Journal of Econometrics 220 (2021) 544–561

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Detecting granular time series in large panels

Christian Brownlees [a], Geert Mesters [a,b,*]

[a] Department of Economics and Business, Universitat Pompeu Fabra and Barcelona GSE, Ramon Trias, Fargas 25-27, Barcelona 08005, Spain
[b] VU Amsterdam, Department of Econometrics, De Boelenlaan 1105, Amsterdam 1081HV, Netherlands

ARTICLE INFO

Article history:
Available online 16 May 2020

JEL classification:
C33
C38

Keywords:
Granularity
Network models
Factor models
Industrial production

ABSTRACT

Large economic and financial panels can include time series that influence the entire cross-section. We name such series *granular*. In this paper we introduce a panel data model that allows to formalize the notion of granular time series. We then propose a methodology, which is inspired by the network literature in statistics and econometrics, to detect the set of granulars when such set is unknown. The influence of the ith series in the panel is measured by the norm of the ith column of the inverse covariance matrix. We show that a detection procedure based on the column norms allows to consistently select granular series when the cross-section and time series dimensions are large. Importantly, the methodology allows to consistently detect granulars also when the series in the panel are influenced by common factors. A simulation study shows that the proposed procedures perform satisfactorily in finite samples. Our empirical study shows the granular influence of the automobile sector in US industrial production.

1. Introduction

Traditionally, theoretical models in economics and finance assume that in large systems the influence of individual entities is negligible. This view has recently been challenged by a number of influential contributions, *inter alia*, Gabaix (2011), Acemoglu et al. (2012) and Acemoglu et al. (2015). The main theme of this strand of the literature is that entity specific shocks – through different mechanisms – impact the entire system. This is called by Gabaix (2011) the granular hypothesis. These models have been applied to explain aggregate fluctuations in macroeconomics and financial stability in finance.

One of the main hurdles in bringing these theories to the data is that in large macroeconomic or financial systems it is often the case that the set of granular entities is unknown. It is natural to ask if it is possible to introduce a methodology to recover the set of granular entities from the data. In this paper we tackle this challenge by (i) introducing a model that allows us to formalize the granular detection problem and (ii) developing a methodology to detect the set of granular series from the data when such set is unknown.[1]

We begin by introducing a model for a panel of stationary time series that formalizes the notion of granularity used in this paper. We assume that the panel is partitioned into a (finite) set of series labeled as granular and a remaining set of

* Corresponding author at: Department of Economics and Business, Universitat Pompeu Fabra and Barcelona GSE, Ramon Trias, Fargas 25-27, Barcelona 08005, Spain.
E-mail addresses: christian.brownlees@upf.edu (C. Brownlees), geert.mesters@upf.edu (G. Mesters).

[1] Our work is complementary to the large macroeconomic literature that uses input–output tables, or other criteria such as firm size, to determine whether a certain series is granular, see among others Gabaix (2011), Acemoglu et al. (2012), Di Giovanni and Levchenko (2012), Carvalho and Gabaix (2013), Di Giovanni et al. (2014), Bernard et al. (2017), Pesaran and Yang (2016), Gaubert and Itskhoki (2016) and Ghysels et al. (2018). Instead of relying on potentially arbitrary criteria for granular selection we detect granular series based on the data.

https://doi.org/10.1016/j.jeconom.2020.04.013
0304-4076/

non-granular series. The granular series coincide with their respective idiosyncratic shocks, which we call granular shocks. Each non-granular series is modeled as a linear combination of the granular shocks and an idiosyncratic non-granular shock. Differently put, we assume that the series in the panel are generated by a factor model in which the granular series are observed factors for the panel of non-granular series. We work under the assumption that set of granulars series and the total number of granulars are *unknown*.

Our granular detection methodology is based on the properties of the inverse covariance matrix of the panel, hereafter concentration matrix, and is inspired by the literature on graphical and network models in statistics, see Lauritzen (1996).[2] As it is well known, the i, j element of the concentration matrix is proportional to the partial correlation between ith and jth series. This motivates us to use as a natural measure of the influence of series i in the panel the norm of the ith column (or row) of the concentration matrix. We show – under appropriate identification assumptions – that the column norms of the concentration matrix that correspond to granular series are larger than the non-granular ones. This implies that ranking series in the panel according to the value of the column norm ranks the granular series higher than the non-granular ones. We then show that the ratio among subsequent ordered column norms is maximized when comparing the column norms of the last granular with the first non-granular series. This implies that the sequential column norm ratio can be used for selecting the number of granulars. This is analogous to the eigenvalue ratio of Ahn and Horenstein (2013) for the selection of the number of factors.

In large panels of time series common factors typically explain a large portion of the total variability, see for example Foerster et al. (2011). To this extent, we consider an extension of the model in which the series in the panel are influenced by an additional set of (unobservable) common factors.[3] We show that the column norms of the concentration matrix retain their detection properties under additional assumptions, yet under more stringent assumptions. These, in particular, require the signal-to-noise ratio of the granular shocks in the granular series to be sufficiently large.

We operationalize our identification results by estimating the column norms of the concentration matrix using the sample covariance matrix of the panel. Using tools from the literature on the estimation of large dimensional covariances (see Fan et al. (2011)) we show that the sample column norms lead to consistent ranking and selection of the granulars when the cross-sectional dimension and the number of observations are large.

Since the granular model that we introduce can be interpreted as a factor model with observed factors, it is natural to resort to principal components or maximum likelihood based methods to carry out granular detection. In brief, such methods are typically based on two steps: first they estimate the space spanned by the granular shocks and the common factors simultaneously and second such estimates are used to determine which series are granular, either by regression analysis or hypothesis testing.[4] The success of these methods depends crucially on the consistency of the first step, see for example the discussion in Onatski (2012). We show that our one-step methods, that are based on the concentration matrix, compare favorable both in theory and in finite samples.

A simulation study assesses the performance of our methodology in finite samples. Results show that the granular detection procedure performs satisfactorily when the strength of the granulars is sufficiently large and compares favorably to factor based methods.

We apply our methodology for detecting granular series in a large panel of industrial production series that was previously considered in Foerster et al. (2011). The documented granular series are mostly related to the automobile industry and secondly to the production of aluminum, plastic and paper products. These findings correspond with conjectures concerning granular sectors made in Acemoglu et al. (2012). At the same time the set of granular series is different from the set that is detected by conventional methods. The number of granulars ranges between two and five and is increasing for more recent samples.

The remainder of this paper is organized as follows. Section 2 formalizes the granular detection problem. Section 3 introduces our granular detection methodology and it establishes its large sample properties. Section 4 compares our methodology to alternative methods. Section 5 carries out a simulation study. Section 6 presents the results of the empirical study and concluding remarks follow in Section 7.

2. The granular detection problem

Let y_t be an n-dimensional time series observed from period $t = 1$ to T. We use $y_{i,t}$ to denote the ith component of y_t and $y_{i:j,t}$ with $i < j$ to denote the $(j - i + 1)$-dimensional time series containing the ith to jth components of y_t. We assume that there are k (fixed) time series whose idiosyncratic shocks g_t influence the entire panel. We label these time series as granular and the $k \times 1$ vector of shocks g_t as granular shocks.[5] For simplicity and without loss of generality we

[2] See also research by Meinshausen and Bühlmann (2006), Diebold and Yılmaz (2014) and Hautsch et al. (2015).

[3] See Long and Plosser (1987) and Forni and Reichlin (1998) for earlier work on the trade-off between idiosyncratic and aggregate shocks in macroeconomics.

[4] Examples of such methods based on principal components analysis are developed in Stock and Watson (2002a), Bai and Ng (2006), Parker and Sul (2016) and Siavash (2016). We provide a detailed comparison in Section 4. We emphasize that none of these methods are developed for granular detection. Instead their objective is to provide interpretation for the otherwise hard to pin down common factors in an approximate factor model.

[5] It is important to emphasize that in this work the term shock refers to reduced form innovations that may have structural interpretation depending on further identification restrictions.

assume that the granular series are the first k series. The remaining $n-k$ time series are called non-granular series and their idiosyncratic shocks are denoted by ϵ_t. All series in the panel are influenced by a set of r common shocks, or factors, f_t. The granular panel data model with common factors is defined as

$$\begin{aligned} y_{1:k,t} &= \Lambda_1 f_t + g_t, \\ y_{k+1:n,t} &= \Lambda_2 f_t + \beta g_t + \epsilon_t, \end{aligned} \tag{1}$$

where β is the $(n-k)\times k$ granular loading matrix and Λ_1 and Λ_2 are the $k\times r$ and $(n-k)\times r$ loading matrices for the common factors. Precise assumptions are given below.

In this paper we work under the assumption that the data is generated according to model (1) and that the researcher does not know (i) which series are granular and (ii) the number of granular series k. Our objective is to recover this information from the data.[6]

We emphasize that the model in (1) is a factor model in which the series $y_{1:k,t}$ play the role of observed factors (conditional on the common factors f_t). The techniques that we introduce however are different from the standard methodology that is adopted in the factor model literature. Specifically, our detection strategy is based on the partial correlation properties of the panel. In Section 4 our methodology is compared to these alternative approaches.

3. Methodology

3.1. Granular panel model

We first study granular detection when no common factors are present. In this case we have

$$\begin{aligned} y_{1:k,t} &= g_t, \\ y_{k+1:n,t} &= \beta g_t + \epsilon_t, \end{aligned} \tag{2}$$

where all components are defined as in (1). We propose a granular detection strategy that is based on the properties of the concentration matrix of the panel. It is straightforward to check that when g_t and ϵ_t are uncorrelated the covariance matrix $\Sigma = \mathrm{Var}(y_t)$ and the concentration matrix $\mathbf{K} = \Sigma^{-1}$ of the panel become

$$\Sigma = \begin{bmatrix} \Sigma_g & \Sigma_g\beta' \\ \beta\Sigma_g & \beta\Sigma_g\beta' + \Sigma_\epsilon \end{bmatrix}, \qquad \mathbf{K} = \begin{bmatrix} \Sigma_g^{-1} + \beta'\Sigma_\epsilon^{-1}\beta & -\beta'\Sigma_\epsilon^{-1} \\ -\Sigma_\epsilon^{-1}\beta & \Sigma_\epsilon^{-1} \end{bmatrix}. \tag{3}$$

As an example, assume that the norms of the columns of the β matrix are larger than one and that Σ_ϵ is the identity matrix. Then, it is straightforward to verify that the norms of the first k columns (or rows) of the concentration matrix $\mathbf{K}$ are larger than the norms of the last $(n-k)$ columns (or rows). Thus, the set of granular series can be identified simply by checking which series are associated with the largest column (or row) norms of $\mathbf{K}$.

The example above suggests that if the elements of the granular loading matrix β are sufficiently large relative to the covariance matrix of the non-granular shocks Σ_ϵ, then the column (or row) norms of the concentration matrix $\mathbf{K}$ can be used to identify the granular series. This motivates us to build a granular detection strategy on the columns of $\mathbf{K}$, that is

$$\|\mathbf{K}_i\| \quad \text{for} \quad i = 1, \dots, n, \tag{4}$$

where $\mathbf{K}_i$ denotes the ith column of $\mathbf{K}$.[7,8]

Our detection strategy has a natural interpretation in terms of a partial correlation network model (e.g. Pourahmadi (2013, Chapter 5)). The partial correlation network of the panel consists of a graph defined over n vertices where each series corresponds to a vertex and vertices i and j are connected by an edge if i and j are correlated given the remaining series in the panel. The concentration matrix embeds the partial dependence structure of the panel: Series i and j are partially uncorrelated if the (i,j) element of the concentration matrix $\mathbf{K}$ is zero.[9] Thus, heuristically, granular time series can be though of as hubs in a partial correlation network of the panel and the granular detection parameter $\|\mathbf{K}_i\|$ can be though as a parameter proportional to the number of connections, or degree, of each vertex.

We impose a number of assumptions on model (2) for identification.[10]

---

[6] If the set of granular series is known, then model (1) is equivalent to a factor model with a subset of the necessary identification restrictions fixed. See Stock and Watson (2016) for a detailed overview of identification restrictions for (structural) factor models and note that – pending further identifying restrictions – the model can be viewed as a structural factor model or factor augmented vector autoregressive model. Also, the specification in (1) implies that, conditional on the factors, the series in the panel have a block triangular "Cholesky" structure in which granular series influence the non-granular ones but not vice-versa.

[7] For an arbitrary vector $v = (v_1, \dots, v_n)'$ the norm $\|v\|$ is defined as $\sqrt{\sum_{i=1}^n v_i^2}$.

[8] Notice that the column norm is not the only function of the concentration matrix that can be used for granular detection.

[9] More precisely, we have that the partial correlation between series i and j ρ^{ij} is related to the concentration matrix $\mathbf{K}$ through the identity $\rho^{ij} = -\mathbf{K}_{ij}/\sqrt{\mathbf{K}_{ii}\mathbf{K}_{jj}}$.

[10] The following notation is adopted. The kth largest eigenvalue of an $N\times N$ matrix $\mathbf{A}$ is denoted as $\mu_k(\mathbf{A})$, the kth largest singular value of an $M\times N$ matrix $\mathbf{A}$ is denoted as $\sigma_k(\mathbf{A})$, $\mathbf{A}\succ 0$ indicates that $\mathbf{A}$ is positive definite and $\mathbf{A}\succeq 0$ indicates that $\mathbf{A}$ is positive semi-definite. The conditioning number of the matrix $\mathbf{A}$ is defined as the ratio of the largest and smallest eigenvalues: $\kappa_A = \mu_1(\mathbf{A})/\mu_N(\mathbf{A})$.

Assumption 1.

