Detecting Granular Time Series in Large Panels

Christian Brownlees and *Geert Mesters*[*]

Universitat Pompeu Fabra and Barcelona GSE

July 12, 2018

Abstract

Large economic and financial panels often contain 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 i-th series in the panel is measured by the norm of the i-th 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 studies demonstrate, among other findings, the granular influence of the automobile sector in US industrial production.

JEL classification: C33, C38.

Keywords: granularity, network models, factor models, panel data, industrial production, CDS spreads.

[*]e-mail: christian.brownlees@upf.edu, geert.mesters@upf.edu.
We would like to thank Dominico 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.

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, Carvalho, Ozdaglar and Tahbaz-Salehi (2012) and Acemoglu, Ozdaglar and Tahbaz-Salehi (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 panel model that allows us to formalize the granular detection problem for a panel of stationary time series 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 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 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. We work under the assumption that the researcher does not know whether a given series belongs to the set of granulars nor the total number of granular series.

[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, Levchenko and Mejean (2014), Bernard, Jensen, Redding and Schott (2017), Pesaran and Yang (2016), Gaubert and Itskhoki (2016) and Ghysels, Liuy and Raymondz (2017). Instead of relying on potentially arbitrary criteria for granular selection we detect granular series based on the covariance properties of the output data directly.

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 (e.g. Lauritzen (1996), Pourahmadi (2013, Chapter 5)).[2] As it is well known, the i, j element of the concentration matrix is proportional to the partial correlation between series i and j. This motivates us to use as a natural measure of the influence of series i in the panel the norm of the i-th 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. Building on this result we 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 we can identify the number of granular series as the index that maximizes the sequential column norm ratio. This criterion is analogous to the eigenvalue ratio criterion proposed by 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, Sarte and Watson (2011). To this extent, we consider an extension of the model in which the series in the panel are additionally influenced by a set of common factors.[3] We show that the column norms of the concentration matrix maintain their detection properties if the signal-to-noise ratio of the granular shocks in the granular series remains sufficiently large. Importantly, the results imply that in order to detect granulars the researcher does not need to know the number of common factors.

We operationalize our identification results by estimating the column norms of the concentration matrix using the sample covariance matrix of the panel. We show that the column norms of the inverse sample covariance matrix consistently estimate their population analog

[2]See also research by Meinshausen and Bühlmann (2006), Peng, Wang, Zhou and Zhu (2009), Diebold and Yılmaz (2014) and Hautsch, Schaumburg and Schienle (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.

when the cross-sectional dimension and the number of time series observations are large. This result allows us to establish that our column norm estimator leads to consistent ranking and selection of the granular series. Our estimation results rely on distributional and dependence assumptions made in the literature on large dimensional covariance estimation, see for example Fan, Liao and Mincheva (2011).

Alternative approaches for granular detection can be based on principal components or maximum likelihood methods. In brief, such methods are 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 is carried out to assess the performance of our methodology in finite samples. In the study we simulate a granular model with common factors and then use our granular detection methodology to recover the granular series. Results show that the granular detection methodology procedures performs satisfactorily in finite samples when the strength of the granulars is sufficiently large.

We apply our methodology in two empirical studies. First, we consider 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

[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.

increasing for more recent sampling periods. In the second study, we use our framework to detect granulars in a panel of CDS spreads of Eurozone financial institutions over a sample covering the 2007–2009 Great Financial Crisis as well as the 2010–2012 European Sovereign Debt Crisis. Our methodology identifies as granulars two of the largest Eurozone periphery banks that suffered severe distress in this period: Banco Santander and BBVA.

The remainder of this paper is organized as follows. Section 2 formalizes the granular detection problem and discusses applications in macroeconomics and finance. Section 3 introduces our granular detection methodology and it establishes its large sample properties. Section 4 compares our methodology to alternative methods based on principal components analysis and maximum likelihood. Section 5 carries out a simulation study to assess the finite sample performance of the proposed methodology. Section 6 presents the results of two empirical studies and concluding remarks follow in Section 7.

