbf{F}}}\| \|\mathbf{V}\mathbf{A}_0\| \|\mathbf{a}_i\|$. Here, (D) yields $\|\mathbf{V}\mathbf{A}_0\|_F = O_p\left(\sqrt{NT}\right)$, and $|B_{i83}| = O_p\left(\sqrt{N}T\right)$. Overall,

$$\begin{aligned} B_{i8} &= \frac{1}{T}\left(B_{i81} + B_{i82} - B_{i83}\right) \\ &= O_p\left(\frac{N}{\min(N,T)}\right) + O_p\left(\frac{\sqrt{N}}{\sqrt{\min(N,T)}}\right) + O_p\left(\sqrt{N}\right) = o_p(N), \end{aligned}$$

as required.

A.2 Proof of Theorem 2

We need to show that

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

It suffices to show that

$$\lim_{N,T\to\infty} \Pr\left(\cap_{i\in I_D}\left\{\hat{\sigma}_{iT}^2 \leq \frac{2\hat{\eta}_{iN}^2 \log T}{N}\right\}\right) = 1,$$

and

$$\lim_{N,T\to\infty} \Pr\left(\cup_{i\in I_{ND}}\left\{\hat{\sigma}_{iT}^2 \leq \frac{2\hat{\eta}_{iN}^2 \log T}{N}\right\}\right) = 0.$$

Let

$$\eta_{iN}^2 = \frac{\mathbf{a}_i'\mathbf{A}_0'\mathbf{\Sigma}_v\mathbf{A}_0\mathbf{a}_i}{N}, \text{ and } C_{iNT} = \frac{2\eta_{iN}^2 \log T}{N}.$$

Then, we need to show equivalently that

$$\lim_{N,T\to\infty} \Pr\left(\cap_{i\in I_D}\left\{\hat{\sigma}_{iT}^2 + \frac{2\log T}{N}\left(\eta_{iN}^2 - \hat{\eta}_{iN}^2\right) \le C_{iNT}\right\}\right) = 1,$$

and

$$\lim_{N,T\to\infty} \Pr\left(\cup_{i\in I_{ND}}\left\{\hat{\sigma}_{iT}^2 + \frac{2\log T}{N}\left(\eta_{iN}^2 - \hat{\eta}_{iN}^2\right) \le C_{iNT}\right\}\right) = 0.$$

The proof of Lemma 1 in the online supplement establishes the expansion $N\hat{\sigma}_{iT}^2 = \sum_{j=1}^6 B_{ij}$ for $i \in I_D$, where all terms apart from $B_{i1} = (NT)^{-1}\mathbf{a}_i'\mathbf{A}_0'\mathbf{V}'\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{V}\mathbf{A}_0\mathbf{a}_i$ converge to zero. More specifically, we have

$$N\hat{\sigma}_{iT}^2 = \frac{\mathbf{a}_i'\mathbf{A}_0'\mathbf{V}'\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{V}\mathbf{A}_0\mathbf{a}_i}{NT} + O_p\left(\frac{1}{\delta_{NT}}\right),$$

as long as $\frac{\sqrt{N}}{T} \to 0$. Moreover, the terms B_{ij}, $j = 1, \ldots, 6$ depend on i only through a_i and by assumption $\sup_i \|\mathbf{a}_i\|^2 < C < \infty$. Therefore, it follows immediately that

$$\lim_{N,T\to\infty} \Pr\left(\sup_i |B_{ij}| \le D_{NT}\right) = 1, \;\; j = 1, 2, \ldots, 6,$$

for any sequence D_{NT} bounded away from zero.

If $i \in I_{ND}$ then $N\hat{\sigma}_{iT}^2 = \sum_{j=1}^8 B_{ij}$, with

$$B_{i7} = \frac{N\mathbf{v}_i'\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{v}_i}{T}, \quad \text{and} \quad B_{i8} = \frac{N\mathbf{a}_i'\mathbf{F}_0'\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{v}_i}{T},$$

and as shown in Appendix A.1, $B_{i7} = O_p(N)$ and $B_{i8} = o_p(N)$, respectively. Further, we need to show that

$$\lim_{N,T\to\infty} \Pr\left(\cap_{i\in I_{ND}}\left\{|B_{i7} + B_{i8}| > C_{iNT}\right\}\right) = 1, \tag{47}$$

and it will be sufficient (assuming $NC_{iNT} = o\left(\min(\sqrt{N}, T)^a\right)$) to show that

$$\lim_{N,T\to\infty} \sum_{i\in I_{ND}} \Pr\left(|B_{i7}| < \min(\sqrt{N}, T)^a\right) = 0,$$

and

$$\lim_{N,T\to\infty} \sum_{i\in I_{ND}} \Pr\left(|B_{i8}| > \min(\sqrt{N}, T)^a\right) = 0,$$

for some $0 < a < 1$. This result follows straightforwardly by noting from a direct application of Lemma A7 of Chudik, Kapetanios, and Pesaran (2018) that

$$\Pr\left(\left|\mathbf{v}_i'\mathbf{M}_{\mathbf{F}}\mathbf{v}_i - T\sigma_{v_i}^2\right| > TNC_{NT}\right) \le \exp\left(-CTN^2C_{NT}^2\right) = \exp\left[-CT\eta_{iN}^4 \left(\log T\right)^2\right],$$

for some $C > 0$. It is easily seen that $N\exp\left(-CT\eta_{iN}^4 \left(\log T\right)^2\right) = o(1)$, noting that $sup_i\left(\eta_{iN}^4\right) > 0$. A similar result obtains for $\Pr\left(|\mathbf{v}_i'\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{v}_i - \mathbf{v}_i'\mathbf{M}_{\mathbf{F}}\mathbf{v}_i| > TNC_{NT}\right)$, along the lines of our analysis below for η_{iN} starting with (50).

To complete the proof it now suffices to show that

$$\lim_{N,T\to\infty} \Pr\left(\cap_{i=1,2,\ldots,N}\left\{|B_{i1}| \le NC_{iNT}\right\}\right) = 1, \tag{48}$$

and

$$\lim_{N,T\to\infty} \Pr\left(\cap_{i=1,2,\ldots,N}\left\{\left|\eta_{iN}^2-\hat{\eta}_{iN}^2\right|\le\eta_{iN}^2\right\}\right)=1, \tag{49}$$

or $\lim_{N,T\to\infty}\Pr\left(\sup_i\left|\eta_{iN}^2-\hat{\eta}_{iN}^2\right|\le C\right)=1$,for some finite $C>0$, since η_{iN}^2 is uniformly bounded away from zero and infinity. (48) follows from auxiliary Lemmas 3-6.

Consider now (49), and note that by equations (33) and (38), we have

$$\eta_{iN}^2-\hat{\eta}_{iN}^2=\frac{\mathbf{a}_i'\mathbf{A}_0'\mathbf{\Sigma}_v\mathbf{A}_0\mathbf{a}_i}{N}-\frac{\hat{\mathbf{a}}_i'\hat{\mathbf{A}}'\tilde{\mathbf{\Sigma}}_v\hat{\mathbf{A}}\hat{\mathbf{a}}_i}{N}. \tag{50}$$

Then,

$$\begin{aligned}
&\left|\eta_{iN}^2-\hat{\eta}_{iN}^2\right|\le\\
&\quad C_1\left|\frac{\mathbf{a}_i'\mathbf{A}_0'\mathbf{\Sigma}_v\mathbf{A}_0\left(\mathbf{a}_i-\hat{\mathbf{a}}_i\right)}{N}\right|+C_2\left|\frac{\mathbf{a}_i'\mathbf{A}_0'\left(\tilde{\mathbf{\Sigma}}_v-\mathbf{\Sigma}_v\right)\mathbf{A}_0\mathbf{a}_i}{N}\right|+C_3\left|\frac{\mathbf{a}_i'\mathbf{A}_0'\mathbf{\Sigma}_v\left(\hat{\mathbf{A}}-\mathbf{A}_0\right)\mathbf{a}_i}{N}\right|,\\
&\quad=A_{i1}+A_{i2}+A_{i3},
\end{aligned}$$

where A_{i2} and A_{i3} depend on i only via $\mathbf{a}_i$. By the boundedness of $\mathbf{a}_i$, auxiliary Lemmas 3-6, and Theorem 1 of Bailey, Pesaran, and Smith (2019), $A_{i2}=o_p(1)$ and $A_{i3}=o_p(1)$. Hence

$$\lim_{N,T\to\infty}\Pr\left(\sup_i A_{i2}\le C\right)=1,\text{ and }\lim_{N,T\to\infty}\Pr\left(\sup_i A_{i3}\le C\right)=1.$$

Now consider A_{i1}, and note all elements of $N^{-1}\mathbf{a}_i'\mathbf{A}_0'\mathbf{\Sigma}_v\mathbf{A}_0$ are uniformly bounded (recall that $\sup_i\left\|\mathbf{a}_i\right\|^2<C<\infty$, and $\mathbf{\Sigma}_v$ is row bounded). Therefore, it suffices to show that $\lim_{N,T\to\infty}\Pr\left(\sup_i\left\|\mathbf{a}_i-\hat{\mathbf{a}}_i\right\|\le C\right)=1$.We have

$$\mathbf{a}_i-\hat{\mathbf{a}}_i=\frac{1}{T}\sum_{t=1}^{T}\mathbf{f}_t v_{it}+\frac{1}{T}\sum_{t=1}^{T}x_{it}\left(\hat{\mathbf{f}}_t-\mathbf{f}_t\right).$$

So we need to show that

$$\lim_{N,T\to\infty}\Pr\left(\sup_i\left\|\sum_{t=1}^{T}\mathbf{f}_t v_{it}\right\|\le TC\right)=1, \tag{51}$$

and

$$\lim_{N,T\to\infty}\Pr\left[\sup_i\left\|\sum_{t=1}^{T}x_{it}\left(\hat{\mathbf{f}}_t-\mathbf{f}_t\right)\right\|\le TC\right]=1. \tag{52}$$

(51) follows easily. We focus on (52). For example, by (A1) of Bai (2003) we note that

$$\hat{f}_{jt}-f_{jt}=\frac{1}{T}\sum_{l=1}^{T}\hat{f}_{jl}\gamma_{lt}+\frac{1}{T}\sum_{l=1}^{T}\hat{f}_{jl}\zeta_{lt}+\frac{1}{T}\sum_{l=1}^{T}\hat{f}_{jl}\varkappa_{lt}+\frac{1}{T}\sum_{l=1}^{T}\hat{f}_{jl}\xi_{lt}, \tag{53}$$

where $\gamma_{lt}=\gamma_{N,lt}=N^{-1}\sum_{i=1}^{N}E(v_{it}v_{il})$, $\zeta_{lt}=N^{-1}\mathbf{v}_l'\mathbf{v}_t-\gamma_{lt}$, $\varkappa_{lt}=N^{-1}\mathbf{f}_l'\mathbf{A}_0'\mathbf{v}_t$, and $\xi_{lt}=\varkappa_{tl}$. So we need to show the following (C changes from instance to instance).

$$\lim_{N,T\to\infty}\Pr\left[\sup_i\left|\sum_{t=1}^{T}x_{it}\left(\frac{1}{T}\sum_{l=1}^{T}\hat{f}_{jl}\gamma_{lt}\right)\right|\le TC\right]=1, \tag{54}$$

$$\lim_{N,T\to\infty} \Pr\left[\sup_i \left|\sum_{t=1}^{T} x_{it}\left(\frac{1}{T}\sum_{l=1}^{T}\hat{f}_{jl}\zeta_{lt}\right)\right| \leq TC\right] = 1, \tag{55}$$

and

$$\lim_{N,T\to\infty} \Pr\left[\sup_i \left|\sum_{t=1}^{T} x_{it}\left(\frac{1}{T}\sum_{l=1}^{T}\hat{f}_{jl}\varkappa_{lt}\right)\right| \leq TC\right] = 1. \tag{56}$$

We proceed in turn.

$$\Pr\left[\sup_i \left|\sum_{t=1}^{T} x_{it}\left(\frac{1}{T}\sum_{l=1}^{T}\hat{f}_{jl}\gamma_{lt}\right)\right| > TC\right] \leq \Pr\left[\sup_i \left|\frac{1}{T}\sum_{t=1}^{T}\sum_{l=1}^{T} x_{it}\left(\hat{f}_{jl}-f_{jl}\right)\gamma_{lt}\right| > TC\right] +$$
$$\Pr\left(\sup_i \left|\frac{1}{T}\sum_{t=1}^{T}\sum_{l=1}^{T} x_{it}f_{jl}\gamma_{lt}\right| > TC\right) = A_{11i} + A_{12i}.$$

We have, for some $0 < a < 1$,

$$A_{11i} = \Pr\left(\sup_i \left|\frac{1}{T}\sum_{t=1}^{T}\sum_{l=1}^{T} x_{it}\left(\hat{f}_{jl}-f_{jl}\right)\gamma_{lt}\right| > TC\right) \leq$$
$$\Pr\left\{\begin{array}{c} \frac{1}{T^{1/2}}\left[\frac{\min(N,T)^a}{T}\sum_{l=1}^{T}\left(\hat{f}_{jl}-f_{jl}\right)^2\right]^{1/2}\left[\frac{1}{T}\sum_{t=1}^{T}\sum_{l=1}^{T}\gamma_{lt}^2\right]^{1/2} \\ \sup_i\left[\frac{\min(N,T)^{-a}}{T}\sum_{t=1}^{T}x_{it}^2\right]^{1/2} > C \end{array}\right\}$$
$$\leq \Pr\left\{\frac{1}{T^{1/2}}\left[\frac{\min(N,T)^a}{T}\sum_{l=1}^{T}\left(\hat{f}_{jl}-f_{jl}\right)^2\right]^{1/2}\left[\frac{1}{T}\sum_{t=1}^{T}\sum_{l=1}^{T}\gamma_{lt}^2\right]^{1/2} > C\right\} +$$
$$\Pr\left[\sup_i\left(\frac{\min(N,T)^{-a}}{T}\sum_{t=1}^{T}x_{it}^2\right)^{1/2} > C\right].$$

But using Theorem 1 in Bai and Ng (2002),

$$\frac{1}{T^{1/2}}\left[\frac{\min(N,T)^a}{T}\sum_{l=1}^{T}\left(\hat{f}_{jl}-f_{jl}\right)^2\right]^{1/2}\left[\frac{1}{T}\sum_{t=1}^{T}\sum_{l=1}^{T}\gamma_{lt}^2\right]^{1/2} = o_p(1),$$

and using Lemma 2 to show

$$\Pr\left[\sup_i\left(\frac{\min(N,T)^{-a}}{T}\sum_{t=1}^{T}x_{it}^2\right)^{1/2} > C\right] = o(1).$$

Hence, it follows that $A_{11i} = o(1)$. Next

$$A_{12i} = \Pr\left(\sup_i \left|\frac{1}{T^2}\sum_{t=1}^{T}\sum_{l=1}^{T} x_{it}f_{jl}\gamma_{lt}\right| > C\right) \leq \Pr\left(\left\|\frac{1}{T^2}\sum_{t=1}^{T}\sum_{l=1}^{T}\mathbf{f}_t f_{jl}\gamma_{lt}\right\| > C\right) +$$
$$\Pr\left(\sup_i \left|\frac{1}{T^2}\sum_{t=1}^{T}\sum_{l=1}^{T} v_{it}f_{jl}\gamma_{lt}\right| > C\right).$$

But

$$\Pr\left(\left\|\frac{1}{T^2}\sum_{t=1}^{T}\sum_{l=1}^{T}\mathbf{f}_t f_{jl}\gamma_{lt}\right\| > C\right) = o(1),$$

We have

$$\Pr\left[\sup_i\left|\frac{1}{T}\sum_{t=1}^{T}v_{it}\left(\frac{1}{T}\sum_{l=1}^{T}f_{jl}\gamma_{lt}\right)\right| > C\right].$$

By the independence of f_{jt} and v_{it} and the martingale difference (m.d.) property of v_{it}, $\left(\frac{1}{T}\sum_{l=1}^{T}f_{jl}\gamma_{lt}\right)v_{it}$ is also m.d., and by the martingale difference exponential inequality of Lemma A3 of Chudik, Kapetanios, and Pesaran (2018),

$$\Pr\left[\sup_i\left|\frac{1}{T}\sum_{t=1}^{T}v_{it}\left(\frac{1}{T}\sum_{l=1}^{T}f_{jl}\gamma_{lt}\right)\right| > C\right] = o(1).$$

Next, for (55),

$$\Pr\left[\sup_i\left|\sum_{t=1}^{T}x_{it}\left(\frac{1}{T}\sum_{l=1}^{T}\hat{f}_{jl}\zeta_{lt}\right)\right| > TC\right] \leq \Pr\left[\sup_i\left|\frac{1}{T}\sum_{t=1}^{T}\sum_{l=1}^{T}x_{it}\left(\hat{f}_{jl}-f_{jl}\right)\zeta_{lt}\right| > TC\right] +$$
$$\Pr\left(\sup_i\left|\frac{1}{T}\sum_{t=1}^{T}\sum_{l=1}^{T}x_{it}f_{jl}\zeta_{lt}\right| > TC\right) = A_{21i} + A_{22i}.$$

But

$$A_{21i} = \Pr\left[\sup_i\left|\frac{1}{T^2}\sum_{t=1}^{T}\sum_{l=1}^{T}x_{it}\left(\hat{f}_{jl}-f_{jl}\right)\zeta_{lt}\right| > C\right] \leq$$
$$\Pr\left\{\left[\frac{\min(N,T)^a}{T}\sum_{l=1}^{T}\left(\hat{f}_{jl}-f_{jl}\right)^2\right]^{1/2}\sup_i\left[\frac{\min(N,T)^{-a}}{T^2}\sum_{t=1}^{T}\sum_{l=1}^{T}x_{it}^2\zeta_{lt}^2\right]^{1/2} > C\right\} \leq$$
$$\Pr\left\{\left[\frac{\min(N,T)^a}{T}\sum_{l=1}^{T}\left(\hat{f}_{jl}-f_{jl}\right)^2\right]^{1/2} > C\right\} + \Pr\left\{\sup_i\left[\frac{\min(N,T)^{-a}}{T^2}\sum_{t=1}^{T}\sum_{l=1}^{T}x_{it}^2\zeta_{lt}^2\right] > C\right\}.$$

As before

$$\Pr\left\{\left[\frac{\min(N,T)^a}{T}\sum_{l=1}^{T}\left(\hat{f}_{jl}-f_{jl}\right)^2\right]^{1/2} > C\right\} = o(1).$$

Then,

$$
\begin{aligned}
&\Pr\left\{\sup_i \left[\frac{\min(N,T)^{-a}}{T^2}\sum_{t=1}^{T}\sum_{l=1}^{T} x_{it}^2\zeta_{lt}^2\right] \leq C\right\} \\
&\leq \Pr\left\{\sup_i \left[\left(\frac{\min(N,T)^{-2a}}{T^2}\sum_{t=1}^{T}\sum_{l=1}^{T} x_{it}^4\right)^{1/2}\left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{l=1}^{T}\zeta_{lt}^4\right)^{1/2}\right] \leq C\right\} \leq \\
&\Pr\left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{l=1}^{T}\zeta_{lt}^4 \leq C\right) + \\
&\Pr\left[\sup_i \left(\frac{1}{T}\sum_{t=1}^{T} x_{it}^4\right) \leq C\min(N,T)^{2a}\right].
\end{aligned}
$$

But since $T^{-2}\sum_{t=1}^{T}\sum_{l=1}^{T}\zeta_{lt}^4 = o_p(1)$, then $\Pr\left(T^{-2}\sum_{t=1}^{T}\sum_{l=1}^{T}\zeta_{lt}^4 > C\right) = o(1)$, and using Lemma 2 we obtain

$$
\Pr\left[\sup_i \left(\frac{1}{T}\sum_{t=1}^{T} x_{it}^4\right) > C\min(N,T)^{2a}\right] = o(1).
$$

A very similar analysis can be applied to (56), proving the required result.

A.3 Proof of Theorem 3

We start with Algorithm 2. It is sufficient to show that adding an extra set of regressors, of finite size, to the regressions underlying (14), and using σ^2 thresholding to select dominant units, still implies that a dominant unit will be selected with probability approaching one and a non-dominant one with probability approaching zero. This is shown in Appendix A.4. Then, consistency follows by noting that at every step of the sequential algorithm some dominant unit will be selected with probability approaching one and once all dominant units are selected no further unit will be selected again with probability approaching one.

Moving on to Algorithm 3, we need to show that if we have a potential dominant unit, x_{i^*t}, then a set of t-type test statistics of the form given in (42), will reject the null hypothesis that $\gamma_j^* = 0$, for all j, in

$$
x_{jt} = \mu_j^* + x_{i^*t}\gamma_j^* + \mathbf{f}_t^{*\prime}\mathbf{a}_j^* + \mathbf{x}_{at}^{*\prime}\mathbf{b}_{aj}^* + v_{jt},\ t = 1, 2, \ldots, T, \tag{57}
$$

with probability approaching one, exponentially fast, if x_{i^*t} is a dominant unit and do so with probability approaching zero, exponentially fast, if x_{i^*t} is not. This is shown in Appendix A.5.

A.4 Analysis of sequential σ^2 thresholding

Consider the extension to a model of the form

$$
\begin{aligned}
x_{it} &= \mathbf{a}_i'\mathbf{f}_t + \mathbf{b}'\mathbf{z}_t, \ \text{ for } i = 1, 2, \ldots, m, \\
x_{it} &= \mathbf{a}_i'\mathbf{f}_t + \mathbf{b}'\mathbf{z}_t + u_{it}, \text{ for } i = m+1, m+2, \ldots, N,
\end{aligned}
$$

where $\mathbf{z}_t$ is a known and observed vector of variables. We wish to repeat the analysis of the threshold based selection, for $x_{it} - \mathbf{b}'\mathbf{z}_t$ but use OLS regression of x_{it} on $\mathbf{z}_t$ to obtain the OLS coefficient $\hat{\mathbf{b}}$ and construct $x_{it} - \hat{\mathbf{b}}'\mathbf{z}_t$. Repeating our earlier analysis without $\mathbf{z}_t$, we note that $N\hat{\sigma}^2_{iT}$ contains now a further term that potentially dominates other previously analyzed terms. This term is given by $\frac{N\left(\hat{\mathbf{b}}-\mathbf{b}\right)'\mathbf{Z}'\mathbf{Z}\left(\hat{\mathbf{b}}-\mathbf{b}\right)}{T}$. A possibility is to modify $N\hat{\sigma}^2_{iT}$ and consider $\min(N,T)\hat{\sigma}^2_{iT}$ instead. So we consider $\left(\hat{\mathbf{b}}-\mathbf{b}\right)'\mathbf{Z}'\mathbf{Z}\left(\hat{\mathbf{b}}-\mathbf{b}\right)$. We simplify the analysis by using a scalar $\mathbf{z}_t$. We wish to bound $\Pr\left[\left(\hat{\mathbf{b}}-\mathbf{b}\right)'\mathbf{Z}'\mathbf{Z}\left(\hat{\mathbf{b}}-\mathbf{b}\right) > C_T\right]$. We have

$$\Pr\left[\left(\hat{\mathbf{b}}-\mathbf{b}\right)'\mathbf{Z}'\mathbf{Z}\left(\hat{\mathbf{b}}-\mathbf{b}\right) > C_T\right] = \Pr\left(\left|\left(\frac{\sum_{t=1}^T z_t v_{it}}{\sum_{t=1}^T z_t^2}\right)^2 \sum_{t=1}^T z_t^2\right| > C_T\right) \leq$$
$$\Pr\left(\left|\left(\frac{\sum_{t=1}^T z_t v_{it}}{\sum_{t=1}^T z_t^2}\right)^2 \sum_{t=1}^T z_t^2\right| > C_T\right).$$

Using our derivations in the previous sections of the appendix, we have

$$\Pr\left(\left|\left(\frac{\sum_{t=1}^T z_t v_{it}}{\sum_{t=1}^T z_t^2}\right)^2 \sum_{t=1}^T z_t^2\right| > C_T\right) \leq \Pr\left(\left|\sum_{t=1}^T z_t^2 - \sigma_z^2\right| > C/C_T\right) + \Pr\left(\left|\frac{1}{\sqrt{T}}\sum_{t=1}^T z_t v_{it}\right| > C_T^{1/2}\right). \tag{58}$$

The right hand side of (58) can be bounded using a martingale difference exponential inequality, as before, thus providing justification for a criterion of the following form. Select unit i to be pervasive if $\hat{\sigma}^2_{iT} \leq \frac{2\hat{\eta}^2_{iN}\log T}{N}$, for $T \geq N$ and $\hat{\sigma}^2_{iT} \leq \frac{2\hat{\sigma}^2_{iu}\log T}{T}$, for $T < N$, where $\hat{\sigma}^2_{iu} = \frac{1}{T}\sum_{t=1}^T \left(x_{it} - \hat{\mathbf{b}}'\mathbf{z}_t\right)^2$.

A.5 Analysis of the t-type statistic in Algorithm 3

Recall from Section 3 that the model is given by

$$\mathbf{x}_{at} = \mathbf{\Lambda}_a \mathbf{g}_t + \mathbf{u}_{at}, \tag{59}$$
$$\mathbf{x}_{bt} = \mathbf{\Lambda}_b \mathbf{g}_t + \mathbf{B}\mathbf{x}_{at} + \mathbf{u}_{bt}, \tag{60}$$

for $t = 1, 2, \ldots, T$, where $\mathbf{x}_{at}$ and $\mathbf{x}_{bt}$ are $m \times 1$ and $n \times 1$ vectors of observations at time t on the pervasive and non-pervasive units, respectively. For simplicity, let $u_{it} \sim iidN\left(0, \sigma_u^2\right)$. Assume we want to apply the MT hurdle to an element of $\mathbf{x}_{at}$ or $\mathbf{x}_{bt}$, denoted by z_t. The auxiliary regressions considered here are

$$x_{it} = \gamma_i^* z_t + \lambda_i \hat{\boldsymbol{f}}_t + v_{it},\ t = 1, 2, \ldots, T,$$

for each $i = 1, 2, \ldots, N$ where $\hat{\boldsymbol{f}}_t$ denotes other variables included in the regression. This amounts to collecting both estimated factors $\hat{\mathbf{f}}_t^*$ and previously selected pervasive units $\mathbf{x}^*_{at}$, as specified in equation (40), into a single vector and this vector $\hat{\boldsymbol{f}}_t$, which should not be confounded with the factor estimate $\hat{\mathbf{f}}_t$ in Sections 4-5. In the first instance, we will not specify how factors are estimated for extra generality. In vector form, we can write the model as $\mathbf{x}_i = \mathbf{z}\gamma_i^* + \hat{\boldsymbol{F}}\boldsymbol{\lambda}_i + \mathbf{v}_i$, or $\mathbf{M}_{\hat{\boldsymbol{F}}}\mathbf{x}_i = \mathbf{M}_{\hat{\boldsymbol{F}}}\mathbf{z}\gamma_i^* + \mathbf{M}_{\hat{\boldsymbol{F}}}\mathbf{v}_i$, where $\mathbf{M}_{\hat{\boldsymbol{F}}} = \mathbf{I}_T - \mathbf{P}_{\hat{\boldsymbol{F}}}$ and $\mathbf{P}_{\hat{\boldsymbol{F}}} = \hat{\boldsymbol{F}}\left(\hat{\boldsymbol{F}}'\hat{\boldsymbol{F}}\right)^{-1}\hat{\boldsymbol{F}}'$. The OLS estimator of the slope coefficient γ_i^* is given by

$$\hat{\gamma}_{iz}^* = \left(T^{-1}\mathbf{z}'\mathbf{M}_{\hat{\boldsymbol{F}}}\mathbf{z}\right)^{-1} T^{-1}\mathbf{z}'\mathbf{M}_{\hat{\boldsymbol{F}}}\mathbf{x}_i, \tag{61}$$

and we are interested in finding the limit expression of both terms on the right-hand side. Both numerator and denominator are of the form $T^{-1}\mathbf{z}'\mathbf{M}_{\hat{\boldsymbol{F}}}\mathbf{s} = T^{-1}\mathbf{z}'\left(\mathbf{I}_T - \mathbf{P}_{\hat{\boldsymbol{F}}}\right)\mathbf{s}$ for some variable $\mathbf{s}$. We consider two cases: The case where $\mathbf{z}$ is an element of $\mathbf{x}_{at}$ and the case where it is an element of $\mathbf{x}_{bt}$. In both cases we assume that $\hat{\boldsymbol{f}}_t$ spans $(\mathbf{u}_{at}', \mathbf{g}_t')'$ asymptotically, i.e. that there exists some matrix $\boldsymbol{C}$ such that $p\lim\left\|\boldsymbol{C}\hat{\boldsymbol{f}}_t - (\mathbf{u}_{at}', \mathbf{g}_t')'\right\| = 0$. We start by considering the case where $\mathbf{z}$ is pervasive. In that case there exists $\boldsymbol{H}$ such that $\mathbf{z} = \boldsymbol{F}\boldsymbol{H}$ where $p\lim\hat{\boldsymbol{F}} = \boldsymbol{F}$.