(i) $\mathrm{E}(g_t) = 0$ and $\mathrm{E}(g_t g_t') = \boldsymbol{\Sigma}_g$ with $\boldsymbol{\Sigma}_g \succ 0$.
(ii) $\mathrm{E}(\epsilon_t) = 0$ and $\mathrm{E}(\epsilon_t \epsilon_t') = \boldsymbol{\Sigma}_\epsilon$ with $\boldsymbol{\Sigma}_\epsilon \succ 0$.
(iii) $\mathrm{E}(g_t \epsilon_{i,t}) = 0$ for all i, t.
(iv) Let $\mathbf{D}_\beta = \boldsymbol{\beta}'\boldsymbol{\beta}$, then there exists an integer N and a constant M_g such that for each $n > N$ we have $\kappa_\beta \kappa_\epsilon < \sigma_k(\mathbf{D}_\beta) \leqslant \sigma_1(\mathbf{D}_\beta) < M_g$, where κ_ϵ and κ_β are the condition numbers of the matrices $\boldsymbol{\Sigma}_\epsilon$ and $\mathbf{D}_\beta$ respectively.

Assumptions (i), (ii) and (iii) are standard and ensure that $\boldsymbol{\Sigma}_g$ and $\boldsymbol{\Sigma}_\epsilon$ are invertible, and that g_t and $\epsilon_{i,t}$ are uncorrelated, which is standard for regression models, e.g. White (2000, Chapter 2). Assumption (iv) requires $\boldsymbol{\beta}'\boldsymbol{\beta}$ to be non-vanishing and lower bounded for large n. The lower bound depends on the degree of collinearity among the non-granular shocks and the granular loadings as it is measured by the condition numbers. The bound is such that, the larger the degree of collinearity the larger the column norms of the loading matrix $\boldsymbol{\beta}$. This assumption is key to establish the identification results of this paper. When the series in the panel are generated by model (2) but the assumption is violated then the influence of the granular series is not sufficiently strong and granular series are not guaranteed to be identified. Hence, there might be economically relevant cases for which this happens.

To understand exactly how weak the granular influence can be while still satisfying Assumption (iv) it is useful to consider an example. Let $k = 1$ such that $\kappa_\beta = 1$. In this setting the elements of $\boldsymbol{\beta}$ can be local to zero, in the sense that $\beta_i = \delta/\sqrt{n}$ with $\delta > \kappa_\epsilon$, and still satisfy assumption (iv). A boundary case occurs when $\boldsymbol{\Sigma}_\epsilon$ is proportional to the identity matrix which requires $\delta > 1$.

Assumption 1 is sufficient to rank the granular series higher than the non-granular ones when ordering series on the basis of the column norms of the concentration matrix of y_t.

Lemma 1. *Let y_t be generated by model (2). Under Assumption 1(i)–(iv) we have that $\mathbf{K}$ exists and for $n > N$ we have that*

$$\|\mathbf{K}_i\| > \|\mathbf{K}_j\| \qquad \text{for all} \quad i = 1, \ldots, k, \quad \text{and} \quad j = k+1, \ldots, n.$$

All proofs of this section are collected in the Appendix A.

In order to select the number of granular time series in the panel we use a strategy that is inspired by the eigenvalue ratio criterion proposed by Ahn and Horenstein (2013) for the selection of the number of factors. Let $\mathbf{K}_{(s)}$ denote the s–th largest column of the concentration matrix.[11] Consider the ratio between two subsequent ordered column norms, that is

$$\|\mathbf{K}_{(s)}\| \;/\; \|\mathbf{K}_{(s+1)}\|\,, \tag{5}$$

for $s = 1, \ldots, n-1$. Heuristically, the column norms are large for granular series and small otherwise. Thus, the ratio ought to be largest when comparing the last column norm corresponding to the granular series with the first column norm corresponding to the non-granular series. This suggests that the sequential column norm ratio is maximized when s is equal to k.[12]

We need to strengthen Assumption 1-(iv) to identify the number of granulars using the sequential column norm ratio.

(iv*) Let $\mathbf{D}_\beta = \boldsymbol{\beta}'\boldsymbol{\beta}$, then there exists an integer N and a constant M_g such that for each $n > N$ we have $\kappa_\beta^2 \kappa_\epsilon \left(\kappa_\epsilon + \frac{\mu_1(\boldsymbol{\Sigma}_\epsilon)}{\mu_k(\boldsymbol{\Sigma}_g)}\right) < \sigma_k(\mathbf{D}_\beta) \leqslant \sigma_1(\mathbf{D}_\beta) < M_g$ where κ_ϵ and κ_β are the condition numbers of the matrices $\boldsymbol{\Sigma}_\epsilon$ and $\mathbf{D}_\beta$ respectively.

We emphasize that in practical situations the smallest eigenvalue of the granular variance $\mu_k(\boldsymbol{\Sigma}_g)$ is likely to be larger than the smallest eigenvalue of the non-granular variance $\mu_{n-k}(\boldsymbol{\Sigma}_\epsilon)$ in which case the bound would be $\sigma_k(\mathbf{D}_\beta) > 2\kappa_\beta^2 \kappa_\epsilon^2$. The interpretation of the bound is the similar as for Assumption 1-(iv). Given the stronger condition on the loading matrix we obtain the following lemma.

Lemma 2. *Let y_t be generated by model (2) under Assumption 1(i)-(iii) and (iv*). Then we have for $n > N$, when $k > 0$ that*

$$k = \underset{s=1,\ldots,n-1}{\arg\max}\; \|\mathbf{K}_{(s)}\| \;/\; \|\mathbf{K}_{(s+1)}\|.$$

Jointly Lemmas 1 and 2 are sufficient for the identification of the set of granulars.

We briefly compare our assumptions and identification results to the factor model literature. Two main differences can be noted in our setup. First, assumptions (iv)/(iv*) reflect that the granular loadings are not orthogonal to each other. Second, it is important to stress that assumptions (iv)/(iv*) are satisfied when we would impose the stronger assumption that the loadings average out proportional to n and that the largest eigenvalue of $\boldsymbol{\Sigma}_\epsilon$ is bounded, e.g. $n^{-\alpha}\boldsymbol{\beta}'\boldsymbol{\beta} \to \mathbf{D}_\beta$ for $\alpha > 0$ and $\mu_1(\boldsymbol{\Sigma}_\epsilon) < \infty$.[13] Instead, we impose lower bounds on the norms of the columns of the granular loading

[11] Columns are ordered on the basis of their norms.

[12] We point out that the column norm ratio in Eq. (5) is not the only function of the concentration matrix $\mathbf{K}$ that identifies k and other criteria may be used.

[13] Such assumptions are common in the factor model literature, see for example Bai and Ng (2002), Bai (2003), and Ahn and Horenstein (2013).

matrix that are sufficient for detection. It is important to stress that in our framework granular series do not necessarily maximize the explained variance in the panel. The granular series are merely those that have a non-vanishing influence in the cross-section. Last, we point out that assumptions (iv)/(iv*) are comparable to weak factor assumptions considered in Onatski (2009, 2010, 2012), Pesaran (2012) and Chudik et al. (2011).

3.2. Granular panel model with common factors

We now consider the granular panel data model (2) with common factors, that is

$$\begin{aligned} y_{1:k,t} &= \mathbf{\Lambda}_1 f_t + g_t \\ y_{k+1:n,t} &= \mathbf{\Lambda}_2 f_t + \boldsymbol{\beta} g_t + \epsilon_t, \end{aligned} \tag{6}$$

where all components are defined as in (1).

Without additional restrictions model (6) is not identified. The problem is similar when compared to the factor-augmented vector autoregressive model, see Bernanke et al. (2005) and Bai et al. (2016). In total we require $r^2 + rk$ additional identification restrictions (spelled out below) to pin down a rotation of the common factors.[14]

The identification results of this section rely on two key features of the granular model. First, they depend on the strength of the signal of the granular shocks g_t for the granular series $y_{1:k,t}$. To this extent, we define the following two signal-to-noise ratios

$$s_g \equiv \frac{\mu_k(\mathbf{\Sigma}_g)}{\mu_1(\mathbf{\Sigma}_g + \mathbf{\Lambda}_1\mathbf{\Lambda}_1')} \qquad \text{and} \qquad s_{\Lambda_1} \equiv \frac{\|\mathbf{\Lambda}_1\|}{\mu_k(\mathbf{\Sigma}_g + \mathbf{\Lambda}_1\mathbf{\Lambda}_1')}. \tag{7}$$

Under our normalization assumptions, the ratio s_g is a lower bound on the signal of the granular shocks relative to the total variance in the granular series whereas the ratio s_{Λ_1} is an upper bound on the contamination induced by the common factors on the granular series. The signal-to-noise ratio of the granular shocks s_g is key for granular detection. When this is sufficiently large the granular series can be detected when common factors are present.

Second, our identification results rely on the amount of variation of the non-granular series that can be explained by the variation of the granular series. To this extent, define $\mathbf{L}_1 = [\mathbf{\Sigma}_g^{1/2} \ \mathbf{\Lambda}_1]$ and $\mathbf{L}_2 = [\boldsymbol{\beta}\mathbf{\Sigma}_g^{1/2} \ \mathbf{\Lambda}_2]$, which determine the variation in the granular series $y_{1:k,t}$ and the non-granular series $y_{k+1:n,t}$, respectively.[15] We decompose $\mathbf{L}_2$ in two orthogonal parts

$$\mathbf{L}_2 = \hat{\boldsymbol{\gamma}}\mathbf{L}_1 + \hat{\mathbf{U}}, \tag{9}$$

where $\hat{\boldsymbol{\gamma}}\mathbf{L}_1$ is the part that can be explained by $\mathbf{L}_1$ and $\hat{\mathbf{U}} = \mathbf{L}_2\mathbf{M}_{L_1}$ is the residual, with $\mathbf{M}_{L_1} = \mathbf{I}_{k+r} - \mathbf{L}_1'(\mathbf{L}_1\mathbf{L}_1')^{-1}\mathbf{L}_1$. Further, $\hat{\boldsymbol{\gamma}} = \mathbf{L}_2\mathbf{L}_1'(\mathbf{L}_1\mathbf{L}_1')^{-1}$ is the $(n-k)\times k$ projection coefficient which is interpretable as the effective granular loading matrix after taking into account that the measurement of the granular shocks is contaminated by $\mathbf{\Lambda}_1$. Indeed we can easily verify that if $\mathbf{\Lambda}_1 = 0$ we have that $\hat{\boldsymbol{\gamma}} = \boldsymbol{\beta}$. Intuitively, when $\hat{\mathbf{U}}$ is small this implies that the loadings of the granular series ($\mathbf{L}_1$) point in the same direction as the loadings of the non-granular series ($\mathbf{L}_2$) and the granular series remain straightforward to identify.

The following set of assumptions formalizes these insights.

Assumption 2.

(i) $\mathrm{E}(f_t) = 0$, $\mathrm{E}(f_t f_t') = \mathbf{I}_r$, $\mathrm{E}(f_t g_t') = 0$ and $\mathbf{D}_\lambda = \mathbf{\Lambda}_2'\mathbf{\Lambda}_2$ with $\mathbf{D}_\lambda$ diagonal.

(ii) $\mathrm{E}(f_t\epsilon_{i,t}) = 0$ for all i, t.

(iii) There exists an N and a constant M_f such that for each $n > N$ we have that $2\|\boldsymbol{\beta}'\mathbf{\Lambda}_2\|s_g s_{\Lambda_1}^{-1} < \mu_r(\mathbf{D}_\lambda) \leqslant \mu_1(\mathbf{D}_\lambda) < M_f$

(iv) Let $\mathbf{D}_\beta = \boldsymbol{\beta}'\boldsymbol{\beta}$ there exists an N and constant a constant M_g such that for each $n > N$ we have that $\kappa_{u,\epsilon}\kappa_{\hat{\boldsymbol{\gamma}}}s_g^{-1} < \sigma_k(\mathbf{D}_\beta) \leqslant \sigma_1(\mathbf{D}_\beta) < M_g$ where $\kappa_{u,\epsilon}$ and $\kappa_{\hat{\boldsymbol{\gamma}}}$ are the condition numbers of the matrices $\hat{\mathbf{U}}\hat{\mathbf{U}}' + \mathbf{\Sigma}_\epsilon$ and $\hat{\boldsymbol{\gamma}}'\hat{\boldsymbol{\gamma}}$ respectively.

---

[14] This follows when we substitute $g_t = y_{1:k,t} - \mathbf{\Lambda}_1 f_t$ in the model for $y_{k+1:n,t}$ to obtain

$$y_{k+1:n,t} = \tilde{\mathbf{\Lambda}}_2 f_t + \boldsymbol{\beta} y_{1:k,t} + \epsilon_t,$$

where $\tilde{\mathbf{\Lambda}}_2 = \mathbf{\Lambda}_2 - \boldsymbol{\beta}\mathbf{\Lambda}_1$. The model for $y_{k+1:n,t}$ is equivalent to a FAVAR model and we may apply proposition 1 of Bai et al. (2016) to find that we need $r^2 + rk$ additional restrictions.

[15] To see this consider the alternative factor representation of the model

$$\begin{aligned} y_{1:k,t} &= \mathbf{L}_1 h_t \\ y_{k+1:n,t} &= \mathbf{L}_2 h_t + \epsilon_t, \end{aligned} \tag{8}$$

where $h_t = (\tilde{g}_t', f_t')'$ is the $(r+k)\times 1$ vector of standardized granular shocks $\mathbf{\Sigma}_g^{-1/2}g_t$ and factor shocks, $\mathbf{L}_1 = [\mathbf{\Sigma}_g^{1/2} \ \mathbf{\Lambda}_1]$ is the $k\times(r+k)$ matrix of granular and factor loadings on the granular series and $\mathbf{L}_2 = [\boldsymbol{\beta}\mathbf{\Sigma}_g^{1/2} \ \mathbf{\Lambda}_2]$ is the $(n-k)\times(r+k)$ matrix of granular and factor loadings on the non-granular series.