2 The granular detection problem

In this section we formalize the granular detection problem and discuss its application for empirical studies in economics and finance. Let y_t be an n-dimensional time series observed from period $t = 1$ to T. We use $y_{i,t}$ to denote the i-th component of y_t and $y_{i:j,t}$ with $i < j$ to denote the $(j - i + 1)$-dimensional time series containing the i-th to j-th 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 assume that the granular series are the first k series in the panel. The other $n - k$ time series are the non-granular series whose 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

[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.

factors is defined as

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

where $\boldsymbol{\beta}$ is the $(n-k)\times k$ granular loading matrix and $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$ are the $k \times r$ and $(n-k)\times r$ loading matrices for the common factors. Precise assumptions on the model are spelled out in the following sections.

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 introduce a methodology that allows to consistently recover this information from the data.[6] We point out that in our framework the researcher does not need to know the number of common factors r in order to detect granular series.[7] Further, it is important to clarify that while model (1) has a factor model representation, the methodology that we introduce in this paper is different from the standard techniques that are adopted in the factor model literature, like maximum likelihood and principal components. Specifically, our detection strategy is based on the partial correlation properties of the panel. In Section 4 we compare our methodology to alternative methods.

We discuss two leading examples that illustrate how our framework can be used to formalize different problems in macroeconomics and finance. We return to these examples in the empirical part of the paper.

[6]We notice that 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] However, notice that in order to carry out inference on the model knowledge of r is required. Existing methods for factor selection introduced in the literature (e.g. Bai and Ng (2002) and Onatski (2010)) may be applied for this task.

2.1 Granular sectors in industrial production

The industrial production index in the United States is constructed as a weighted average of production indices across many sectors. Yet the aggregate volatility of the index is large. This implies that much of the variability in the index does not average out across different sectors.

Two leading explanations for this phenomenon have been proposed, see Foerster et al. (2011). First, aggregate shocks may exist that influence many sectors at the same time. Examples include, monetary policy shocks, exchange rate shocks and technology shocks. Second, sector specific idiosyncratic shocks may affect a large number of other sectors. For example, this may be a consequence of the interconnectedness in the production network, as in Acemoglu et al. (2012). In reality, it is reasonable to assume that a mixture of both aggregate and idiosyncratic shocks influence aggregate volatility.

Model (1) can disentangle both explanations. When we define y_t as the vector of sector specific industrial production outcomes, model (1) implies that aggregate volatility is determined by the k granular shocks g_t and the r aggregate shocks f_t. Both have influence over the entire panel. Our methodology may be used to determine which sectors are granular and how many there are.

2.2 Granular institutions in the financial system

One of the lessons from the financial crisis is that the distress of few yet highly influential financial entities may impair the entire system. The model of Acemoglu et al. (2015) formalizes this insight and shows that a highly interconnected financial system may be vulnerable to the idiosyncratic shocks of the most interconnected institutions.

These ideas have motivated a large literature that aims at detecting and ranking institutions in the financial system according to their "systemicness", see for instance Adrian and Brunnermeier (2016) and Brownlees and Engle (2017). A number of papers in this area have proposed to measure systemic risk on the basis of network models like Billio, Getmansky, Lo and Pellizzon (2012) and Diebold and Yılmaz (2014). Broadly speaking, these contributions measure how systemic an institution is on the basis of the number and magnitude of spillovers effects of that institution on the rest of the financial system. Despite the intuitive appeal of these proposals, these papers do not introduce a model that precisely defines when an institution is indeed systemic, and, consequently, they do not establish the properties of their ranking/selection procedures.

We can cast the problem of detecting systemic institutions as yet another instance of a granular detection problem. Following Diebold and Yılmaz (2014) we may define y_t as a vector of risk measures such as volatilities or credit default swap (CDS) spreads for a set of financial institutions. Assuming that the panel is generated by model (1), the methodology introduced in this work may be used detect granular/systemic institutions while controlling for system wide sources of risk through the common factors.

3 Methodology

In this section we introduce the granular detection methodology. We first consider a simplified version of the granular panel data model (1) where there are no common factors. We extend the identification results to allow for common factors in section 3.2 and section 3.3 discusses the estimation of the granular detection statistics based on large data panels.