We have

$$\begin{aligned}
T^{-1}\mathbf{z}'\left(\mathbf{I}_T - \mathbf{P}_{\hat{\boldsymbol{F}}}\right)\mathbf{s} &= T^{-1}\mathbf{z}'\left(\mathbf{I}_T - \mathbf{P}_{\hat{\boldsymbol{F}}}\right)\mathbf{s} - T^{-1}\mathbf{z}'\left(\mathbf{I}_T - \mathbf{P}_{\hat{\boldsymbol{F}}}\right)\mathbf{s} \\
&= T^{-1}\mathbf{z}'\left(\mathbf{P}_{\boldsymbol{F}} - \mathbf{P}_{\hat{\boldsymbol{F}}}\right)\mathbf{s} = T^{-1}\mathbf{z}'\left(\boldsymbol{F} - \hat{\boldsymbol{F}}\right)\left(\boldsymbol{F}'\boldsymbol{F}\right)^{-1}\boldsymbol{F}'\mathbf{s} + \\
&\quad T^{-1}\mathbf{z}'\hat{\boldsymbol{F}}\left[\left(\boldsymbol{F}'\boldsymbol{F}\right)^{-1} - \left(\hat{\boldsymbol{F}}'\hat{\boldsymbol{F}}\right)^{-1}\right]\boldsymbol{F}'\mathbf{s} + T^{-1}\mathbf{z}'\hat{\boldsymbol{F}}\left(\hat{\boldsymbol{F}}'\hat{\boldsymbol{F}}\right)^{-1}\left(\boldsymbol{F} - \hat{\boldsymbol{F}}\right)'\mathbf{s} \\
&= A_1 + A_2 + A_3.
\end{aligned}$$

Let $T^{-1}\mathbf{z}'\left(\boldsymbol{F} - \hat{\boldsymbol{F}}\right) = O_p\left(c_{1,N,T}\right)$ where $c_{1,N,T} \to 0$ and depends on the factor estimation method. Further assuming that $\boldsymbol{F}'\mathbf{s} = O_p(T)$, we obtain $A_1 = T^{-1}\mathbf{z}'\left(\boldsymbol{F} - \hat{\boldsymbol{F}}\right)\left(\boldsymbol{F}'\boldsymbol{F}\right)^{-1}\boldsymbol{F}'\mathbf{s} = O_p\left(c_{1,N,T}\right)$. Similarly, $A_3 = T^{-1}\mathbf{z}'\hat{\boldsymbol{F}}\left(\hat{\boldsymbol{F}}'\hat{\boldsymbol{F}}\right)^{-1}\left(\boldsymbol{F} - \hat{\boldsymbol{F}}\right)'\mathbf{s} = O_p\left(c_{1,N,T}\right)$. The order in probability of A_2 differs slightly since this expression is a function of $\hat{\boldsymbol{F}}'\left(\hat{\boldsymbol{F}} - \boldsymbol{F}\right) = \boldsymbol{F}'\left(\hat{\boldsymbol{F}} - \boldsymbol{F}\right) - \left(\hat{\boldsymbol{F}} - \boldsymbol{F}\right)'\left(\hat{\boldsymbol{F}} - \boldsymbol{F}\right)$. Let $T^{-1}\left(\hat{\boldsymbol{F}} - \boldsymbol{F}\right)'\left(\hat{\boldsymbol{F}} - \boldsymbol{F}\right) = O_p\left(c_{2,N,T}\right)$. More specifically, we have

$$\begin{aligned}
A_2 &= T^{-1}\mathbf{z}'\hat{\boldsymbol{F}}\left[\left(\boldsymbol{F}'\boldsymbol{F}\right)^{-1} - \left(\hat{\boldsymbol{F}}'\hat{\boldsymbol{F}}\right)^{-1}\right]\boldsymbol{F}'\mathbf{s} = T^{-1}\mathbf{z}'\hat{\boldsymbol{F}}\left(\boldsymbol{F}'\boldsymbol{F}\right)^{-1}\hat{\boldsymbol{F}}'\left(\hat{\boldsymbol{F}} - \boldsymbol{F}\right)\left(\hat{\boldsymbol{F}}'\hat{\boldsymbol{F}}\right)^{-1}\boldsymbol{F}'\mathbf{s} + \\
&\quad T^{-1}\mathbf{z}'\hat{\boldsymbol{F}}\left(\boldsymbol{F}'\boldsymbol{F}\right)^{-1}\left(\hat{\boldsymbol{F}} - \boldsymbol{F}\right)'\boldsymbol{F}\left(\hat{\boldsymbol{F}}'\hat{\boldsymbol{F}}\right)^{-1}\boldsymbol{F}'\mathbf{s} = O_p\left(c_{1,N,T}\right) + O_p\left(c_{2,N,T}\right).
\end{aligned}$$

Note that by Lemma A1 of Bai and Ng (2006), we have that for PC factor estimation $T^{-1}\mathbf{z}'\left(\boldsymbol{F} - \hat{\boldsymbol{F}}\right) = O_p\left(\min\left(N,T\right)^{-1}\right)$ and $T^{-1}\left(\hat{\boldsymbol{F}} - \boldsymbol{F}\right)'\left(\hat{\boldsymbol{F}} - \boldsymbol{F}\right) = O_p\left(\min\left(N,T\right)^{-1}\right)$, so that for this factor estimator $c_{1,N,T} = c_{2,N,T} = \min\left(N,T\right)^{-1}$. It follows that both $T^{-1}\mathbf{z}'\mathbf{M}_{\hat{\boldsymbol{F}}}\mathbf{x}_i$ and $T^{-1}\mathbf{z}'\mathbf{M}_{\hat{\boldsymbol{F}}}\mathbf{x}_i$ are $O_p\left(\min\left(N,T\right)^{-1}\right)$, and using (61) we have $\hat{\gamma}_{iz}^* = O_p\left(1\right)$.

Next consider $\mathfrak{t}_{iz} = \left[\widehat{\mathbb{V}\mathrm{ar}}\left(\hat{\gamma}_{iz}^*\right)\right]^{-1/2}\hat{\gamma}_{iz}^*$, where $\widehat{\mathbb{V}\mathrm{ar}}\left(\hat{\gamma}_{iz}^*\right) = \left(\mathbf{z}'\mathbf{M}_{\hat{\boldsymbol{F}}}\mathbf{z}\right)^{-1}T^{-1}\sum_{t=1}^T\hat{v}_{it}^2$, and hence $\mathfrak{t}_{iz} = \left(T^{-1}\sum_{t=1}^T\hat{v}_{it}^2\right)^{-1/2}\left(\mathbf{z}'\mathbf{M}_{\hat{\boldsymbol{F}}}\mathbf{z}\right)^{-1/2}\mathbf{z}'\mathbf{M}_{\hat{\boldsymbol{F}}}\mathbf{x}_i$. It is also easily seen that $T^{-1}\sum_{t=1}^T\hat{v}_{it}^2 = O_p\left(1\right)$, and

$$T^{-1}\mathbf{z}'\mathbf{M}_{\hat{\boldsymbol{F}}}\mathbf{z} = O_p\left(\min\left(N,T\right)^{-1}\right),\ T^{-1}\mathbf{z}'\mathbf{M}_{\hat{\boldsymbol{F}}}\mathbf{x}_i = O_p\left(\min\left(N,T\right)^{-1}\right).$$

So overall $\mathfrak{t}_{iz} = O_p\left(\min(N,T)^{-1/2} T^{1/2}\right)$, implying that a standard t-statistic to test the statistical significance of γ_i^* need not diverge if z_i is a pervasive unit.

Now consider the t-type statistic of (42), given by

$$t_{iz}^* = \frac{\hat{\gamma}_i}{\sqrt{\frac{T^{-1}\sum_{t=1}^{T}\hat{v}_{it}^2}{\mathbf{z}'\mathbf{z}}}} = (\mathbf{z}'\mathbf{z})^{1/2}\left(T^{-1}\sum_{t=1}^{T}\hat{v}_{it}^2\right)^{-1/2}\left(\mathbf{z}'\mathbf{M}_{\hat{\boldsymbol{F}}}\mathbf{z}\right)^{-1}\mathbf{z}'\mathbf{M}_{\hat{\boldsymbol{F}}}\mathbf{x}_i = O_p\left(T^{1/2}\right).$$

We therefore note that, unlike the standard t statistic which does not necessarily diverge if $\mathbf{z}$ is pervasive, t_{iz}^* does diverge at the usual rate.

Under the case where the z is not pervasive and assuming that $p_{\max}$ is large enough to span the true factors, it is obvious that $\mathfrak{t}_{iz} = O_p(1)$ and $t_{iz}^* = O_p(1)$. Using arguments similar to the rest of the paper (consider, e.g., the proofs of (54)-(56) and Lemma 2) it follows that, using standard multiple testing critical values $c_N = O\left(\ln(N)^{1/2}\right)$, we can make $1 - \Pr(t_{iz}^* > c_N)$ exponentially small if $\mathbf{z}$ denotes a pervasive unit and $\Pr(t_{iz}^* > c_N)$ exponentially small if $\mathbf{z}$ denotes a non-pervasive unit.

References

Acemoglu, D., U. Akcigit, and W. Kerr (2016): "Networks and the macroeconomy: An empirical exploration," *NBER Macroeconomics Annual*, 30(1), 273–335.

Acemoglu, D., D. Autor, D. Dorn, G. H. Hanson, and B. Price (2016): "Import competition and the great US employment sag of the 2000s," *Journal of Labor Economics*, 34(S1), S141–S198.

Acemoglu, D., V. M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi (2012): "The network origins of aggregate fluctuations," *Econometrica*, 80(5), 1977–2016.

Ahn, S. C., and A. R. Horenstein (2013): "Eigenvalue ratio test for the number of factors," *Econometrica*, 81(3), 1203–1227.

Bai, J. (2003): "Inferential theory for factor models of large dimensions," *Econometrica*, 71(1), 135–171.

Bai, J., and S. Ng (2002): "Determining the number of factors in approximate factor models," *Econometrica*, 70(1), 191–221.

——— (2006): "Evaluating latent and observed factors in macroeconomics and finance," *Journal of Econometrics*, 131(1-2), 507–537.

——— (2008): "Large dimensional factor analysis," *Foundations and Trends® in Econometrics*, 3(2), 89–163.

Bailey, N., S. Holly, and M. H. Pesaran (2016): "A Two-Stage Approach to Spatio-Temporal Analysis with Strong and Weak Cross-Sectional Dependence," *Journal of Applied Econometrics*, 31(1), 249–280.

Bailey, N., G. Kapetanios, and M. H. Pesaran (2016): "Exponent of Cross Sectional Dependence: Estimation and Inference," *Journal of Applied Econometrics*, 31(6), 929–960.

——— (2019): "Exponent of Cross-sectional Dependence for Residuals," *Sankhya: The Indian Journal of Statistics (forthcoming)*, `https://doi.org/10.1007/s13571-019-00196-9`.

Bailey, N., M. H. Pesaran, and L. V. Smith (2019): "A Multiple Testing Approach to the Regularisation of Large Sample Correlation Matrices," *Journal of Econometrics*, 208(2), 507–534.

Bickel, P. J., and E. Levina (2008): "Covariance regularization by thresholding," *The Annals of Statistics*, 36(6), 2577–2604.

Brownlees, C., and G. Mesters (2019): "Detecting Granular Time Series in Large Panels," *Journal of Econometrics (forthcoming)*.

Cai, T., and W. Liu (2011): "Adaptive Thresholding for Sparse Covariance Matrix Estimation," *Journal of American Statistical Association*, 106(494), 672–684.

Cesa-Bianchi, A., M. H. Pesaran, and A. Rebucci (2019): "Uncertainty and Economic Activity: A Multicountry Perspective," *The Review of Financial Studies (forthcoming)*, `https://doi.org/10.1093/rfs/hhz098`.

Chudik, A., G. Kapetanios, and M. H. Pesaran (2018): "A One Covariate at a Time, Multiple Testing Approach to Variable Selection in High-Dimensional Linear Regression Models," *Econometrica*, 86(4), 1479–1512.

Chudik, A., M. H. Pesaran, and E. Tosetti (2011): "Weak and strong cross-section dependence and estimation of large panels," *The Econometrics Journal*, 14(1), C45–C90.

Dungey, M., and V. Volkov (2018): "R&D and wholesale trade are critical to the economy: Identifying dominant sectors from economic networks," *Economics Letters*, 162, 81–85.

El Karoui, N. (2008): "Operator norm consistent estimation of large-dimensional sparse covariance matrices," *The Annals of Statistics*, 36(6), 2717–2756.

Foerster, A. T., P.-D. G. Sarte, and M. W. Watson (2011): "Sectoral versus aggregate shocks: A structural factor analysis of industrial production," *Journal of Political Economy*, 119(1), 1–38.

Gaibulloev, K., T. Sandler, and D. Sul (2013): "Common drivers of transnational terrorism: Principal component analysis," *Economic Inquiry*, 51(1), 707–721.

Greenaway-McGrevy, R., N. C. Mark, D. Sul, and J.-L. Wu (2018): "Identifying exchange rate common factors," *International Economic Review*, 59(4), 2193–2218.

Holly, S., M. H. Pesaran, and T. Yamagata (2010): "A spatio-temporal model of house prices in the USA," *Journal of Econometrics*, 158(1), 160–173.

Mohaddes, K., and M. Raissi (2018): "Compilation, Revision and Updating of the Global VAR (GVAR) Database, 1979Q2-2016Q4," (mimeo).

Onatski, A. (2012): "Asymptotics of the principal components estimator of large factor models with weakly influential factors," *Journal of Econometrics*, 168(2), 244–258.

Parker, J., and D. Sul (2016): "Identification of Unknown Common Factors: Leaders and Followers," *Journal of Business & Economic Statistics*, 34(2), 227–239.

Pesaran, M. H., and C. F. Yang (2019): "Econometric Analysis of Production Networks with Dominant Units," *Journal of Econometrics (forthcoming)*.

Soofi-Siavash, S. (2018): "Dominant U.S. Manufacturing Sectors: A Factor Model Analysis," (mimeo). Available at `https://sites.google.com/site/sorooshsoofisiavash/research`.

Yang, C. F. (2018): "Common Factors and Spatial Dependence: An Application to US House Prices," Available at `https://sites.google.com/site/cynthiafanyang/research`.

Online Supplement to

Detection of Units with Pervasive Effects in Large Panel Data Models

by

G. Kapetanios, M. H. Pesaran and S. Reese

November 2019

This online supplement contains additional theoretical, simulation and empirical results that complement the main paper. It is composed of six sections. Section S1 provides the statement and proofs of a number of auxiliary lemmas used in the proofs. Section S2 gives a more detailed description of the steps required to implement the various variants of the basic σ^2 thresholding method proposed in the paper. A summary of other approaches proposed in the literature for the detection of pervasive units is given in Section S3. The finite sample performance of the different variants of the σ^2 thresholding (that are not considered in Section 6 of the paper) is discussed in Section S4. The robustness of our detection method to the choice of the constant C (which is set to $C = 1$ in our MC and empirical analyses) that enters the threshold rule is investigated in Section S5. Two additional σ^2 thresholding schemes, based on the difference and the ratio of two successive ordered error variance estimates, are considered in Section S6 and their small sample properties investigated using Monte Carlo simulations. Finally, Sections S7 and S8 report simulation and empirical results using an unmodified version of (Brownlees and Mesters, 2019, BM in the following) procedure.

S1 Proof of Proposition 1 and Auxiliary Lemmas

This section provides proofs of Proposition 1 and the lemmas used in the paper.

S1.1 Proof of Proposition 1

Let $\mathcal{G}_1 \subset \sigma(x_{it}) \subset \mathcal{F}$ and $\mathcal{G}_2 \subset \sigma(x_{it}) \subset \mathcal{F}$ be any σ-fields for which

$$E_P\left[\left|E\left(u_{it}|\mathcal{G}_1\right) - E\left(u_{it}|\mathcal{G}_2\right)\right|\right] > 0.$$

and, therefore,

$$E_P\left[\left|E\left(x_{it}|\mathcal{G}_1\right) - E\left(x_{it}|\mathcal{G}_2\right)\right|\right] > 0.$$

By independence across i,

$$E_P\left[\left|E\left(u_{jt}|\mathcal{G}_1\right) - E\left(u_{jt}|\mathcal{G}_2\right)\right|\right] = 0,\ j \neq i.$$

Then, as long as $|b_{ij}| > 0$, for $i = m+1, m+2, ..., N$, and $\sup_{ij} |b_{ij}| < K < \infty$,

$$E_P\left[\left|E\left(x_{jt}|\mathcal{G}_1\right) - E\left(x_{jt}|\mathcal{G}_2\right)\right|\right] = E_P\left[\left|E\left(b_{ij}u_{it}|\mathcal{G}_1\right) - E\left(b_{ij}u_{it}|\mathcal{G}_2\right)\right|\right] > 0,\ j \neq i.$$

Therefore, (1) holds, proving the result.

S1.2 Auxiliary Lemmas

We next present and prove auxiliary lemmas. First we provide a lemma handling the remainder terms of $N\hat{\sigma}_{iT}^2$. We have

Lemma 1 *Let i denote a pervasive unit, Assumptions 1-4 hold and $\frac{\sqrt{N}}{T} \to 0$, as $N, T \to \infty$. Then,*

$$N\hat{\sigma}_{iT}^2 = \frac{\mathbf{a}_i'\mathbf{A}_0\mathbf{V}'\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{V}\mathbf{A}_0\mathbf{a}_i}{NT} + O_p\left(\frac{1}{\delta_{NT}}\right) + O_p\left(\frac{\sqrt{N}}{\delta_{NT}^2}\right),$$

uniformly over i, where $\delta_{NT}^2 = \min(N, T)$.

Proof. Since unit i is pervasive, we have $N\hat{\sigma}_{iT}^2 = (N/T)\,\mathbf{a}_i'\mathbf{F}_0\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{F}_0\mathbf{a}_i$. Now using (29) we obtain $N\hat{\sigma}_{iT}^2 = \sum_{j=1}^{6} B_{ij}$, where

$$\begin{aligned}
B_{i1} &= \frac{\mathbf{a}_i'\mathbf{A}_0\mathbf{V}'\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{V}\mathbf{A}_0\mathbf{a}_i,}{NT} \\
B_{i2} &= 2\frac{\mathbf{a}_i'\mathbf{A}_0'\mathbf{V}'\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{V}\left(\mathbf{A}_0 - \hat{\mathbf{A}}\right)\mathbf{a}_i}{NT}, \\
B_{i3} &= 2\frac{\mathbf{a}_i'\mathbf{A}_0'\mathbf{V}'\mathbf{M}_{\hat{\mathbf{F}}}\left(\mathbf{F}_0 - \hat{\mathbf{F}}\right)\mathbf{A}_0'\left(\mathbf{A}_0 - \hat{\mathbf{A}}\right)\mathbf{a}_i}{NT}, \\
B_{i4} &= \frac{\mathbf{a}_i'\left(\mathbf{A_0} - \hat{\mathbf{A}}\right)'\mathbf{V}'\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{V}\left(\mathbf{A}_0 - \hat{\mathbf{A}}\right)\mathbf{a}_i}{NT}, \\
B_{i5} &= 2\frac{\mathbf{a}_i'\left(\mathbf{A_0} - \hat{\mathbf{A}}\right)'\mathbf{V}'\mathbf{M}_{\hat{\mathbf{F}}}\left(\mathbf{F}_0 - \hat{\mathbf{F}}\right)\mathbf{A}_0'\left(\mathbf{A}_0 - \hat{\mathbf{A}}\right)\mathbf{a}_i}{NT}, \\
B_{i6} &= \frac{\mathbf{a}_i'\left(\mathbf{A_0} - \hat{\mathbf{A}}\right)'\mathbf{A}_0\left(\mathbf{F}_0 - \hat{\mathbf{F}}\right)'\mathbf{M}_{\hat{\mathbf{F}}}\left(\mathbf{F}_0 - \hat{\mathbf{F}}\right)\mathbf{A}_0'\left(\mathbf{A}_0 - \hat{\mathbf{A}}\right)\mathbf{a}_i}{NT}.
\end{aligned}$$

First, note that

$$\|B_{i2}\| \leq \frac{2}{NT}\|\mathbf{a}_i\|^2 \|\mathbf{A}_0'\mathbf{V}'\| \|\mathbf{M}_{\hat{\mathbf{F}}}\| \left\|\mathbf{V}\left(\mathbf{A_0} - \hat{\mathbf{A}}\right)\right\|.$$

But $\|\mathbf{M}_{\hat{\mathbf{F}}}\| = 1$, since $\mathbf{M}_{\hat{\mathbf{F}}}$ is an idempotent matrix. Furthermore, $\sup_i \|\mathbf{a}_i\|^2 < C$, by Assumption 2. Together with (C) and (D) of Proposition 2, these two results imply

$$\|B_{i2}\| = \frac{1}{NT}O_p\left(\sqrt{NT}\right)O_p\left(\frac{\sqrt{NT}}{\delta_{NT}}\right) = O_p\left(\frac{1}{\delta_{NT}}\right).$$

Similarly, using (A), (D) and (E) of Proposition 2,

$$\begin{aligned}
\|B_{i3}\| &\leq \frac{C}{NT}\|\mathbf{A}_0'\mathbf{V}'\| \left\|\mathbf{F}_0 - \hat{\mathbf{F}}\right\| \left\|\mathbf{A}_0'\left(\mathbf{A_0} - \hat{\mathbf{A}}\right)\right\| \\
&= \frac{1}{NT}O_p\left(\sqrt{NT}\right)O_p\left(\frac{\sqrt{T}}{\delta_{NT}}\right)O_p\left(\frac{N}{\delta_{NT}}\right) \\
&= O_p\left(\frac{\sqrt{N}}{\delta_{NT}^2}\right).
\end{aligned}$$

Next,

$$\|B_{i4}\| \leq \frac{C}{NT}\left\|\mathbf{V}\left(\mathbf{A_0}-\hat{\mathbf{A}}\right)\right\|^2 = O_p\left(\frac{1}{\delta_{NT}^2}\right),$$

follows from (C). Using this latter result, as well as (A) and (E), we also obtain

$$\begin{aligned}\|B_{i5}\| &\leq \frac{C}{NT}\left\|\mathbf{V}\left(\mathbf{A_0}-\hat{\mathbf{A}}\right)\right\|\left\|\mathbf{F}_0-\hat{\mathbf{F}}\right\|\left\|\mathbf{A}_0'\left(\mathbf{A_0}-\hat{\mathbf{A}}\right)\right\| \\ &= \frac{1}{NT}O_p\left(\frac{\sqrt{NT}}{\delta_{NT}}\right)O_p\left(\frac{\sqrt{T}}{\delta_{NT}}\right)O_p\left(\frac{N}{\delta_{NT}}\right) \\ &= O_p\left(\frac{\sqrt{N}}{\delta_{NT}^3}\right).\end{aligned}$$

Finally,

$$\begin{aligned}\|B_{i6}\| &\leq \frac{C}{NT}\left\|\mathbf{A}_0'\left(\mathbf{A_0}-\hat{\mathbf{A}}\right)\right\|^2\left\|\mathbf{F}_0-\hat{\mathbf{F}}\right\|^2 \\ &= \frac{1}{NT}O_p\left(\frac{N^2}{\delta_{NT}^2}\right)O_p\left(\frac{T}{\delta_{NT}^2}\right) = O_p\left(\frac{N}{\delta_{NT}^4}\right),\end{aligned}$$

by the same intermediate results. Summarizing the order results above and noting that $\frac{\sqrt{N}}{T} \to 0$, we have

$$N\hat{\sigma}_{iT}^2 = \frac{\mathbf{a}_i'\mathbf{A}_0'\mathbf{V}'\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{V}\mathbf{A}_0\mathbf{a}_i}{NT} + O_p\left(\frac{1}{\delta_{NT}}\right) + O_p\left(\frac{\sqrt{N}}{\delta_{NT}^2}\right),$$

proving the required result. ■

Lemma 2 *Let Assumptions 1-4 hold. Then,*

$$\Pr\left[\sup_i\left(\frac{1}{T}\sum_{t=1}^{T}x_{it}^j\right) > C\right] = o(1),\ \ j = 1,2,3,4.$$

Proof. We will prove the case for $j=4$ only. The cases for $j=1,2,3$ follow straightforwardly. We have

$$x_{it} = \mathbf{a}_i'\mathbf{f}_t + v_{it} = \varphi_{it} + v_{it},$$

So

$$x_{it}^4 = \varphi_{it}^4 + 4\varphi_{it}^3 v_{it} + 6\varphi_{it}^2 v_{it}^2 + 4\varphi_{it}v_{it}^3 + v_{it}^4 = \sum_j^5 A_{jit}.$$

So

$$\begin{aligned}\Pr\left[\sup_i\left(\frac{1}{T}\sum_{t=1}^{T}x_{it}^4\right) > C\right] &= \Pr\left[\sup_i\left[\sum_{j=1}^{5}\left(\frac{1}{T}\sum_{t=1}^{T}A_{jit}\right)\right] > C\right] \\ &\leq \Pr\left[\sum_{j=1}^{5}\sup_i\left(\frac{1}{T}\sum_{t=1}^{T}A_{jit}\right) > C\right] \\ &\leq \sum_{j=1}^{5}\Pr\left[\sup_i\left(\frac{1}{T}\sum_{t=1}^{T}A_{ijt}\right) > \pi_j C\right] = \sum_{j=1}^{5}B_j,\end{aligned}$$

where $\pi_j > 0$, and $\sum_{j=1}^{5} \pi_j = 1$. We examine each B_j in turn. We have that for sufficiently large finite constant C, there exists some constant C_1 such that

$$\begin{aligned}
\Pr\left[\sup_i \left(\frac{1}{T}\sum_{t=1}^{T} A_{i1t}\right) > \pi_1 C\right] &\leq \Pr\left[\sup_i \left|\frac{1}{T}\sum_{t=1}^{T} (\mathbf{a}_i'\mathbf{f}_t)^4\right| > \pi_1 C\right] \\
&= \Pr\left[\left(\sup_i \|\mathbf{a}_i\|^4\right)\left|\frac{1}{T}\sum_{t=1}^{T} \|\mathbf{f}_t\|^4\right| > \pi_1 C\right] \\
&\leq \Pr\left[\left|\sum_{t=1}^{T} \left[\|\mathbf{f}_t\|^4 - E\left(\|\mathbf{f}_t\|^4\right)\right]\right| > T\pi_1 C_1\right].
\end{aligned}$$

However, since by Assumption 1, $\frac{1}{T}\sum_{t=1}^{T} \left[\|\mathbf{f}_t\|^4 - E\left(\|\mathbf{f}_t\|^4\right)\right] = o_p(1)$,

$$\Pr\left[\left|\sum_{t=1}^{T} \|\mathbf{f}_t\|^4 - E\left(\|\mathbf{f}_t\|^4\right)\right| > T\pi_1 C_1\right] = o(1),$$

for any finite $C_1 > 0$. For $B_2 - B_5$ it is sufficient to note that A_{jit}, $j = 2, \ldots, 5$ are martingale difference processes since φ_{it} and v_{it} are independent and $v_{it}^j - E\left(v_{it}^j\right)$, for $j = 1, 2, 3, 4$ are martingale difference processes by the serial independence of $\boldsymbol{\varepsilon}_t$ (see Assumption 3). Therefore, by the martingale difference exponential inequality Lemma A3 of Chudik, Kapetanios, and Pesaran (2018), we have that for $j = 1, \ldots, 4$, and for a sufficiently large finite constant, C, there exist some constants C_1 and C_2 such that

$$\Pr\left[\sup_i \left(\frac{1}{T}\sum_{t=1}^{T} A_{ijt}\right) > \pi_j C\right] \leq \Pr\left[\sup_i \left|\frac{1}{T}\sum_{t=1}^{T} A_{ijt} - E\left(A_{ijt}\right)\right| > \pi_j C_1\right] \leq \exp\left(-C_2 T\right),$$

proving the result. ■

The rest of the lemmas in this section prove the results of Proposition 2 in the main text. The five results A–E are analyzed in separate lemmas due to the length of the proofs. It is also important to note that the required assumptions for the subsequent lemmas are considerably weaker than those needed for consistency of the σ^2 thresholding procedure. The minimal conditions needed, which are satisfied by Assumptions 1–4 in the main text, as noted in Remark 2, are as follows:

1. $E||\mathbf{f}_t||^4 \leq C < \infty$, $T^{-1}\sum_{t=1}^{T} \mathbf{f}_t\mathbf{f}_t' \xrightarrow{p} \Sigma_f$ for some $m \times m$ positive definite matrix $\boldsymbol{\Sigma}_f$. $\mathbf{A}_0$ has bounded elements. Further $||N^{-1}\mathbf{A}_0'\mathbf{A}_0 - \mathbf{D}|| \to 0$, as $N \to \infty$, where $\mathbf{D}$ is a positive definite matrix.