Assumption (i) imposes the identification restrictions that identify the common factors. We get $r(r+1)/2$ restrictions by imposing $\mathbf{\Sigma}_f = \mathbf{I}_r$, $r(r-1)/2$ restrictions from $\mathbf{D}_\lambda$ diagonal and rk restrictions by imposing $\mathrm{E}(f_t g_t') = 0$. This yields a total of $r^2 + rk$ restrictions which is sufficient, see Bai et al. (2016, Proposition 1). Notice that these restrictions do not identify additional granular series nor do they trivialize in any way the identification of the granular shocks g_t. Assumption (ii) imposes that the factors, similar as the granular shocks, are uncorrelated with the non-granular shocks. Assumption (iii) imposes that the correlation between the granular loadings and the common factor loadings $\|\boldsymbol{\beta}'\mathbf{\Lambda}_2\|$ cannot be larger then the common factor variance $\mu_r(\mathbf{D}_\lambda)$.[16] Assumption (iv) is the key assumption for granular identification in this setting. It imposes a lower bound on the smallest singular value of the granular loadings $\boldsymbol{\beta}'\boldsymbol{\beta}$. The bound is similar to the bound imposed for the model without common factors – Assumption 1-(iv) – with the difference that we need to account for the signal-to-noise ratio of the granular shocks s_g, the condition number of $\boldsymbol{\beta}$ is replaced by that of the effective granular loadings $\hat{\boldsymbol{\gamma}}$ and finally $\mathbf{\Sigma}_\epsilon$ is replaced by $\hat{\mathbf{U}}\hat{\mathbf{U}}' + \mathbf{\Sigma}_\epsilon$.

We provide lemmas that extend the identification results of the previous section.

Lemma 3. *Let y_t be generated by model (6) under Assumptions 1(i)–(iii) and 2(i)–(iv). Then $\mathbf{K} = \mathbf{\Sigma}^{-1}$, where $\mathbf{\Sigma} = \mathrm{Var}(y_t)$, exists and we have for $n > N$ that*

$$\|\mathbf{K}_i\| > \|\mathbf{K}_j\| \qquad \text{for all} \quad i = 1, \dots, k, \quad \text{and} \quad j = k+1, \dots, n\,.$$

Analogously to the analysis of the previous section, we require a stronger version of Assumption 2(iv) for the selection of the number of granulars.

(iv*) Let $\mathbf{D}_\beta = \boldsymbol{\beta}'\boldsymbol{\beta}$ there exists an N and constant a constant M_g such that for each $n > N$ we have that $\kappa_{u,\epsilon}\kappa^2_{\hat{\boldsymbol{\gamma}}}s_g^{-1}\left(\kappa_{u,\epsilon} + \frac{\mu_1(\hat{\mathbf{U}}\hat{\mathbf{U}}'+\mathbf{\Sigma}_\epsilon)}{\mu_k(\mathbf{\Sigma}_g+\mathbf{\Lambda}_1\mathbf{\Lambda}_1')}\right) < \sigma_k(\mathbf{D}_\beta) \leqslant \sigma_1(\mathbf{D}_\beta) < M_g$, where $\kappa_{u,\epsilon}$ and $\kappa_{\hat{\boldsymbol{\gamma}}}$ are the condition numbers of the matrices $\hat{\mathbf{U}}\hat{\mathbf{U}}' + \mathbf{\Sigma}_\epsilon$ and $\hat{\boldsymbol{\gamma}}'\hat{\boldsymbol{\gamma}}$ respectively.

Let us point out that, analogously to the case where no common factors are present, the order of magnitude of the lower bound in assumption (iv*) is the square of the lower bound required by Assumption 2-(iv). The interpretation of the bound remains the same as discussed above.

We establish the identification of the number of granular series in the following lemma.

Lemma 4. *Let y_t be generated by model (6) under Assumptions 1(i)–(iii) and 2(i)–(iii) and (iv*). Then we have for $n > N$ and $k > 0$ that*

$$k = \underset{s=1,\dots,n-1}{\arg\max}\ \|\mathbf{K}_{(s)}\| \ / \ \|\mathbf{K}_{(s+1)}\|\,.$$

The results of this section show that granular detection is more challenging when unobserved common factors are present. The identification conditions that we impose are more stringent. It is important to emphasize that, for a given variance of the granular shocks the granular series can be identified only if the factors do not load excessively on the granular series relative to their influence on the non-granular series. Therefore, in practice, it is important to question whether the common factors may have distinctively different influences on the granular and non-granular series. For these cases granular detection via the column norms is compromised.

Provided that our identification assumptions hold, the presence of common factors does not alter our detection strategy. In particular, it is not necessary to know the number of factors r to carry out granular detection. Last, note that the identification results of this section can also be used to establish identification for the case where the granulars are contaminated by measurement error. In fact, if we think of the factors as measurement error and we restrict $\mathbf{\Lambda}_2 = 0$, then the results of this section immediately establish identification conditions for this case. The dependence of Assumption 2-(iv) on the signal-to-noise ratio s_g then shows that the measurement error, as caused by $\mathbf{\Lambda}_1$, cannot be too large.

3.3. Estimation

A natural estimator of the norm of ith column of the concentration matrix $\|\mathbf{K}_i\|$ is its sample analog, that is $\|\hat{\mathbf{K}}_i\|$ where $\hat{\mathbf{K}} = \hat{\mathbf{\Sigma}}^{-1}$ with $\hat{\mathbf{\Sigma}} = T^{-1}\sum_{t=1}^T y_t y_t'$. We need to impose appropriate dependence and distributional assumptions on y_t in order to establish the large sample properties of this estimator. Notice that in what follows we formulate assumptions on the sequence of isotropic random vectors $\mathbf{\Sigma}^{-1/2}y_t$.[17] Let $\mathcal{F}^0_{-\infty}$ and $\mathcal{F}^\infty_m$ denote the σ-algebras generated by $\{\mathbf{\Sigma}^{-1/2}y_m : -\infty \leqslant m \leqslant 0\}$ and $\{\mathbf{\Sigma}^{-1/2}y_m : t \leqslant m \leqslant \infty\}$, respectively. We define the α-mixing coefficients of the $\mathbf{\Sigma}^{-1/2}y_t$ process as

$$\alpha(m) = \sup_{A\in\mathcal{F}^0_{-\infty}, B\in\mathcal{F}^\infty_m} |P(A)P(B) - P(AB)|\,.$$

[16] In the proof of Lemma 3 we show that this assumption can be dropped at the expense of a stronger condition on the granular loadings.

[17] Notice that $\mathbf{\Sigma}^{-1/2}y_t$ is a sequence of isotropic random vectors as $\mathrm{E}(\mathbf{\Sigma}^{-1/2}y_t y_t'\mathbf{\Sigma}^{-1/2'}) = \mathbf{I}$.

We make the following assumptions.

Assumption 3. Let y_t be an n-dimensional time series process.

(i) $\{\mathbf{\Sigma}^{-1/2}y_t\}$ is a stationary and ergodic α-mixing process. There exists positive constants γ_1 and C_1 such that for all positive integers m we have that the α-mixing coefficients satisfy $\alpha(m) \leqslant \exp(-C_1 m^{\gamma_1})$.

(ii) There exists positive constants γ_2 and C_2 such that for any $s > 0$ and any vector x with $\|x\| = 1$

$$\sup_t \Pr(|x'\mathbf{\Sigma}^{-1/2}y_t| > s) \leqslant \exp(1 - (s/C_2)^{\gamma_2}) .$$

(iii) Let γ be defined as $\gamma^{-1} = \gamma_1^{-1} + 2\gamma_2^{-1}$. Then, $\gamma < 1$.

The conditions in Assumption 3 are analogous to the ones used in Fan et al. (2011). For convenience we impose assumptions on y_t directly instead of on f_t, g_t and ϵ_t. In particular, (i) states that $\mathbf{\Sigma}^{-1/2}y_t$ is strongly mixing and assumption (ii) states that the isotropic random vectors $\mathbf{\Sigma}^{-1/2}y_t$ have generalized-exponential tails. This condition is satisfied by the multivariate Gaussian but also allows for heavier tails. The parameter γ in assumption (iii) measures the degree of dependence and tail thickness of the data: The smaller the parameter the more dependent and thick tailed the data are.

We establish the following result concerning the sample covariance matrix.

Lemma 5. *Let y_t be generated by model (6) under Assumptions 1–3. Then, for any $\eta > 0$ there exists positive constants $C_1, \ldots, C_5$ such that for n sufficiently large and $T = O(n^{2/\gamma - 1})$ we have*

(i) $\mu_n(\mathbf{\Sigma}) - C_1\sqrt{\frac{n}{T}} \leqslant \mu_n(\hat{\mathbf{\Sigma}}) \leqslant \mu_1(\hat{\mathbf{\Sigma}}) \leqslant \mu_1(\mathbf{\Sigma}) + C_2\sqrt{\frac{n}{T}}$

(ii) $\left\|\hat{\mathbf{\Sigma}} - \mathbf{\Sigma}\right\| \leqslant C_3\sqrt{\frac{n}{T}}$

(iii) $\left\|\hat{\mathbf{K}} - \mathbf{K}\right\| \leqslant C_4\sqrt{\frac{n}{T}}$

(iv) $\|\hat{\mathbf{K}}_i\| - \|\mathbf{K}_i\| \leqslant C_5\sqrt{\frac{n}{T}}$

at least with probability $1 - O(n^{-\eta})$.

All proofs of this section are collected in the web-Appendix A.

Lemma 5(*i*) states that the eigenvalues of the sample covariance matrix are bounded away from zero and infinity when n and T are large. The proof follows the arguments laid out in Vershynin (2012). Let us emphasize that the Lemma extends the results of Vershynin (2012) under a set of assumptions that are more convenient for economic and financial applications. Parts (*ii*) to (*iv*) then follow from (*i*).

We use Lemma 5 to establish the selection properties of the granular statistic $\|\hat{\mathbf{K}}_i\|$. In light of Lemma 3, it is natural to rank the series on the basis of $\|\hat{\mathbf{K}}_i\|$. Define

$$\mathcal{E}_{\mathsf{R}} = \left\{ \|\hat{\mathbf{K}}_i\| > \|\hat{\mathbf{K}}_j\| \;\text{ for all }\; i = 1, \ldots, k \text{ and } j = k+1, \ldots, n \right\}, \tag{10}$$

that is the event that the granular statistics of the granular series are larger than the ones of the non-granular series. Then, we establish the ranking consistency of our procedure.

Corollary 1. *Let y_t be generated by model (6) under Assumptions 1–3. Consider the event $\mathcal{E}_{\mathsf{R}}$ defined in Eq. (10). Then, for any $\eta > 0$, n sufficiently large and $T = O(n^{2/\gamma - 1})$ we have $\mathrm{P}(\mathcal{E}_{\mathsf{R}}) \geqslant 1 - O(n^{-\eta})$.*

The corollary shows that $\|\hat{\mathbf{K}}_i\|$ consistently ranks the granular series ahead of the non-granulars. Next, in light of Lemma 4, we estimate the number of granulars by

$$\hat{k} = \arg \max_{s=1,\ldots,n-1} \|\hat{\mathbf{K}}_{(s)}\| / \|\hat{\mathbf{K}}_{(s+1)}\| , \tag{11}$$

where $\hat{\mathbf{K}}_{(s)}$ denotes the s–th sample concentration matrix column when the columns are ordered on the basis of their norm in decreasing order. Define the event

$$\mathcal{E}_{\mathsf{S}} = \left\{ \hat{k} = k \right\}, \tag{12}$$

that is the event that the correct numbers of granular series are selected. The following corollary establishes that the probability of the event $\mathcal{E}_{\mathsf{S}}$ approaches one.

Corollary 2. *Let y_t be generated by model (6) under Assumptions 1, 2(i)-(iii) and (iv*), and 3. Consider the event $\mathcal{E}_{\mathsf{S}}$ defined in Eq. (12). Then, for any $\eta > 0$, n sufficiently large and $T = O(n^{2/\gamma - 1})$ we have $\mathrm{P}(\mathcal{E}_{\mathsf{S}}) \geqslant 1 - O(n^{-\eta})$.*

It is important to emphasize that Corollary 2 allows to consistently select k provided that k is larger than zero. We do not address the important problem of choosing between $k > 0$ and $k = 0$, which is more challenging. Drawing an analogy with the problem of the selection of the number of factors in factor models, we note that papers such as

Ahn and Horenstein (2013) have introduced criteria to select between zero or more than one factors. These criteria however are typically based on additional assumptions on the model. In our setting this would require the granular loadings to increase proportional to n.[18] For the sample concentration matrix $\hat{\mathbf{K}}$ to be well defined we require $n \leqslant T$. This can possibly be circumvented by employing some regularized estimator for $\mathbf{K}$ (see Fan et al. (2016)). However, regularized estimators would typically require to make additional assumptions on the model specification (for instance, sparsity assumptions). Further, it is unclear what the properties of such estimators are in the presence of an unknown number of common factors. We leave this extension for further research. Finally, it is important to highlight that the results of this section can also be obtained by making assumptions similar to those in Stock and Watson (2002a), Bai and Ng (2002) and Doz et al. (2012). Such results rely on weaker distributional assumptions than the ones spelled out in Assumption 3. However they rely on stronger dependence assumptions.

4. Comparison to other methods

In this section we compare our detection methodology with principal components and maximum likelihood based methods. Note that these alternatives do not exploit the partial correlation structure of the model, but given that the granular panel model can be thought of as a factor model with observed factors, it is natural to consider them as alternatives.

4.1. Principal components based methods

Stock and Watson (2002b), Bai and Ng (2006) and Parker and Sul (2016) propose methods based on principal components to give meaning to the otherwise hard to interpret estimated common factors. They estimate factors in a factor model using principal components and use regression analysis to find the series that correlate most with the estimated factors.