3.1 Granular panel model

The granular panel data model without common factors is given by

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

where $y_{1:k,t}$ denotes the k granular series, $y_{k+1:n,t}$ denotes the $n-k$ non-granular series, g_t is the $k \times 1$ vector of granular shocks, $\boldsymbol{\beta}$ is the $(n-k) \times k$ granular loading matrix and ϵ_t is the $(n-k) \times 1$ vector of non-granular shocks.

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, if g_t and ϵ_t are uncorrelated, the covariance matrix $\mathbf{\Sigma} = \text{Var}(y_t)$ and the concentration matrix $\mathbf{K} = \mathbf{\Sigma}^{-1}$ of the panel are given by

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

As an example, assume that the norms of the columns of the $\boldsymbol{\beta}$ matrix are larger than one and that $\mathbf{\Sigma}_\epsilon$ is the identity matrix. Then, it is straightforward to verify that the norms of the first k columns (or rows) of the concentration matrix 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 the concentration matrix.

The example above suggests that if the elements of the granular loading matrix $\boldsymbol{\beta}$ are sufficiently large relative to the covariance matrix of the non-granular shocks $\mathbf{\Sigma}_\epsilon$, then the column (or row) norms of the concentration matrix $\mathbf{K}$ can be used to identify the granular series. This motivates us to base our granular detection methodology on the column norms of the concentration matrix, that is

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

where $\mathbf{K}_i$ denotes the i-th column of $\mathbf{K}$.[8,9]

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 representation 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

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

[9] Notice that the decomposition (3) implies that the column norm is not the only function of the concentration matrix that can be used for granular detection.

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.[10]

Thus, heuristically, granular time series can be though of as hubs in a partial correlation network representation 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. This interpretation is illustrated in Figure 1 where we show the partial correlation network representation of the panel when $n = 6$, $k = 1$ and $\mathbf{\Sigma}_\epsilon$ is a diagonal matrix. The granular series corresponds to the node with the largest number of connections.[11]

We impose a number of assumptions on the components of model (2) to establish the identification results.[12]

Assumption 1.

(i) $\mathrm{E}(g_t) = 0$ *and* $\mathrm{E}(g_t g_t') = \mathbf{\Sigma}_g$ *with* $\mathbf{\Sigma}_g \succ 0$.

(ii) $\mathrm{E}(\epsilon_t) = 0$ *and* $\mathrm{E}(\epsilon_t \epsilon_t') = \mathbf{\Sigma}_\epsilon$ *with* $\mathbf{\Sigma}_\epsilon \succ 0$.

(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* $\mathbf{\Sigma}_\epsilon$ *and* $\mathbf{D}_\beta$ *respectively.*

Assumptions *(i)*, *(ii)* and *(iii)* are standard and ensure that $\mathbf{\Sigma}_g$ and $\mathbf{\Sigma}_\epsilon$ are invertible, and that g_t and $\epsilon_{i,t}$ are uncorrelated, which is standard for regression models, e.g. White

[10]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

$$\rho^{ij} = -\frac{\mathbf{K}_{ij}}{\sqrt{\mathbf{K}_{ii}\mathbf{K}_{jj}}} \ .$$

[11]We emphasize that in our framework standardizing the elements of $\mathbf{K}$ by rescaling, e.g. $\mathbf{K}_{ij}/\sqrt{\mathbf{K}_{ii}\mathbf{K}_{jj}}$, is inappropriate as this distorts the ordering of the column norms.

[12]The following notation is adopted. The k-th largest eigenvalue of an $N \times N$ matrix $\mathbf{A}$ is denoted as $\mu_k(\mathbf{A})$, the k-th 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 $\mathbf{A} \geq 0$ 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: $\kappa_A = \mu_1(\mathbf{A})/\mu_N(\mathbf{A})$.

(2000, Chapter 2). Assumption (iv) is important as it characterizes the granular model. It 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 $\boldsymbol{\beta}$.

Assumption (iv) is key to establish the identification results of this paper. When the series in the panel are generated by model (2) but Assumption (iv) 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. The following lemma establishes the population ranking result formally.

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*

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

All proofs are collected in the appendix.