2. $E(v_{it}) = 0$, $E|v_{it}|^8 \leq C$ where $\mathbf{v}_t = (v_{1t}, \ldots, v_{Nt})'$ The variance of $\mathbf{v}_t$ is denoted by $\boldsymbol{\Sigma}_v$. $\mathbf{f}_s$ and $\mathbf{v}_t$ are independent for all s, t.

3. For $\tau_{i,j,t,s} \equiv E(v_{it}v_{js})$ the following hold

    - $(NT)^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T} |\sum_{i=1}^{N} \tau_{i,i,t,s}| \leq C$.
    - $\sum_{l=1}^{T} |1/N\sum_{i=1}^{N} \tau_{i,i,s,l}| \leq C$ for all s.

- $N^{-1} \sum_{i=1}^{N} \sum_{j=1}^{N} |\tau_{i,j,s,s}| \leq C$.
- $(NT)^{-1} \sum_{s=1}^{T} \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{j=1}^{N} |\tau_{i,j,t,s}| \leq C$.
- For every (t, s), $E|(N)^{-1/2} \sum_{i=1}^{N} (v_{is} v_{it} - \tau_{i,i,s,t})|^4 \leq C$.
- For each t, $\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \mathbf{a}_i v_{it} \rightarrow^d N(0, \mathbf{\Gamma}_t)$ where $\mathbf{\Gamma}_t = \lim_{N\to\infty} \sum_{i=1}^{N} \sum_{j=1}^{N} E\left(\mathbf{a}_i \mathbf{a}_j' v_{it} v_{jt}\right)$.

The above list is essentially the set of assumptions in Bai (2003). Analogous to the definition in Appendix A.2, let $\gamma_{st} = \gamma_{N,st} = \frac{1}{N} \sum_{i=1}^{N} \tau_{i,i,t,s}$.

Lemma 3 *Under Assumptions 1–4*

$$\left\| \frac{\mathbf{V}\left(\hat{\mathbf{A}} - \mathbf{A}_0\right)}{N} \right\|_F = O_p\left(\frac{\sqrt{T}}{\sqrt{N \min(N,T)}}\right) = O_p\left(\frac{\sqrt{T}}{N}\right) + O_p\left(\frac{1}{\sqrt{N}}\right).$$

Proof.

We have by the proof of Theorem 2 of Bai (2003, expression above (B.2)) that

$$\hat{\mathbf{a}}_i - \mathbf{a}_i = \frac{1}{T} \sum_{t=1}^{T} \mathbf{f}_t v_{it} + \frac{1}{T} \sum_{t=1}^{T} x_{it} \left(\hat{\mathbf{f}}_t - \mathbf{f}_t\right). \tag{62}$$

We have

$$\frac{1}{N^2}\left\|\mathbf{V}\left(\hat{\mathbf{A}}-\mathbf{A}_0\right)\right\|_F^2 = \frac{1}{N^2}\sum_{t=1}^{T}\sum_{i=1}^{N} v_i^2\left(\hat{\mathbf{a}}_i-\mathbf{a}_i\right)'\left(\hat{\mathbf{a}}_i-\mathbf{a}_i\right)$$

$$\leq \frac{1}{N^2}\sum_{t=1}^{T}\sum_{i=1}^{N} v_{it}^2\left(\frac{1}{T^2}\sum_{s=1}^{T}\mathbf{f}_s'\mathbf{f}_s v_{is}^2+\frac{1}{T^2}\sum_{s=1}^{T}\sum_{s',s\neq s'}^{T}\mathbf{f}_s'\mathbf{f}_{s'}v_{is}v_{is'}\right)+$$

$$\frac{1}{N^2}\sum_{t=1}^{T}\sum_{i=1}^{N} v_{it}^2\left(\begin{array}{c}\frac{1}{T^2}\sum_{s=1}^{T}x_{is}^2\left\|\hat{\mathbf{f}}_s-\mathbf{f}_s\right\|_F^2+\\ \frac{1}{T^2}\sum_{s=1}^{T}\sum_{s',s\neq s'}^{T}\left|x_{is}x_{is'}\right|\left\|\hat{\mathbf{f}}_s-\mathbf{f}_s\right\|_F\left\|\hat{\mathbf{f}}_{s'}-\mathbf{f}_{s'}\right\|_F\end{array}\right)+$$

$$\frac{1}{N^2}\sum_{t=1}^{T}\sum_{i=1}^{N} v_{it}^2\left[\frac{1}{T^2}\sum_{s=1}^{T}\sum_{s'=1}^{T}x_{is'}v_{is}\mathbf{f}_s'\left(\hat{\mathbf{f}}_{s'}-\mathbf{f}_{s'}\right)\right]=$$

$$\frac{1}{N^2}\sum_{t=1}^{T}\sum_{i=1}^{N} v_{it}^2\left(\frac{1}{T^2}\sum_{s=1}^{T}\mathbf{f}_s'\mathbf{f}_s v_{is}^2\right)+$$

$$\frac{1}{N^2}\sum_{t=1}^{T}\sum_{i=1}^{N} v_{it}^2\left(\frac{1}{T^2}\sum_{s=1}^{T}\sum_{s',s\neq s'}^{T}\mathbf{f}_s'\mathbf{f}_{s'}v_{is}v_{is'}\right)+$$

$$\frac{1}{N^2}\sum_{t=1}^{T}\sum_{i=1}^{N} v_{it}^2\left(\frac{1}{T^2}\sum_{s=1}^{T}x_{is}^2\left\|\hat{\mathbf{f}}_s-\mathbf{f}_s\right\|_F^2\right)+$$

$$\frac{1}{N^2}\sum_{t=1}^{T}\sum_{i=1}^{N} v_{it}^2\left(\frac{1}{T^2}\sum_{s=1}^{T}\sum_{s',s\neq s'}^{T}\left|x_{is}x_{is'}\right|\left\|\hat{\mathbf{f}}_s-\mathbf{f}_s\right\|_F\left\|\hat{\mathbf{f}}_{s'}-\mathbf{f}_{s'}\right\|_F\right)+$$

$$\frac{1}{N^2}\sum_{t=1}^{T}\sum_{i=1}^{N} v_{it}^2\left[\frac{1}{T^2}\sum_{s=1}^{T}\sum_{s'=1}^{T}x_{is'}v_{is}\mathbf{f}_s'\left(\hat{\mathbf{f}}_{s'}-\mathbf{f}_{s'}\right)\right]$$

$$=\sum_{i=1}^{5}C_i.$$

We have

$$C_1=\frac{1}{N^2}\sum_{t=1}^{T}\sum_{i=1}^{N} v_{it}^2\left(\frac{1}{T}\sum_{s=1}^{T}\mathbf{f}_s'\mathbf{f}_s v_{is}^2\right)=\frac{1}{N^2}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{s=1}^{T}\mathbf{f}_s'\mathbf{f}_s v_{is}^2\right)\left(\frac{1}{T}\sum_{t=1}^{T}v_{it}^2\right)=O_p\left(N^{-1}\right).$$

Also

$$\begin{aligned}C_2&=\frac{1}{N^2}\sum_{t=1}^{T}\sum_{i=1}^{N} v_{it}^2\left(\frac{1}{T^2}\sum_{s=1}^{T}\sum_{s',s\neq s'}^{T}\mathbf{f}_s'\mathbf{f}_{s'}v_{is}v_{is'}\right)\\ &=\frac{1}{N^2}\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N} v_{it}^2\left[\frac{1}{\sqrt{T}}\sum_{s=1}^{T}v_{is}\mathbf{f}_s'\left(\frac{1}{\sqrt{T}}\sum_{s',s\neq s'}^{T}\mathbf{f}_{s'}v_{is'}\right)\right]=O_p\left(N^{-1}\right).\end{aligned}$$

Next, we have

$$
\begin{aligned}
C_3 &= \frac{1}{T^2N^2}\sum_{t=1}^{T}\sum_{i=1}^{N} v_{it}^2 \left(\sum_{s=1}^{T} x_{is}^2 \left\|\hat{\mathbf{f}}_s - \mathbf{f}_s\right\|_F^2\right) \\
&= \sum_{j=1}^{m}\left[\frac{1}{T^2N^2}\sum_{t=1}^{T}\sum_{i=1}^{N} v_{it}^2 \left(\sum_{s=1}^{T} x_{is}^2 \left(\hat{f}_{js} - f_{js}\right)^2\right)\right] \leq \\
&\max_j \frac{1}{T^2N^2}\sum_{t=1}^{T}\sum_{i=1}^{N} v_{it}^2 \left(\sum_{s=1}^{T} x_{is}^2 \left(\hat{f}_{js} - f_{js}\right)^2\right).
\end{aligned}
$$

But

$$
\frac{1}{T^2N^2}\sum_{t=1}^{T}\sum_{i=1}^{N} v_{it}^2 \left(\sum_{s=1}^{T} x_{is}^2 \left(\hat{f}_{js} - f_{js}\right)^2\right) =
$$
$$
\frac{1}{TN^2}\sum_{s=1}^{T}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T} v_{it}^2\right) x_{is}^2 \left(\hat{f}_{js} - f_{js}\right)^2 = \frac{1}{TN^2}\sum_{s=1}^{T}\sum_{i=1}^{N} z_{is}\left(\hat{f}_{js} - f_{js}\right)^2,
$$

where $z_{it} = \left(\frac{1}{T}\sum_{s=1}^{T} v_{is}^2\right) x_{it}^2$. Note that $\sup_{i,t} E\left(z_{it}^2\right) < \infty$. Then, by a similar analysis to term A_1 in (63) of Lemma 5,

$$
C_3 = O_p\left(N^{-1}\min(N,T)^{-1}\right).
$$

Further,

$$
\begin{aligned}
C_4 &= \frac{1}{N^2}\sum_{t=1}^{T}\sum_{i=1}^{N} v_{it}^2 \left(\frac{1}{T^2}\sum_{s=1}^{T}\sum_{s',s\neq s'}^{T} \left|x_{is}x_{is'}\right| \left\|\hat{\mathbf{f}}_s - \mathbf{f}_s\right\|_F \left\|\hat{\mathbf{f}}_{s'} - \mathbf{f}_{s'}\right\|_F\right) \\
&\leq \frac{1}{N^2}\sum_{t=1}^{T}\sum_{i=1}^{N} v_{it}^2 \left[\begin{array}{c} \left(\frac{1}{T^2}\sum_{s=1}^{T}\sum_{s',s\neq s'}^{T}\left(x_{is}x_{is'}\right)^2\right)^{1/2} \\ \left(\frac{1}{T^2}\sum_{s=1}^{T}\sum_{s',s\neq s'}^{T}\left\|\hat{\mathbf{f}}_s - \mathbf{f}_s\right\|_F^2 \left\|\hat{\mathbf{f}}_{s'} - \mathbf{f}_{s'}\right\|_F^2\right)^{1/2} \end{array}\right] \\
&\leq \frac{1}{N^2}\sum_{t=1}^{T}\sum_{i=1}^{N} v_{it}^2 \left[\begin{array}{c} \left(\frac{1}{T^2}\sum_{s=1}^{T}\sum_{s',s\neq s'}^{T}\left(x_{is}x_{is'}\right)^2\right)^{1/2} \\ \left\{\frac{1}{T}\sum_{s=1}^{T}\left[\left\|\hat{\mathbf{f}}_s - \mathbf{f}_s\right\|_F^2 \left(\frac{1}{T}\sum_{s',s\neq s'}^{T}\left\|\hat{\mathbf{f}}_s - \mathbf{f}_s\right\|_F^2\right)\right]\right\}^{1/2} \end{array}\right] \\
&\leq \left(\frac{1}{T}\sum_{s=1}^{T}\left\|\hat{\mathbf{f}}_s - \mathbf{f}_s\right\|_F^2\right) \\
&\left\{\frac{1}{N^2}\sum_{t=1}^{T}\sum_{i=1}^{N} v_{it}^2\left[\left(\frac{1}{T^2}\sum_{s=1}^{T}\sum_{s',s\neq s'}^{T}\left|x_{is}x_{is'}\right|\right)^{1/2}\right]\right\}.
\end{aligned}
$$

But

$$
\sup_i E\left(\frac{1}{T^2}\sum_{s=1}^{T}\sum_{s',s\neq s'}^{T}\left(x_{is}x_{is'}\right)^2\right)^{1/2} = O(1),
$$

and therefore

$$\left\{\frac{1}{N^2}\sum_{t=1}^{T}\sum_{i=1}^{N}v_{it}^2\left[\left(\frac{1}{T^2}\sum_{s=1}^{T}\sum_{s',s\neq s'}^{T}\left(x_{is}x_{is'}\right)^2\right)^{1/2}\right]\right\}=O_p\left(TN^{-1}\right).$$

Further,

$$E\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\hat{\mathbf{f}}_s-\mathbf{f}_s\right\|_F^2\right)=O\left[\min(N,T)^{-1}\right],$$

and overall we have $C_4=O_p\left(\frac{T}{N\min(N,T)}\right)$. Finally,

$$\begin{aligned}&\frac{1}{N^2}\sum_{t=1}^{T}\sum_{i=1}^{N}v_{it}^2\left[\frac{1}{T^2}\sum_{s=1}^{T}\sum_{s'=1}^{T}x_{is'}v_{is}\mathbf{f}_s'\left(\hat{\mathbf{f}}_{s'}-\mathbf{f}_{s'}\right)\right]\\&=\left[\frac{1}{N^2}\sum_{t=1}^{T}\sum_{i=1}^{N}v_{it}^2\left(\frac{1}{T}\sum_{s=1}^{T}v_{is}\mathbf{f}_s'\right)\right]\left[\frac{1}{T}\sum_{s=1}^{T}x_{is'}\left(\hat{\mathbf{f}}_{s'}-\mathbf{f}_{s'}\right)\right].\end{aligned}$$

But using Lemma A.1 of Bai (2003) and $\sup_{i,t}E\left(x_{it}^2\right)<\infty$,

$$\left\|\frac{1}{T}\sum_{s=1}^{T}x_{is'}\left(\hat{\mathbf{f}}_{s'}-\mathbf{f}_{s'}\right)\right\|_F=O_p\left[\min(N,T)^{-1}\right],$$

and

$$\frac{1}{N^2}\sum_{t=1}^{T}\sum_{i=1}^{N}v_{it}^2\left(\frac{1}{T}\sum_{s=1}^{T}v_{is}\mathbf{f}_s'\right)=O_p\left(TN^{-1}\right).$$

So, we have

$$\frac{1}{N^2}\sum_{t=1}^{T}\sum_{i=1}^{N}v_{it}^2\left[\frac{1}{T^2}\sum_{s=1}^{T}\sum_{s'=1}^{T}x_{is'}v_{is}\mathbf{f}_s'\left(\hat{\mathbf{f}}_{s'}-\mathbf{f}_{s'}\right)\right]=O_p\left[TN^{-1}\min(N,T)^{-1}\right],$$

and hence

$$\left\|\frac{\mathbf{V}\left(\hat{\mathbf{A}}-\mathbf{A}_0\right)}{N}\right\|_F=O_p\left(N^{-1/2}\right)+O_p\left(\frac{\sqrt{T}}{\sqrt{N\min(N,T)}}\right)=O_p\left(\frac{\sqrt{T}}{\sqrt{N\min(N,T)}}\right).$$

■

Lemma 4 *Under Assumptions 1–4,*

$$\frac{\left\|\hat{\mathbf{F}}-\mathbf{F}_0\right\|_F^2}{T}=O_p\left(\frac{1}{\min(N,T)}\right).$$

Proof.

Since

$$\frac{1}{T}\left\|\hat{\mathbf{F}}-\mathbf{F}_0\right\|_F^2=\frac{1}{T}\sum_{t=1}^{T}\left\|\hat{\mathbf{f}}_t-\mathbf{f}_t\right\|_F^2,$$

then the required result follows immediately from Theorem 1 of Bai and Ng (2002). ■

Lemma 5 *Under Assumptions 1–4,*

$$\left\|\frac{\mathbf{V}'\left(\hat{\mathbf{F}}-\mathbf{F}_0\right)}{T}\right\|_F = O_p\left(\frac{\sqrt{N}}{T}\right) + O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{N^{1/4}}{T^{3/4}}\right).$$

Proof. We have that

$$\left\|\frac{\mathbf{V}'\left(\hat{\mathbf{F}}-\mathbf{F}_0\right)}{T}\right\|_F^2 = \sum_{i=1}^{N}\sum_{j=1}^{m}\left[\frac{1}{T}\sum_{t=1}^{T} v_{it}\left(\hat{f}_{jt}-f_{jt}\right)\right]^2 \leq \max_j \sum_{i=1}^{N}\left[\frac{1}{T}\sum_{t=1}^{T} v_{it}\left(\hat{f}_{jt}-f_{jt}\right)\right]^2.$$

But,

$$\sum_{i=1}^{N}\left[\frac{1}{T}\sum_{t=1}^{T} v_{it}\left(\hat{f}_{jt}-f_{jt}\right)\right]^2 = \frac{1}{T^2}\sum_{i=1}^{N}\sum_{t=1}^{T} v_{it}^2\left(\hat{f}_{jt}-f_{jt}\right)^2 + \frac{2}{T^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} v_{it}v_{is}\left(\hat{f}_{jt}-f_{jt}\right)\left(\hat{f}_{js}-f_{js}\right) = A_1 + A_2. \tag{63}$$

By equation (53) we can write

$$\hat{f}_{jt}-f_{jt} = \frac{1}{T}\sum_{l=1}^{T}\hat{f}_{jl}\gamma_{lt} + \frac{1}{T}\sum_{l=1}^{T}\hat{f}_{jl}\zeta_{lt} + \frac{1}{T}\sum_{l=1}^{T}\hat{f}_{jl}\varkappa_{lt} + \frac{1}{T}\sum_{l=1}^{T}\hat{f}_{jl}\xi_{lt},$$

where $\zeta_{lt} = N^{-1}\mathbf{v}_l'\mathbf{v}_t - \gamma_{lt}$, $\varkappa_{lt} = N^{-1}\mathbf{f}_l'\mathbf{A}_0'\mathbf{v}_t$, and $\xi_{lt} = \varkappa_{tl}$. We have

$$\begin{aligned}\frac{1}{T^2}\sum_{i=1}^{N}\sum_{t=1}^{T} v_{it}^2\left(\hat{f}_{jt}-f_{jt}\right)^2 \leq{}& \frac{4}{T^2}\sum_{i=1}^{N}\sum_{t=1}^{T} v_{it}^2\left(\frac{1}{T}\sum_{l=1}^{T}\hat{f}_{jl}\gamma_{lt}\right)^2 + \\ & \frac{4}{T^2}\sum_{i=1}^{N}\sum_{t=1}^{T} v_{it}^2\left(\frac{1}{T}\sum_{l=1}^{T}\hat{f}_{jl}\zeta_{lt}\right)^2 + \\ & \frac{4}{T^2}\sum_{i=1}^{N}\sum_{t=1}^{T} v_{it}^2\left(\frac{1}{T}\sum_{l=1}^{T}\hat{f}_{jl}\varkappa_{lt}\right)^2 + \\ & \frac{4}{T^2}\sum_{i=1}^{N}\sum_{t=1}^{T} v_{it}^2\left(\frac{1}{T}\sum_{l=1}^{T}\hat{f}_{jl}\xi_{lt}\right)^2 \\ ={}& A_{11} + A_{12} + A_{13} + A_{14}.\end{aligned}$$

Now,

$$A_{11} = \frac{1}{T^4}\sum_{i=1}^{N}\sum_{t=1}^{T} v_{it}^2\left(\sum_{l=1}^{T}\hat{f}_{jl}\gamma_{lt}\right)^2 \leq \left(\frac{1}{T}\sum_{l=1}^{T}\hat{f}_{jl}^2\right)\frac{1}{T^3}\sum_{i=1}^{N}\sum_{t=1}^{T} v_{it}^2\left(\sum_{l=1}^{T}\gamma_{lt}^2\right).$$

But $T^{-1}\sum_{l=1}^{T}\hat{f}_{jl}^2 = O_p(1)$, and $\sum_{l=1}^{T}\gamma_{lt}^2 < C$. Hence

$$\sum_{i=1}^{N}\sum_{t=1}^{T} v_{it}^2\left(\sum_{l=1}^{T}\gamma_{lt}^2\right) \leq C\sum_{i=1}^{N}\sum_{t=1}^{T} v_{it}^2 = O_p\left(TN\right).$$

So

$$A_{11} = \frac{1}{T^4}\sum_{i=1}^{N}\sum_{t=1}^{T} v_{it}^2 \left(\sum_{l=1}^{T} \hat{f}_{jl}\gamma_{lt}\right)^2 = O_p\left(NT^{-2}\right).$$

Next

$$\begin{aligned}
A_{12} &= \frac{1}{T^2}\sum_{i=1}^{N}\sum_{t=1}^{T} v_{it}^2 \left(\frac{1}{T}\sum_{l=1}^{T} \hat{f}_{jl}\zeta_{lt}\right)^2 = \frac{1}{T^4}\sum_{i=1}^{N}\sum_{t=1}^{T} v_{it}^2 \left(\sum_{l=1}^{T} \hat{f}_{jl}\zeta_{lt}\right)^2 \\
&= \frac{1}{T^4}\left(\sum_{l=1}^{T}\sum_{u=1}^{T} \hat{f}_{jl}\hat{f}_{ju}\sum_{t=1}^{T}\zeta_{lt}\zeta_{ut}\left(\sum_{i=1}^{N} v_{it}^2\right)\right) \leq \\
&\frac{N}{T^2}\left[\frac{1}{T^2}\sum_{l=1}^{T}\sum_{u=1}^{T}\left(\hat{f}_{jl}\hat{f}_{ju}\right)^2\right]^{1/2}\left\{\frac{1}{T^2}\sum_{l=1}^{T}\sum_{u=1}^{T}\left[\sum_{t=1}^{T}\zeta_{lt}\zeta_{ut}\left(\frac{1}{N}\sum_{i=1}^{N} v_{it}^2\right)\right]^2\right\}^{1/2} \\
&\leq \frac{N}{T^2}\left(\frac{1}{T}\sum_{l=1}^{T}\hat{f}_{jl}^2\right)\left\{\frac{1}{T^2}\sum_{l=1}^{T}\sum_{u=1}^{T}\left[\sum_{t=1}^{T}\zeta_{lt}\zeta_{ut}\left(\frac{1}{N}\sum_{i=1}^{N} v_{it}^2\right)\right]^2\right\}^{1/2}.
\end{aligned}$$

But

$$E\left[\left(\sum_{t=1}^{T}\zeta_{lt}\zeta_{ut}\left(\frac{1}{N}\sum_{i=1}^{N} v_{it}^2\right)\right)^2\right] \leq T^2N^{-2}.$$

So

$$A_{12} = \frac{N}{T^2}\cdot O_p(1)\cdot\sqrt{T^2N^{-2}} = O_p\left(T^{-1}\right).$$

Next,

$$\begin{aligned}
A_{13} &= \frac{4}{T^2}\sum_{i=1}^{N}\sum_{t=1}^{T} v_{it}^2 \left(\frac{1}{T}\sum_{l=1}^{T} \hat{f}_{jl}\varkappa_{lt}\right)^2 = \frac{4}{N^2T^2}\sum_{i=1}^{N}\sum_{t=1}^{T} v_{it}^2\frac{1}{T^2}\left(\sum_{l=1}^{T}\hat{f}_{jl}\mathbf{f}_l'\mathbf{A}_0'\mathbf{v}_t\right)^2 \leq \\
&\left(\frac{1}{T}\sum_{l=1}^{T}\hat{f}_{jl}^2\right)\left(\frac{1}{T}\sum_{l=1}^{T}\|\mathbf{f}_l\|^2\right)\frac{4}{NT^2}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N} v_{it}^2\right)\|\mathbf{A}_0'\mathbf{v}_t\|^2 \\
&= \left(\frac{1}{T}\sum_{l=1}^{T}\hat{f}_{jl}^2\right)\left(\frac{1}{T}\sum_{l=1}^{T}\|\mathbf{f}_l\|^2\right)\frac{4}{T^2}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N} v_{it}^2\right)\left\|\frac{\mathbf{A}_0'\mathbf{v}_t}{\sqrt{N}}\right\|^2 = O_p\left(T^{-1}\right).
\end{aligned}$$

Similarly for A_{14}. So, overall

$$A_1 = O_p\left(NT^{-2}\right) + O_p\left(T^{-1}\right) = O_p\left(\frac{N}{T\min(N,T)}\right).$$

Next we consider A_2, and note that

$$
\begin{aligned}
\left(\hat{f}_{jt}-f_{jt}\right)\left(\hat{f}_{js}-f_{js}\right) &= \left(\frac{1}{T}\sum_{l=1}^{T}\hat{f}_{jl}\gamma_{lt}+\frac{1}{T}\sum_{l=1}^{T}\hat{f}_{jl}\zeta_{lt}+\frac{1}{T}\sum_{l=1}^{T}\hat{f}_{jl}\varkappa_{lt}+\frac{1}{T}\sum_{l=1}^{T}\hat{f}_{jl}\xi_{lt}\right) \\
&\times\left(\frac{1}{T}\sum_{l=1}^{T}\hat{f}_{jl}\gamma_{ls}+\frac{1}{T}\sum_{l=1}^{T}\hat{f}_{jl}\zeta_{ls}+\frac{1}{T}\sum_{l=1}^{T}\hat{f}_{jl}\varkappa_{ls}+\frac{1}{T}\sum_{l=1}^{T}\hat{f}_{jl}\xi_{ls}\right) \\
&= \frac{1}{T^2}\sum_{l=1}^{T}\sum_{u=1}^{T}\hat{f}_{jl}\hat{f}_{ju}\gamma_{lt}\gamma_{us}+\frac{2}{T^2}\sum_{l=1}^{T}\sum_{u=1}^{T}\hat{f}_{jl}\hat{f}_{ju}\gamma_{lt}\zeta_{us}+ \\
&\frac{2}{T^2}\sum_{l=1}^{T}\sum_{u=1}^{T}\hat{f}_{jl}\hat{f}_{ju}\gamma_{lt}\varkappa_{us}+\frac{2}{T^2}\sum_{l=1}^{T}\sum_{u=1}^{T}\hat{f}_{jl}\hat{f}_{ju}\gamma_{lt}\xi_{us}+ \\
&\frac{1}{T^2}\sum_{l=1}^{T}\sum_{u=1}^{T}\hat{f}_{jl}\hat{f}_{ju}\zeta_{lt}\zeta_{us}+\frac{2}{T^2}\sum_{l=1}^{T}\sum_{u=1}^{T}\hat{f}_{jl}\hat{f}_{ju}\zeta_{lt}\varkappa_{us}+\frac{2}{T^2}\sum_{l=1}^{T}\sum_{u=1}^{T}\hat{f}_{jl}\hat{f}_{ju}\zeta_{lt}\xi_{us}+ \\
&\frac{1}{T^2}\sum_{l=1}^{T}\sum_{u=1}^{T}\hat{f}_{jl}\hat{f}_{ju}\varkappa_{lt}\varkappa_{us}+\frac{2}{T^2}\sum_{l=1}^{T}\sum_{u=1}^{T}\hat{f}_{jl}\hat{f}_{ju}\varkappa_{lt}\xi_{us}+\frac{1}{T^2}\sum_{l=1}^{T}\sum_{u=1}^{T}\hat{f}_{jl}\hat{f}_{ju}\xi_{lt}\xi_{us}.
\end{aligned}
$$