There are important cases in which principal components based methods are not able to detect the granulars. First, we note that these methods perform well when the granulars explain a large portion of the variance. In contrast, when the granular shocks have low signal-to-noise ratios principal components methods have low detection power. Consider the baseline granular model (2) with $k = 1$ and the following parametrization

$$\begin{array}{rclcrcl} y_{1,t} & = & g_t & & \mathrm{Var}(g_t) & = & 1 \\ y_{2:n,t} & = & \frac{\delta}{\sqrt{n-1}}\iota_{n-1}g_t + \epsilon_t & & \mathrm{Var}(\epsilon_t) & = & \mathbf{I}_{n-1}c, \end{array}$$

where ι_n is an $n \times 1$ vector of ones and δ and c are constants. For this model the condition number κ_ϵ is 1 and the identification Assumption 1-(iv) is satisfied for $\delta > 1$. In contrast, principal components methods fail when c is large as the first principal component does not correlate with the granular series.[19] Moreover, estimators for the number of factors (as in Bai and Ng (2002) and Ahn and Horenstein (2013)) will not detect any factors. This problem becomes more prominent when common factors f_t are present in the model. When these explain a large portion of the variance, principal components methods primarily detect series that correlate with the factors. Notice, however, that common factors are also a challenge for our column norm method. In particular, Assumption 2-(iv) implies that in order to successfully recover the granulars our detection strategy requires the signal-to-noise ratio of the granular shocks in the granular series to be sufficiently large.[20]

Recovering the number of granular series using principal components based methods can also be challenging. To see this consider again the example above for the case where c is small relative to δ. Clearly for large n and T the R^2 from the regression of the first principal component on the first series tends to one. However, for small c the R^2's from the other regressions also become arbitrarily close to one. This makes selecting the number of granular series challenging, as the same principal component can be generated by multiple correlated granular series. More generally, principal components based methods cannot distinguish between (i) a model with one granular series that implies one large eigenvalue in $\mathbf{\Sigma}$ and (ii) a model with $k > 1$ correlated granular series that also implies one large eigenvalue in $\mathbf{\Sigma}$. On the other hand, the column norm method is able to distinguish between these two scenarios.

These observations are verified in the Monte Carlo study in the next section, where we confirm that the detection power of the principal components based methods does not perform satisfactorily (i) under a weak factor settings and (ii) when additional factors are present. Also, in the empirical section we show that the documented rankings of granular series can be vastly different for principal components and column norm methods.

[18] Without allowing the loadings $\boldsymbol{\beta}$ to increase proportional to n we can only distinguish between $k > 0$ and $k = 0$ using hypothesis tests. These we aim to develop in future work.

[19] We refer to Onatski (2012) for a detailed discussion regarding the behavior of the principal components estimates in the weak factor setting.

[20] Note, however, that when Assumption 2 does not hold principal component methods are also not guaranteed to work, see Onatski (2012).

4.2. Likelihood based methods

Doz et al. (2012), Bai and Li (2012, 2016) and Jungbacker and Koopman (2015) show that, if the granular series are known, the parameters of the granular model (6) can be estimated consistently under mild assumptions using the maximum likelihood method. Hence, a granular detection strategy could be based on estimating the granular model for different orderings of the variables in y_t and then comparing an appropriate goodness-of-fit statistic across the different orderings. The challenge here lies in the fact that this would involve estimating $\binom{n}{k}$ possible models, which is only feasible for small values of n and k. Also, different orderings for different k may lead to equivalent goodness-of-fit statistics.

Alternatively, it is possible to think about granular detection as testing for zero measurement error. This approach is explored for small scale models in Kolenikov and Bollen (2012) who build on earlier work of Heywood (1931).[21] To outline the difficulties with this approach for large panels consider the factor representation of the granular panel model

$$y_t = \mathbf{L}h_t + \zeta_t$$

where $h_t = (f_t', g_t')'$. A likelihood based detection method proceeds by first estimating $\mathbf{L}$ and the variance matrix of ζ_t, say $\mathbf{\Sigma}_\zeta$, and in the second step testing whether the diagonal elements of $\mathbf{\Sigma}_\zeta$ are zero. The main technical difficulty of such approach is that one needs to solve a multiple hypothesis testing problem on the boundary of the parameter space. In particular, under the null of zero measurement error the assumptions for consistency in Doz et al. (2012) and Bai and Li (2012, 2016) are violated and the limiting distribution is unknown.

5. Simulation study

We perform a simulation study to assess the finite sample performance of our methodology. We evaluate the performance of the detection methods based on the granular statistic $\|\hat{\mathbf{K}}_i\|$ under different data generating processes. We first evaluate the fraction of true granular series that correspond to the k largest granular statistics and then we consider the frequency by which we correctly select the number of granular series. Finally, we compare our procedure against methods based on principal components.

We generate data from the granular panel data model with common factors of Eq. (6). We consider panel dimensions $n = 50, 100$ and $T = 200, 400$. The number of granular series that is equal to $k = 3, 5$ and the number of common factors is equal to $r = 0, 5$. The granular shocks and common factors follow the vector autoregressive process

$$\begin{bmatrix} f_t \\ g_t \end{bmatrix} = \begin{bmatrix} \mathbf{\Phi}_{ff} & \mathbf{\Phi}_{fg} \\ \mathbf{\Phi}_{gf} & \mathbf{\Phi}_{gg} \end{bmatrix} \begin{bmatrix} f_{t-1} \\ g_{t-1} \end{bmatrix} + \begin{bmatrix} \eta_{f,t} \\ \eta_{g,t} \end{bmatrix}, \tag{13}$$

where the variance matrix $\mathbf{\Sigma}_\eta = \mathrm{Var}(\eta_{f,t}, \eta_{g,t})$ has ones on the main diagonal and correlation coefficient 0.5 on the off-diagonal elements. Note that in this setting we allow for correlation between granulars and common factors. The elements for the diagonal of $\mathbf{\Phi} = [\mathbf{\Phi}_{ff}, \mathbf{\Phi}_{fg} : \mathbf{\Phi}_{gf}, \mathbf{\Phi}_{gg}]$ are drawn uniformly for each panel over the range (0.5,0.95). The off-diagonal elements are drawn from $N(0, 0.1)$. The transformations of Ansley and Kohn (1986) are applied to ensure that (13) admits a stationary vector autoregressive process. We generate the non-granular idiosyncratic shocks from

$$e_t = \mathbf{\Gamma} e_{t-1} + \eta_{e,t} \qquad \eta_{e,t} \sim NID\left(0, \mathbf{I}_{n-k} - \mathbf{\Gamma}\mathbf{\Gamma}'\right),$$

where $\mathbf{\Gamma}$ is diagonal with elements $\mathbf{\Gamma}_{ii} \sim U(0.5, 0.95)$ and $U(a, b)$ indicates the uniform distribution over the range (a, b). The specification ensures that e_t follows a stationary vector autoregressive process with variance $\mathbf{I}_{n-k}$. From this we generate $\epsilon_t = \mathbf{\Sigma}_\epsilon^{1/2} e_t$ so that $\mathrm{Var}(\epsilon_t) = \mathbf{\Sigma}_\epsilon$. We assume that $\mathbf{\Sigma}_\epsilon$ has a banded structure similar to the one considered in Stock and Watson (2002a) and Bai and Ng (2002). In particular, we have that $\mathbf{\Sigma}_{\epsilon,i,j} \sim U(0.5, 1.5)$ if $i = j$, $\mathbf{\Sigma}_{\epsilon,i,j}^{1/2} = c_\epsilon$ with $c_\epsilon = 0.2$ if $i > j$ and $i - j < 10$ and zero otherwise. Finally, the loadings of the granular socks are drawn from a normal with mean zero and variance σ_b^2, where $\sigma_b^2 = 0.01, 0.05, 0.1, 0.25, 0.5, 0.75, 1$. The loadings of the common factors are drawn from a standard normal distribution. Note that we vary the variance of the granular loadings in order to change the magnitude of their effect and that small values of σ_b^2 reflect the scenario where the common factors explain more variance in the observations than the granular shocks. Such settings are argued to be empirically relevant by Foerster et al. (2011).

Overall, we vary the granular panel model in five dimensions: (i) panel dimensions, (ii) number of granulars, (iii) number of factors, (iv) strength of the granulars, (v) correlation among the granulars and factors.[22] For each possible combination we draw 1000 data panels. For each simulated panel we rank the series according to the column norms of the concentration matrix and we then select the number of granulars using the column ratio statistic given in Eq. (11). When selecting the number of granulars, we set the maximum number of possible granular series to $n/2$, see also Ahn and Horenstein (2013).

[21] Note that the context of these works is completely different to the setting in which we work.

[22] We point out that in the working paper version of this paper we also considered different specifications for the covariance matrix of the non-granular shocks $\mathbf{\Sigma}_\epsilon$. Results across the different designs we originally considered for $\mathbf{\Sigma}_\epsilon$ are in-line with the results for the banded design reported here.

Table 1
Simulation study.

n	T	k	r	0.01	0.05	0.10	0.25	0.50	0.75	1.00
Granular ranking probabilities: $\|\mathbf{K}_i\|$										
50	200	3	0	0.140	0.800	0.967	1.000	1.000	1.000	1.000
100	200	3	0	0.140	0.926	0.996	1.000	1.000	1.000	1.000
50	400	3	0	0.207	0.941	0.998	1.000	1.000	1.000	1.000
100	400	3	0	0.266	0.992	1.000	1.000	1.000	1.000	1.000
50	200	5	0	0.206	0.813	0.954	0.997	1.000	1.000	1.000
100	200	5	0	0.252	0.932	0.997	1.000	1.000	1.000	1.000
50	400	5	0	0.187	0.905	0.992	1.000	1.000	1.000	1.000
100	400	5	0	0.334	0.991	1.000	1.000	1.000	1.000	1.000
50	200	3	5	0.103	0.676	0.876	0.969	0.983	0.985	0.989
100	200	3	5	0.148	0.865	0.972	0.998	0.998	0.998	0.999
50	400	3	5	0.104	0.800	0.953	0.989	0.992	0.994	0.994
100	400	3	5	0.184	0.967	0.996	0.999	1.000	0.999	1.000
50	200	5	5	0.156	0.721	0.884	0.970	0.988	0.988	0.993
100	200	5	5	0.216	0.878	0.976	0.998	0.999	0.998	0.999
50	400	5	5	0.150	0.822	0.950	0.990	0.995	0.997	0.998
100	400	5	5	0.273	0.966	0.997	1.000	0.999	0.999	1.000
Granular selection probabilities: $\|\mathbf{K}_i\|$										
50	200	3	0	0.098	0.190	0.520	0.883	0.979	0.992	0.994
100	200	3	0	0.104	0.377	0.773	0.969	0.995	0.999	1.000
50	400	3	0	0.093	0.404	0.833	0.988	0.998	1.000	1.000
100	400	3	0	0.090	0.732	0.974	0.999	1.000	1.000	1.000
50	200	5	0	0.044	0.101	0.357	0.832	0.959	0.969	0.986
100	200	5	0	0.048	0.265	0.782	0.981	0.995	0.998	0.999
50	400	5	0	0.042	0.207	0.661	0.963	0.993	0.997	0.999
100	400	5	0	0.047	0.652	0.941	0.998	1.000	1.000	1.000
50	200	3	5	0.117	0.136	0.310	0.482	0.594	0.587	0.615
100	200	3	5	0.115	0.273	0.556	0.771	0.789	0.781	0.795
50	400	3	5	0.098	0.188	0.418	0.656	0.704	0.708	0.723
100	400	3	5	0.135	0.525	0.785	0.882	0.884	0.875	0.878
50	200	5	5	0.036	0.068	0.141	0.428	0.575	0.626	0.641
100	200	5	5	0.042	0.159	0.507	0.784	0.825	0.838	0.852
50	400	5	5	0.044	0.105	0.296	0.608	0.708	0.746	0.759
100	400	5	5	0.047	0.425	0.787	0.902	0.916	0.922	0.917
Granular ranking probabilities: R_i^2 vs. $\|\mathbf{K}_i\|$										
50	200	3	0	0.062	0.582	0.835	0.977	0.993	0.995	0.995
100	200	3	0	0.000	0.561	0.860	0.991	0.997	0.999	0.999
50	400	3	0	0.056	0.701	0.949	0.999	1.000	1.000	1.000
100	400	3	0	0.006	0.855	0.993	1.000	1.000	1.000	1.000
50	200	5	0	0.044	0.636	0.843	0.954	0.977	0.979	0.976
100	200	5	0	0.036	0.655	0.899	0.992	0.997	0.998	0.999
50	400	5	0	0.000	0.618	0.874	0.981	0.992	0.992	0.994
100	400	5	0	0.002	0.754	0.966	0.999	1.000	1.000	1.000
50	200	3	5	0.419	0.331	0.544	0.830	0.935	0.952	0.954
100	200	3	5	0.153	0.218	0.591	0.938	0.991	0.999	0.996
50	400	3	5	0.365	0.299	0.554	0.876	0.963	0.980	0.979
100	400	3	5	0.136	0.201	0.655	0.983	0.999	1.000	1.000
50	200	5	5	0.446	0.347	0.502	0.736	0.859	0.884	0.891
100	200	5	5	0.207	0.250	0.536	0.904	0.983	0.991	0.992
50	400	5	5	0.461	0.312	0.492	0.762	0.899	0.936	0.942
100	400	5	5	0.160	0.242	0.604	0.967	0.998	1.000	0.999

The table reports the results of the simulation exercise for the banded error design. The first panel reports the average proportion of correctly ranked granulars. The second panel reports the average proportion of times the correct number of granulars is selected. The third panel reports the ratio between the average proportion of correctly ranked granulars based on the R^2 statistic and the column norm statistic.

We summarize the performance of the procedure in Table 1 by reporting the average proportion of correctly ranked granular series (first panel) and the proportions of correctly selected number of granulars (second panel). The table reveals some interesting patterns.