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.[13] Consider the ratio between two subsequent ordered column norms,

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

that is

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

for $s = 1, ..., n-1$. 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 ought to be maximized when s is equal to k. In order to identify the number of granulars using the sequential column norm ratio we need strengthen assumption 1-(iv).

(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 can be simplified to $\sigma_k(\mathbf{D}_\beta) > 2\kappa_\beta^2 \kappa_\epsilon^2$. 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*

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

Note that jointly lemmas 1 and 2 are sufficient for the identification of the set of granular series.

A number of comments are in order. Clearly, the column norm ratio in equation (5) is not the only function of the concentration matrix $\mathbf{K}$ that identifies k. In fact, several other functions of the elements of the concentration matrix can be used to identify the number of granular series. For instance, one could consider appropriate variants of selection criteria introduced in the factor model literature, see, among others, the criteria in Onatski (2009) and Cavicchioli, Forni, Lippi and Zaffaroni (2016).

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 $\mathbf{\Sigma}_\epsilon$ is bounded, e.g. $n^{-\alpha}\boldsymbol{\beta}'\boldsymbol{\beta} \to \mathbf{D}_\beta$ for $\alpha > 0$ and $\mu_1(\mathbf{\Sigma}_\epsilon) < \infty$.[14] Instead, we impose lower bounds on the norms of the columns of the granular loading matrix that are sufficient to carry out granular 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), Onatski (2010) and Onatski (2012). See also Pesaran (2012) and Chudik, Pesaran and Tosetti (2011) for more discussion on the distinction between weak and strong factors.

3.2 Granular panel model with common factors

We now consider the general version of the granular panel data model (2) in which the series are influenced by a set of common factors. The complete granular panel data model is given by

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

where f_t is an $r \times 1$ vector of common dynamic factor shocks and $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$ are the $k \times r$ and $(n-k) \times r$ loadings matrices. All other components are the same as in the previous section.

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

Without additional restrictions model (6) is not identified. The problem is similar when compared to the factor-augmented vector autoregressive model, see Bernanke, Boivin and Eliasz (2005) and Bai, Li and Lu (2016). In total we require $r^2 + rk$ additional identification restrictions (which are spelled out explicitly below) to pin down a rotation of the common factors.[15]

Additionally, the identification results of this section rely on two features of the granular model. First, they depend on the strength of the signal of the granular shocks g_t in the model 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 even when the series in the panel are influenced by common factors.

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 $\mathbf{L}_2 = [\boldsymbol{\beta}\ \mathbf{\Lambda}_2]$, which determine the variation in the granular series $y_{1:k,t}$

[15] 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.

and the non-granular series $y_{k+1:n,t}$, respectively.[16] 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 $\boldsymbol{\Lambda}_1$. Indeed we can easily verify that if $\boldsymbol{\Lambda}_1 = 0$ 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 easy to identify.

Based on these insights the following assumptions formalize the conditions that are sufficient for granular identification in this setting.

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* $\mathbf{D}_\lambda = \boldsymbol{\Lambda}_2'\boldsymbol{\Lambda}_2$ *with* $\mathbf{D}_\lambda$ *diagonal.*

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

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

(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* $\kappa_{u,\epsilon}\kappa_{\hat{\gamma}} s_g^{-1} < \sigma_k(\mathbf{D}_\beta) \leqslant \sigma_1(\mathbf{D}_\beta) < M_g$ *where* $\kappa_{u,\epsilon}$ *and* $\kappa_{\hat{\gamma}}$ *are the condition numbers of the matrices* $\hat{\mathbf{U}}\hat{\mathbf{U}}' + \boldsymbol{\Sigma}_\epsilon$ *and* $\hat{\boldsymbol{\gamma}}'\hat{\boldsymbol{\gamma}}$ *respectively.*

[16] 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 $\boldsymbol{\Sigma}_g^{-1/2} g_t$ 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 $\mathbf{L}_2 = [\boldsymbol{\beta}\ \boldsymbol{\Lambda}_2]$ 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 $\mu_r(\mathbf{D}_\lambda)$.[17] This implies that there can be correlation between the loadings of f_t and g_t, but the correlation cannot be too large. 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 $\hat{\gamma}$ and finally the idiosyncratic matrix $\mathbf{\Sigma}_\epsilon$ is replaced by $\hat{\mathbf{U}}\hat{\mathbf{U}}' + \mathbf{\Sigma}_\epsilon$.