Therefore

$$
A_2=\frac{2}{T^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}v_{it}v_{is}\left(\hat{f}_{jt}-f_{jt}\right)\left(\hat{f}_{js}-f_{js}\right)=\sum_{i=1}^{10}A_{2i}.
$$

Denoting equality in order of probability by $A\sim B$, we proceed term by term noting that $A_{23}\sim A_{24}$, $A_{26}\sim A_{27}$ and $A_{28}\sim A_{210}$. So the terms to consider are A_{21}, A_{22}, A_{23}, A_{25}, A_{26}, A_{28} and A_{29}. Starting with A_{21} we have

$$
\begin{aligned}
A_{21} &= \frac{2}{T^4}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}v_{it}v_{is}\left(\sum_{l=1}^{T}\sum_{u=1}^{T}\hat{f}_{jl}\hat{f}_{ju}\gamma_{lt}\gamma_{us}\right) \\
&= \frac{2}{T^4}\sum_{i=1}^{N}\sum_{l=1}^{T}\sum_{u=1}^{T}\hat{f}_{jl}\hat{f}_{ju}\left(\sum_{t=1}^{T}\sum_{s=1}^{T}v_{it}v_{is}\gamma_{lt}\gamma_{us}\right) \\
&\le \frac{2}{T^2}\sum_{i=1}^{N}\left\{\begin{array}{c}\left(\frac{1}{T^2}\sum_{l=1}^{T}\sum_{u=1}^{T}\left(\hat{f}_{jl}\hat{f}_{ju}\right)^2\right)^{1/2} \\ \left[\frac{1}{T^3}\sum_{l=1}^{T}\sum_{u=1}^{T}\left(\sum_{t=1}^{T}\sum_{s=1}^{T}v_{it}v_{is}\gamma_{lt}\gamma_{us}\right)^2\right]^{1/2}\end{array}\right\} \\
&= \left\{\frac{2}{T^2}\sum_{i=1}^{N}\left[\frac{1}{T^2}\sum_{l=1}^{T}\sum_{u=1}^{T}\left(\sum_{t=1}^{T}\sum_{s=1}^{T}v_{it}v_{is}\gamma_{lt}\gamma_{us}\right)^2\right]^{1/2}\right\}\left(\frac{1}{T}\sum_{u=1}^{T}\hat{f}_{ju}^2\right).
\end{aligned}
$$

But, due to summability of γ_{lt}

$$
E\left[\left(\sum_{t=1}^{T}\sum_{s=1}^{T}v_{it}v_{is}\gamma_{lt}\gamma_{us}\right)^2\right]\le T\ C.
$$

Noting further that, again due to summability of γ_{lt}, the double sum over l and u will only have terms bounded away from zero if l and u are close we obtain

$$E\left[\frac{1}{T^2}\sum_{l=1}^{T}\sum_{u=1}^{T}\left(\sum_{t=1}^{T}\sum_{s=1}^{T}v_{it}v_{is}\gamma_{lt}\gamma_{us}\right)^2\right]=O\left(1\right),$$

and hence $A_{21}=O_p\left(NT^{-2}\right)$. Consider now

$$\begin{aligned}
A_{22}&=\frac{2}{T^4}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}v_{it}v_{is}\left(\sum_{l=1}^{T}\sum_{u=1}^{T}\hat{f}_{jl}\hat{f}_{ju}\gamma_{lt}\zeta_{us}\right)\\
&=\frac{2}{T^4}\sum_{i=1}^{N}\sum_{l=1}^{T}\sum_{u=1}^{T}\hat{f}_{jl}\hat{f}_{ju}\left(\sum_{t=1}^{T}\sum_{s=1}^{T}v_{it}v_{is}\gamma_{lt}\zeta_{us}\right)\\
&\leq\frac{2}{T^{3/2}}\sum_{i=1}^{N}\left[\begin{array}{c}\left(\frac{1}{T^2}\sum_{l=1}^{T}\sum_{u=1}^{T}\left(\hat{f}_{jl}\hat{f}_{ju}\right)^2\right)^{1/2}\\ \left[\frac{1}{T^3}\sum_{l=1}^{T}\sum_{u=1}^{T}\left(\sum_{t=1}^{T}\sum_{s=1}^{T}v_{it}v_{is}\gamma_{lt}\zeta_{us}\right)^2\right]^{1/2}\end{array}\right]\\
&\leq\left[\frac{2}{T^{3/2}}\sum_{i=1}^{N}\left[\left[\frac{1}{T^3}\sum_{l=1}^{T}\sum_{u=1}^{T}\left(\sum_{t=1}^{T}\sum_{s=1}^{T}v_{it}v_{is}\gamma_{lt}\zeta_{us}\right)^2\right]^{1/2}\right]\right]\times\left(\frac{1}{T}\sum_{u=1}^{T}\hat{f}_{ju}^2\right).
\end{aligned}$$

But

$$E\left[\left(\sum_{t=1}^{T}\sum_{s=1}^{T}v_{it}v_{is}\gamma_{lt}\zeta_{us}\right)^2\right]\leq T^2N^{-1}.$$

Further, due to summability of γ_{lt} the double sum over l and t will only have terms bounded away from zero if l and t are close so

$$E\left[\frac{1}{T^3}\sum_{l=1}^{T}\sum_{u=1}^{T}\left(\sum_{t=1}^{T}\sum_{s=1}^{T}v_{it}v_{is}\gamma_{lt}\zeta_{us}\right)^2\right]=O\left(N^{-1}\right), \tag{64}$$

and as a result $A_{22}=O_p\left(N^{1/2}T^{-3/2}\right)$. Next, and similarly to the previous terms

$$\begin{aligned}
A_{23}&=\frac{2}{T^4}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}v_{it}v_{is}\left(\sum_{l=1}^{T}\sum_{u=1}^{T}\hat{f}_{jl}\hat{f}_{ju}\gamma_{lt}\varkappa_{us}\right)\leq\\
&\leq\left(\frac{2}{T^{3/2}}\sum_{i=1}^{N}\left\{\left[\frac{1}{T^3}\sum_{l=1}^{T}\sum_{u=1}^{T}\left(\sum_{t=1}^{T}\sum_{s=1}^{T}v_{it}v_{is}\gamma_{lt}\varkappa_{us}\right)^2\right]^{1/2}\right\}\right)\left(\frac{1}{T}\sum_{u=1}^{T}\hat{f}_{ju}^2\right),
\end{aligned}$$

which again, by a manipulation similar to that used for (64), yields $A_{23} = O_p\left(N^{1/2}T^{-3/2}\right)$. Next,

$$
\begin{aligned}
A_{25} &= \frac{2}{T^4}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} v_{it}v_{is}\left(\sum_{l=1}^{T}\sum_{u=1}^{T}\hat{f}_{jl}\hat{f}_{ju}\zeta_{lt}\zeta_{us}\right) \\
&= \frac{2}{T^4}\sum_{i=1}^{N}\left[\sum_{l=1}^{T}\sum_{u=1}^{T}\hat{f}_{jl}\hat{f}_{ju}\left(\sum_{t=1}^{T}\sum_{s=1}^{T}\zeta_{lt}\zeta_{us}v_{it}v_{is}\right)\right] \\
&\leq \frac{2}{T}\sum_{i=1}^{N}\left\{\begin{array}{c}\left(\frac{1}{T^2}\sum_{l=1}^{T}\sum_{u=1}^{T}\left(\hat{f}_{jl}\hat{f}_{ju}\right)^2\right)^{1/2} \\ \left[\frac{1}{T^4}\sum_{l=1}^{T}\sum_{u=1}^{T}\left(\sum_{t=1}^{T}\sum_{s=1}^{T}\zeta_{lt}\zeta_{us}v_{it}v_{is}\right)^2\right]^{1/2}\end{array}\right\} \\
&\leq \frac{2}{T}\sum_{i=1}^{N}\left\{\left(\frac{1}{T}\sum_{u=1}^{T}\hat{f}_{ju}^2\right)\left[\frac{1}{T^4}\sum_{l=1}^{T}\sum_{u=1}^{T}\left(\sum_{t=1}^{T}\sum_{s=1}^{T}\zeta_{lt}\zeta_{us}v_{it}v_{is}\right)^2\right]^{1/2}\right\}.
\end{aligned}
$$

But, by absolute summability of the autocovariance of v_{it},

$$
E\left[\left(\sum_{t=1}^{T}\sum_{s=1}^{T}\zeta_{lt}\zeta_{us}v_{it}v_{is}\right)^2\right] \leq CT^2E\left(\zeta_{lt}^4\right) \leq CT^4N^{-2}.
$$

So

$$
\left[\frac{1}{T^4}\sum_{l=1}^{T}\sum_{u=1}^{T}\left(\sum_{t=1}^{T}\sum_{s=1}^{T}\zeta_{lt}\zeta_{us}v_{it}v_{is}\right)^2\right]^{1/2} = O_p\left(N^{-1}\right),
$$

$$
\frac{2}{T}\sum_{i=1}^{N}\left\{\left(\frac{1}{T}\sum_{u=1}^{T}\hat{f}_{ju}^2\right)\left[\frac{1}{T^4}\sum_{l=1}^{T}\sum_{u=1}^{T}\left(\sum_{t=1}^{T}\sum_{s=1}^{T}\zeta_{lt}\zeta_{us}v_{it}v_{is}\right)^2\right]^{1/2}\right\} = O_p\left(T^{-1}\right),
$$

and $A_{25} = O_p\left(T^{-1}\right)$. Next

$$
\begin{aligned}
A_{26} &= \frac{2}{T^4}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} v_{it}v_{is}\left(\sum_{l=1}^{T}\sum_{u=1}^{T}\hat{f}_{jl}\hat{f}_{ju}\zeta_{lt}\varkappa_{us}\right) \leq \\
&\leq \left(\frac{2}{T}\sum_{i=1}^{N}\left\{\left[\frac{1}{T^4}\sum_{l=1}^{T}\sum_{u=1}^{T}\left(\sum_{t=1}^{T}\sum_{s=1}^{T}v_{it}v_{is}\zeta_{lt}\varkappa_{us}\right)^2\right]^{1/2}\right\}\right)\left(\frac{1}{T}\sum_{u=1}^{T}\hat{f}_{ju}^2\right).
\end{aligned}
$$

But

$$
E\left[\left(\sum_{t=1}^{T}\sum_{s=1}^{T}v_{it}v_{is}\zeta_{lt}\varkappa_{us}\right)^2\right] \leq T^2N^{-2},
$$

and

$$
E\left[\frac{1}{T^3}\sum_{l=1}^{T}\sum_{u=1}^{T}\left(\sum_{t=1}^{T}\sum_{s=1}^{T}v_{it}v_{is}\zeta_{lt}\varkappa_{us}\right)^2\right] = O\left(N^{-2}\right).
$$

So $A_{26} = O_p\left(T^{-1}\right)$. Similarly, we obtain $A_{28} = O_p\left(T^{-1}\right)$ and $A_{29} = O_p\left(T^{-1}\right)$. Collecting the terms, we have

$$A_2 = O_p\left(NT^{-2}\right) + O_p\left(N^{1/2}T^{-3/2}\right) + O_p\left(T^{-1}\right).$$

Thus

$$\left\| \frac{\mathbf{V}'\left(\hat{\mathbf{F}} - \mathbf{F}_0\right)}{T} \right\|_F^2 = O_p\left(NT^{-2}\right) + O_p\left(T^{-1}\right) + O_p\left(N^{1/2}T^{-3/2}\right),$$

and hence

$$\left\| \frac{\mathbf{V}'\left(\hat{\mathbf{F}} - \mathbf{F}_0\right)}{T} \right\|_F = O_p\left(N^{1/2}T^{-1}\right) + O_p\left(T^{-1/2}\right) + O_p\left(N^{1/4}T^{-3/4}\right).$$

■

Lemma 6 *Under Assumptions 1–4,*

$$\frac{\left\| \hat{\mathbf{A}} - \mathbf{A}_0 \right\|_F^2}{N} = O_p\left(\frac{1}{\min(N,T)}\right),$$

and

$$\left\| \mathbf{A}_0'\left(\hat{\mathbf{A}} - \mathbf{A}_0\right) \right\|_F = O_p\left(\frac{N}{\sqrt{\min(N,T)}}\right). \tag{65}$$

Proof. We have by the proof of Theorem 2 of Bai (2003, expression above (B.2)) that

$$\hat{\mathbf{a}}_i - \mathbf{a}_i = \frac{1}{T}\sum_{t=1}^{T} \mathbf{f}_t v_{it} + \frac{1}{T}\sum_{t=1}^{T} x_{it}\left(\hat{\mathbf{f}}_t - \mathbf{f}_t\right).$$

This result can be used to obtain

$$\begin{aligned}
\frac{1}{N}\left\|\left(\hat{\mathbf{A}}-\mathbf{A}_0\right)\right\|_F^2 &= \frac{1}{N}\sum_{i=1}^{N}\left(\hat{\mathbf{a}}_i-\mathbf{a}_i\right)'\left(\hat{\mathbf{a}}_i-\mathbf{a}_i\right) \\
&\leq \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T^2}\sum_{s=1}^{T}\mathbf{f}_s'\mathbf{f}_s v_{is}^2+\frac{1}{T^2}\sum_{s=1}^{T}\sum_{s',s\neq s'}^{T}\mathbf{f}_s'\mathbf{f}_{s'}v_{is}v_{is'}\right)+ \\
&\frac{1}{N}\sum_{i=1}^{N}\left(\begin{array}{c}\frac{1}{T^2}\sum_{s=1}^{T}x_{is}^2\left\|\hat{\mathbf{f}}_s-\mathbf{f}_s\right\|_F^2+ \\ \frac{1}{T^2}\sum_{s=1}^{T}\sum_{s',s\neq s'}^{T}\left|x_{is}x_{is'}\right|\left\|\hat{\mathbf{f}}_s-\mathbf{f}_s\right\|_F\left\|\hat{\mathbf{f}}_{s'}-\mathbf{f}_{s'}\right\|_F\end{array}\right)+ \\
&\frac{2}{N}\sum_{i=1}^{N}\left(\frac{1}{T^2}\sum_{s=1}^{T}\sum_{s'=1}^{T}x_{is'}v_{is}\mathbf{f}_s'\left(\hat{\mathbf{f}}_{s'}-\mathbf{f}_{s'}\right)\right)= \\
&\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T^2}\sum_{s=1}^{T}\mathbf{f}_s'\mathbf{f}_s v_{is}^2\right)+ \\
&\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T^2}\sum_{s=1}^{T}\sum_{s',s\neq s'}^{T}\mathbf{f}_s'\mathbf{f}_{s'}v_{is}v_{is'}\right)+ \\
&\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T^2}\sum_{s=1}^{T}x_{is}^2\left\|\hat{\mathbf{f}}_s-\mathbf{f}_s\right\|_F^2\right)+ \\
&\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T^2}\sum_{s=1}^{T}\sum_{s',s\neq s'}^{T}\left|x_{is}x_{is'}\right|\left\|\hat{\mathbf{f}}_s-\mathbf{f}_s\right\|_F\left\|\hat{\mathbf{f}}_{s'}-\mathbf{f}_{s'}\right\|_F\right)+ \\
&\frac{2}{N}\sum_{i=1}^{N}\left(\frac{1}{T^2}\sum_{s=1}^{T}\sum_{s'=1}^{T}x_{is'}v_{is}\mathbf{f}_s'\left(\hat{\mathbf{f}}_{s'}-\mathbf{f}_{s'}\right)\right) \\
&=\sum_{i=1}^{5}C_i.
\end{aligned}$$

We have

$$C_1=\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T^2}\sum_{s=1}^{T}\mathbf{f}_s'\mathbf{f}_s v_{is}^2\right)=O_p\left(T^{-1}\right).$$

Also

$$\begin{aligned}
C_2 &= \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T^2}\sum_{s=1}^{T}\sum_{s',s\neq s'}^{T}\mathbf{f}_s'\mathbf{f}_{s'}v_{is}v_{is'}\right) \\
&=\frac{1}{N}\frac{1}{T}\sum_{i=1}^{N}\left[\frac{1}{\sqrt{T}}\sum_{s=1}^{T}v_{is}\mathbf{f}_s'\left(\frac{1}{\sqrt{T}}\sum_{s',s\neq s'}^{T}\mathbf{f}_{s'}v_{is'}\right)\right]=O_p\left(T^{-1}\right),
\end{aligned}$$

and

$$C_3 = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T^2}\sum_{s=1}^{T}x_{is}^2\left\|\hat{\mathbf{f}}_s-\mathbf{f}_s\right\|_F^2\right)$$
$$= \frac{1}{\min(N,T)NT}\sum_{i=1}^{N}\left(\frac{\min(N,T)}{T}\sum_{s=1}^{T}x_{is}^2\left\|\hat{\mathbf{f}}_s-\mathbf{f}_s\right\|_F^2\right) = O_p\left(T^{-1}\min(N,T)^{-1}\right),$$

noting that by Lemma A.1 of Bai (2003) and $\sup_i \sup_t E\left(x_{it}^2\right) < \infty$,

$$\sup_i E\left(\frac{\min(N,T)}{T}\sum_{s=1}^{T}x_{is}^2\left\|\hat{\mathbf{f}}_s-\mathbf{f}_s\right\|_F^2\right)^2 = O(1).$$

Further,

$$C_4 = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T^2}\sum_{s=1}^{T}\sum_{s',s\neq s'}^{T}|x_{is}x_{is'}|\left\|\hat{\mathbf{f}}_s-\mathbf{f}_s\right\|_F\left\|\hat{\mathbf{f}}_{s'}-\mathbf{f}_{s'}\right\|_F\right)$$
$$\leq \frac{1}{N}\sum_{i=1}^{N}\left[\begin{array}{c}\left(\frac{1}{T^2}\sum_{s=1}^{T}\sum_{s',s\neq s'}^{T}\left(x_{is}x_{is'}\right)^2\right)^{1/2} \\ \left(\frac{1}{T^2}\sum_{s=1}^{T}\sum_{s',s\neq s'}^{T}\left\|\hat{\mathbf{f}}_s-\mathbf{f}_s\right\|_F^2\left\|\hat{\mathbf{f}}_{s'}-\mathbf{f}_{s'}\right\|_F^2\right)^{1/2}\end{array}\right]$$
$$\leq \frac{1}{N}\sum_{i=1}^{N}\left[\begin{array}{c}\left(\frac{1}{T^2}\sum_{s=1}^{T}\sum_{s',s\neq s'}^{T}\left(x_{is}x_{is'}\right)^2\right)^{1/2} \\ \left\{\frac{1}{T}\sum_{s=1}^{T}\left[\left\|\hat{\mathbf{f}}_s-\mathbf{f}_s\right\|_F^2\left(\frac{1}{T}\sum_{s',s\neq s'}^{T}\left\|\hat{\mathbf{f}}_s-\mathbf{f}_s\right\|_F^2\right)\right]\right\}^{1/2}\end{array}\right]$$
$$\leq \left(\frac{1}{T}\sum_{s=1}^{T}\left\|\hat{\mathbf{f}}_s-\mathbf{f}_s\right\|_F^2\right)\left\{\frac{1}{N}\sum_{i=1}^{N}\left[\left(\frac{1}{T^2}\sum_{s=1}^{T}\sum_{s',s\neq s'}^{T}\left(x_{is}x_{is'}\right)^2\right)^{1/2}\right]\right\}.$$

But

$$\sup_i E\left(\frac{1}{T^2}\sum_{s=1}^{T}\sum_{s',s\neq s'}^{T}\left(x_{is}x_{is'}\right)^2\right)^{1/2} = O\left(1\right),$$

and so

$$\frac{1}{N}\sum_{i=1}^{N}\left[\left(\frac{1}{T^2}\sum_{s=1}^{T}\sum_{s',s\neq s'}^{T}\left(x_{is}x_{is'}\right)^2\right)^{1/2}\right] = O_p\left(1\right).$$

Further,

$$E\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\hat{\mathbf{f}}_s-\mathbf{f}_s\right\|_F^2\right) = O\left[\min(N,T)^{-1}\right].$$

So, overall $C_4 = O_p\left(\frac{1}{\min(N,T)}\right)$. Finally, noting by Lemma A.1 of Bai (2003) (or can be proven by first principles) that

$$\sup_i\left\{E\left[\frac{1}{T^2}\sum_{s=1}^{T}\sum_{s'=1}^{T}x_{is'}v_{is}\mathbf{f}_s'\left(\hat{\mathbf{f}}_{s'}-\mathbf{f}_{s'}\right)\right]^2\right\}^{1/2} = O_p\left[\min(N,T)^{-1}\right],$$

we have

$$C_5 = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{T^2}\sum_{s=1}^{T}\sum_{s'=1}^{T} x_{is'}v_{is}\mathbf{f}_s'\left(\hat{\mathbf{f}}_{s'} - \mathbf{f}_{s'}\right)\right] = O_p\left[\min(N,T)^{-1}\right],$$

So overall

$$\frac{1}{N}\left\|\hat{\mathbf{A}} - \mathbf{A}_0\right\|_F^2 = O_p\left(\frac{1}{\min(N,T)}\right).$$

To prove (65), recall that by equation (62),

$$\hat{\mathbf{a}}_i - \mathbf{a}_i = \frac{1}{T}\sum_{t=1}^{T}\mathbf{f}_t v_{it} + \frac{1}{T}\sum_{t=1}^{T} x_{it}\left(\hat{\mathbf{f}}_t - \mathbf{f}_t\right).$$

Define $\mathbf{B} = \mathbf{A}_0'\mathbf{A}_0$. Note that every element of $\mathbf{B}$ is $O(N)$ and so every element of $N^{-1}\mathbf{B}$ is bounded. We have

$$\begin{aligned}
\left\|\mathbf{A}_0'\left(\hat{\mathbf{A}} - \mathbf{A}_0\right)\right\|_F^2 &= Tr\left[\left(\hat{\mathbf{A}} - \mathbf{A}_0\right)\mathbf{A}_0'\mathbf{A}_0\left(\hat{\mathbf{A}} - \mathbf{A}_0\right)'\right] = \sum_{i=1}^{N}\left(\hat{\mathbf{a}}_i - \mathbf{a}_i\right)'\mathbf{B}\left(\hat{\mathbf{a}}_i - \mathbf{a}_i\right) \\
&\leq \sum_{i=1}^{N}\left(\frac{1}{T^2}\sum_{s=1}^{T}\mathbf{f}_s'\mathbf{B}\mathbf{f}_s v_{is}^2 + \frac{1}{T^2}\sum_{s=1}^{T}\sum_{s',s\neq s'}^{T}\mathbf{f}_s'\mathbf{B}\mathbf{f}_{s'}v_{is}v_{is'}\right) + \\
\|\mathbf{B}\|_F &\left[\sum_{i=1}^{N}\left(\begin{array}{c}\frac{1}{T^2}\sum_{s=1}^{T} x_{is}^2\left\|\hat{\mathbf{f}}_s - \mathbf{f}_s\right\|_F^2 + \\ \frac{1}{T^2}\sum_{s=1}^{T}\sum_{s',s\neq s'}^{T}|x_{is}x_{is'}|\left\|\hat{\mathbf{f}}_s - \mathbf{f}_s\right\|_F\left\|\hat{\mathbf{f}}_{s'} - \mathbf{f}_{s'}\right\|_F\end{array}\right)\right] + \\
&\sum_{i=1}^{N}\left[\frac{1}{T^2}\sum_{s=1}^{T}\sum_{s'=1}^{T} x_{is'}v_{is}\mathbf{f}_s'\mathbf{B}\left(\hat{\mathbf{f}}_{s'} - \mathbf{f}_{s'}\right)\right] = \\
&\sum_{i=1}^{N}\left(\frac{1}{T^2}\sum_{s=1}^{T}\mathbf{f}_s'\mathbf{B}\mathbf{f}_s v_{is}^2\right) + \sum_{i=1}^{N}\left(\frac{1}{T^2}\sum_{s=1}^{T}\sum_{s',s\neq s'}^{T}\mathbf{f}_s'\mathbf{B}\mathbf{f}_{s'}v_{is}v_{is'}\right) + \\
&\|\mathbf{B}\|_F\left[\sum_{i=1}^{N}\left(\frac{1}{T^2}\sum_{s=1}^{T} x_{is}^2\left\|\hat{\mathbf{f}}_s - \mathbf{f}_s\right\|_F^2\right)\right] + \\
&\|\mathbf{B}\|_F\left[\sum_{i=1}^{N}\left(\begin{array}{c}\frac{1}{T^2}\sum_{s=1}^{T}\sum_{s',s\neq s'}^{T}|x_{is}x_{is'}| \times \\ \left\|\hat{\mathbf{f}}_s - \mathbf{f}_s\right\|_F\left\|\hat{\mathbf{f}}_{s'} - \mathbf{f}_{s'}\right\|_F\end{array}\right)\right] + \\
&\sum_{i=1}^{N}\left[\frac{1}{T^2}\sum_{s=1}^{T}\sum_{s'=1}^{T} x_{is'}v_{is}\mathbf{f}_s'\mathbf{B}\left(\hat{\mathbf{f}}_{s'} - \mathbf{f}_{s'}\right)\right] \\
&= \sum_{i=1}^{5}\tilde{C}_i.
\end{aligned}$$

We have

$$\tilde{C}_1 = \sum_{i=1}^{N}\left(\frac{1}{T^2}\sum_{s=1}^{T}\mathbf{f}_s'\mathbf{B}\mathbf{f}_s v_{is}^2\right) = O_p\left(NT^{-1}\right).$$

Also

$$
\begin{aligned}
\tilde{C}_2 &= \sum_{i=1}^{N}\left(\frac{1}{T^2}\sum_{s=1}^{T}\sum_{s',s\neq s'}^{T}\mathbf{f}_s'\mathbf{B}\mathbf{f}_{s'}v_{is}v_{is'}\right) \\
&= \frac{1}{T}\sum_{i=1}^{N}\left[\frac{1}{\sqrt{T}}\sum_{s=1}^{T}v_{is}\mathbf{f}_s'\mathbf{B}\left(\frac{1}{\sqrt{T}}\sum_{s',s\neq s'}^{T}\mathbf{f}_{s'}v_{is'}\right)\right] = O_p\left(NT^{-1}\right)
\end{aligned}
$$

and

$$
\begin{aligned}
\tilde{C}_3 &= \|\mathbf{B}\|_F\sum_{i=1}^{N}\left(\frac{1}{T^2}\sum_{s=1}^{T}x_{is}^2\left\|\hat{\mathbf{f}}_s-\mathbf{f}_s\right\|_F^2\right) \\
&= \|\mathbf{B}\|_F\frac{1}{\min(N,T)T}\sum_{i=1}^{N}\left(\frac{\min(N,T)}{T}\sum_{s=1}^{T}x_{is}^2\left\|\hat{\mathbf{f}}_s-\mathbf{f}_s\right\|_F^2\right) \\
&= O_p\left[N^2T^{-1}\min(N,T)^{-1}\right].
\end{aligned}
$$