First, the key parameter for which the outcomes fluctuate the most is the magnitude of the granular loadings as captured by the standard deviation coefficients σ_b^2. When this variance is close to zero this implies that the granular loadings are close to zero and by result ranking the granulars correctly becomes more challenging. When the variance

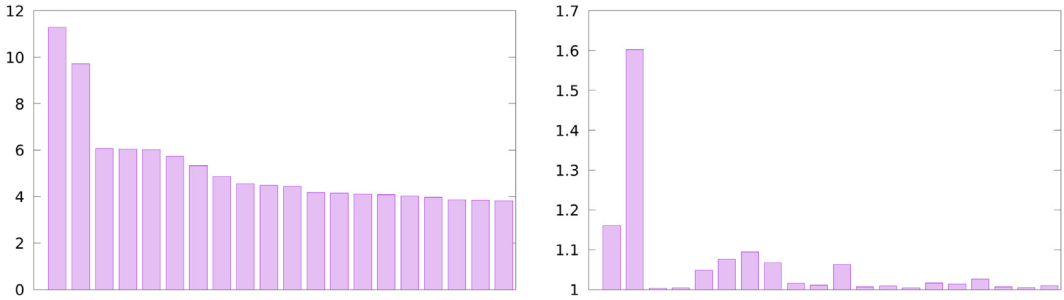

Fig. 1. Granular detection results for the IP series.

increases the percentage of correctly ranked granulars increases rapidly. Notice that when $\sigma_b^2 = 0.1$, which still implies that the coefficients are on average local-to-zero for $n = 100$ the detection rate is close to one for most cases. Hence reasonably connected granular series are expected to be detected easily. For estimating the correct number of granular series a similar pattern is detected. However, as obtaining the correct number of granulars requires a stronger identification condition, see Lemma 2, we see that the percentages are overall lower for this statistic. Second, the panel dimensions imply that larger panels n, T improve the detection results. Both increases in n and T improve the ranking and the estimation of the number of granular series. In line with our theoretical results the increases in accuracy are larger when increasing T when compared to n. Third, differences between $k = 3$ and $k = 5$ granular series are small. Also, the performance of the methodology is only mildly affected by the numbers of factors in the specification. Only, the selection of the number of granular series suffers slightly when including additional common factors. Overall, the results convey that the granular detection methodology performs satisfactorily in finite samples.

Last, we compare the performance of our methods with granular identification procedures that are based on principal components analysis. Here we consider a straightforward implementation of such methods: (i) estimate the number of factors, (ii) regress each time series on the common factors and (iii) rank according to the R^2 of this regression. For estimating the number of factors we use the IC2 criteria from Bai and Ng (2002) which gave slightly better results when compared to the eigenvalue ratio estimator from Ahn and Horenstein (2013). Rankings based on the R^2 are commonly reported in the factor model literature (e.g. Stock and Watson (2002b) and Foerster et al. (2011)) and we emphasize that they are not designed for granular detection but for interpreting the common factors. Nevertheless, it is interesting to compare our methodology to this procedure.

In third panel of Table 1 we show the ratios between the percentage of correctly ranked granulars based on the R^2's and the column norm statistic. We find that in all cases the column norm statistic determines a better ranking when compared to the R^2's. On average – across all specifications – we find that the column norm ranking method performs 25% better. The differences are large for small values of σ_b^2 and tend to zero when the influence of the granular series becomes larger. This is in line with the discussion in Section 4 where we illustrated that the factor based methods become difficult for "weak" granular series. A comparison for the granular selection statistic $\hat{k}$, as specified in (11), is omitted as it is unclear how to choose the number of granular series based on the principal components method. Finally, we comment that more refined principal components based procedures, such as Parker and Sul (2016), did not lead to different results.

6. Granular series in US industrial production

We study the presence of granular series in US industrial production, see also Forni and Reichlin (1998), Foerster et al. (2011), Pesaran and Yang (2016), Siavash (2016) and Atalay (2017). We consider a panel of sector specific industrial production monthly growth rates from 1972 until 2007.[23] The panel covers $n = 138$ sectors for a total of $T = 431$ periods.[24] We standardize each series to have mean zero and unit variance. A preliminary factor analysis conveys evidence of factors in the panel. Using the criteria of Bai and Ng (2002) (IC2), Onatski (2010) and Ahn and Horenstein (2013) we find evidence of one common factor. This is in line with Foerster et al. (2011) who find one or two factors for a similar panel.

In the left panel of Fig. 1 we show the ordered column norms $\|\hat{\mathbf{K}}_i\|$ of the concentration matrix. We find that there are two series that are clearly distinct from the others: "Motor Vehicle Parts" and "Automobiles and Light Duty Motor Vehicles". Both sectors fall within the transportation, or automobile industry which was signaled as a potentially granular industry during the financial crises by Alan R. Mulally, the chief executive of Ford, see Mulally (2008) and the discussion in Acemoglu et al. (2012). The importance of the automobile industry is further confirmed by a more detailed inspection

[23] The data is taken from Mark Watson's website: https://www.princeton.edu/~mwatson/.

[24] We point out that rather than using a quarterly panel as in Foerster et al. (2011) we employ a monthly panel in order to take advantage of a large sample size.

Table 2
Granular series for the US industrial production.

Sample	$\|\hat{\mathbf{K}}_i\|$ rankings			R^2 rankings	
	Sector	$\|\hat{\mathbf{K}}_{(i)}\|$	$\frac{\|\hat{\mathbf{K}}_{(i)}\|}{\|\hat{\mathbf{K}}_{(i+1)}\|}$	Sector	R^2
1972–2007	Motor Vehicle Parts	11.284	1.161	Plastics Products	0.651
	Automobiles and Light Duty Motor Vehicles	9.718	1.603	Household and Institutional Furniture and Kitchen Cabinets	0.520
	Aluminum Extruded Products	6.062	1.004	Metal Valves Except Ball and Roller Bearings	0.462
	Plastics Products	6.039	1.004	Architectural and Structural Metal Products	0.448
	Miscellaneous Aluminum Materials	6.013	1.049	Other Miscellaneous Manufacturing	0.441
	Motor Vehicle Bodies	5.732	1.076	Fabricated Metals Spring and Wire Products	0.422
	Paper and Paperboard Mills	5.328	1.095	Commercial and Service Industry Machines	0.405
	Household and Institutional Furniture and Kitchen Cabinets	4.867	1.067	Fabricated Metals Forging and Stamping	0.402
	Commercial and Service Industry Machines	4.561	1.016	Coating Engraving Heat Treating and Allied Activities	0.355
	Motor Homes	4.490	1.011	Other Textile Product Mills	0.332
1972–1983	Motor Vehicle Parts	4546.027	1.150	Plastics Products	0.733
	Household and Institutional Furniture and Kitchen Cabinets	3953.440	1.340	Household and Institutional Furniture and Kitchen Cabinets	0.634
	Plastics Products	2949.915	1.083	Metal Valves Except Ball and Roller Bearings	0.626
	Commercial and Service Industry Machines	2722.886	1.069	Architectural and Structural Metal Products	0.597
	Automobiles and Light Duty Motor Vehicles	2547.430	1.004	Fabricated Metals Spring and Wire Products	0.588
	Foundries	2538.000	1.057	Other Miscellaneous Manufacturing	0.534
	Organic Chemicals	2401.269	1.170	Other Textile Product Mills	0.486
	Semiconductors and Other Electronic Components	2053.029	1.027	Fabricated Metals Forging and Stamping	0.486
	Farm Machinery and Equipment	1998.120	1.022	Plastics Materials and Resins	0.485
	Animal Slaughtering and Meat Processing Ex Poultry	1954.887	1.036	Communication and Energy Wires and Cables	0.481
1984–2007	Motor Vehicle Parts	26.253	1.039	Plastics Products	0.407
	Automobiles and Light Duty Motor Vehicles	25.261	1.356	Commercial and Service Industry Machines	0.368
	Aluminum Extruded Products	18.626	1.014	Architectural and Structural Metal Products	0.328
	Miscellaneous Aluminum Materials	18.365	1.183	Other Miscellaneous Manufacturing	0.327
	Motor Vehicle Bodies	15.530	1.455	Household and Institutional Furniture and Kitchen Cabinets	0.301
	Truck Trailers	10.675	1.009	Coating Engraving Heat Treating and Allied Activities	0.294
	Carpet and Rug Mills	10.575	1.010	Fabricated Metals Forging and Stamping	0.265
	Paper and Paperboard Mills	10.474	1.064	Metalworking Machinery	0.232
	Motor Homes	9.843	1.034	Metal Valves Except Ball and Roller Bearings	0.217
	Concrete and Products	9.519	1.035	Fabricated Metals Spring and Wire Products	0.215

The table reports the ranking of granular series for US industrial production panel based on the granular statistic and based on the R^2.

of the granular rankings which we report in the top left panel of Table 2. We find that in the top ten there are four series directly related to the automobile industry. Other potentially granular sectors that we find are related to aluminum, plastics and paper products. We emphasize that after the first six or seven sectors the differences in the column norms become small. In the right panel of Fig. 1 we show the column norm ratios. The estimator $\hat{k}$ selects two granulars. Thus, in summary, our granular detection methodology identifies a model with two granular series: "Motor Vehicle Parts" and "Automobiles and Light Duty Motor Vehicles".

Next, we consider the stability of the granular detection method for different sampling periods. In particular, we follow Foerster et al. (2011) and split the sample into two different periods, 1972–1983 and 1984–2007, and repeat the previous analysis. We report the rankings for the two sub-samples in the middle and bottom panels of Table 2. For the 1972–1983 period we still find "Motor Vehicle Parts" as the top granular sector. That said, for this sampling period the top ten granular series displays more heterogeneity and the role of the automobile industry is not as prominent as in the full sample. For the 1984–2007 sampling period we find a similar ranking as for the full sample. In particular, nine of the top ten series are also in the top twenty for the full sample and the top five series are practically unchanged. The automobile industry is even more visible in this sub-sample with half of the top ten granular series being directly related. The estimator for k now indicates that there are five granular series in the model. This is in-line with the finding in Foerster et al. (2011) who find that idiosyncratic shocks have become more important in recent years.

Last, we compare our granular detection method to methods based on principal components. Like in the simulation study we consider a ranking based on the R^2 of the regression of the ith series on the principal components. A similar

ranking is also presented in Foerster et al. (2011) and we follow their construction by using two principal components.[25] In the right panels of Table 2 we show the selected granular series that result from the R^2 ranking (over the full sample as well as the two sub-periods). It is interesting to point out that in this case find a quite different set of granular series and in particular automobile industry related sectors do not show up in the rankings. A possible explanation for this is that the sectors related to the automobile industry do not explain much of the variance in the panel, hence principal components have difficulty detecting these, see the discussion in Section 4.

7. Conclusion

In this work we introduce a panel model in which the idiosyncratic shocks of a subset of time series influence the entire cross-section. We call these series granular in the sense that the influence of such series does not vanish when the system dimension is large. We work under the assumption that the set of granular series is unknown and our objective is to introduce a selection methodology that consistently detects the set of granular series from the data. A key property of the model that we introduce is that the column norms of the concentration matrix of the panel are large for the granular series. This motivates us to introduce a granular detection framework based on the norms of the sample concentration matrix. In particular, we use this statistic to construct indices to rank granulars as well as selecting their number. The large sample properties of the proposed procedures are analyzed and we establish that when the time series and cross-sectional dimensions are sufficiently large our procedure consistently detects the set of granulars. A simulation study is used to show that our proposed procedure performs satisfactorily in finite samples. In our empirical application we analyze granularity in US industrial production. Results show that sectors in the automobile industry are granular.

Acknowledgments

We would like to thank the editor and the referee for insightful comments which helped improve our paper. We would also like to thank Domenico Giannone, Frank Kleibergen, Alexei Onatski, Andrew Patton, Barbara Rossi as well as seminar and conference participants. Special thanks goes to Kevin Sheppard and Majid Al Sadoon for discussing the paper and providing numerous helpful comments. Brownlees and Mesters acknowledge support from the Spanish Ministry of Science and Technology (Brownlees: Grant MTM2012-37195) (Mesters: Grant ECO2015-68136-P), Marie Sklodowska-Curie Actions (Mesters: FP7-PEOPLE-2012-COFUND Action grant agreement no: 600387), the Spanish Ministry of Economy and Competitiveness through the Severo Ochoa Programme for Centres of Excellence in R&D (SEV-2011-0075) and the Analysis of Big Data in Economics and Empirical Applications - 2016 BBVA Foundation Grants for Scientific Research, Spain (PR16_DAT_0043).

Appendix A. Proofs

In this section we provide the proofs for the identification Lemmas 1-4. The proof for the consistency of the concentration matrix estimate is deferred to the web-appendix that accompanies this paper.

We adopt the following notation. For an arbitrary vector $v = (v_1, \ldots, v_n)'$ we have $\|v\| = \sqrt{\sum_{i=1}^n v_i^2}$. For an $N \times N$ matrix $\mathbf{B}$ the kth largest eigenvalue of $\mathbf{B}$ is denoted as $\mu_k(\mathbf{B})$. For an $M \times N$ matrix $\mathbf{A}$ the kth largest singular value of $\mathbf{A}$ is denoted as $\sigma_k(\mathbf{A})$. As a matrix norm we generally adopt the spectral norm is given by $\|\mathbf{A}\|_2 = \sqrt{\mu_1(\mathbf{A}'\mathbf{A})}$. We drop the index when no confusion can arise and write $\|\mathbf{A}\|_2 = \|\mathbf{A}\|$. The Frobenius norm is given by $\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^N \sum_{j=1}^N a_{i,j}^2} = \sqrt{\text{Trace}(\mathbf{A}'\mathbf{A})}$. We have $\|\mathbf{A}\|_2 \leqslant \|\mathbf{A}\|_F \leqslant \text{rank}(\mathbf{A})\|\mathbf{A}\|_2$. For a square matrix $\mathbf{B}$ we let $\mathbf{B} \succ 0$ indicate that $\mathbf{B}$ is positive definite. The selection vector $e_{m,i}$ has length m and entries that are equal to zero except for entry i which is equal to one.