We provide lemmas that extend the identification results established in the previous section for the baseline granular model to the case of a granular model with common factors.

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*

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

Analogously to the case of the granular model with no common factors, in order to establish the identification lemmas for the selection of the number of granulars we require a stronger version of Assumption 2 (iv).

(iv) Let $\mathbf{D}_\beta = \boldsymbol{\beta}'\boldsymbol{\beta}$ there exists an N and constant a constant M_g such that for each $n > N$*

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

we have that $\kappa_{u,\epsilon}\kappa_{\hat{\gamma}}^2 s_g^{-1}\left(\kappa_{u,\epsilon}+\frac{\mu_1(\hat{\mathbf{U}}\hat{\mathbf{U}}'+\boldsymbol{\Sigma}_\epsilon)}{\mu_k(\boldsymbol{\Sigma}_g+\boldsymbol{\Lambda}_1\boldsymbol{\Lambda}_1')}\right)<\sigma_k(\mathbf{D}_\beta)\leqslant\sigma_1(\mathbf{D}_\beta)<M_g$, *where* $\kappa_{u,\epsilon}$ *and* $\kappa_{\hat{\gamma}}$ *are the condition numbers of the matrices* $\hat{\mathbf{U}}\hat{\mathbf{U}}'+\boldsymbol{\Sigma}_\epsilon$ *and* $\hat{\gamma}'\hat{\gamma}$ *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 Assumptions 2-(*iv*). The interpretation of the bound remains the same as discussed above. Under the likely scenario where the condition number of $\hat{\mathbf{U}}\hat{\mathbf{U}}'+\boldsymbol{\Sigma}_\epsilon$ is larger than $\mu_1(\hat{\mathbf{U}}\hat{\mathbf{U}}'+\boldsymbol{\Sigma}_\epsilon)/\mu_k(\boldsymbol{\Sigma}_g+\boldsymbol{\Lambda}_1\boldsymbol{\Lambda}_1')$ the condition becomes $\sigma_k(\mathbf{D}_\beta)>2\kappa_{u,\epsilon}^2\kappa_{\hat{\gamma}}^2 s_g^{-1}$.

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 for* $n>N$ *and* $k>0$ *that*

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

A number of comments are in order. It is important to emphasize that in our framework the presence of factors does not alter the detection properties of the column norms of the concentration matrix provided that the identification assumptions hold. It is also important to emphasize that the results imply that it is not necessary to know the number of factors r to carry out granular detection. Last, note that the identification results established in this section can also be used to establish identification in case the granulars are contaminated by measurement error. In fact, if we think of the factors as measurement error and we restrict $\boldsymbol{\Lambda}_2=0$, then the results of this section immediately establish identification conditions in case of a granular model with the addition of measurement error on the granular series.

3.3 Estimation

We estimate the column norms of the concentration matrix $\|\mathbf{K}_i\|$ for each of the n series in the panel using a sample of T observations from the process y_t. Let $\hat{\boldsymbol{\Sigma}}$ denote the sample covariance matrix $T^{-1}\sum_{t=1}^T y_t y_t'$ and let $\hat{\mathbf{K}}$ denote the sample concentration matrix $\hat{\boldsymbol{\Sigma}}^{-1}$. A natural estimator of the granular statistic of series i is the norm of the i-th column of the sample concentration matrix, that is $\|\hat{\mathbf{K}}_i\|$.