Further,

$$
\begin{aligned}
\tilde{C}_4 &= \|\mathbf{B}\|_F\sum_{i=1}^{N}\left(\frac{1}{T^2}\sum_{s=1}^{T}\sum_{s',s\neq s'}^{T}|x_{is}x_{is'}|\left\|\hat{\mathbf{f}}_s-\mathbf{f}_s\right\|_F\left\|\hat{\mathbf{f}}_{s'}-\mathbf{f}_{s'}\right\|_F\right) \\
&\leq \|\mathbf{B}\|_F\sum_{i=1}^{N}\left[\begin{array}{c}\left(\frac{1}{T^2}\sum_{s=1}^{T}\sum_{s',s\neq s'}^{T}|x_{is}x_{is'}|\right)^{1/2} \\ \left(\frac{1}{T^2}\sum_{s=1}^{T}\sum_{s',s\neq s'}^{T}\left\|\hat{\mathbf{f}}_s-\mathbf{f}_s\right\|_F^2\left\|\hat{\mathbf{f}}_{s'}-\mathbf{f}_{s'}\right\|_F^2\right)^{1/2}\end{array}\right] \\
&\leq \|\mathbf{B}\|_F\sum_{i=1}^{N}\left[\begin{array}{c}\left(\frac{1}{T^2}\sum_{s=1}^{T}\sum_{s',s\neq s'}^{T}|x_{is}x_{is'}|\right)^{1/2} \\ \left\{\frac{1}{T}\sum_{s=1}^{T}\left[\left\|\hat{\mathbf{f}}_s-\mathbf{f}_s\right\|_F^2\left(\frac{1}{T}\sum_{s',s\neq s'}^{T}\left\|\hat{\mathbf{f}}_s-\mathbf{f}_s\right\|_F^2\right)\right]\right\}^{1/2}\end{array}\right] \\
&\leq \|\mathbf{B}\|_F\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\hat{\mathbf{f}}_s-\mathbf{f}_s\right\|_F^2\right)\left\{\sum_{i=1}^{N}\left[\left(\frac{1}{T^2}\sum_{s=1}^{T}\sum_{s',s\neq s'}^{T}|x_{is}x_{is'}|\right)^{1/2}\right]\right\},
\end{aligned}
$$

and it follows that $\tilde{C}_4 = O_p\left(\frac{N^2}{\min(N,T)}\right)$. Finally, since

$$
\sup_i\left\{E\left[\frac{1}{T^2}\sum_{s=1}^{T}\sum_{s'=1}^{T}x_{is'}v_{is}\mathbf{f}_s'\left(\hat{\mathbf{f}}_{s'}-\mathbf{f}_{s'}\right)\right]^2\right\}^{1/2} = O_p\left(\min(N,T)^{-1}\right),
$$

we have

$$
\tilde{C}_5 = \|\mathbf{B}\|_F\sum_{i=1}^{N}\left[\frac{1}{T^2}\sum_{s=1}^{T}\sum_{s'=1}^{T}x_{is'}v_{is}\mathbf{f}_s'\left(\hat{\mathbf{f}}_{s'}-\mathbf{f}_{s'}\right)\right] = O_p\left[N^2\min(N,T)^{-1}\right].
$$

So overall

$$
\left\|\mathbf{A}_0'\left(\hat{\mathbf{A}}-\mathbf{A}_0\right)\right\|_F^2 = O_p\left(\frac{N^2}{\min(N,T)}\right).
$$

and

$$\left\| \mathbf{A}_0' \left(\hat{\mathbf{A}} - \mathbf{A}_0 \right) \right\|_F = O_p \left(\frac{N}{\sqrt{\min(N,T)}} \right),$$

proving the required result. ■

S2 Variants of the basic σ^2 thresholding Methods

This section provides a step by step description of the various refinements of the basic σ^2 thresholding advanced in Section 5 of the paper. Let $\mathbf{x}_i$ be the $T \times 1$ vector of observations on the i-th unit in the panel, and $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$ 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. Denote by $\mathbf{X}_a^*$ the $T \times r$ matrix containing all pervasive units that have been identified at a given step of the two algorithms described below. Analogously, let the $T \times N_1$ matrix $\mathbf{X}_b^* = (\mathbf{x}_{b,1}; \ldots; \mathbf{x}_{b,N_1})$ contain observations for the $N_1 = N - r$ remaining cross-section units that have not been identified as pervasive. Furthermore, let

$$\mathbf{M}_{\mathbf{X}_a^*} = \begin{cases} \mathbf{I}_T, & \text{if } r = 0, \\ \mathbf{I}_T - \mathbf{X}_a^* (\mathbf{X}_a^{*\prime} \mathbf{X}_a^*)^{-1} \mathbf{X}_a^{*\prime}, & \text{if } r > 0. \end{cases}.$$

Given the sequential nature of the two algorithms described below, the values of $r, N_1, \mathbf{X}_a^*$ and $\mathbf{X}_b^*$ and the dimensions of the latter two matrices change as the algorithm proceeds. Furthermore, $\mathbf{X}_a^*$ and $\mathbf{X}_b^*$ represent an estimated partition of the data into pervasive and non-pervasive units which is to be distinguish from the true partition $\mathbf{X} = (\mathbf{X}_a; \mathbf{X}_b)$.

Algorithm 4 (Sequential σ^2 thresholding)

1. *Set* $r = 0$.

2. *Compute* $\hat{\mathbf{F}} = \frac{1}{\sqrt{N}} \mathbf{M}_{\mathbf{X}_a^*} \mathbf{X}_b^* \hat{\mathbf{Q}}$, *where* $\hat{\mathbf{Q}}$ *is the* $N \times (p_{\max} - r)$ *matrix whose columns are the orthonormal eigenvectors of* $\mathbf{X}_b^{*\prime} \mathbf{M}_{\mathbf{X}_a^*} \mathbf{X}_b^*$, *such that* $N^{-1} \hat{\mathbf{Q}}' \hat{\mathbf{Q}} = \mathbf{I}_{p_{\max}}$. *For each* $i = 1$, *compute* $\hat{\mathbf{a}}_i$, $\hat{v}_{it}$ *and* $\hat{\sigma}_{iT}^2$ *to be the OLS estimator, residual and residual variance of the regression of* $\mathbf{x}_{b,i}^*$ *on* $\hat{\mathbf{F}}$, *namely*

$$\begin{aligned} \hat{\mathbf{a}}_i &= \left(\hat{\mathbf{F}}' \hat{\mathbf{F}} \right)^{-1} \hat{\mathbf{F}}' \mathbf{x}_{b,i}^*, \\ \hat{\mathbf{v}}_i &= (\hat{v}_{i1}, \hat{v}_{i2}, \ldots, \hat{v}_{iT})' = \mathbf{M}_{\hat{F}} \mathbf{x}_{b,i}^* = \left[\mathbf{I}_T - \hat{\mathbf{F}} \left(\hat{\mathbf{F}}' \hat{\mathbf{F}} \right)^{-1} \hat{\mathbf{F}}' \right] \mathbf{x}_{b,i}^*, \\ \hat{\sigma}_{iT}^2 &= T^{-1} \mathbf{x}_{b,i}^{*\prime} \mathbf{M}_{\hat{\mathbf{F}}} \mathbf{x}_{b,i}^*. \end{aligned}$$

3. *Sort* $\hat{\sigma}_{iT}^2$ *in ascending order and denote the sorted series* $\hat{\sigma}_{(1)T}^2, \hat{\sigma}_{(2)T}^2, \ldots, \hat{\sigma}_{(N)T}^2$ *with* $\hat{\sigma}_{(i)T}^2$ *being the i-th smallest value. Consider the cross-section indexes* $i_1, i_2, \ldots, i_{p_{\max}-r}$ *corresponding to* $\hat{\sigma}_{(1)T}^2, \hat{\sigma}_{(2)T}^2, \ldots, \hat{\sigma}_{(p_{\max}-r)T}^2$. *Compute*

$$\hat{\eta}_{iN}^2 = \frac{\hat{\mathbf{a}}_i' \hat{\mathbf{A}}' \tilde{\mathbf{\Sigma}}_v \hat{\mathbf{A}} \hat{\mathbf{a}}_i}{N},$$

for every $j \in \{i_1, i_2, \dots i_{,p_{\max}-r}\}$ *where* $\tilde{\mathbf{\Sigma}}_v$ *is the multiple testing estimator of* $\mathbf{\Sigma}_v$ *by Bailey, Pesaran, and Smith (2019), as described in Section 4.2 of the main paper. If for all* j,

$$\hat{\sigma}^2_{jT} > \frac{2\hat{\eta}^2_{jN}\log T}{N},$$

then stop the algorithm and conclude that there are $\hat{m} = r$ *pervasive units whose identities are given by the indexes of the columns in* $\mathbf{X}$ *that coincide with columns in* $\mathbf{X}^*_a$*. Otherwise, proceed to step 4.*

4. *Let* $i^* = \arg\min_i \hat{\sigma}^2_i$*. Update* $\mathbf{X}^*_a = (\mathbf{X}^*_a; \mathbf{x}_{b,i^*})$ *and eliminate* $\mathbf{x}_{b,i^*}$ *from* $\mathbf{X}^*_b$*. Update,* $r := r+1$ *and* $N_1 := N_1 - 1$ *and return to step 2.*

Algorithm 5 (Sequential-MT σ^2 thresholding)

1. *Set* $r = 0$.

2. *Compute* $\hat{\mathbf{F}} = \frac{1}{\sqrt{N}}\mathbf{M}_{\mathbf{X}^*_a}\mathbf{X}^*_b\hat{\mathbf{Q}}$*, where* $\hat{\mathbf{Q}}$ *is the* $N \times (p_{\max} - r)$ *matrix whose columns are the orthonormal eigenvectors of* $\mathbf{X}^{*\prime}_b\mathbf{M}_{\mathbf{X}^*_a}\mathbf{X}^*_b$*, such that* $N^{-1}\hat{\mathbf{Q}}'\hat{\mathbf{Q}} = \mathbf{I}_{p_{\max}}$*. For each* $i = 1$*, compute* $\hat{\mathbf{a}}_i$*,* $\hat{v}_{it}$ *and* $\hat{\sigma}^2_{iT}$ *to be the OLS estimator, residual and residual variance of the regression of* $\mathbf{x}^*_{b,i}$ *on* $\hat{\mathbf{F}}$*, namely*

$$\begin{aligned}\hat{\mathbf{a}}_i &= \left(\hat{\mathbf{F}}'\hat{\mathbf{F}}\right)^{-1}\hat{\mathbf{F}}'\mathbf{x}^*_{b,i},\\ \hat{\mathbf{v}}_i &= (\hat{v}_{i1}, \hat{v}_{i2}, \dots, \hat{v}_{iT})' = \mathbf{M}_{\hat{\mathbf{F}}}\mathbf{x}^*_{b,i} = \left[\mathbf{I}_T - \hat{\mathbf{F}}\left(\hat{\mathbf{F}}'\hat{\mathbf{F}}\right)^{-1}\hat{\mathbf{F}}'\right]\mathbf{x}^*_{b,i},\\ \hat{\sigma}^2_{iT} &= T^{-1}\mathbf{x}^{*\prime}_{b,i}\mathbf{M}_{\hat{\mathbf{F}}}\mathbf{x}^*_{b,i}.\end{aligned}$$

3. *Sort* $\hat{\sigma}^2_{iT}$ *in ascending order and denote the sorted series* $\hat{\sigma}^2_{(1)T}, \hat{\sigma}^2_{(2)T}, \dots, \hat{\sigma}^2_{(N)T}$ *with* $\hat{\sigma}^2_{(i)T}$ *being the* i*th smallest value. Consider the cross-section indexes* $i_1, i_2, \dots, i_{p_{\max}-r}$ *corresponding to* $\hat{\sigma}^2_{(1)T}, \hat{\sigma}^2_{(2)T}, \dots, \hat{\sigma}^2_{(p_{\max}-r)T}$ *. Compute*

$$\hat{\eta}^2_{iN} = \frac{\hat{\mathbf{a}}'_i\hat{\mathbf{A}}'\tilde{\mathbf{\Sigma}}_v\hat{\mathbf{A}}\hat{\mathbf{a}}_i}{N},$$

for every $j \in \{i_1, i_2, \dots i_{,p_{\max}-r}\}$ *where* $\tilde{\mathbf{\Sigma}}_v$ *is the multiple testing estimator of* $\mathbf{\Sigma}_v$ *by Bailey, Pesaran, and Smith (2019), as described in Section 4.2 of the main paper. If for all* j,

$$\hat{\sigma}^2_{jT} > \frac{2\hat{\eta}^2_{jN}\log T}{N},$$

then stop the algorithm and conclude that there are $\hat{m} = r$ *pervasive units whose identities are given by the indexes of the columns in* $\mathbf{X}$ *that coincide with columns in* $\mathbf{X}^*_a$*. Otherwise, proceed to step 4.*

4. *Let* $i^* = \arg\min_i \hat{\sigma}^2_i$*. For each* $j = 1, \dots i^* - 1, i^* + 1, \dots, N_1$ *estimate the model*

$$\mathbf{M}_{\mathbf{X}^*_a}\mathbf{x}_{b,j} = \mathbf{M}_{\mathbf{X}^*_a}\mathbf{x}_{b,i^*}\gamma^*_j + \mathbf{f}^{*\prime}_t\mathbf{a}^*_j + \mathbf{v}_j,$$

where $\mathbf{f}^*_t$ *is a* $p_{max} - r - 1$ *vector of unobserved factors which we estimate as in step 2 but using* $\mathbf{M}_{\mathbf{X}^*_a}\mathbf{X}^*_{b,-i^*}$ *with* $\mathbf{X}_{b,-i^*} = (\mathbf{x}_{b,1}; \dots; \mathbf{x}_{b,i^*-1}; \mathbf{x}_{b,i^*+1}; \dots; \mathbf{x}_{b,N_1})$ *instead of* $\mathbf{M}_{\mathbf{X}^*_a}\mathbf{X}^*_b$*.*

5. *Apply individual significance tests to the* N_1-1 *estimated slope parameters* $\hat{\gamma}_1^*, \ldots, \hat{\gamma}_{i^*-1}^*, \hat{\gamma}_{i^*+1}^*, \ldots, \hat{\gamma}_{N_1}^*$ *using the critical value* $\Phi^{-1}\left[1-\frac{\pi}{2(N_1-1)}\right]$ *with* $\Phi^{-1}(\cdot)$ *denoting the inverse normal CDF, and* π *is set to* 0.01.

6. *Let* M *denote the number of rejections among these* N_1-1 *tests. If* $\log(M)/\log(N) \leq 1/2$, *stop and conclude that there are* $\hat{m} = r$ *pervasive units whose identities are given by the indices of the columns in* $\mathbf{X}$ *that coincide with the columns of* $\mathbf{X}_a^*$. *Otherwise proceed to step 7.*

7. *Update* $\mathbf{X}_a^* = (\mathbf{X}_a^*; \mathbf{x}_{b,i^*})$ *and eliminate* $\mathbf{x}_{b,i^*}$ *from* $\mathbf{X}_b^*$. *Update,* $r := r+1$ *and* $N_1 := N_1 - 1$ *and return to step 2.*

S3 Pervasive unit detection procedures proposed in the literature

S3.1 Brownlees and Mesters (BM) procedure

The model considered in Brownlees and Mesters (2019, BM in the following) has an equivalent reformulation of our pervasive unit model, formally given by

$$\underset{m\times 1}{\mathbf{x}_{ta}} = \mathbf{f}_t, \tag{66}$$

$$\underset{n\times 1}{\mathbf{x}_{tb}} = \mathbf{B}\mathbf{x}_{ta} + \mathbf{u}_t, \tag{67}$$

where the covariance matrix of $\mathbf{f}_t$ may be any positive definite matrix. Brownlees and Mesters (2019) also allow for the presence of unobserved common factors, but we will be abstracting from such factors to simplify the exposition. The model described by equations (66)–(67) is very similar to the pervasive units model expressed by equations (4)–(5) in the paper. Differences between what Brownlees and Mesters (2019) call granular shocks and innovations u_{at} to our pervasive units are purely notational. More substantial differences arise with respect to the properties of $\mathbf{B}$, where Brownlees and Mesters prove the validity of their method in the case of weak factors, assuming that restrictions on deviations from sphericity of $E\left(\mathbf{u}_t\mathbf{u}_t'\right)$ are even stronger than restrictions on the pervasiveness of the least pervasive granular unit. By contrast, our results on the consistency of σ-thresholding are derived in the case where pervasive units affect (nearly) all non-pervasive units since this scenario is far more relevant for policy questions related to systemically important entities. However, the validity of our approach relies predominantly on the consistency of PC estimates of the unobserved factors in equation (9). As argued by Onatski (2012, p.248), consistency of both factor estimates and estimated factor loadings is warranted if factors are only slightly stronger than assumed by Brownlees and Mesters (2019), i.e. if the diagonal elements of $\mathbf{B}'\mathbf{B}$ diverge as $N \to \infty$. Consequently, we can confidently conjecture that the properties of σ^2 thresholding established in Theorem 2 continue to hold for cross-section units that are close to being strongly pervasive.

Brownlees and Mesters (2019) estimate the number of pervasive units[23] and their identities from the precision matrix (i.e. the inverse covariance matrix) of the observed data $\mathbf{X}$. Formally,

[23] For notational consistency with the remainder of this article, we will use this term rather than the notion of granular shocks employed by Brownlees and Mesters (2019).

let

$$\hat{\mathbf{K}} = \left(T^{-1}\mathbf{X}'\mathbf{X} - \overline{\mathbf{x}}\overline{\mathbf{x}}'\right)^{-1},$$

where $\overline{\mathbf{x}} = (\overline{x}_1; \ldots; \overline{x}_N)'$, and $\overline{x}_i = T^{-1}\sum_{t=1}^{T} x_{it}$. Additionally, let $\hat{\mathbf{K}} = \begin{pmatrix}\hat{\mathbf{k}}_1 & \ldots & \hat{\mathbf{k}}_N\end{pmatrix}$. BM then compute $\hat{\kappa}_i = ||\hat{\mathbf{k}}_i||$, $i = 1, 2, \ldots, N$, where $||\hat{\mathbf{k}}_i|| = \sqrt{\hat{\mathbf{k}}_i'\hat{\mathbf{k}}_i}$. These N vector norms are then ordered in a *descending* manner, denoted as $\hat{\kappa}_{(1)}, \hat{\kappa}_{(2)}, \ldots, \hat{\kappa}_{(N)}$. The estimated number of pervasive units is then

$$\hat{m} = \underset{j=1,2,\ldots,N-1}{\arg\max} \frac{\hat{\kappa}_{(j)}}{\hat{\kappa}_{(j+1)}},$$

and the pervasive units are determined as columns with the norms $\hat{\kappa}_{(1)}, \hat{\kappa}_{(2)}, \ldots, \hat{\kappa}_{(\hat{m})}$. Monte Carlo simulations and empirical applications in the main paper employ a slight modification of this procedure, also used in Section 6 of Brownlees and Mesters (2019), whereby the above maximization problem is solved with respect to the first $N/2$ ratios instead of all $N-1$ ratios. Supplementary simulation results obtained without this modification are reported in Sections S7 and S8.

BM detection method is subject to two main shortcomings. First, estimation of the precision matrix requires $T > N$. Second, by construction the estimated number of pervasive units is at least one. Consequently, it is impossible to use the BM procedure to investigate *whether* there is in fact any pervasive unit in the panel data set under consideration. As an illustration consider the simple factor specification

$$x_{it} = \beta_i f_t + u_{it}, \tag{68}$$

where $f_t \sim (0, 1)$ is the common factor, β_i is the factor loading with $\sup_i |\beta_i| < K$, and u_{it} is the unit-specific component which we assume to be $IID\,(0, \sigma^2)$ over all i and t, $i = 1, 2, \ldots, N; t = 1, 2, \ldots, T$. Assuming that $\sigma^2 > 0$ ensures that there is no pervasive unit in this model. Let $\mathbf{x}_t = (x_{1t}, x_{2t}, \ldots, x_{Nt})'$, $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_N)$, and $\mathbf{u}_t = (u_{1t}, u_{2t}, \ldots, u_{Nt})$, and write (68) as

$$\mathbf{x}_t = \boldsymbol{\beta} f_t + \mathbf{u}_t, \tag{69}$$

and note that

$$Cov\,(\mathbf{x}_t) = \boldsymbol{\Sigma} = \boldsymbol{\beta}\boldsymbol{\beta}' + \sigma^2\mathbf{I}_N. \tag{70}$$

Then,

$$\begin{aligned}\mathbf{K} =& \begin{pmatrix}\mathbf{k}_1 & \mathbf{k}_2 & \cdots & \mathbf{k}_N\end{pmatrix} \\ =& \boldsymbol{\Sigma}^{-1} = \left(\sigma^2\mathbf{I}_N + \boldsymbol{\beta}\boldsymbol{\beta}'\right)^{-1} = \frac{1}{\sigma^2}\left(\mathbf{I}_N - \frac{\boldsymbol{\delta}\boldsymbol{\delta}'}{1+\boldsymbol{\delta}'\boldsymbol{\delta}}\right),\end{aligned} \tag{71}$$

where $\boldsymbol{\delta} = (\delta_1, \delta_2, \ldots, \delta_N)'$ and $\delta_i = \beta_i/\sigma$. Then, it is easily seen that

$$||\mathbf{k}_i||^2 = \frac{1}{\sigma^4}\left[\left(1 - \frac{\boldsymbol{\delta}\boldsymbol{\delta}'}{1+\boldsymbol{\delta}'\boldsymbol{\delta}}\right)^2 + \frac{\delta_i^2\sum_{j\neq i}^{N}\delta_j^2}{\left(1+\boldsymbol{\delta}'\boldsymbol{\delta}\right)^2}\right]. \tag{72}$$

Suppose that $\boldsymbol{\beta}'\boldsymbol{\beta} = \sigma^2\boldsymbol{\delta}'\boldsymbol{\delta} = \Theta\left(N^{\alpha}\right)$, with $\alpha = 1$, signifying f_t to be strong. Then, $||\mathbf{k}_i||^2 = \frac{1}{\sigma^4}\delta_i^2\frac{\boldsymbol{\delta}\boldsymbol{\delta}'}{(1+\boldsymbol{\delta}'\boldsymbol{\delta})^2}$, and hence, as $N \to \infty$,

$$\lim_{N\to\infty}\frac{||\mathbf{k}_i||}{||\mathbf{k}_s||} = \frac{|\delta_i|}{|\delta_s|} = \frac{|\beta_i|}{|\beta_s|}, \tag{73}$$

and $\|\mathbf{k}_i\|$ is maximized for the unit with the largest factor loading in absolute value.

The same result holds if we allow the variance of u_{it} to vary over i. In such a case the relevant measure is $|\beta_i|/\sigma_i^2$, where $\sigma_i^2 = V(u_{it})$, with $0 < \sigma_i^2 < K$. Thus, the column norms of the concentration matrix measure the relative importance of the common factors for the units in the panel, and is not informative about the importance of the unit for the rest of the units in the panel.

S3.2 Parker and Sul (PS) procedure

The *pervasive leader* framework of (Parker and Sul, 2016, henceforth PS) is primarily aimed at investigating whether a time series external to the dataset at hand is one of the latent factors driving the observed data. However, this framework can be represented in terms of the model (66)–(67) by simply including the potential pervasive unit(s) into the dataset (see also Parker and Sul, 2016, p.229). The pervasive leader framework also deals with *approximate pervasive leaders* which will not be considered here.

The key idea of PS is whether a known potentially pervasive unit can replace one of the factor estimates obtained from the factor model representation of the pervasive unit model. If so, then this candidate unit is identified as pervasive.

PS assume *a priori* knowledge of a fixed number r of potential pervasive units, denoted as $\mathbf{G} = (\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_r)$. Each time series in the dataset is standardized and the true number of factors in the data is determined. In order to avoid a subjective choice, we let $p_{max} = \hat{\#}(\mathbf{X})$ where $\hat{\#}(\mathbf{X})$ denotes the number of factors in $\mathbf{X}$ minimizing the the IC_{p2} criterion of Bai and Ng (2002).[24] Subsequently, the factor estimates $\hat{\mathbf{F}}$ are obtained as $\sqrt{T}$ times the eigenvectors corresponding to the p_{max} largest eigenvalues of $N^{-1}\mathbf{X}\mathbf{X}'$. Now, for each potential pervasive unit $\mathbf{g}_\ell, \ell = 1, \ldots, p_{max}$, Parker and Sul consider the p_{max} regression models

$$
\begin{aligned}
x_{it} &= \gamma_{i,1} g_{t,\ell} + \alpha_{i,2} \hat{f}_{t,2} + \ldots + \alpha_{i,p_{max}} \hat{f}_{t,p_{max}} + \eta_{it}^{(1)}, \\
x_{it} &= \alpha_{i,1} \hat{f}_{t,1} + \gamma_{i,2} g_{t,\ell} + \ldots + \alpha_{i,p_{max}} \hat{f}_{t,p_{max}} + \eta_{it}^{(2)}, \\
&\vdots \\
x_{it} &= \alpha_{i,1} \hat{f}_{t,1} + \alpha_{i,2} \hat{f}_{t,2} + \ldots + \gamma_{i,p_{max}} g_{t,\ell} + \eta_{it}^{(p_{max})},
\end{aligned}
$$

for $i = 1, 2, \ldots, N$. Let $\hat{\mathbf{H}}^{(1)} = \left(\hat{\eta}_{1t}^{(1)}, \ldots, \hat{\eta}_{Nt}^{(1)}\right), \ldots, \hat{\mathbf{H}}^{(p_{max})} = \left(\hat{\eta}_{1t}^{(\hat{m}_0)}, \ldots, \hat{\eta}_{Nt}^{(p_{max})}\right)$ denote the OLS residuals of the p_{max} regression models above. If at least one among $\hat{\#}\left(\hat{\mathbf{H}}^{(1)}\right), \ldots, \hat{\#}\left(\hat{\mathbf{H}}^{(p_{max})}\right)$ is equal to zero then $\mathbf{g}_\ell$ is considered as a pervasive unit.

PS suggest a further step if any of the units in the dataset is selected as pervasive. For each unit $\mathbf{x}_i$, the authors consider the p_{max} regression models

$$
\begin{aligned}
\hat{f}_{t,1} &= c_{1,i}^{(1)} x_{it} + c_{2,i}^{(1)} \hat{f}_{t,2} + \ldots + c_{p_{max},i}^{(1)} \hat{f}_{t,p_{max}} + \zeta_t^{(1)}, \\
\hat{f}_{t,2} &= c_{1,i}^{(2)} \hat{f}_{t,1} + c_{2,i}^{(2)} x_{it} + \ldots + c_{p_{max},i}^{(2)} \hat{f}_{t,p_{max}} + \zeta_t^{(2)}, \\
&\vdots \\
\hat{f}_{t,\hat{m}_0} &= c_{1,i}^{(p_{max})} \hat{f}_{t,1} + c_{2,i}^{(p_{max})} \hat{f}_{t,2} + \ldots + c_{p_{max},i}^{(p_{max})} x_{it} + \zeta_t^{(p_{max})}.
\end{aligned}
$$

[24] In application of the Bai-Ng selection procedure, we set the maximum number of factors to 10.

The coefficients of determination $R^2_{1,i}, \ldots, R^2_{p_{max},i}$ for these p_{max} regression equations are obtained. Having done this for every $i = 1, 2, \ldots, N$, the R^2 values for the first model above, denoted by $R^2_{1,1}, \ldots, R^2_{1,N}$, are ordered in a descending manner. The units with the coefficient of determination $R^2_{1,(1)}, \ldots, R^2_{1,(r^*)}$ are chosen as r^* potential pervasive units. This procedure is repeated for the remaining $p_{max} - 1$ models as set out above, providing in total $r = r^* p_{max}$ potential pervasive units (duplicates included). A guideline for the choice of r^* is "[...] around 10% of the size of N." (Parker and Sul, 2016, p.232).

The PS procedure is subject to two limitations. First, Parker and Sul (2016, p.230) acknowledge that treating all units in the sample as potentially pervasive may lead to a non-negligible probability of making a Type I error. However, this problem is not solved by restricting the number of potential pervasive units to 10% of the number of cross-sections. Second, the performance of the procedure depends crucially on the choice of m, the number of factors, and how well it is estimated. If m is underestimated not all true pervasive units may be chosen. If it is overestimated, non-pervasive units may falsely be identified as pervasive.