Before we present the proofs for the identification lemmas we give two useful propositions.

Proposition 1. *Let $\mathbf{A}$ be an $M \times N$ matrix and let $\mathbf{A}_i$ be the ith column of $\mathbf{A}$. Also, let $\mathbf{D}_A = \mathbf{A}'\mathbf{A}$. Then,*

(i) $\mu_N(\mathbf{D}_A) \leqslant \|\mathbf{A}_i\|^2 \leqslant \mu_1(\mathbf{D}_A)$
(ii) $\mu_N(\mathbf{D}_A)\|\mathbf{A}_i\|^2 \leqslant \|\mathbf{A}'\mathbf{A}_i\|^2 \leqslant \mu_1(\mathbf{D}_A)\|\mathbf{A}_i\|^2$

Proof of Proposition 1. (*i*) The first inequality follows from the fact that

$$\|\mathbf{A}_i\|^2 = \|\mathbf{A}e_{N,i}\|^2 \geqslant \min_{\substack{x \in \mathbb{R}^N \\ \|x\|=1}} \|\mathbf{A}x\|^2 = \mu_N(\mathbf{D}_A) \, .$$

The second inequality follows from the fact that

$$\|\mathbf{A}_i\|^2 = \|\mathbf{A}e_{N,i}\|^2 \leqslant \max_{\substack{x \in \mathbb{R}^N \\ \|x\|=1}} \|\mathbf{A}x\|^2 = \mu_1(\mathbf{D}_A) \, .$$

[25] The differences with Foerster et al. (2011) stem from the fact that we use monthly growth rates whereas they consider quarterly growth rates.

(*ii*) The first inequality follows from the fact that

$$\|\mathbf{A}'\mathbf{A}_i\|^2 = \sum_{j=1}^{N}(\mathbf{A}_j'\mathbf{A}_i)^2 \geqslant (\mathbf{A}_i'\mathbf{A}_i)^2 = \|\mathbf{A}_i\|^2\|\mathbf{A}_i\|^2 \geqslant \min_{j=1,\dots,N}\|\mathbf{A}_j\|^2\|\mathbf{A}_i\|^2 \geqslant \mu_N(\mathbf{D}_A)\|\mathbf{A}_i\|^2 .$$

The second inequality follows from the Cauchy–Schwarz inequality. □

Proposition 2. *Let* $\mathbf{A}$ *be a positive definite* $k \times k$ *matrix,* $\mathbf{B}$ *a positive definite* $(n-k)\times(n-k)$ *matrix and* $\boldsymbol{d}$ *an* $(n-k)\times k$ *matrix with full column rank. The following inequalities hold.*

(i) $\|(\mathbf{A}^{-1} + \boldsymbol{d}'\mathbf{B}^{-1}\boldsymbol{d})e_{k,i}\|^2 \geqslant \mu^2_{n-k}(\mathbf{B}^{-1})\mu_k(\boldsymbol{d}'\boldsymbol{d})\|\boldsymbol{d}_i\|^2$
(ii) $\|(\mathbf{A}^{-1} + \boldsymbol{d}'\mathbf{B}^{-1}\boldsymbol{d})e_{k,i}\|^2 \leqslant 2\mu^2_1(\mathbf{A}^{-1}) + 2\mu^2_1(\mathbf{B}^{-1})\mu_1(\boldsymbol{d}'\boldsymbol{d})\|\boldsymbol{d}_i\|^2$
(iii) $\|\mathbf{B}^{-1}\boldsymbol{d}e_{k,i}\|^2 \geqslant \mu^2_{n-k}(\mathbf{B}^{-1})\|\boldsymbol{d}_i\|^2$
(iv) $\|\mathbf{B}^{-1}\boldsymbol{d}e_{k,i}\|^2 \leqslant \mu^2_1(\mathbf{B}^{-1})\|\boldsymbol{d}_i\|^2$
(v) $\|\boldsymbol{d}'\mathbf{B}^{-1}e_{n-k,j-k}\|^2 \geqslant \mu^2_{n-k}(\mathbf{B}^{-1})\|\boldsymbol{d}\|^2$
(vi) $\|\boldsymbol{d}'\mathbf{B}^{-1}e_{n-k,j-k}\|^2 \leqslant \mu^2_1(\mathbf{B}^{-1})\|\boldsymbol{d}\|^2$
(vii) $\|\mathbf{B}^{-1}e_{n-k,j-k}\|^2 \geqslant \mu^2_{n-k}(\mathbf{B}^{-1})$
(viii) $\|\mathbf{B}^{-1}e_{n-k,j-k}\|^2 \leqslant \mu^2_1(\mathbf{B}^{-1})$

Proof of Proposition 2. Follows directly from Proposition 1, the Cauchy–Schwartz inequality and the fact that for vectors $u, v \in \mathbb{R}^n$ $\|u+v\|^2 \leqslant 2\|u\|^2 + 2\|v\|^2$. □

Proof of Lemma 1. First, we show that $\|\boldsymbol{K}\|$ exists. Under Assumption 1(i), (ii) and (iii) we can write $\boldsymbol{K}$, as given in Eq. (3), as

$$\mathbf{K} = \begin{bmatrix} \mathbf{I}_k & -\boldsymbol{\beta}' \\ \mathbf{0} & \mathbf{I}_{n-k} \end{bmatrix}\begin{bmatrix} \boldsymbol{\Sigma}_g^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_\epsilon^{-1} \end{bmatrix}\begin{bmatrix} \mathbf{I}_k & \mathbf{0} \\ -\boldsymbol{\beta} & \mathbf{I}_{n-k} \end{bmatrix},$$

which implies that

$$\|\mathbf{K}\| \leqslant \left\|\begin{bmatrix} \mathbf{I}_k & -\boldsymbol{\beta}' \\ \mathbf{0} & \mathbf{I}_{n-k} \end{bmatrix}\right\|^2 \left\|\begin{bmatrix} \boldsymbol{\Sigma}_g^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_\epsilon^{-1} \end{bmatrix}\right\| .$$

For the first term for all $n > N$ we have that

$$\begin{aligned}\left\|\begin{bmatrix} \mathbf{I}_k & -\boldsymbol{\beta}' \\ \mathbf{0} & \mathbf{I}_{n-k} \end{bmatrix}\right\|^2 &\leqslant \left\|\begin{bmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n-k} \end{bmatrix}\right\|^2 + \left\|\begin{bmatrix} \mathbf{0} & -\boldsymbol{\beta}' \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\right\|^2 + 2\left\|\begin{bmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n-k} \end{bmatrix}\right\|\left\|\begin{bmatrix} \mathbf{0} & -\boldsymbol{\beta}' \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\right\| \\ &= 1 + \mu_1(\boldsymbol{\beta}'\boldsymbol{\beta}) + 2\sigma_1(\boldsymbol{\beta}'\boldsymbol{\beta}) \leqslant 1 + M_g^2 + 2M_g < \infty\end{aligned}$$

where the final bound follow from Assumption 1-(iv). For the second term we have

$$\begin{aligned}\left\|\begin{bmatrix} \boldsymbol{\Sigma}_g^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_\epsilon^{-1} \end{bmatrix}\right\| &= \max\{\mu_1(\boldsymbol{\Sigma}_g^{-1}), \mu_1(\boldsymbol{\Sigma}_\epsilon^{-1})\} \\ &= \left(\min\{\mu_k(\boldsymbol{\Sigma}_g), \mu_{n-k}(\boldsymbol{\Sigma}_\epsilon)\}\right)^{-1} < \infty\end{aligned}$$

Since Assumption 1-(i) implies $\mu_k(\boldsymbol{\Sigma}_g) > 0$ and 1-(ii) requires $\mu_{n-k}(\boldsymbol{\Sigma}_\epsilon) > 0$. The latter is preserved for any $n > N$ by 1-(iv) which requires $\kappa_\epsilon < \infty$ which implies $\mu_{n-k}(\boldsymbol{\Sigma}_\epsilon) > 0$ for all $n > N$.

Second, we show that $\|\mathbf{K}_i\|^2 > \|\mathbf{K}_j\|^2$ for any $i = 1,\dots,k$ and $j = k+1,\dots,n$, which implies the claim of the lemma. Note that Eq. (3) implies that for $i = 1,\dots,k$ we have that $\|\mathbf{K}_i\|^2 = \|(\boldsymbol{\Sigma}_g^{-1} + \boldsymbol{\beta}'\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta})e_{k,i}\|^2 + \|\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta}e_{k,i}\|^2$ and $\|\mathbf{K}_j\|^2 = \|\boldsymbol{\beta}'\boldsymbol{\Sigma}_\epsilon^{-1}e_{n-k,j-k}\|^2 + \|\boldsymbol{\Sigma}_\epsilon^{-1}e_{n-k,j-k}\|^2$. Sufficient conditions for $\|\mathbf{K}_i\|^2 > \|\mathbf{K}_j\|^2$ are given by $\|(\boldsymbol{\Sigma}_g^{-1} + \boldsymbol{\beta}'\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta})e_{k,i}\|^2 > \|\boldsymbol{\beta}'\boldsymbol{\Sigma}_\epsilon^{-1}e_{n-k,j-k}\|^2$ and $\|\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta}e_{k,i}\|^2 > \|\boldsymbol{\Sigma}_\epsilon^{-1}e_{n-k,j-k}\|^2$. Using the inequalities (*i*) and (*vi*) from Proposition 2 the first condition immediately gives $\|\boldsymbol{\beta}_i\| > \kappa_\beta\kappa_\epsilon$ and from parts (*iii*) and (*viii*) of Proposition 2 it follows that the second condition gives $\|\boldsymbol{\beta}_i\| > \kappa_\epsilon$. Both conditions are satisfied by Assumption 1-(iv). □

Proof of Lemma 2. Assume, without loss of generality that the columns of $\boldsymbol{K}$ are ordered in decreasing order by their norms. We show that Assumption 1-(iv*) is sufficient to prove the lemma after the structure of the covariance matrix is imposed by Assumption 1(i), (ii) and (iii). We require that

$$\frac{\|\boldsymbol{K}e_{n,k}\|^2}{\|\boldsymbol{K}e_{n,k+1}\|^2} > \frac{\|\boldsymbol{K}e_{n,s}\|^2}{\|\boldsymbol{K}e_{n,s+1}\|^2} \quad \forall \quad s = 1,\dots,k-1,k+1,\dots,n-1.$$

When $s < k$ and $s > k$ the condition can be expressed as, respectively,

$$\frac{\|(\boldsymbol{\Sigma}_g^{-1} + \boldsymbol{\beta}'\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta})e_{k,k}\|^2 + \|\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta}e_{k,k}\|^2}{\|\boldsymbol{\beta}'\boldsymbol{\Sigma}_\epsilon^{-1}e_{n-k,1}\|^2 + \|\boldsymbol{\Sigma}_\epsilon^{-1}e_{n-k,1}\|^2} > \frac{\|\boldsymbol{\beta}'\boldsymbol{\Sigma}_\epsilon^{-1}e_{n-k,s}\|^2 + \|\boldsymbol{\Sigma}_\epsilon^{-1}e_{n-k,s}\|^2}{\|\boldsymbol{\beta}'\boldsymbol{\Sigma}_\epsilon^{-1}e_{n-k,s+1}\|^2 + \|\boldsymbol{\Sigma}_\epsilon^{-1}e_{n-k,s+1}\|^2}, \tag{14}$$

$$\frac{\|(\boldsymbol{\Sigma}_g^{-1} + \boldsymbol{\beta}'\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta})e_{k,k}\|^2 + \|\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta}e_{k,k}\|^2}{\|\boldsymbol{\beta}'\boldsymbol{\Sigma}_\epsilon^{-1}e_{n-k,1}\|^2 + \|\boldsymbol{\Sigma}_\epsilon^{-1}e_{n-k,1}\|^2} > \frac{\|(\boldsymbol{\Sigma}_g^{-1} + \boldsymbol{\beta}'\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta})e_{k,s}\|^2 + \|\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta}e_{k,s}\|^2}{\|(\boldsymbol{\Sigma}_g^{-1} + \boldsymbol{\beta}'\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta})e_{k,s+1}\|^2 + \|\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta}e_{k,s+1}\|^2}. \tag{15}$$