S4 Finite sample performance of alternative σ^2 thresholding methods

As discussed in Section 5 of the paper, it is possible to apply certain refinements to the σ^2 thresholding method in order to improve its finite sample properties. Our preference for the sequential-MT σ^2 thresholding is based on its finite sample performance relative to a number of other modified versions of the basic method. This section provides simulation results to support our choice.

The σ^2 thresholding variations considered are as follows:

1. σ^2 thresholding, as described by Algorithm 1 in the paper.
2. S$-\sigma^2$ thresholding, as described by Algorithm 4 given above, or Algorithm 2 in the paper.
3. Sequential-MT σ^2 thresholding with an alternative threshold. This method coincides with Algorithm 5 except for the application of the threshold specified in Appendix A.4 for the σ^2 thresholding step.

We conduct simulation experiments identical to those in Section 6 of the paper, and report the performance of σ^2 thresholding, as discussed in Section 4.2, as well as S$-\sigma^2$ thresholding, and the $SMT - \sigma^2$ thresholding with an alternative scaling, as set out above. As before, our performance measures are (a) the percent probability of correctly determining only the true pervasive units, and (b) the average number of units falsely selected as pervasive.

Tables S4.1 and S4.2 report results for the case where there is no pervasive unit. The performance of the four measures considered differs only with respect to whether they involve a multiple testing hurdle or not. Algorithms that include this extra step perform better, especially when the DGP includes an external factor. This observation suggests that the multiple testing hurdle makes a noticeable contribution to minimizing the probability of falsely discovering a pervasive unit.

Noticeable differences between all four algorithms begin to emerge when the number of pervasive units is at least equal to one. As reported in Table S4.3, the performance of σ^2

thresholding declines considerably when the total number of factors, both pervasive units and external factors, is larger than one. This problem is somewhat mitigated if one considers S$-\sigma^2$ thresholding. However, this method often fails to correctly detect the true pervasive units when $T > N$, and there are external factors affecting the observations. The multiple testing hurdle in $SMT - \sigma^2$ thresholding addresses this problem and leads to substantial performance gains, thus making it our method of choice. Finally, considering the alternative scaling of the threshold value (variant 3 above) leads to ambiguous results: improved performance is obtained when N is much larger than T , in the case where there are two pervasive units and at least one external factor. However, the opposite result is obtained if N is only twice as large as T. For this reason, we discard the alternative thresholding even though it certainly has benefits in samples where $N - T$ is sufficiently large.

Summary results for the number of units falsely detected as pervasive are reported in Table S4.4, and suggest that all the four methods generally perform well in this respect and do not severely overestimate the number of pervasive units. However, there is some evidence of false discovery when $k_0 = 2$, and N and T are relatively small.

Qualitatively similar results are obtained when we consider Monte Carlo designs with weakly pervasive units. Table S4.5 summarizes the results when $\alpha = 0.8$. As can be seen these results are comparable to those reported in S4.3 for $\alpha = 1$, the main difference being that with weakly pervasive units the probability of correctly determining the true pervasive units is lower. Additionally, all σ^2 thresholding versions suffer from performance losses if N is too large relative to T. This is to be expected since the fraction of cross section units that are unaffected by pervasive units increases in N. Finally, Table S4.6 reports the empirical frequency of false discoveries in the case of weakly pervasive units. Once again the results are similar to those obtained for $\alpha = 1$.

Table S4.1: Empirical frequency of correctly identifying the absence of a pervasive unit

σ^2 thresholding

$k_0 = 0$

N \T	60	110	210	250
50	96.1	93.5	93.6	92.6
100	99.0	98.1	96.3	96.6
200	99.8	99.6	99.5	98.9
500	100	100	100	100

$k_0 = 1$

N \T	60	110	210	250
50	60.9	47.8	41.5	40.4
100	89.8	80.8	71.4	68.9
200	99.3	98.6	97.6	97.0
500	100	100	100	99.9

$k_0 = 2$

N \T	60	110	210	250
50	35.1	24.0	17.1	15.8
100	77.3	60.4	43.9	39.7
200	98.3	95.7	90.9	89.3
500	100	100	100	99.8

S$-\sigma^2$

$k_0 = 0$

N \T	60	110	210	250
50	96.1	93.5	93.6	92.6
100	99.0	98.1	96.3	96.6
200	99.8	99.6	99.5	98.9
500	100	100	100	100

$k_0 = 1$

N \T	60	110	210	250
50	60.9	47.8	41.5	40.4
100	89.8	80.8	71.4	68.9
200	99.3	98.6	97.6	97.0
500	100	100	100	99.9

$k_0 = 2$

N \T	60	110	210	250
50	35.1	24.0	17.1	15.8
100	77.3	60.4	43.9	39.7
200	98.3	95.7	90.9	89.3
500	100	100	100	99.8

$SMT - \sigma^2$

$k_0 = 0$

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

$k_0 = 1$

N \T	60	110	210	250
50	88.4	86.4	82.7	80.3
100	94.1	92.3	90.7	88.9
200	99.8	99.2	99.4	99.2
500	100	100	100	100

$k_0 = 2$

N \T	60	110	210	250
50	61.6	55.9	47.7	44.3
100	84.0	74.5	64.2	60.9
200	98.6	97.7	94.2	94.1
500	100	100	100	99.9

$SMT - \sigma^2$, alternative scaling

$k_0 = 0$

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

$k_0 = 1$

N \T	60	110	210	250
50	88.4	86.4	82.7	80.3
100	94.1	92.3	90.7	88.9
200	99.8	99.2	99.4	99.2
500	100	100	100	100

$k_0 = 2$

N \T	60	110	210	250
50	61.6	55.9	47.7	44.3
100	84.0	74.5	64.2	60.9
200	98.6	97.7	94.2	94.1
500	100	100	100	99.9

Notes: σ^2 thresholding is implemented using Algorithm 1 in the main article, with $p_{max} = m_0 + k_0 + 1$, where m_0 is the true number of pervasive units (if any) and k_0 is the number of external factors. S$-\sigma^2$ and $SMT - \sigma^2$ refer to Sequential σ^2 thresholding and Sequential-MT σ^2 thresholding, as implemented using Algorithms 2 and 3 in the main article, respectively. Threshold in the σ^2 thresholding step of all three algorithms is given by $\hat{\sigma}^2_{iT} \leq 2\hat{\eta}^2_{iN} N^{-1} \log(T)$. The threshold chosen for $N > T$ in the alternative version of $SMT - \sigma^2$ is given by $\hat{\sigma}^2_{iT} \leq 2\hat{\sigma}^2_{ui} T^{-1} \log(T)$. See Appendix A.4 for further details.

Table S4.2: Average number of non-pervasive units falsely selected as pervasive ($m_0 = 0$)

σ^2 thresholding

$k_0 = 0$

N \T	60	110	210	250
50	0	0.1	0.1	0.1
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	0.4	0.6	0.7	0.7
100	0.1	0.2	0.3	0.3
200	0	0	0	0
500	0	0	0	0

$k_0 = 2$

N \T	60	110	210	250
50	0.9	1.1	1.4	1.4
100	0.2	0.5	0.7	0.8
200	0	0	0.1	0.1
500	0	0	0	0

S$-\sigma^2$

$k_0 = 0$

N \T	60	110	210	250
50	0	0.1	0.1	0.1
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	0.5	0.7	0.8	0.8
100	0.1	0.2	0.3	0.4
200	0	0	0	0
500	0	0	0	0

$k_0 = 2$

N \T	60	110	210	250
50	0.9	1.2	1.5	1.5
100	0.3	0.5	0.8	0.9
200	0	0	0.1	0.1
500	0	0	0	0

$SMT - \sigma^2$

$k_0 = 0$

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	0.1	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	0.4	0.5	0.6	0.7
100	0.2	0.3	0.4	0.4
200	0	0	0.1	0.1
500	0	0	0	0

$SMT - \sigma^2$, alternative scaling

$k_0 = 0$

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	0.1	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	0.4	0.5	0.6	0.7
100	0.2	0.3	0.4	0.4
200	0	0	0.1	0.1
500	0	0	0	0

Notes: See the notes to Table S4.1.

Table S4.3: Empirical frequency of correctly identifying only the true strongly pervasive units ($m_0 > 0, \alpha = 1$)

Part A: $m_0 = 1$

		σ^2 thresholding				S$-\sigma^2$				$SMT-\sigma^2$				$SMT-\sigma^2$, alternative scaling			
	N \ T	60	110	210	250	60	110	210	250	60	110	210	250	60	110	210	250
$k_0 = 0$	50	76.7	67.4	64.1	63.0	95.1	94.5	92.9	92.6	97.7	99.9	100	100	97.7	99.9	100	100
	100	93.0	89.5	85.0	81.9	99.2	98.5	96.9	96.3	100	100	100	100	100	100	100	100
	200	99.4	99.3	98.2	97.2	99.8	99.7	99.5	98.7	100	100	100	100	100	100	100	100
	500	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
$k_0 = 1$	50	22.3	25.8	22.3	23.4	37.5	42.4	39.1	40.5	58.9	80.6	82.9	82.3	58.9	80.6	82.9	82.3
	100	54.2	59.2	49.4	49.9	64.7	74.9	68.6	69.3	68.1	88.4	93.3	93.0	71.7	88.4	93.3	93.0
	200	77.4	95.2	93.2	91.8	79.0	97.0	97.1	95.0	79.1	97.8	99.6	99.5	78.3	98.2	99.6	99.5
	500	82.1	99.9	100	99.8	82.1	99.9	100	99.8	82.1	99.9	100	100	81.1	99.9	100	100
$k_0 = 2$	50	9.8	8.5	7.8	7.6	22.4	20.4	17.0	16.9	52.5	61.7	61.1	55.5	52.5	61.7	61.1	55.5
	100	40.0	36.1	22.4	21.0	57.8	56.6	42.8	41.2	65.3	75.9	74.7	74.2	68.8	75.9	74.7	74.2
	200	68.0	86.9	80.1	78.1	72.0	93.3	91.1	88.8	72.7	95.6	97.1	96.0	65.6	97.3	97.1	96.0
	500	77.0	99.4	99.9	99.7	77.1	99.4	100	99.8	77.1	99.4	100	100	64.2	99.4	100	100

Part B: $m_0 = 2$

		σ^2 thresholding				S$-\sigma^2$				$SMT-\sigma^2$				$SMT-\sigma^2$, alternative scaling			
	N \ T	60	110	210	250	60	110	210	250	60	110	210	250	60	110	210	250
$k_0 = 0$	50	17.3	28.8	35.6	35.2	60.0	82.7	88.9	90.4	55.8	85.0	96.0	97.7	55.8	85.0	96.0	97.7
	100	37.3	56.1	63.4	66.6	58.8	85.8	95.0	95.2	58.9	87.3	98.2	98.6	47.2	87.3	98.2	98.6
	200	50.8	82.1	92.6	91.7	59.0	88.6	97.4	97.5	59.0	88.8	98.4	98.9	58.5	79.8	98.4	98.9
	500	54.3	93.3	100	99.8	60.9	94.8	100	99.9	60.9	94.8	100	100	60.9	94.8	99.1	97.6
$k_0 = 1$	50	5.4	7.2	7.4	7.8	27.6	35.9	34.7	35.5	36.2	67.3	79.1	79.5	36.2	67.3	79.1	79.5
	100	20.0	32.4	29.3	27.0	38.6	65.5	68.6	67.8	41.7	78.5	91.5	92.4	29.0	78.5	91.5	92.4
	200	35.2	71.1	81.1	79.4	43.1	86.6	95.7	95.9	43.5	87.6	98.3	99.3	54.1	43.5	98.3	99.3
	500	41.7	90.0	99.6	99.8	45.9	96.2	100	100	46.0	96.2	100	100	74.2	90.7	91.1	89.2
$k_0 = 2$	50	1.9	2.1	2.3	2.5	19.0	20.5	16.6	16.9	38.9	61.3	63.0	60.5	38.9	61.3	63.0	60.5
	100	17.4	20.3	11.0	9.5	41.8	51.4	42.9	40.1	48.4	73.3	79.6	79.6	47.5	73.3	79.6	79.6
	200	36.4	67.8	63.5	60.5	46.7	84.0	89.0	88.8	47.5	86.9	96.8	97.1	65.3	49.0	96.8	97.1
	500	38.4	90.9	99.4	99.5	41.0	94.6	99.8	99.8	41.0	94.6	99.9	100	73.8	88.8	90.6	88.9

Notes: See the notes to Table S4.1 where references to the descriptions of all four methods are given.

Table S4.4: Average number of non-pervasive units falsely selected as pervasive units ($m_0 > 0$ and $\alpha = 1$)

Part A: $m_0 = 1$

	σ^2 thresholding				$S{-}\sigma^2$				$SMT - \sigma^2$				$SMT - \sigma^2$, alternative scaling			
	$k_0 = 0$				$k_0 = 0$				$k_0 = 0$				$k_0 = 0$			
N \ T	60	110	210	250	60	110	210	250	60	110	210	250	60	110	210	250
50	0.2	0.3	0.4	0.4	0	0.1	0.1	0.1	0	0	0	0	0	0	0	0
100	0.1	0.1	0.2	0.2	0	0	0	0	0	0	0	0	0	0	0	0
200	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
500	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	$k_0 = 1$				$k_0 = 1$				$k_0 = 1$				$k_0 = 1$			
N \ T	60	110	210	250	60	110	210	250	60	110	210	250	60	110	210	250
50	0.7	0.9	1.0	1.0	0.5	0.6	0.7	0.7	0.2	0.1	0.2	0.2	0.2	0.1	0.2	0.2
100	0.2	0.4	0.6	0.6	0.1	0.2	0.4	0.3	0.1	0.1	0.1	0.1	0	0.1	0.1	0.1
200	0	0	0.1	0.1	0	0	0	0.1	0	0	0	0	0	0	0	0
500	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	$k_0 = 2$				$k_0 = 2$				$k_0 = 2$				$k_0 = 2$			
N \ T	60	110	210	250	60	110	210	250	60	110	210	250	60	110	210	250
50	1.3	1.6	1.7	1.7	1.0	1.2	1.4	1.4	0.4	0.3	0.4	0.5	0.4	0.3	0.4	0.5
100	0.5	0.7	1.1	1.2	0.2	0.5	0.8	0.9	0.1	0.2	0.3	0.3	0.1	0.2	0.3	0.3
200	0.1	0.1	0.2	0.2	0	0	0.1	0.1	0	0	0	0	0.1	0	0	0
500	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0

Part B: $m_0 = 2$

	σ^2 thresholding				$S{-}\sigma^2$				$SMT - \sigma^2$				$SMT - \sigma^2$, alternative scaling			
	$k_0 = 0$				$k_0 = 0$				$k_0 = 0$				$k_0 = 0$			
N \ T	60	110	210	250	60	110	210	250	60	110	210	250	60	110	210	250
50	0.4	0.4	0.5	0.5	0.1	0.1	0.1	0.1	0	0	0	0	0	0	0	0
100	0.1	0.2	0.3	0.3	0	0	0	0	0	0	0	0	0	0	0	0
200	0	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0	0
500	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	$k_0 = 1$				$k_0 = 1$				$k_0 = 1$				$k_0 = 1$			
N \ T	60	110	210	250	60	110	210	250	60	110	210	250	60	110	210	250
50	1.0	1.1	1.2	1.2	0.5	0.6	0.7	0.7	0.2	0.1	0.1	0.1	0.2	0.1	0.1	0.1
100	0.3	0.6	0.8	0.9	0.1	0.2	0.3	0.3	0	0	0	0	0	0	0	0
200	0.1	0.1	0.1	0.2	0	0	0	0	0	0	0	0	0	0	0	0
500	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	$k_0 = 2$				$k_0 = 2$				$k_0 = 2$				$k_0 = 2$			
N \ T	60	110	210	250	60	110	210	250	60	110	210	250	60	110	210	250
50	1.8	1.9	2.0	2.0	1.0	1.2	1.4	1.4	0.3	0.3	0.3	0.4	0.3	0.3	0.3	0.4
100	0.7	1.0	1.5	1.6	0.2	0.5	0.8	0.8	0.1	0.1	0.2	0.2	0.1	0.1	0.2	0.2
200	0.1	0.2	0.4	0.4	0	0.1	0.1	0.1	0	0	0	0	0	0	0	0
500	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0

Notes: See the notes to Table S4.1.

Table S4.5: Empirical frequency of correctly identifying only the true weakly pervasive (influential) units ($m_0 > 0, \alpha = 0.8$)

Part A: $m_0 = 1$

$k_0 = 0$

	σ^2 thresholding				S$-\sigma^2$				$SMT-\sigma^2$				$SMT-\sigma^2$, alternative scaling			
N \T	60	110	210	250	60	110	210	250	60	110	210	250	60	110	210	250
50	60.8	67.5	66.9	65.4	61.5	84.5	90.4	91.1	51.2	80.6	95.4	97.5	51.2	80.6	95.4	97.5
100	85.7	85.3	81.9	77.9	89.7	98.2	96.9	96.3	87.2	98.9	100	100	87.2	98.9	100	100
200	95.7	94.6	90.0	89.0	97.7	99.7	99.5	98.7	97.5	100	100	100	97.5	100	100	100
500	97.7	99.7	99.4	99.4	97.7	100	100	100	97.7	100	100	100	97.7	100	100	100

$k_0 = 1$

	σ^2 thresholding				S$-\sigma^2$				$SMT-\sigma^2$				$SMT-\sigma^2$, alternative scaling			
N \T	60	110	210	250	60	110	210	250	60	110	210	250	60	110	210	250
50	23.0	26.8	24.2	25.2	26.3	36.8	37.8	39.7	37.4	65.2	79.1	78.9	37.4	65.2	79.1	78.9
100	58.1	57.0	45.5	44.0	62.9	77.5	69.0	69.8	65.1	90.5	93.7	93.5	67.8	90.5	93.7	93.5
200	81.2	87.4	80.9	79.7	84.5	98.6	97.1	95.0	84.6	99.4	99.6	99.5	83.9	99.8	99.6	99.5
500	82.6	99.4	99.3	98.1	82.7	99.9	100	99.8	82.7	99.9	100	100	81.8	99.9	100	100

$k_0 = 2$

	σ^2 thresholding				S$-\sigma^2$				$SMT-\sigma^2$				$SMT-\sigma^2$, alternative scaling			
N \T	60	110	210	250	60	110	210	250	60	110	210	250	60	110	210	250
50	10.4	9.5	8.7	8.7	18.5	18.7	16.7	16.5	37.7	53.3	58.2	54.5	37.7	53.3	58.2	54.5
100	44.6	37.3	20.9	19.3	58.1	60.3	43.2	41.4	64.2	79.7	75.3	74.4	67.4	79.7	75.3	74.4
200	75.4	79.1	65.4	61.9	82.0	95.9	91.1	88.8	82.7	98.2	97.1	96.0	75.7	99.8	97.1	96.0
500	80.9	98.9	98.4	97.4	80.9	100	100	99.8	80.9	100	100	100	68.7	100	100	100

Part B: $m_0 = 2$

$k_0 = 0$

	σ^2 thresholding				S$-\sigma^2$				$SMT-\sigma^2$				$SMT-\sigma^2$, alternative scaling			
N \T	60	110	210	250	60	110	210	250	60	110	210	250	60	110	210	250
50	4.4	16.1	25.2	22.8	18.2	47.8	66.2	67.1	6.6	31.6	63.7	67.3	6.6	31.6	63.7	67.3
100	4.6	22.0	46.7	51.1	17.7	62.7	87.4	90.3	13.7	57.4	89.2	92.6	2.2	57.4	89.2	92.6
200	2.8	20.4	59.1	63.9	7.8	50	87.5	90.7	7.7	48.2	88.1	92.0	4.1	3.8	88.1	92.0
500	0.4	16.2	58.8	67.4	0.9	23.0	71.0	79.2	0.9	23.0	71.0	79.3	0.9	17.7	9.0	2.9

$k_0 = 1$

	σ^2 thresholding				S$-\sigma^2$				$SMT-\sigma^2$				$SMT-\sigma^2$, alternative scaling			
N \T	60	110	210	250	60	110	210	250	60	110	210	250	60	110	210	250
50	1.8	4.6	7.3	7.3	11.0	23.1	28.8	29.4	8.9	32.0	60.6	63.7	8.9	32.0	60.6	63.7
100	3.5	15.3	26.8	26.3	18.2	55.0	67.4	67.7	16.5	61.3	88.8	91.5	3.5	61.3	88.8	91.5
200	2.9	20.8	55.0	57.2	11.4	61.8	92.2	93.7	11.4	61.8	94.7	97.1	6.1	4.4	94.7	97.1
500	0.6	17.8	62.7	70.7	1.8	32.3	84.5	91.7	1.8	32.3	84.6	91.7	2.1	22.1	15.8	9.9

$k_0 = 2$

	σ^2 thresholding				S$-\sigma^2$				$SMT-\sigma^2$				$SMT-\sigma^2$, alternative scaling			
N \T	60	110	210	250	60	110	210	250	60	110	210	250	60	110	210	250
50	1.0	1.8	1.9	2.4	9.9	14.7	14.2	14.7	13.3	33.5	50.1	50.4	13.3	33.5	50.1	50.4
100	5.4	11.8	12.3	11.7	24.8	48.9	43.8	40.8	26.6	65.3	79.3	80.2	11.7	65.3	79.3	80.2
200	4.0	29.3	45.0	47.0	17.6	73.7	88.3	88.3	17.6	75.7	96.1	96.6	13.8	12.9	96.1	96.6
500	0.6	17.9	66.0	73.7	2.4	36.5	89.1	93.6	2.4	36.5	89.2	93.8	3.0	25.5	24.3	18.7

Notes: See notes to Table S4.3.

Table S4.6: Average number of non-pervasive units falsely selected as pervasive units ($m_0 > 0$ and $\alpha = 0.8$)

Part A: $m_0 = 1$

$k_0 = 0$

	σ^2 thresholding				S$-\sigma^2$				$SMT-\sigma^2$				$SMT-\sigma^2$, alternative scaling			
N \ T	60	110	210	250	60	110	210	250	60	110	210	250	60	110	210	250
50	0.2	0.3	0.3	0.3	0.3	0.1	0.1	0.1	0	0	0	0	0	0	0	0
100	0.1	0.1	0.2	0.2	0.1	0	0	0	0	0	0	0	0	0	0	0
200	0	0.1	0.1	0.1	0	0	0	0	0	0	0	0	0	0	0	0
500	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

$k_0 = 1$

	σ^2 thresholding				S$-\sigma^2$				$SMT-\sigma^2$				$SMT-\sigma^2$, alternative scaling			
N \ T	60	110	210	250	60	110	210	250	60	110	210	250	60	110	210	250
50	0.8	0.9	1.0	1.0	0.8	0.8	0.8	0.8	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
100	0.3	0.5	0.7	0.7	0.3	0.2	0.4	0.3	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
200	0.1	0.1	0.2	0.2	0	0	0	0.1	0	0	0	0	0	0	0	0
500	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

$k_0 = 2$

	σ^2 thresholding				S$-\sigma^2$				$SMT-\sigma^2$				$SMT-\sigma^2$, alternative scaling			
N \ T	60	110	210	250	60	110	210	250	60	110	210	250	60	110	210	250
50	1.4	1.6	1.7	1.7	1.3	1.4	1.5	1.5	0.5	0.4	0.5	0.5	0.5	0.4	0.5	0.5
100	0.5	0.9	1.2	1.3	0.4	0.5	0.8	0.9	0.2	0.2	0.3	0.3	0.2	0.2	0.3	0.3
200	0.1	0.2	0.4	0.4	0.1	0	0.1	0.1	0	0	0	0	0.1	0	0	0
500	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0

Part B: $m_0 = 2$

$k_0 = 0$

	σ^2 thresholding				S$-\sigma^2$				$SMT-\sigma^2$				$SMT-\sigma^2$, alternative scaling			
N \ T	60	110	210	250	60	110	210	250	60	110	210	250	60	110	210	250
50	0.5	0.5	0.4	0.4	0.4	0.3	0.2	0.2	0.1	0.1	0	0	0.1	0.1	0	0
100	0.2	0.2	0.3	0.3	0.2	0.1	0.1	0.1	0	0	0	0	0	0	0	0
200	0	0.1	0.1	0.1	0	0	0	0	0	0	0	0	0	0	0	0
500	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

$k_0 = 1$

	σ^2 thresholding				S$-\sigma^2$				$SMT-\sigma^2$				$SMT-\sigma^2$, alternative scaling			
N \ T	60	110	210	250	60	110	210	250	60	110	210	250	60	110	210	250
50	1.2	1.2	1.2	1.2	1.0	0.9	0.9	0.9	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
100	0.5	0.6	0.7	0.8	0.4	0.3	0.4	0.4	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1
200	0.1	0.1	0.2	0.3	0.1	0.1	0	0	0	0	0	0	0	0	0	0
500	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

$k_0 = 2$

	σ^2 thresholding				S$-\sigma^2$				$SMT-\sigma^2$				$SMT-\sigma^2$, alternative scaling			
N \ T	60	110	210	250	60	110	210	250	60	110	210	250	60	110	210	250
50	1.5	1.5	1.5	1.4	1.9	1.9	1.9	1.9	0.5	0.4	0.4	0.5	0.5	0.4	0.4	0.5
100	0.8	1.1	1.4	1.5	0.6	0.6	0.8	0.9	0.3	0.2	0.2	0.2	0.2	0.2	0.2	0.2
200	0.2	0.3	0.5	0.6	0.1	0.1	0.1	0.1	0.1	0	0	0	0.1	0	0	0
500	0	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0	0

Notes: See notes to Table S4.3.

S5 Sensitivity of σ^2 thresholding to different normalizing constants

Section 4.2 in the main paper established the expression $2C\eta_{in}^2 \log(T)/N$ as a threshold for cross-section specific residual variances from a static factor model in order to detect pervasive units in the data. σ^2 thresholding and its refinements were based on this threshold with $C = 1$. The purpose of this section is to illustrate the sensitivity of σ^2 thresholding as well as $SMT - \sigma^2$ thresholding to the value of C. In particular we consider the values $C = 1, 1.25, 1.5$ and iterate the set of simulation experiments leading to Tables 1-4 in Section 6 in the main paper. The results of these simulations are reported in Tables S5.1–S5.3. Generally, increasing the constant C will render σ^2 thresholding less conservative. That is, the number of detected pervasive units will generally be higher. As one can observe from Tables S5.1 and S5.3, a direct consequence is a slight increase in the average number of falsely detected units. The effect on the ability of σ^2 thresholding to detect the correct pervasive units (and no others) is ambiguous: While the performance of σ^2 thresholding, judged by this measure, decreases in C for $m_0 = 0$, performance improvements can be observed for $m_0 > 0$ in large-N datasets. Corresponding results for $SMT - \sigma^2$ thresholding can be observed in Tables S5.4–S5.6. While the general patterns observed in these three tables are in line with those seen for σ^2 thresholding, the magnitude of performance changes due to changed in C is smaller. This is a reasonable result, given that $SMT - \sigma^2$ thresholding complements the thresholding procedure on residual variances with a second criterion. This set-up allows eliminating some of the units falsely detected by σ^2 thresholding and can even lead to noticeable performance improvements in the case $m_0 = 2$. However, overall the performance of $SMT - \sigma^2$ thresholding is not drastically affected by the choice of C.

Table S5.1: $m_0 = 0$: Performance of σ^2 thresholding

Part A: Empirical frequency of correctly identifying the absence of pervasive units

$C = 1$, $k_0 = 0$

N \ T	60	110	210	250
50	96.1	93.5	93.6	92.6
100	99.0	98.1	96.3	96.6
200	99.8	99.6	99.5	98.9
500	100	100	100	100

$C = 1.25$, $k_0 = 0$

N \ T	60	110	210	250
50	93.9	90.8	90.3	89.2
100	98.4	96.8	94.2	94.7
200	99.6	99.3	98.6	98.2
500	100	100	99.8	99.9

$C = 1.5$, $k_0 = 0$

N \ T	60	110	210	250
50	91.4	87.9	87.3	85.6
100	96.9	95.4	91.9	92.9
200	99.3	98.8	97.9	97.4
500	100	100	99.8	99.9

$C = 1$, $k_0 = 1$

N \ T	60	110	210	250
50	60.9	47.8	41.5	40.4
100	89.8	80.8	71.4	68.9
200	99.3	98.6	97.6	97.0
500	100	100	100	99.9

$C = 1.25$, $k_0 = 1$

N \ T	60	110	210	250
50	45.0	34.4	29.3	28.9
100	77.4	64.3	55.3	52.3
200	98.2	96.1	92.3	91.8
500	100	100	99.8	99.7

$C = 1.5$, $k_0 = 1$

N \ T	60	110	210	250
50	34.9	26.6	22.1	20.7
100	65.2	53.0	44.2	42.1
200	95.7	89.4	82.6	80.2
500	99.9	99.8	99.5	99.3

$C = 1$, $k_0 = 2$

N \ T	60	110	210	250
50	35.1	24.0	17.1	15.8
100	77.3	60.4	43.9	39.7
200	98.3	95.7	90.9	89.3
500	100	100	100	99.8

$C = 1.25$, $k_0 = 2$

N \ T	60	110	210	250
50	18.0	12.1	8.4	7.0
100	58.6	37.9	23.5	21.3
200	92.5	85.7	76.5	73.3
500	99.9	99.9	99.9	99.6

$C = 1.5$, $k_0 = 2$

N \ T	60	110	210	250
50	10.4	7.0	4.1	4.1
100	40.4	23.3	13.3	12.2
200	84.9	71.4	57.4	52.3
500	99.6	99.7	99.0	99.0

Part B: Average number of non-pervasive units falsely selected as pervasive

$C = 1$, $k_0 = 0$

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

$C = 1.25$, $k_0 = 0$

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

$C = 1.5$, $k_0 = 0$

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

$C = 1$, $k_0 = 1$

N \ T	60	110	210	250
50	0.4	0.6	0.7	0.7
100	0.1	0.2	0.3	0.3
200	0	0	0	0
500	0	0	0	0

$C = 1.25$, $k_0 = 1$

N \ T	60	110	210	250
50	0.7	0.8	0.9	0.9
100	0.2	0.4	0.5	0.5
200	0	0	0.1	0.1
500	0	0	0	0

$C = 1.5$, $k_0 = 1$

N \ T	60	110	210	250
50	0.8	1.0	1.0	1.1
100	0.4	0.6	0.7	0.7
200	0	0.1	0.2	0.2
500	0	0	0	0

$C = 1$, $k_0 = 2$

N \ T	60	110	210	250
50	0.9	1.1	1.4	1.4
100	0.2	0.5	0.7	0.8
200	0	0	0.1	0.1
500	0	0	0	0

$C = 1.25$, $k_0 = 2$

N \ T	60	110	210	250
50	1.3	1.5	1.7	1.7
100	0.5	0.8	1.2	1.2
200	0.1	0.2	0.3	0.3
500	0	0	0	0

$C = 1.5$, $k_0 = 2$

N \ T	60	110	210	250
50	1.6	1.8	2.0	2.0
100	0.8	1.2	1.5	1.5
200	0.2	0.3	0.5	0.6
500	0	0	0	0

Notes: σ^2 thresholding is implemented using Algorithm 1 in the main article, with $p_{max} = m_0 + k_0 + 1$, where m_0 is the true number of pervasive units (if any) and k_0 is the number of external factors. The threshold is given by $\hat{\sigma}^2_{iT} \leq 2C\hat{\eta}^2_{iN} N^{-1} \log(T)$, where different values of C are specified in this table.

Table S5.2: σ^2 thresholding: Empirical frequency of correctly identifying only the true pervasive units for $m_0 > 0$

Part A: $m_0 = 1$

$C = 1$

$k_0 = 0$

N \ T	60	110	210	250
50	76.5	67.4	64.1	63.0
100	93.0	89.5	85.0	81.9
200	99.4	99.3	98.2	97.2
500	100	100	100	100

$k_0 = 1$

N \ T	60	110	210	250
50	22.3	25.8	22.3	23.4
100	54.2	59.2	49.4	49.9
200	77.4	95.2	93.2	91.8
500	82.1	99.9	100	99.8

$k_0 = 2$

N \ T	60	110	210	250
50	9.8	8.5	7.8	7.6
100	40.0	36.1	22.4	21.0
200	68.0	86.9	80.1	78.1
500	77.0	99.4	99.9	99.7

$C = 1.25$

$k_0 = 0$

N \ T	60	110	210	250
50	65.2	56.2	53.8	53.8
100	86.3	79.4	74.7	71.3
200	98.5	97.1	95.3	94.7
500	100	100	99.9	99.8

$k_0 = 1$

N \ T	60	110	210	250
50	17.2	17.3	14.6	14.7
100	47.2	42.5	31.8	31.9
200	79.1	87.0	81.4	79.6
500	88.2	100	99.8	99.3

$k_0 = 2$

N \ T	60	110	210	250
50	4.6	4.0	2.9	3.2
100	27.1	18.0	9.7	8.5
200	66.7	70.9	55.9	52.2
500	83.5	99.5	99.4	99.3

$C = 1.5$

$k_0 = 0$

N \ T	60	110	210	250
50	56.4	48.9	47.0	45.4
100	79.3	71.6	65.7	63.1
200	96.8	93.8	89.1	88.7
500	99.9	100	99.7	99.7

$k_0 = 1$

N \ T	60	110	210	250
50	13.6	12.7	9.4	9.0
100	38.8	30.4	21.2	20.9
200	75.3	74.0	64.2	63.0
500	91.0	99.7	99.5	98.9

$k_0 = 2$

N \ T	60	110	210	250
50	2.3	1.5	1.0	0.9
100	17.0	7.6	4.1	4.0
200	57.1	50.5	32.0	28.6
500	87.7	98.9	98.0	98.0

Part B: $m_0 = 2$

$C = 1$

$k_0 = 0$

N \ T	60	110	210	250
50	17.3	28.8	35.6	35.2
100	37.3	56.1	63.4	66.6
200	50.8	82.1	92.6	91.7
500	54.3	93.3	100	99.8

$k_0 = 1$

N \ T	60	110	210	250
50	5.4	7.2	7.4	7.8
100	20.0	32.4	29.3	27.0
200	35.2	71.1	81.1	79.4
500	41.7	90.0	99.6	99.8

$k_0 = 2$

N \ T	60	110	210	250
50	1.9	2.1	2.3	2.5
100	17.4	20.3	11.0	9.5
200	36.4	67.8	63.5	60.5
500	38.4	90.9	99.4	99.5

$C = 1.25$

$k_0 = 0$

N \ T	60	110	210	250
50	19.8	30.8	32.5	31.6
100	39.9	53.7	54.0	55.1
200	57.1	83.2	87.0	85.8
500	62.7	95.2	99.8	99.7

$k_0 = 1$

N \ T	60	110	210	250
50	3.6	5.2	5.0	5.1
100	18.9	22.1	15.6	14.9
200	39.5	65.0	63.4	60.8
500	51.3	92.0	99.5	99.4

$k_0 = 2$

N \ T	60	110	210	250
50	0.8	0.9	0.8	1.2
100	12.7	8.9	3.5	3.4
200	37.4	48.6	33.8	31.9
500	49.5	93.1	98.7	98.6

$C = 1.5$

$k_0 = 0$

N \ T	60	110	210	250
50	20.1	27.2	29.6	27.6
100	38.4	49.4	46.6	44.2
200	59.5	79.4	78.4	76.6
500	68.6	96.5	99.5	99.5

$k_0 = 1$

N \ T	60	110	210	250
50	2.6	3.3	3.1	3.8
100	15.5	15.1	9.1	9.1
200	38.0	51.8	42.9	40.2
500	57.6	93.8	98.1	98.2

$k_0 = 2$

N \ T	60	110	210	250
50	0.2	0.4	0.2	0.6
100	6.5	4.2	1.0	1.3
200	31.9	29.5	15.4	14.0
500	58.2	92.7	94.7	95.4

Notes: See the notes to Table S5.1

Table S5.3: σ^2 thresholding: Average number of non-pervasive units falsely selected as pervasive for $m_0 > 0$

Part A: $m_0 = 1$

$C = 1$

$k_0 = 0$

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

$k_0 = 1$

N \ T	60	110	210	250
50	0.7	0.9	1.0	1.0
100	0.2	0.4	0.6	0.6
200	0	0	0.1	0.1
500	0	0	0	0

$k_0 = 2$

N \ T	60	110	210	250
50	1.3	1.6	1.7	1.7
100	0.5	0.7	1.1	1.2
200	0.1	0.1	0.2	0.2
500	0	0	0	0

$C = 1.25$

$k_0 = 0$

N \ T	60	110	210	250
50	0.3	0.4	0.5	0.5
100	0.1	0.2	0.3	0.3
200	0	0	0	0.1
500	0	0	0	0

$k_0 = 1$

N \ T	60	110	210	250
50	1.0	1.1	1.2	1.2
100	0.4	0.7	0.9	0.9
200	0.1	0.1	0.2	0.2
500	0	0	0	0

$k_0 = 2$

N \ T	60	110	210	250
50	1.8	2.0	2.1	2.1
100	0.8	1.2	1.6	1.7
200	0.2	0.3	0.5	0.6
500	0	0	0	0

$C = 1.5$

$k_0 = 0$

N \ T	60	110	210	250
50	0.4	0.5	0.5	0.5
100	0.2	0.3	0.3	0.4
200	0	0.1	0.1	0.1
500	0	0	0	0

$k_0 = 1$

N \ T	60	110	210	250
50	1.2	1.3	1.4	1.4
100	0.6	0.9	1.1	1.1
200	0.1	0.3	0.4	0.4
500	0	0	0	0

$k_0 = 2$

N \ T	60	110	210	250
50	2.1	2.2	2.3	2.3
100	1.2	1.6	1.9	2.0
200	0.4	0.6	1.0	1.0
500	0	0	0	0

Part B: $m_0 = 2$

$C = 1$

$k_0 = 0$

N \ T	60	110	210	250
50	0.4	0.4	0.5	0.5
100	0.1	0.2	0.3	0.3
200	0	0	0	0.1
500	0	0	0	0

$k_0 = 1$

N \ T	60	110	210	250
50	1.0	1.1	1.2	1.2
100	0.3	0.6	0.8	0.9
200	0.1	0.1	0.1	0.2
500	0	0	0	0

$k_0 = 2$

N \ T	60	110	210	250
50	1.8	1.9	2.0	2.0
100	0.7	1.0	1.5	1.6
200	0.1	0.2	0.4	0.4
500	0	0	0	0

$C = 1.25$

$k_0 = 0$

N \ T	60	110	210	250
50	0.5	0.6	0.6	0.6
100	0.2	0.3	0.4	0.4
200	0	0.1	0.1	0.1
500	0	0	0	0

$k_0 = 1$

N \ T	60	110	210	250
50	1.3	1.4	1.5	1.4
100	0.6	0.9	1.1	1.2
200	0.2	0.2	0.4	0.4
500	0	0	0	0

$k_0 = 2$

N \ T	60	110	210	250
50	2.2	2.3	2.4	2.3
100	1.2	1.6	2.0	2.1
200	0.3	0.6	0.9	0.9
500	0	0	0	0

$C = 1.5$

$k_0 = 0$

N \ T	60	110	210	250
50	0.7	0.7	0.7	0.7
100	0.3	0.4	0.5	0.6
200	0.1	0.1	0.2	0.2
500	0	0	0	0

$k_0 = 1$

N \ T	60	110	210	250
50	1.5	1.6	1.6	1.6
100	0.9	1.1	1.3	1.4
200	0.3	0.5	0.7	0.7
500	0	0	0	0

$k_0 = 2$

N \ T	60	110	210	250
50	2.5	2.5	2.6	2.6
100	1.6	2.0	2.3	2.3
200	0.6	1.0	1.4	1.4
500	0	0	0.1	0

Notes: See the notes to Table S5.1

Table S5.4: $m_0 = 0$: Performance of $SMT - \sigma^2$ thresholding

Part A: Empirical frequency of correctly identifying the absence of pervasive units

$k_0 = 0$

	$C = 1$				$C = 1.25$				$C = 1.5$			
N \ T	60	110	210	250	60	110	210	250	60	110	210	250
50	100	100	100	100	100	100	100	100	100	100	100	100
100	100	100	100	100	100	100	100	100	100	100	100	100
200	100	100	100	100	100	100	100	100	100	100	100	100
500	100	100	100	100	100	100	100	100	100	100	100	100

$k_0 = 1$

	$C = 1$				$C = 1.25$				$C = 1.5$			
N \ T	60	110	210	250	60	110	210	250	60	110	210	250
50	88.4	86.4	82.7	80.3	86.5	85.4	81.9	79.2	85.8	85.2	81.7	78.8
100	94.1	92.3	90.7	88.9	88.3	88.8	88.6	86.6	84.5	86.8	87.8	85.8
200	99.8	99.2	99.4	99.2	99.0	97.8	97.6	98.0	98.0	94.9	94.8	95.0
500	100	100	100	100	100	100	100	99.9	100	100	99.7	99.9

$k_0 = 2$

	$C = 1$				$C = 1.25$				$C = 1.5$			
N \ T	60	110	210	250	60	110	210	250	60	110	210	250
50	61.6	55.9	47.7	44.3	53.8	52.0	44.5	39.8	51.5	50.6	43.2	38.6
100	84.0	74.5	64.2	60.9	71.6	62.8	53.9	52.3	60.5	55.9	48.8	48.3
200	98.6	97.7	94.2	94.1	94.2	91.3	85.9	84.2	88.5	82.6	75.5	73.1
500	100	100	100	99.9	99.9	99.9	99.9	99.7	99.6	99.7	99.3	99.5

Part B: Average number of non-pervasive units falsely selected as pervasive

$k_0 = 0$

	$C = 1$				$C = 1.25$				$C = 1.5$			
N \ T	60	110	210	250	60	110	210	250	60	110	210	250
50	0	0	0	0	0	0	0	0	0	0	0	0
100	0	0	0	0	0	0	0	0	0	0	0	0
200	0	0	0	0	0	0	0	0	0	0	0	0
500	0	0	0	0	0	0	0	0	0	0	0	0

$k_0 = 1$

	$C = 1$				$C = 1.25$				$C = 1.5$			
N \ T	60	110	210	250	60	110	210	250	60	110	210	250
50	0.1	0.2	0.2	0.2	0.1	0.2	0.2	0.3	0.2	0.2	0.2	0.3
100	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.2	0.1	0.1	0.2
200	0	0	0	0	0	0	0	0	0	0.1	0.1	0.1
500	0	0	0	0	0	0	0	0	0	0	0	0

$k_0 = 2$

	$C = 1$				$C = 1.25$				$C = 1.5$			
N \ T	60	110	210	250	60	110	210	250	60	110	210	250
50	0.4	0.5	0.6	0.7	0.5	0.5	0.7	0.8	0.5	0.5	0.7	0.8
100	0.2	0.3	0.4	0.4	0.3	0.4	0.5	0.5	0.4	0.5	0.6	0.6
200	0	0	0.1	0.1	0.1	0.1	0.2	0.2	0.1	0.2	0.3	0.3
500	0	0	0	0	0	0	0	0	0	0	0	0

Notes: $SMT - \sigma^2$ refers to Sequential-MT σ^2 thresholding, as implemented using Algorithm 3 in the main article. The threshold in the σ^2 thresholding step of this algorithm is given by $\hat{\sigma}_{iT}^2 \leq 2C\hat{\eta}_{iN}^2 N^{-1}\log(T)$, where different values of C are considered in this table.

Table S5.5: $SMT-\sigma^2$ thresholding: Empirical frequency of correctly identifying only the true pervasive units for $m_0 > 0$

Part A: $m_0 = 1$

$C = 1$

$k_0 = 0$ N \ T	60	110	210	250
50	97.7	99.9	100	100
100	100	100	100	100
200	100	100	100	100
500	100	100	100	100

$k_0 = 1$ N \ T	60	110	210	250
50	58.9	80.6	82.9	82.3
100	68.1	88.4	93.3	93.0
200	79.1	97.8	99.6	99.5
500	82.1	99.9	100	100

$k_0 = 2$ N \ T	60	110	210	250
50	52.5	61.7	61.1	55.5
100	65.3	75.9	74.7	74.2
200	72.7	95.6	97.1	96.0
500	77.1	99.4	100	100

$C = 1.25$

$k_0 = 0$ N \ T	60	110	210	250
50	97.7	99.9	100	100
100	100	100	100	100
200	100	100	100	100
500	100	100	100	100

$k_0 = 1$ N \ T	60	110	210	250
50	65.0	84.1	85.1	83.3
100	74.9	89.8	92.5	92.4
200	84.7	97.5	99.0	98.9
500	88.2	100	100	100

$k_0 = 2$ N \ T	60	110	210	250
50	53.1	62.1	60.1	54.0
100	65.9	70.9	69.4	68.4
200	79.4	93.1	91.6	90.3
500	83.8	99.6	99.9	99.9

$C = 1.5$

$k_0 = 0$ N \ T	60	110	210	250
50	97.7	99.9	100	100
100	100	100	100	100
200	100	100	100	100
500	100	100	100	100

$k_0 = 1$ N \ T	60	110	210	250
50	68.6	85.5	85.3	83.5
100	78.1	90.5	92.2	92.3
200	87.4	96.3	97.9	97.7
500	91.3	99.9	100	99.9

$k_0 = 2$ N \ T	60	110	210	250
50	53.3	62.0	59.2	53.5
100	63.5	68.1	67.1	66.5
200	80.8	88.5	85.8	83.7
500	88.3	99.7	99.8	99.8

Part B: $m_0 = 2$

$C = 1$

$k_0 = 0$ N \ T	60	110	210	250
50	55.8	85.0	96.0	97.7
100	58.9	87.3	98.2	98.6
200	59.0	88.8	98.4	98.9
500	60.9	94.8	100	100

$k_0 = 1$ N \ T	60	110	210	250
50	36.2	67.3	79.1	79.5
100	41.7	78.5	91.5	92.4
200	43.5	87.6	98.3	99.3
500	46.0	96.2	100	100

$k_0 = 2$ N \ T	60	110	210	250
50	38.9	61.3	63.0	60.5
100	48.4	73.3	79.6	79.6
200	47.5	86.9	96.8	97.1
500	41.0	94.6	99.9	100

$C = 1.25$

$k_0 = 0$ N \ T	60	110	210	250
50	68.8	92.3	98.8	99.1
100	72.7	95.3	99.6	99.8
200	67.9	93.3	99.5	99.7
500	68.7	96.3	100	100

$k_0 = 1$ N \ T	60	110	210	250
50	43.8	75.1	84.4	83.8
100	56.6	86.7	93.7	94.2
200	56.7	92.0	98.1	99.6
500	57.3	97.4	100	100

$k_0 = 2$ N \ T	60	110	210	250
50	41.2	63.8	64.6	61.1
100	57.3	75.9	77.7	78.4
200	59.6	88.2	94.3	93.8
500	53.6	96.9	99.9	99.9

$C = 1.5$

$k_0 = 0$ N \ T	60	110	210	250
50	73.6	94.7	99.4	99.5
100	83.3	98.3	100	100
200	75.1	96.9	99.9	100
500	73.4	97.5	100	100

$k_0 = 1$ N \ T	60	110	210	250
50	48.0	78.0	86.4	85.6
100	65.6	90.4	94.5	94.4
200	64.7	94.3	97.6	98.9
500	66.3	98.4	99.9	100

$k_0 = 2$ N \ T	60	110	210	250
50	42.7	65.1	64.9	61.1
100	61.0	76.1	76.8	78.2
200	67.3	87.1	90.7	90.2
500	62.8	97.6	99.9	99.8

Notes: See the notes to Table S5.4.

Table S5.6: $SMT-\sigma^2$ thresholding: Average number of non-pervasive units falsely selected as pervasive for $m_0 > 0$

Part A: $m_0 = 1$

$C = 1$, $k_0 = 0$

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

$C = 1.25$, $k_0 = 0$

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

$C = 1.5$, $k_0 = 0$

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

$C = 1$, $k_0 = 1$

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

$C = 1.25$, $k_0 = 1$

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

$C = 1.5$, $k_0 = 1$

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

$C = 1$, $k_0 = 2$

N \ T	60	110	210	250
50	0.4	0.3	0.4	0.5
100	0.1	0.2	0.3	0.3
200	0	0	0	0
500	0	0	0	0

$C = 1.25$, $k_0 = 2$

N \ T	60	110	210	250
50	0.4	0.4	0.5	0.5
100	0.2	0.3	0.3	0.4
200	0	0.1	0.1	0.1
500	0	0	0	0

$C = 1.5$, $k_0 = 2$

N \ T	60	110	210	250
50	0.5	0.4	0.5	0.5
100	0.3	0.3	0.4	0.4
200	0.1	0.1	0.2	0.2
500	0	0	0	0

Part B: $m_0 = 2$

$C = 1$, $k_0 = 0$

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

$C = 1.25$, $k_0 = 0$

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

$C = 1.5$, $k_0 = 0$

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

$C = 1$, $k_0 = 1$

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

$C = 1.25$, $k_0 = 1$

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

$C = 1.5$, $k_0 = 1$

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

$C = 1$, $k_0 = 2$

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

$C = 1.25$, $k_0 = 2$

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

$C = 1.5$, $k_0 = 2$

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

Notes: See the notes to Table S5.4.

S6 Maximum difference and maximum ratio thresholding

The idea of considering the maximum difference or the maximum ratio between two ordered statistics, has been recently suggested by Ahn and Horenstein (2013) and used by Brownlees and Mesters (2019) in the context of detecting pervasive units, can also be applied to σ^2 thresholding. Denote by $\hat{\sigma}^2_{(1)T}, \hat{\sigma}^2_{(2)T}, \ldots, \hat{\sigma}^2_{(N)T}$ the ordered estimated error variances in *ascending* order for a dataset with N cross section units and T time periods. Then the following two simple algorithms can be considered:

Algorithm 6 (Max σ^2-difference algorithm)

1. *Conduct σ^2 thresholding using p_{max} estimated factors. If the estimated number of pervasive units, denoted by $\widetilde{m}$, is zero, stop and conclude that there is no pervasive unit. Otherwise, proceed with step 2.*

2. *Let the estimated number of pervasive units be given by*

$$\hat{m} = \underset{j=1,2,\ldots,p_{max}}{\arg\max} \left(\hat{\sigma}^2_{(j+1)T} - \hat{\sigma}^2_{(j)T}\right),$$

and the estimated identities by the indices of the units whose estimated error variances are $\hat{\sigma}^2_{(1)T}, \hat{\sigma}^2_{(2)T}, \ldots, \hat{\sigma}^2_{(\hat{m})T}$.

Algorithm 7 (Max σ^2-ratio algorithm)

1. *Conduct σ^2 thresholding using p_{max} estimated factors. If $\widetilde{m} = 0$, stop and conclude that there is no pervasive unit. Otherwise, proceed to step 2.*

2. *Let the estimated number of pervasive units be given by*

$$\hat{m} = \underset{j=1,\ldots,p_{max}}{\arg\max} \left(\frac{\hat{\sigma}^2_{(j+1)T}}{\hat{\sigma}^2_{(j)T}}\right),$$

and the estimated identities by the indices of the units whose estimated error variances are $\hat{\sigma}^2_{(1)T}, \hat{\sigma}^2_{(2)T}, \ldots, \hat{\sigma}^2_{(\hat{m})T}$.

In Table S6.1 we report the performance of the two approaches described above using the Monte Carlo set up described in Section 6 of the paper. The case $m = 0$ is left out since the probability of correctly detecting the absence of pervasive units is entirely determined by the initial σ^2 thresholding step of max difference and max ratio thresholding methods. Results for models with at least one pervasive unit show that the two algorithms, based on either the maximum difference or the maximum ratio, perform quite similarly to the $SMT-\sigma^2$ thresholding. However, the former two methods exhibit inferior performance in samples where N is small. This comparative disadvantage is compensated by a superior performance in cases where there are both external common factors and more than one pervasive units. However, empirical evidence for the presence of at least one pervasive unit in the existing applied literature is rather limited.[25] Furthermore, the relative advantage of max difference or max ratio thresholding disappears when weakly pervasive units are considered. As reported in Table S6.3, $SMT - \sigma^2$ thresholding has a performance comparable to that of the two new algorithms considered here, even when N is large and the number of pervasive units is larger than 1.

Tables S6.2 and S6.4 report the average numbers of falsely selected pervasive units, and show that the max difference and max ratio thresholding procedures perform reasonably well. But as compared to $SMT - \sigma^2$ thresholding, the max thresholding approaches tend to show a higher proportion of false discoveries, and overall we are led to favor $SMT - \sigma^2$ thresholding over the max difference and the max ratio thresholding.

[25] see e.g. Pesaran and Yang (2019) or Dungey and Volkov (2018) who find that the degree of dominance of the most influential unit in their datasets is quite far from the value of 1 that would indicate a pervasive unit in the sense of a factor common to all cross-section units.

Table S6.1: Empirical frequency of correctly identifying only the true strongly pervasive units ($m > 0$, $\alpha = 1$)

Part A: $m_0 = 1$

$SMT-\sigma^2$

$k_0 = 0$

N \T	60	110	210	250
50	97.7	99.9	100	100
100	100	100	100	100
200	100	100	100	100
500	100	100	100	100

$k_0 = 1$

N \T	60	110	210	250
50	58.9	80.6	82.9	82.3
100	68.1	88.4	93.3	93.0
200	79.1	97.8	99.6	99.5
500	82.1	99.9	100	100

$k_0 = 2$

N \T	60	110	210	250
50	52.5	61.7	61.1	55.5
100	65.3	75.9	74.7	74.2
200	72.7	95.6	97.1	96.0
500	77.1	99.4	100	100

max σ^2−diff

$k_0 = 0$

N \T	60	110	210	250
50	89.9	97.4	99.1	99.4
100	97.9	99.5	100	100
200	99.5	100	100	100
500	100	100	100	100

$k_0 = 1$

N \T	60	110	210	250
50	45.3	74.1	88.1	91.7
100	62.6	90.4	99.2	99.1
200	76.3	98.1	100	100
500	81.8	99.9	100	100

$k_0 = 2$

N \T	60	110	210	250
50	45.9	71.1	85.4	87.7
100	62.4	89.8	98.8	99.2
200	71.7	97.0	100	100
500	76.3	99.4	100	100

max σ^2−ratio

$k_0 = 0$

N \T	60	110	210	250
50	94.6	99.2	99.8	99.8
100	99.7	100	100	100
200	100	100	100	100
500	100	100	100	100

$k_0 = 1$

N \T	60	110	210	250
50	49.9	78.9	90.9	93.7
100	66.4	92.4	99.4	99.3
200	78.6	98.2	100	100
500	82.1	99.9	100	100

$k_0 = 2$

N \T	60	110	210	250
50	51.7	75.7	89.7	90.4
100	67.0	91.5	99.1	99.6
200	73.3	97.3	100	100
500	77.1	99.4	100	100

Part B: $m_0 = 2$

$SMT-\sigma^2$

$k_0 = 0$

N \T	60	110	210	250
50	55.8	85.0	96.0	97.7
100	58.9	87.3	98.2	98.6
200	59.0	88.8	98.4	98.9
500	60.9	94.8	100	100

$k_0 = 1$

N \T	60	110	210	250
50	36.2	67.3	79.1	79.5
100	41.7	78.5	91.5	92.4
200	43.5	87.6	98.3	99.3
500	46.0	96.2	100	100

$k_0 = 2$

N \T	60	110	210	250
50	38.9	61.3	63.0	60.5
100	48.4	73.3	79.6	79.6
200	47.5	86.9	96.8	97.1
500	41.0	94.6	99.9	100

max σ^2−diff

$k_0 = 0$

N \T	60	110	210	250
50	37.8	67.3	86.6	89.1
100	49.5	83.9	97.7	98.2
200	57.1	88.5	98.4	98.9
500	60.5	94.7	100	100

$k_0 = 1$

N \T	60	110	210	250
50	30.7	56.1	72.2	76.5
100	49.1	84.9	97.4	98.2
200	64.4	97.0	99.9	100
500	75.1	98.3	100	100

$k_0 = 2$

N \T	60	110	210	250
50	29.4	53.2	73.0	75.8
100	54.2	87.5	97.8	98.5
200	71.0	98.5	100	99.9
500	77.6	99.3	100	100

max σ^2−ratio

$k_0 = 0$

N \T	60	110	210	250
50	36.6	66.2	86.5	88.6
100	48.4	84.9	97.8	98.5
200	58.0	88.8	98.4	98.9
500	60.9	94.8	100	100

$k_0 = 1$

N \T	60	110	210	250
50	28.2	52.6	67.3	70.3
100	44.3	77.6	91.6	92.7
200	56.8	91.1	99.3	99.5
500	65.4	97.6	100	100

$k_0 = 2$

N \T	60	110	210	250
50	28.7	50.1	65.5	69.8
100	49.8	78.3	91.9	93.0
200	61.4	92.8	99.2	99.7
500	66.0	97.8	100	100

Notes: $SMT-\sigma^2$ refers to $SMT-\sigma^2$ thresholding, implemented with $pmax = m_0 + k_0 + 1$ as described in Algorithm 5. max σ^2−diff and max σ^2−ratio denote detection of pervasive units via algorithms 6 and 7, conducted with $pmax = m_0 + k_0 + 1$.