Both expressions are of the form $\frac{a+b}{c+d} > \frac{e+f}{g+h}$ with $a, \ldots, h > 0$. We use that $\frac{a}{c} > \frac{e}{g}$, $\frac{a}{c} > \frac{f}{h}$, $\frac{b}{d} > \frac{e}{g}$ and $\frac{b}{d} > \frac{f}{h}$ are sufficient for this condition to hold. We obtain a total of 8 sufficient conditions. For condition (14) we obtain by direct calculation – using Proposition 2 – the bounds

$$\begin{aligned}
\frac{\|(\boldsymbol{\Sigma}_g^{-1}+\boldsymbol{\beta}'\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta})e_{k,k}\|^2}{\|\boldsymbol{\beta}'\boldsymbol{\Sigma}_\epsilon^{-1}e_{n-k,1}\|^2} &> \frac{\|\boldsymbol{\beta}'\boldsymbol{\Sigma}_\epsilon^{-1}e_{n-k,s}\|^2}{\|\boldsymbol{\beta}'\boldsymbol{\Sigma}_\epsilon^{-1}e_{n-k,s+1}\|^2} \quad \Rightarrow \quad \|\boldsymbol{\beta}_k\| > \kappa_\beta\kappa_\epsilon^2 \\
\frac{\|(\boldsymbol{\Sigma}_g^{-1}+\boldsymbol{\beta}'\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta})e_{k,k}\|^2}{\|\boldsymbol{\beta}'\boldsymbol{\Sigma}_\epsilon^{-1}e_{n-k,1}\|^2} &> \frac{\|\boldsymbol{\Sigma}_\epsilon^{-1}e_{n-k,s}\|^2}{\|\boldsymbol{\Sigma}_\epsilon^{-1}e_{n-k,s+1}\|^2} \quad \Rightarrow \quad \|\boldsymbol{\beta}_k\| > \kappa_\beta\kappa_\epsilon^2 \\
\frac{\|\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta}e_{k,k}\|^2}{\|\boldsymbol{\Sigma}_\epsilon^{-1}e_{n-k,1}\|^2} &> \frac{\|\boldsymbol{\beta}'\boldsymbol{\Sigma}_\epsilon^{-1}e_{n-k,s}\|^2}{\|\boldsymbol{\beta}'\boldsymbol{\Sigma}_\epsilon^{-1}e_{n-k,s+1}\|^2} \quad \Rightarrow \quad \|\boldsymbol{\beta}_k\| > \kappa_\epsilon^2 \\
\frac{\|\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta}e_{k,k}\|^2}{\|\boldsymbol{\Sigma}_\epsilon^{-1}e_{n-k,1}\|^2} &> \frac{\|\boldsymbol{\Sigma}_\epsilon^{-1}e_{n-k,s}\|^2}{\|\boldsymbol{\Sigma}_\epsilon^{-1}e_{n-k,s+1}\|^2} \quad \Rightarrow \quad \|\boldsymbol{\beta}_k\| > \kappa_\epsilon^2
\end{aligned}$$

and for (20) we have

$$\begin{aligned}
\frac{\|(\boldsymbol{\Sigma}_g^{-1}+\boldsymbol{\beta}'\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta})e_{k,k}\|^2}{\|\boldsymbol{\beta}'\boldsymbol{\Sigma}_\epsilon^{-1}e_{n-k,1}\|^2} &> \frac{\|(\boldsymbol{\Sigma}_g^{-1}+\boldsymbol{\beta}'\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta})e_{k,i}\|^2}{\|(\boldsymbol{\Sigma}_g^{-1}+\boldsymbol{\beta}'\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta})e_{k,i+1}\|^2} \quad \Rightarrow \quad \|\boldsymbol{\beta}_k\| > \kappa_\beta^2\kappa_\epsilon\left(\kappa_\epsilon + \tfrac{\mu_1(\boldsymbol{\Sigma}_\epsilon)}{\mu_k(\boldsymbol{\Sigma}_g)}\right) \\
\frac{\|(\boldsymbol{\Sigma}_g^{-1}+\boldsymbol{\beta}'\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta})e_{k,k}\|^2}{\|\boldsymbol{\beta}'\boldsymbol{\Sigma}_\epsilon^{-1}e_{n-k,1}\|^2} &> \frac{\|\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta}e_{k,i}\|^2}{\|\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta}e_{k,i+1}\|^2} \quad \Rightarrow \quad \|\boldsymbol{\beta}_k\| > \kappa_\beta^2\kappa_\epsilon^2 \\
\frac{\|\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta}e_{k,k}\|^2}{\|\boldsymbol{\Sigma}_\epsilon^{-1}e_{n-k,1}\|^2} &> \frac{\|(\boldsymbol{\Sigma}_g^{-1}+\boldsymbol{\beta}'\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta})e_{k,i}\|^2}{\|(\boldsymbol{\Sigma}_g^{-1}+\boldsymbol{\beta}'\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta})e_{k,i+1}\|^2} \quad \Rightarrow \quad \|\boldsymbol{\beta}_k\| > \kappa_\beta\kappa_\epsilon\left(\kappa_\epsilon + \tfrac{\mu_1(\boldsymbol{\Sigma}_\epsilon)}{\mu_k(\boldsymbol{\Sigma}_g)}\right) \\
\frac{\|\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta}e_{k,k}\|^2}{\|\boldsymbol{\Sigma}_\epsilon^{-1}e_{n-k,1}\|^2} &> \frac{\|\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta}e_{k,i}\|^2}{\|\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta}e_{k,i+1}\|^2} \quad \Rightarrow \quad \|\boldsymbol{\beta}_k\| > \kappa_\beta^2\kappa_\epsilon^2.
\end{aligned}$$

These are all implied by Assumption 1-(iv*). □

Proof of Lemma 3. It is convenient to write the model in the factor representation (see Eq. (8)):

$$\begin{aligned}
y_{1:k,t} &= \mathbf{L}_1 h_t \\
y_{k+1:n,t} &= \mathbf{L}_2 h_t + \epsilon_t,
\end{aligned}$$

where $h_t = (\tilde{g}_t', f_t')'$ is the $(r+k)\times 1$ vector of standardized granular shocks and factor shocks, $\mathbf{L}_1 = [\boldsymbol{\Sigma}_g^{1/2} \ \boldsymbol{\Lambda}_1]$ is the $k \times (r+k)$ matrix of granular and factor loadings on the granular series and $\mathbf{L}_2 = [\boldsymbol{\beta}\boldsymbol{\Sigma}_g^{1/2} \ \boldsymbol{\Lambda}_2]$ is the $(n-k)\times(r+k)$ matrix of granular and factor loadings on the non-granular series. Notice that from Assumptions 1-(i) and 2-(i) it follows that $\mathrm{Var}(h_t) = \mathbf{I}_{k+r}$. From Assumptions 1(i)–(iii) and 2(i)–(ii) it follows that the variance matrix of y_t can be written as

$$\boldsymbol{\Sigma} = \begin{bmatrix} \mathbf{L}_1\mathbf{L}_1' & \mathbf{L}_1\mathbf{L}_2' \\ \mathbf{L}_2\mathbf{L}_1' & \mathbf{L}_2\mathbf{L}_2' + \boldsymbol{\Sigma}_\epsilon \end{bmatrix}$$

From the inverse for block matrices we find that

$$\mathbf{K} = \begin{bmatrix} (\mathbf{L}_1\mathbf{L}_1')^{-1} + (\mathbf{L}_1\mathbf{L}_1')^{-1}\mathbf{L}_1\mathbf{L}_2'\mathbf{X}^{-1}\mathbf{L}_2\mathbf{L}_1'(\mathbf{L}_1\mathbf{L}_1')^{-1} & -(\mathbf{L}_1\mathbf{L}_1')^{-1}\mathbf{L}_1\mathbf{L}_2'\mathbf{X}^{-1} \\ -\mathbf{X}^{-1}\mathbf{L}_2\mathbf{L}_1'(\mathbf{L}_1\mathbf{L}_1')^{-1} & \mathbf{X}^{-1} \end{bmatrix}$$

where $\mathbf{X} = \mathbf{L}_2\mathbf{L}_2' + \boldsymbol{\Sigma}_\epsilon - \mathbf{L}_2\mathbf{L}_1'(\mathbf{L}_1\mathbf{L}_1')^{-1}\mathbf{L}_1\mathbf{L}_2'$. Next, we define $\mathbf{M}_{L_1} = \mathbf{I}_{k+r} - \mathbf{L}_1'(\mathbf{L}_1\mathbf{L}_1')^{-1}\mathbf{L}_1$. We have that $\mathbf{X} = \mathbf{L}_2\mathbf{M}_{L_1}\mathbf{L}_2' + \boldsymbol{\Sigma}_\epsilon = \hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon$ where $\hat{\mathbf{U}} = \mathbf{L}_2\mathbf{M}_{L_1}$. Also, define $\hat{\boldsymbol{\gamma}} = \mathbf{L}_2\mathbf{L}_1'(\mathbf{L}_1\mathbf{L}_1')^{-1}$ which is the $(n-k)\times k$ projection coefficient of the regression that explains $\mathbf{L}_2$ in terms of $\mathbf{L}_1$. The resulting concentration matrix becomes

$$\mathbf{K} = \begin{bmatrix} (\mathbf{L}_1\mathbf{L}_1')^{-1} + \hat{\boldsymbol{\gamma}}'(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)^{-1}\hat{\boldsymbol{\gamma}} & -\hat{\boldsymbol{\gamma}}'(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)^{-1} \\ -(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)^{-1}\hat{\boldsymbol{\gamma}} & (\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)^{-1} \end{bmatrix} \tag{16}$$

We have that $\|\mathbf{K}_i\|^2 = \left\| \left((\mathbf{L}_1\mathbf{L}_1')^{-1} + \hat{\boldsymbol{\gamma}}'(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)^{-1}\hat{\boldsymbol{\gamma}} \right) e_{k,i} \right\|^2 + \|(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)^{-1}\hat{\boldsymbol{\gamma}}e_{k,i}\|^2$. From Proposition 2-(*i*) it follows that $\| \left((\mathbf{L}_1\mathbf{L}_1')^{-1} + \hat{\boldsymbol{\gamma}}'(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)^{-1}\hat{\boldsymbol{\gamma}} \right) e_{k,i}\|^2 \geqslant \mu_{n-k}^2((\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)^{-1})\mu_k(\hat{\boldsymbol{\gamma}}'\hat{\boldsymbol{\gamma}})\|\hat{\boldsymbol{\gamma}}_i\|^2$ and from Proposition 2-(*iii*) $\|(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)^{-1}\hat{\boldsymbol{\gamma}}e_{k,i}\|^2 \geqslant \mu_{n-k}^2((\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)^{-1})\|\hat{\boldsymbol{\gamma}}_i\|^2$. Next, we have $\|\mathbf{K}_j\|^2 = \|\hat{\boldsymbol{\gamma}}'(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)^{-1}e_{n-k,j-k}\|^2 + \|(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)^{-1}e_{n-k,j-k}\|^2$. Proposition 2 parts (*vi*) and (*viii*) imply that $\|\hat{\boldsymbol{\gamma}}'(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)^{-1}e_{n-k,j-k}\|^2 \leqslant \mu_1^2((\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)^{-1})\|\hat{\boldsymbol{\gamma}}\|^2$ and $\|(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)^{-1}e_{n-k,j-k}\|^2 \leqslant \mu_1^2((\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)^{-1})$. Combining the inequalities we find that a sufficient condition for $\|\mathbf{K}_i\|^2 > \|\mathbf{K}_j\|^2$

is given by

$$\|\hat{\boldsymbol{\gamma}}_i\|^2 > \frac{\mu_1^2(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)}{\mu_{n-k}^2(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)} \frac{\mu_1^2(\hat{\boldsymbol{\gamma}}'\hat{\boldsymbol{\gamma}})}{\mu_k^2(\hat{\boldsymbol{\gamma}}'\hat{\boldsymbol{\gamma}})}.$$

We can bound $\|\hat{\boldsymbol{\gamma}}_i\|^2$ as follows

$$\begin{aligned} \hat{\boldsymbol{\gamma}}_i'\hat{\boldsymbol{\gamma}}_i &= w_{i,1}'\boldsymbol{\beta}'\boldsymbol{\beta}w_{i,1} + w_{i,2}'\boldsymbol{\Lambda}_2'\boldsymbol{\Lambda}_2 w_{i,2} + 2w_{i,1}'\boldsymbol{\beta}'\boldsymbol{\Lambda}_2 w_{i,2} \\ &\geqslant \mu_k(\boldsymbol{\beta}'\boldsymbol{\beta})\|w_{i,1}\|^2 + \mu_r(\boldsymbol{\Lambda}_2'\boldsymbol{\Lambda}_2)\|w_{i,2}\|^2 + 2w_{i,1}'\boldsymbol{\beta}'\boldsymbol{\Lambda}_2 w_{i,2} \end{aligned} \tag{17}$$

where $w_{i,1} = \mathbf{W}_1 e_{k,i}$, $\mathbf{W}_1 = \boldsymbol{\Sigma}_g(\boldsymbol{\Sigma}_g + \boldsymbol{\Lambda}_1\boldsymbol{\Lambda}_1')^{-1}$ $w_{i,2} = \mathbf{W}_2 e_{k,i}$ and $\mathbf{W}_2 = \boldsymbol{\Lambda}_1'(\boldsymbol{\Sigma}_g + \boldsymbol{\Lambda}_1\boldsymbol{\Lambda}_1')^{-1}$. Further, notice that

$$\|w_{i,1}\|^2 \geqslant \frac{\mu_k^2(\boldsymbol{\Sigma}_g)}{\mu_1^2(\boldsymbol{\Sigma}_g + \boldsymbol{\Lambda}_1\boldsymbol{\Lambda}_1')} \equiv s_g^2 \qquad \text{and} \qquad \|w_{i,2}\|^2 \leqslant \frac{\|\boldsymbol{\Lambda}_1\|^2}{\mu_k^2(\boldsymbol{\Sigma}_g + \boldsymbol{\Lambda}_1\boldsymbol{\Lambda}_1')} \equiv s_{\Lambda_1}^2 \tag{18}$$

The bound of interest becomes

$$\mu_k(\boldsymbol{\beta}'\boldsymbol{\beta}) \geqslant \frac{\mu_1^2(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)}{\mu_{n-k}^2(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)} \frac{\kappa_{\hat{\boldsymbol{\gamma}}}^2}{s_g^2} - \mu_r(\mathbf{D}_\lambda)\frac{s_{\Lambda_1}^2}{s_g^2} - 2\frac{w_{i,1}'\boldsymbol{\beta}'\boldsymbol{\Lambda}_2 w_{i,2}}{s_g^2}.$$

Notice that $-2\frac{w_{i,1}'\boldsymbol{\beta}'\boldsymbol{\Lambda}_2 w_{i,2}}{s_g^2} \leqslant 2\|\boldsymbol{\beta}'\boldsymbol{\Lambda}_2\| s_{\Lambda_1} s_g^{-1}$ such that we obtain

$$\mu_k(\boldsymbol{\beta}'\boldsymbol{\beta}) \geqslant \frac{\mu_1^2(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)}{\mu_{n-k}^2(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)} \frac{\kappa_{\hat{\boldsymbol{\gamma}}}^2}{s_g^2} + 2\|\boldsymbol{\beta}'\boldsymbol{\Lambda}_2\| s_{\Lambda_1} s_g^{-1} - \mu_r(\mathbf{D}_\lambda)\frac{s_{\Lambda_1}^2}{s_g^2}.$$