Table S6.2: Average number of non-pervasive units falsely selected as pervasive units ($m_0 > 0$ and $\alpha = 1$)

Part A: $m_0 = 1$

	$SMT-\sigma^2$				$\max \sigma^2-$diff				$\max \sigma^2-$ratio			
$k_0 = 0$												
N \T	60	110	210	250	60	110	210	250	60	110	210	250
50	0	0	0	0	0.1	0	0	0	0.1	0	0	0
100	0	0	0	0	0	0	0	0	0	0	0	0
200	0	0	0	0	0	0	0	0	0	0	0	0
500	0	0	0	0	0	0	0	0	0	0	0	0
$k_0 = 1$												
N \T	60	110	210	250	60	110	210	250	60	110	210	250
50	0.2	0.1	0.2	0.2	0.5	0.2	0.1	0.1	0.4	0.2	0.1	0
100	0.1	0.1	0.1	0.1	0.1	0	0	0	0.1	0	0	0
200	0	0	0	0	0	0	0	0	0	0	0	0
500	0	0	0	0	0	0	0	0	0	0	0	0
$k_0 = 2$												
N \T	60	110	210	250	60	110	210	250	60	110	210	250
50	0.4	0.3	0.4	0.5	0.7	0.4	0.2	0.1	0.6	0.3	0.1	0.1
100	0.1	0.2	0.3	0.3	0.2	0.1	0	0	0.2	0	0	0
200	0	0	0	0	0	0	0	0	0	0	0	0
500	0	0	0	0	0	0	0	0	0	0	0	0

Part B: $m_0 = 2$

	$SMT-\sigma^2$				$\max \sigma^2-$diff				$\max \sigma^2-$ratio			
$k_0 = 0$												
N \T	60	110	210	250	60	110	210	250	60	110	210	250
50	0	0	0	0	0.2	0.1	0.1	0.1	0.2	0.1	0	0
100	0	0	0	0	0.1	0	0	0	0.1	0	0	0
200	0	0	0	0	0	0	0	0	0	0	0	0
500	0	0	0	0	0	0	0	0	0	0	0	0
$k_0 = 1$												
N \T	60	110	210	250	60	110	210	250	60	110	210	250
50	0.2	0.1	0.1	0.1	0.4	0.2	0.1	0.1	0.3	0.1	0.1	0
100	0	0	0	0	0.2	0.1	0	0	0.1	0	0	0
200	0	0	0	0	0.1	0	0	0	0	0	0	0
500	0	0	0	0	0	0	0	0	0	0	0	0
$k_0 = 2$												
N \T	60	110	210	250	60	110	210	250	60	110	210	250
50	0.3	0.3	0.3	0.4	0.6	0.3	0.1	0.1	0.4	0.2	0.1	0.1
100	0.1	0.1	0.2	0.2	0.2	0.1	0	0	0.1	0	0	0
200	0	0	0	0	0.1	0	0	0	0	0	0	0
500	0	0	0	0	0	0	0	0	0	0	0	0

Notes: See the notes to Table S6.1.

Table S6.3: Empirical frequency of correctly identifying only the true weakly pervasive (influential) units ($m > 0$, $\alpha = 0.8$)

Part A: $m_0 = 1$

	$SMT - \sigma^2$				max σ^2−diff				max σ^2−ratio			
	$k_0 = 0$				$k_0 = 0$				$k_0 = 0$			
N \T	60	110	210	250	60	110	210	250	60	110	210	250
50	51.2	80.6	95.4	97.5	47.4	74.6	88.0	89.9	50.4	77.8	90.8	92.6
100	87.2	98.9	100	100	75.6	95.0	99.2	99.5	79.3	97.3	99.8	99.9
200	97.5	100	100	100	91.6	99.4	100	100	94.1	99.9	100	100
500	97.7	100	100	100	96.3	100	100	100	97.5	100	100	100
	$k_0 = 1$				$k_0 = 1$				$k_0 = 1$			
N \T	60	110	210	250	60	110	210	250	60	110	210	250
50	37.4	65.2	79.1	78.9	26.3	49.6	73.1	77.4	28.9	53.5	76.9	81.0
100	65.1	90.5	93.7	93.5	48.9	84.7	98.0	98.1	52.8	88.2	99.3	99.1
200	84.6	99.4	99.6	99.5	72.0	98.1	100	100	76.0	99.4	100	100
500	82.7	99.9	100	100	80.2	99.7	100	100	81.9	99.9	100	100
	$k_0 = 2$				$k_0 = 2$				$k_0 = 2$			
N \T	60	110	210	250	60	110	210	250	60	110	210	250
50	37.7	53.3	58.2	54.5	28.2	50	70.9	74.1	32.0	54.3	75.5	78.6
100	64.2	79.7	75.3	74.4	50.3	83.1	97.6	98.2	54.5	87.9	98.6	99.4
200	82.7	98.2	97.1	96.0	71.3	98.0	99.8	100	75.8	99.2	100	100
500	80.9	100	100	100	77.5	99.9	100	100	79.6	100	100	100

Part B: $m_0 = 2$

	$SMT - \sigma^2$				max σ^2−diff				max σ^2−ratio			
	$k_0 = 0$				$k_0 = 0$				$k_0 = 0$			
N \T	60	110	210	250	60	110	210	250	60	110	210	250
50	6.6	31.6	63.7	67.3	14.1	27.0	39.2	41.5	13.4	25.7	38.6	40.2
100	13.7	57.4	89.2	92.6	13.0	36.8	62.4	68.1	12.6	35.9	60.4	66.2
200	7.7	48.2	88.1	92.0	5.4	36.0	78.7	85.1	4.8	34.8	77.0	84.0
500	0.9	23.0	71.0	79.3	0.6	20.1	70.6	79.1	0.5	19.9	70.6	79.2
	$k_0 = 1$				$k_0 = 1$				$k_0 = 1$			
N \T	60	110	210	250	60	110	210	250	60	110	210	250
50	8.9	32.0	60.6	63.7	9.6	21.6	32.3	34.1	9.5	20.7	31.6	33.1
100	16.5	61.3	88.8	91.5	13.6	37.8	60.6	65.3	13.3	36.4	58.7	63.9
200	11.4	61.8	94.7	97.1	9.8	44.1	82.5	87.7	8.9	42.7	80.7	86.2
500	1.8	32.3	84.6	91.7	0.9	28.4	83.7	91.5	0.9	27.9	83.7	91.7
	$k_0 = 2$				$k_0 = 2$				$k_0 = 2$			
N \T	60	110	210	250	60	110	210	250	60	110	210	250
50	13.3	33.5	50.1	50.4	9.7	18.7	32.9	36.7	9.4	18.8	32.2	36.3
100	26.6	65.3	79.3	80.2	16.6	39.2	67.5	72.8	15.3	38.3	65.3	70.9
200	17.6	75.7	96.1	96.6	13.1	51.2	88.0	91.8	12.5	50.4	86.4	90
500	2.4	36.5	89.2	93.8	1.6	33.0	88.5	93.6	1.7	31.9	88.4	93.6

Notes: See the notes to Table S6.1.

Table S6.4: Average number of non-pervasive units falsely selected as pervasive units ($m_0 > 0$ and $\alpha = 0.8$)

Part A: $m_0 = 1$

$SMT - \sigma^2$

$k_0 = 0$

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

max σ^2-diff

$k_0 = 0$

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

$k_0 = 1$

N \T	60	110	210	250
50	0.9	0.7	0.3	0.3
100	0.5	0.2	0	0
200	0.2	0	0	0
500	0	0	0	0

$k_0 = 2$

N \T	60	110	210	250
50	1.1	0.8	0.4	0.4
100	0.6	0.2	0	0
200	0.2	0	0	0
500	0	0	0	0

max σ^2-ratio

$k_0 = 0$

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

$k_0 = 1$

N \T	60	110	210	250
50	0.8	0.6	0.3	0.2
100	0.4	0.1	0	0
200	0.1	0	0	0
500	0	0	0	0

$k_0 = 2$

N \T	60	110	210	250
50	1.0	0.7	0.4	0.3
100	0.5	0.2	0	0
200	0.2	0	0	0
500	0	0	0	0

Part B: $m_0 = 2$

$SMT - \sigma^2$

$k_0 = 0$

N \T	60	110	210	250
50	0.1	0.1	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	0.2	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	0.4	0.5
100	0.3	0.2	0.2	0.2
200	0.1	0	0	0
500	0	0	0	0

max σ^2-diff

$k_0 = 0$

N \T	60	110	210	250
50	0.5	0.4	0.3	0.3
100	0.3	0.2	0.1	0.1
200	0.1	0.1	0	0
500	0	0	0	0

$k_0 = 1$

N \T	60	110	210	250
50	0.9	0.7	0.5	0.5
100	0.6	0.4	0.2	0.2
200	0.2	0.2	0.1	0
500	0	0	0	0

$k_0 = 2$

N \T	60	110	210	250
50	1.2	0.9	0.6	0.6
100	0.8	0.5	0.2	0.2
200	0.3	0.3	0.1	0
500	0	0	0	0

max σ^2-ratio

$k_0 = 0$

N \T	60	110	210	250
50	0.5	0.4	0.2	0.2
100	0.2	0.2	0.1	0.1
200	0.1	0.1	0	0
500	0	0	0	0

$k_0 = 1$

N \T	60	110	210	250
50	0.8	0.6	0.4	0.4
100	0.5	0.4	0.2	0.2
200	0.2	0.1	0	0
500	0	0	0	0

$k_0 = 2$

N \T	60	110	210	250
50	1.0	0.7	0.5	0.5
100	0.7	0.4	0.1	0.1
200	0.2	0.2	0	0
500	0	0	0	0

Notes: See the notes to Table S6.1.

S7 Simulation results for unmodified BM

In the paper we have used a modified version of BM's detection method discussed in Section 6 of Brownlees and Mesters (2019), whereby only the $N/2$ most connected cross-section units are considered when determining the number of pervasive units. This section complements the simulations in Section 6 of the paper and report results for BM without this modification (henceforth *unmodified BM*). When implementing this procedure, the number of pervasive units is determined from all N cross section units in the dataset. All other details of the simulation exercise are as described in Section 6 of the paper.

Results on the probability of correctly determining the absence of pervasive units from the data are left out since BM selects at least one unit as pervasive by construction. The results for experiments with $m_0 > 0$ are summarized in Table S7.1. As can be seen the average number of units detected as pervasive turns out to be much larger as compared to the modified BM. In fact, more than half of the cross section units in the sample are, on average, found to be pervasive. In some cases, standardization of the data leads to a considerable decrease in the number of detected units. However, the set of cross section units falsely identified as pervasive continues to be sizeable.

In cases where the data are driven by at least one pervasive unit, unmodified BM method exhibits a reasonable performance if $T - N$ is large enough, and if the data is not standardized (see Table S7.2). By contrast, standardizing individual-specific time series has severe consequences for the probability of correctly detecting the true pervasive units, especially in the presence of external factors. As can be seen from Table S7.4, the same results obtain if the true pervasive units are weakly pervasive. The average number of units falsely detected as pervasive can be substantial. See Table S7.3 and Table S7.5).

Table S7.1: Average number of non-pervasive units falsely selected as pervasive ($m_0 = 0$)

	unmodifed BM				unmodified BM (standardized)			
	$k_0 = 0$				$k_0 = 0$			
N \T	60	110	210	250	60	110	210	250
50	29.7	31.2	36.4	36.9	29.3	23.2	26.0	29.5
100	n/a	61.7	69.0	71.9	n/a	60.7	46.3	49.0
200	n/a	n/a	126.2	123.4	n/a	n/a	126.0	93.8
500	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
	$k_0 = 1$				$k_0 = 1$			
N \T	60	110	210	250	60	110	210	250
50	28.7	31.6	37.1	37.4	26.0	17.2	14.3	14.6
100	n/a	60.6	67.9	73.1	n/a	54.8	29.8	26.2
200	n/a	n/a	123.1	128.4	n/a	n/a	111.1	78.5
500	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
	$k_0 = 2$				$k_0 = 2$			
N \T	60	110	210	250	60	110	210	250
50	30.5	31.4	36.1	37.6	25.8	17.8	13.9	12.6
100	n/a	61.6	69.3	72.1	n/a	52.2	25.4	24.8
200	n/a	n/a	126.8	127.2	n/a	n/a	105.5	72.0
500	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

Notes: unmodified BM refers to the detection method of Brownlees and Mesters (2019) as introduced formally in Section 3 of their paper. unmodified BM (standardized) stands for application of unmodified BM to data that have been recentered and rescaled so that each cross-section specific time-series has an average of zero and a variance of one. BM methods are not applicable (n/a) if $T < N$.

Table S7.2: Empirical frequency of correctly identifying only the true strongly pervasive units ($m_0 > 0$, and $\alpha = 1$)

Part A: $m_0 = 1$

	unmodifed BM				unmodified BM (standardized)			
	$k_0 = 0$				$k_0 = 0$			
N \T	60	110	210	250	60	110	210	250
50	48.4	98.3	100.0	100.0	47.0	98.2	99.7	99.9
100	n/a	73.6	100.0	100.0	n/a	69.3	100.0	100.0
200	n/a	n/a	89.6	100.0	n/a	n/a	87.4	100.0
500	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
	$k_0 = 1$				$k_0 = 1$			
N \T	60	110	210	250	60	110	210	250
50	42.1	96.7	99.9	99.9	25.6	74.7	90.7	93.1
100	n/a	67.3	100.0	100.0	n/a	47.6	99.0	99.8
200	n/a	n/a	85.0	100.0	n/a	n/a	69.4	99.9
500	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
	$k_0 = 2$				$k_0 = 2$			
N \T	60	110	210	250	60	110	210	250
50	36.7	95.4	99.6	99.8	12.8	45.1	62.7	65.4
100	n/a	63.6	100	100	n/a	29.6	94.4	96.0
200	n/a	n/a	83.7	100	n/a	n/a	53.3	98.3
500	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

Part B: $m_0 = 2$

	unmodifed BM				unmodified BM (standardized)			
	$k_0 = 0$				$k_0 = 0$			
N \T	60	110	210	250	60	110	210	250
50	23.6	91.7	99.4	99.7	5.4	40.8	68.7	72.4
100	n/a	46.0	100.0	100.0	n/a	16.7	94.9	98.4
200	n/a	n/a	66.5	100.0	n/a	n/a	38.1	97.8
500	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
	$k_0 = 1$				$k_0 = 1$			
N \T	60	110	210	250	60	110	210	250
50	17.1	85.1	97.9	98.5	1.4	7.5	16.5	17.7
100	n/a	36.9	99.9	100.0	n/a	5.6	58.5	63.1
200	n/a	n/a	55.7	99.9	n/a	n/a	14.8	75.6
500	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
	$k_0 = 2$				$k_0 = 2$			
N \T	60	110	210	250	60	110	210	250
50	15.55	79.5	96.3	97.35	0.2	1.0	2.2	1.6
100	n/a	33.1	99.95	99.85	n/a	1.5	22.8	28.7
200	n/a	n/a	50.35	99.65	n/a	n/a	5.8	46.0
500	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

Notes: See the notes to Table S7.1.

Table S7.3: Average number of non-pervasive units falsely selected as pervasive ($m_0 > 0$, and $\alpha = 1$)

Part A: $m_0 = 1$

	unmodifed BM					unmodified BM (standardized)			
	$k_0 = 0$					$k_0 = 0$			
N \T	60	110	210	250	N \T	60	110	210	250
50	14.7	0.3	0.0	0.0	50	12.5	0.0	0.0	0.0
100	n/a	16.0	0.0	0.0	100	n/a	15.4	0.0	0.0
200	n/a	n/a	12.1	0.0	200	n/a	n/a	12.3	0.0
500	n/a	n/a	n/a	n/a	500	n/a	n/a	n/a	n/a
	$k_0 = 1$					$k_0 = 1$			
N \T	60	110	210	250	N \T	60	110	210	250
50	17.6	0.6	0.0	0.0	50	19.1	2.9	0.5	0.3
100	n/a	19.6	0.0	0.0	100	n/a	26.0	0.0	0.0
200	n/a	n/a	19.6	0.0	200	n/a	n/a	30.0	0.0
500	n/a	n/a	n/a	n/a	500	n/a	n/a	n/a	n/a
	$k_0 = 2$					$k_0 = 2$			
N \T	60	110	210	250	N \T	60	110	210	250
50	18.5	1.1	0.1	0.1	50	23.7	10.0	5.3	4.7
100	n/a	21.6	0.0	0.0	100	n/a	37.6	0.7	0.3
200	n/a	n/a	19.4	0.0	200	n/a	n/a	49.1	0.4
500	n/a	n/a	n/a	n/a	500	n/a	n/a	n/a	n/a

Part B: $m_0 = 2$

	unmodifed BM					unmodified BM (standardized)			
	$k_0 = 0$					$k_0 = 0$			
N \T	60	110	210	250	N \T	60	110	210	250
50	11.7	0.4	0.0	0.0	50	16.9	2.2	0.4	0.5
100	n/a	8.2	0.0	0.0	100	n/a	19.6	0.0	0.0
200	n/a	n/a	6.2	0.0	200	n/a	n/a	18.2	0.0
500	n/a	n/a	n/a	n/a	500	n/a	n/a	n/a	n/a
	$k_0 = 1$					$k_0 = 1$			
N \T	60	110	210	250	N \T	60	110	210	250
50	13.0	1.1	0.1	0.0	50	21.4	10.0	6.8	6.3
100	n/a	13.7	0.0	0.0	100	n/a	33.7	1.0	0.7
200	n/a	n/a	10.0	0.0	200	n/a	n/a	42.5	0.6
500	n/a	n/a	n/a	n/a	500	n/a	n/a	n/a	n/a
	$k_0 = 2$					$k_0 = 2$			
N \T	60	110	210	250	N \T	60	110	210	250
50	13.8	1.1	0.2	0.2	50	26.0	19.2	17.8	18.2
100	n/a	13.7	0.0	0.0	100	n/a	44.0	8.3	6.3
200	n/a	n/a	12.7	0.0	200	n/a	n/a	67.3	4.9
500	n/a	n/a	n/a	n/a	500	n/a	n/a	n/a	n/a

Notes: See the notes to Table S7.1.

Table S7.4: Empirical frequency of correctly identifying only the true weakly pervasive (influential) units ($m_0 > 0$, and $\alpha = 1$)

Part A: $m_0 = 1$

	unmodifed BM					unmodified BM (standardized)			
	$k_0 = 0$					$k_0 = 0$			
N \T	60	110	210	250	N \T	60	110	210	250
50	15.8	63.2	87.6	90.4	50	21.9	69.3	85.3	87.8
100	n/a	30.1	99.0	100.0	100	n/a	37.8	98.4	99.4
200	n/a	n/a	44.8	99.3	200	n/a	n/a	54.9	99.0
500	n/a	n/a	n/a	n/a	500	n/a	n/a	n/a	n/a
	$m_0 = 1, k = 1$					$k_0 = 1$			
N \T	60	110	210	250	N \T	60	110	210	250
50	13.3	59.8	87.1	89.3	50	9.7	31.1	45.4	45.1
100	n/a	27.5	98.8	99.6	100	n/a	21.7	82.0	84.9
200	n/a	n/a	44.3	98.7	200	n/a	n/a	37.5	89.8
500	n/a	n/a	n/a	n/a	500	n/a	n/a	n/a	n/a
	$m_0 = 1, k = 2$					$k_0 = 2$			
N \T	60	110	210	250	N \T	60	110	210	250
50	13.3	59.0	84.6	88.7	50	5.9	13.5	15.9	17.4
100	n/a	25.6	98.8	99.5	100	n/a	12.9	52.8	59.7
200	n/a	n/a	44.6	98.6	200	n/a	n/a	24.3	75.3
500	n/a	n/a	n/a	n/a	500	n/a	n/a	n/a	n/a

Part B: $m_0 = 2$

	unmodifed BM					unmodified BM (standardized)			
	$k_0 = 0$					$k_0 = 0$			
N \T	60	110	210	250	N \T	60	110	210	250
50	4.6	43.8	74.9	79.4	50	1.4	4.9	9.0	11.5
100	n/a	10.8	94.3	98.0	100	n/a	3.1	43.7	50.9
200	n/a	n/a	20.7	94.2	200	n/a	n/a	10.3	64.5
500	n/a	n/a	n/a	n/a	500	n/a	n/a	n/a	n/a
	$m_0 = 2, k = 1$					$k_0 = 1$			
N \T	60	110	210	250	N \T	60	110	210	250
50	4.2	41.6	70.2	76.5	50	0.2	0.6	0.7	0.7
100	n/a	10.2	94.3	96.4	100	n/a	1.3	11.9	11.8
200	n/a	n/a	17.1	92.2	200	n/a	n/a	3.3	29.3
500	n/a	n/a	n/a	n/a	500	n/a	n/a	n/a	n/a
	$m_0 = 2, k = 2$					$k_0 = 2$			
N \T	60	110	210	250	N \T	60	110	210	250
50	3.2	39.1	68.8	73.5	50	0.0	0.1	0.0	0.0
100	n/a	9.1	92.1	95.9	100	n/a	0.4	1.5	1.4
200	n/a	n/a	17.0	91.2	200	n/a	n/a	1.4	9.4
500	n/a	n/a	n/a	n/a	500	n/a	n/a	n/a	n/a

Notes: See the notes to Table S7.1.

Table S7.5: Average number of non-pervasive units falsely selected as pervasive ($m_0 > 0$, and $\alpha = 0.8$)

Part A: $m_0 = 1$

	unmodifed BM					unmodified BM (standardized)			
	$k_0 = 0$					$k_0 = 0$			
N \T	60	110	210	250	N \T	60	110	210	250
50	25.1	10.1	4.4	3.3	50	19.2	2.7	0.3	0.2
100	n/a	43.5	0.3	0.0	100	n/a	32.7	0.0	0.0
200	n/a	n/a	72.9	0.6	200	n/a	n/a	46.5	0.1
500	n/a	n/a	n/a	n/a	500	n/a	n/a	n/a	n/a
	$k_0 = 1$					$k_0 = 1$			
N \T	60	110	210	250	N \T	60	110	210	250
50	26.8	12.0	4.3	3.5	50	21.2	7.3	3.3	2.8
100	n/a	44.1	0.4	0.3	100	n/a	38.7	1.1	0.4
200	n/a	n/a	72.6	1.3	200	n/a	n/a	59.4	2.2
500	n/a	n/a	n/a	n/a	500	n/a	n/a	n/a	n/a
	$k_0 = 2$					$k_0 = 2$			
N \T	60	110	210	250	N \T	60	110	210	250
50	25.5	12.2	5.2	3.9	50	23.1	11.7	8.8	8.1
100	n/a	46.8	0.7	0.3	100	n/a	43.6	7.0	4.4
200	n/a	n/a	69.4	1.5	200	n/a	n/a	72.1	8.6
500	n/a	n/a	n/a	n/a	500	n/a	n/a	n/a	n/a

Part B: $m_0 = 2$

	unmodified BM					unmodified BM (standardized)			
	$k_0 = 0$					$k_0 = 0$			
N \T	60	110	210	250	N \T	60	110	210	250
50	22.5	7.1	2.7	2.2	50	18.6	5.2	3.3	3.2
100	n/a	33.7	0.2	0.1	100	n/a	29.8	0.9	0.6
200	n/a	n/a	48.2	0.4	200	n/a	n/a	43.0	0.9
500	n/a	n/a	n/a	n/a	500	n/a	n/a	n/a	n/a
	$k_0 = 1$					$k_0 = 1$			
N \T	60	110	210	250	N \T	60	110	210	250
50	22.1	8.5	3.3	2.5	50	20.2	9.5	6.5	6.6
100	n/a	35.4	0.4	0.1	100	n/a	35.5	4.6	3.5
200	n/a	n/a	52.3	0.7	200	n/a	n/a	56.8	5.8
500	n/a	n/a	n/a	n/a	500	n/a	n/a	n/a	n/a
	$k_0 = 2$					$k_0 = 2$			
N \T	60	110	210	250	N \T	60	110	210	250
50	23.2	7.9	3.5	2.9	50	21.6	12.8	10.4	10.0
100	n/a	36.9	0.6	0.1	100	n/a	41.3	12.6	10.2
200	n/a	n/a	56.3	0.7	200	n/a	n/a	75.1	18.5
500	n/a	n/a	n/a	n/a	500	n/a	n/a	n/a	n/a

Notes: See the notes to Table S7.1.

Table S8.1: Pervasive units in sector-wise industrial production in the U.S.

	Full sample (1972m1 - 2007m12)	
Approach:	unmodified BM	unmodified BM (standardized)
Number of pervasive units:	137	2
Identities:	*all except* *Iron Ore Mining*	Automobiles and Light Duty Motor Vehicles Motor Vehicle Parts
	Sub-sample A (1972m1 - 1983m12)	
Approach:	unmodified BM	unmodified BM (standardized)
Number of pervasive units:	135	2
Identities:	*all except* *Iron Ore Mining* *Heavy Duty Trucks* *Motor Homes*	Motor Vehicle Parts
	Sub-sample B (1984m1 - 2007m12)	
Approach:	unmodified BM	unmodified BM (standardized)
Number of pervasive units:	137	5
Identities:	*all except* *Audio and Video Equipment*	Motor Vehicle Parts Automobiles and Light Duty Motor Vehicles Aluminum Extruded Products Miscellaneous Aluminum Materials Motor Vehicle Bodies

Notes: Data taken from Foerster, Sarte, and Watson (2011).

S8 Empirical results for unmodified BM

In this section we provide results obtained if the unmodified BM procedure is used in our empirical applications. The data sources and transformations are as described in Section 7 of the paper. Again, unmodified BM method is applied to the data with and without standardization. The results are summarized in Tables S8.1-S8.3, and suggest that unmodified BM grossly overestimates the number of pervasive units in almost all applications, regularly detecting all but one or two cross section units as pervasive. The use of standardized data leads in all but one case to a lower detected number of pervasive. However, while the reduction can be quite substantial, in a number of applications the number of pervasive units detected using standardized data can be quite large (5 or more in some the applications).

Table S8.2: Pervasive countries in terms of quarterly macroeconomic indicators

Variable:	*real GDP growth*		*real equity price growth*	
Approach:	unmodified BM	unmodified BM (std)	unmodified BM	unmodified BM (std)
Number of pervasive units:	2	11	25	1
Identities:	France Spain	*	*all except Argentina*	Netherlands

*: Italy; Spain; France; USA; Germany; Canada; UK; Malaysia; Belgium; Finland; South Africa.

Notes: Data taken from GVAR database (Mohaddes and Raissi, 2018).

Table S8.3: Estimated U.S. states with pervasive housing market

Approach:	unmodified BM	unmodified BM (standardized)	
Number of pervasive units:	47	6	
Identities:	*all except Nevada*	Connecticut New Hampshire Massachusetts	Maryland Virginia Rhode Island

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