This is the bound that one can obtain without any restrictions on the correlation between the granular loadings and the factor loadings for the non-granular series. Now it follows that

$$\mu_r(\mathbf{D}_\lambda)\frac{s_{\Lambda_1}^2}{s_g^2} > 2\|\boldsymbol{\beta}'\boldsymbol{\Lambda}_2\| s_{\Lambda_1} s_g^{-1}$$

as

$$\mu_r(\mathbf{D}_\lambda) > 2\|\boldsymbol{\beta}'\boldsymbol{\Lambda}_2\| s_{\Lambda_1}^{-1} s_g$$

is implied by Assumption 2-(iii). Under this condition we can drop the last two terms from the bound for $\mu_k(\boldsymbol{\beta}'\boldsymbol{\beta})$ and obtain

$$\mu_k(\boldsymbol{\beta}'\boldsymbol{\beta}) \geqslant \frac{\mu_1^2(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)}{\mu_{n-k}^2(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)} \frac{\kappa_{\hat{\boldsymbol{\gamma}}}^2}{s_g^2}$$

After taking the square root we find that this bound is implied by Assumption 2-(iii). □

Proof of Lemma 4. Assume, without loss of generality that the columns of $\boldsymbol{K}$ are ordered in decreasing order by their norms. We show that Assumption 2-(iv*) is sufficient to prove the lemma after the structure of the covariance matrix is imposed by Assumptions 1(i)–(iii) and 2(i)–(iii). We require that

$$\frac{\|\boldsymbol{K}e_{n,k}\|^2}{\|\boldsymbol{K}e_{n,k+1}\|^2} > \frac{\|\boldsymbol{K}e_{n,s}\|^2}{\|\boldsymbol{K}e_{n,s+1}\|^2} \quad \forall \quad s = 1, \ldots, k-1, k+1, \ldots, n-1.$$

Using the representation for the concentration matrix given in (16) we find that for $s < k$ and $s > k$ the conditions can be expressed as, respectively,

$$\mathbf{K} = \begin{bmatrix} (\mathbf{L}_1\mathbf{L}_1')^{-1} + \hat{\boldsymbol{\gamma}}'(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)^{-1}\hat{\boldsymbol{\gamma}} & -\hat{\boldsymbol{\gamma}}'(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)^{-1} \\ -(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)^{-1}\hat{\boldsymbol{\gamma}} & (\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)^{-1} \end{bmatrix}$$

$$\frac{\left\|\left((\mathbf{L}_1\mathbf{L}_1')^{-1} + \hat{\boldsymbol{\gamma}}'(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)^{-1}\hat{\boldsymbol{\gamma}}\right)e_{k,k}\right\|^2 + \|(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)^{-1}\hat{\boldsymbol{\gamma}}e_{k,k}\|^2}{\|\hat{\boldsymbol{\gamma}}'(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)^{-1}\hat{\boldsymbol{\gamma}}e_{n-k,1}\|^2 + \|(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)^{-1}e_{n-k,1}\|^2} > \frac{\|\hat{\boldsymbol{\gamma}}'(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)^{-1}e_{n-k,s}\|^2 + \|(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)^{-1}e_{n-k,s}\|^2}{\|\hat{\boldsymbol{\gamma}}'(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)^{-1}e_{n-k,s+1}\|^2 + \|(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)^{-1}e_{n-k,s+1}\|^2} \tag{19}$$

and

$$\frac{\left\|\left((\mathbf{L}_1\mathbf{L}_1')^{-1} + \hat{\boldsymbol{\gamma}}'(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)^{-1}\hat{\boldsymbol{\gamma}}\right)e_{k,k}\right\|^2 + \|(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)^{-1}\hat{\boldsymbol{\gamma}}e_{k,k}\|^2}{\|\hat{\boldsymbol{\gamma}}'(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)^{-1}\hat{\boldsymbol{\gamma}}e_{n-k,1}\|^2 + \|(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)^{-1}e_{n-k,1}\|^2} > \frac{\left\|\left((\mathbf{L}_1\mathbf{L}_1')^{-1} + \hat{\boldsymbol{\gamma}}'(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)^{-1}\hat{\boldsymbol{\gamma}}\right)e_{k,s}\right\|^2 + \|(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)^{-1}\hat{\boldsymbol{\gamma}}e_{k,s}\|^2}{\left\|\left((\mathbf{L}_1\mathbf{L}_1')^{-1} + \hat{\boldsymbol{\gamma}}'(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)^{-1}\hat{\boldsymbol{\gamma}}\right)e_{k,s+1}\right\|^2 + \|(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)^{-1}\hat{\boldsymbol{\gamma}}e_{k,s+1}\|^2} \tag{20}$$

From Proposition 2 it follows that a sufficient condition for selection of the number granulars is[26]

$$\|\hat{\boldsymbol{\gamma}}_k\|^2 > \kappa^4_{\hat{\boldsymbol{\gamma}}} \frac{\mu_1^2(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)}{\mu_{n-k}^2(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)} \left(\frac{\mu_1(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)}{\mu_{n-k}(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)} + \frac{\mu_1(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)}{\mu_k(\mathbf{L}_1\mathbf{L}_1')} \right)^2$$

From Eqs. (17) and (18) it follows that we may rewrite the bound as

$$\mu_k(\boldsymbol{\beta}'\boldsymbol{\beta}) \geqslant \frac{\kappa^4_{\hat{\boldsymbol{\gamma}}}}{s_g^2} \frac{\mu_1^2(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)}{\mu_{n-k}^2(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)} \left(\frac{\mu_1(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)}{\mu_{n-k}(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)} + \frac{\mu_1(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)}{\mu_k(\mathbf{L}_1\mathbf{L}_1')} \right)^2 - \mu_r(\mathbf{D}_\lambda)\frac{s^2_{\Lambda_1}}{s_g^2} - 2\frac{w_{i,1}'\boldsymbol{\beta}'\boldsymbol{\Lambda}_2 w_{i,2}}{s_g^2}$$

Similar as in Lemma 3 it follows from Assumption 2-(iii) that the sum of the last two terms is negative and hence they can be removed from the bound. Such that we get

$$\mu_k(\boldsymbol{\beta}'\boldsymbol{\beta}) \geqslant \frac{\kappa^4_{\hat{\boldsymbol{\gamma}}}}{s_g^2} \frac{\mu_1^2(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)}{\mu_{n-k}^2(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)} \left(\frac{\mu_1(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)}{\mu_{n-k}(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)} + \frac{\mu_1(\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon)}{\mu_k(\mathbf{L}_1\mathbf{L}_1')} \right)^2$$

After taking the square root we find that this bound is implied by Assumption 2-(iv*) □

Appendix B. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2020.04.013.

References

Acemoglu, D., Carvalho, V.M., Ozdaglar, A., Tahbaz-Salehi, A., 2012. The network origens of aggregate fluctuations. Econometrica 80, 1977–2016.
Acemoglu, D., Ozdaglar, A., Tahbaz-Salehi, A., 2015. Systemic risk and stability in financial networks. Amer. Econ. Rev. 105, 564–608.
Ahn, S.C., Horenstein, A.R., 2013. Eigenvalue ratio test for the number of factors. Econometrica 80, 1203–1227.
Ansley, C.F., Kohn, R., 1986. A note on reparameterizing a vector autoregressive moving average model to enforce stationarity. J. Stat. Comput. Simul. 24, 99–106.
Atalay, E., 2017. How important are sectoral shocks? Am. Econ. J.: Macroecon. 9 (4), 254–280.
Bai, J., 2003. Inferential theory for factor models of large dimensions. Econometrica 71, 135–171.
Bai, J., Li, K., 2012. Statistical analysis of factor models of high dimension. Ann. Statist. 40, 436–465.
Bai, J., Li, K., 2016. Maximum likelihood estimation and inference for approximate factor models of high dimension. Rev. Econ. Stat. 98, 298–309.
Bai, J., Li, K., Lu, L., 2016. Estimation and inference of FAVAR models. J. Bus. Econom. Statist. 34, 620–641.
Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70, 191–221.
Bai, J., Ng, S., 2006. Evaluating latent and observed factors in macroeconomics and finance. J. Econometrics 131, 507–537.
Bernanke, B.S., Boivin, J., Eliasz, P., 2005. Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach. Q. J. Econ. 120, 387–422.
Bernard, A.B., Jensen, J.B., Redding, S.J., Schott, P.K., 2017. Global firms. J. Econ. Lit. 56 (2), 565–619.
Carvalho, V., Gabaix, X., 2013. The great diversification and its undoing. Amer. Econ. Rev. 103, 1697–1727.
Chudik, A., Pesaran, M.H., Tosetti, E., 2011. Testing weak cross-sectional dependence in large panels. Econom. J. 14, 45–90.
Di Giovanni, J., Levchenko, A.A., 2012. International trade and aggregate fluctuations in granular economies. J. Political Econ. 120, 1083–1132.
Di Giovanni, J., Levchenko, A.A., Mejean, I., 2014. Firms, destinations and aggregate fluctuations. Econometrica 82, 1303–1340.
Diebold, F.X., Yılmaz, K., 2014. On the network topology of variance decompositions: Measuring the connectedness of financial firms. J. Econometrics 182, 119–134.
Doz, C., Giannone, D., Reichlin, L., 2012. A quasi maximum likelihood approach for large approximate dynamic factor models. Rev. Econ. Stat. 94, 1014–1024.
Fan, J., Liao, Y., Liu, H., 2016. An overview of the estimation of large covariance and precision matrices. Econom. J. 19, 1–32.
Fan, J., Liao, Y., Mincheva, M., 2011. High-dimensional covariance matrix estimation in approximate factor models. Ann. Statist. 39, 3320–3356.
Foerster, A.T., Sarte, P.D.G., Watson, M.W., 2011. Sectoral versus aggregate shocks: A structural factor analaysis of industrial production. J. Political Econ. 119, 1–38.
Forni, M., Reichlin, L., 1998. Let's get real: A dynamic factor analytical approach to disaggregated business cycle. Rev. Econom. Stud. 65, 453–474.
Gabaix, X., 2011. The granular origens of aggregate fluctuations. Econometrica 79, 733–772.
Gaubert, C., Itskhoki, O., 2016. Granular Comparative Advantage. Working paper.
Ghysels, E., Liu, H., Raymond, S., 2018. Granularity and (Downside) Risk in Equity Markets. Technical report, University of North Carolina at Chapel Hill.
Hautsch, N., Schaumburg, J., Schienle, M., 2015. Financial network systemic risk contributions. Rev. Finance 685–738.
Heywood, H.B., 1931. On finite sequences of real numbers. Proc. R. Soc. A 134, 486–501.
Jungbacker, B., Koopman, S., 2015. Likelihood-based analysis for dynamic factor analysis for measurement and forecasting. Econom. J. 18, C1–C21.
Kolenikov, S., Bollen, K.A., 2012. Testing negative error variances. is a heywood case a symptom of misspecification? Sociol. Methods Res. 41, 124–167.
Lauritzen, S.L., 1996. Graphical Models. Oxford University Press, Oxford.
Long, J., Plosser, C., 1987. Sectoral vs. aggregate shocks in the business cycle. Amer. Econ. Rev. 77, 333–336.
Meinshausen, N., Bühlmann, P., 2006. High dimensional graphs and variable selection with the lasso. Ann. Statist. 34, 1436–1462.
Mulally, A.R., 2008. Examining the State of the Domestic Automobile Industry. Hearing, United States Senate Committee on Banking, Housing and Urban Affairs.
Onatski, A., 2009. Testing hypotheses about the number of factors in large factor models. Econometrica 77, 1447–1479.

[26] Notice that we may apply Proposition 2 with $\hat{\boldsymbol{\gamma}}$ since Assumption 2-(iv*) implies that the conditioning number of $\hat{\boldsymbol{\gamma}}'\hat{\boldsymbol{\gamma}}$ must be finite, which implies that its smallest eigenvalue is greater then zero, which in its turn implies that $\hat{\boldsymbol{\gamma}}$ is full rank.

Onatski, A., 2010. Determining the number of factors from empirical distribution of eigenvalues. Rev. Econ. Stat. 92, 1004–1016.
Onatski, A., 2012. Asymptotics of the principal components estimator of large factor models with weakly influential factors. J. Econometrics 168, 244–258.
Parker, J., Sul, D., 2016. Identification of unknown common factors: Leaders and followers. J. Bus. Econom. Statist. 34, 227–239.
Pesaran, M.H., 2012. Testing weak cross-sectional dependence in large panels. Econometric Rev. 34, 1089–1117.
Pesaran, M.H., Yang, C.F., 2016. Econometric Analysis of Production Networks with Dominant Units. Institute for New Economic Thinking: Working Paper No. 16-25.
Pourahmadi, M., 2013. High Dimensional Covariance Estimation. John Wiley & Sons, Hoboken, New Jersey.
Siavash, S.S., 2016. Dominant Sectors in the US: A Factor Model Analysis of Sectoral Industrial Production. Working paper.
Stock, J.H., Watson, M.W., 2002a. Forecasting using principal components from a large number of predictors. J. Amer. Statist. Assoc. 97, 1167–1179.
Stock, J.H., Watson, M.W., 2002b. Macroeconomic forecasting using diffusion indexes. J. Bus. Econom. Statist. 220, 147–162.
Stock, J.H., Watson, M.W., 2016. Factor models and structural vector autoregressions in macroeconomics. In: Taylor, J.B., Uhlig, H. (Eds.), Handbook of Macroeconomics. Elsevier, Amsterdam, North Holland.
Vershynin, R., 2012. Introduction to the non-asymptotic analysis of random matrices. In: Compressed Sensing. Cambridge University Press, Cambridge, pp. 210–268.
White, H., 2000. Asymptotic Theory for Econometrians - 2nd Edition. Academic Press, San Diego, California.