We need to impose appropriate dependence and distributional assumptions on y_t in order to establish the large sample properties of our estimator. Notice that in what follows we formulate assumptions on the the sequence of isotropic random vectors $\mathbf{\Sigma}^{-1/2}y_t$.[18] Let $\mathcal{F}_{-\infty}^0$ and $\mathcal{F}_m^\infty$ 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

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

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 stated in assumption 3 are analogous to the dependence and distributional assumptions used in Fan et al. (2011). In particular, assumption (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. The latter condition is satisfied, for instance, by the multivariate Gaussian distribution but the assumption also allow for heavier tails. The parameter γ defined in assumption (iii) is a key quantity in this work as it measures the

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

degree of dependence and tail thickness of the data: The smaller the parameter the more dependent and thick tailed the data are. These conditions allow to apply a Bernstein-type inequality for mixing processes derived in Merlevede, Peligrad and Rio (2011), which is one of the main tools needed to establish the results of this section. Notice that we directly impose the assumptions on the observed series y_t instead of on f_t, g_t and ϵ_t separately. This is only for convenience and the results of this section may also be obtained by assuming that f_t, g_t and ϵ_t satisfy assumption 3 after being standardized appropriately.

We establish the following result concerning the sample covariance matrix.

Theorem 1. *Let y_t be generated by model* (6) *under assumptions 1, 2 and 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 $1 - O(n^{-\eta})$.

Theorem 1 (i) establishes a Bai-Yin-type law (Bai and Yin (1993)) for the largest and smallest eigenvalues of the sample covariance matrix of strongly mixing data with generalized-exponential tails. The theorem states that the eigenvalues of the sample covariance matrix are bounded away from zero and infinity when n and T are large. The proof of the theorem follows the arguments laid out in Vershynin (2012), with appropriate modifications for the present setting. Let us emphasize that the theorem extends the results of Vershynin (2012) for a set of assumptions that are more convenient for economic and financial applications. Theorem 1 part (i) facilitates the derivation of the subsequent parts (ii) to (iv).

We use Theorem 1 to establish two important results concerning the selection properties of the granular statistic $\|\hat{\mathbf{K}}_i\|$. First, in light of Lemma 3, it is natural to rank the series in the panel on the basis of the value of the granular statistic $\|\hat{\mathbf{K}}_i\|$. Define the event

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

that is the event that the granular statistics of the granular series are larger than the ones of the non-granular series. Then, the following corollary establishes that when n and T are large the probability of the event $\mathcal{E}_{\mathsf{R}}$ approaches one.

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

$$\mathrm{P}(\mathcal{E}_{\mathsf{R}}) \geqslant 1 - O(n^{-\eta}) \ .$$

In other words, the corollary shows that the granular statistic consistently ranks the granular series ahead of the non-granular ones. Second, in light of Lemma 4, it is natural to estimate the number of granular series 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 when n and T are large the probability of the event $\mathcal{E}_{\mathsf{S}}$ approaches one.

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

$$\mathrm{P}(\mathcal{E}_{\mathsf{S}}) \geqslant 1 - O(n^{-\eta}) \; .$$

A number of additional comments are in order. 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 paper 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.[19]

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, Liao and Liu (2016) for an overview of the relevant literature. 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. Therefore 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, Giannone and Reichlin (2012). Such results rely on weaker distributional assumptions than the ones spelled out in assumption 3. However they rely on stronger dependence assumptions.

[19] 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.

4 Comparison to other methods

In this section we compare our granular detection methodology with methods based on principal components and maximum likelihood, see Stock and Watson (2002b), Bai and Ng (2006), Parker and Sul (2016), Doz et al. (2012), Bai and Li (2016) and Jungbacker and Koopman (2015). We emphasize that none of these methods are specifically designed to detect granular series as they are defined in our setting. However, given that our model has a factor model representation it is not unreasonable to consider such methods for granular detection. It is important to note that none of the alternative methods exploit the partial correlation structure that is imposed by the granular model. The next sections explain in detail the strength and weakness of these alternative approaches.

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 analysis to give meaning to the otherwise hard to interpret estimates for common factors. They estimate the factors in an approximate factor model using principal components and subsequently use regression analysis to find the series that correlate most with the factors.

Upon first sight, it seems that these methods could be adopted to detect granular series as well and it is true that in some settings these methods will yield the same set of granular series when compared to detection based on the column norms. However, there are several important scenarios in which a principal components based method will not be able to detect the granular series.

First, we outline some obvious differences. Principal components based methods are two-step procedures whereas our column norm method requires only a single step. Also, any principal components method amounts to detecting the series that explain the most variance in the panel. This follows as the common factors from principal components are estimated to maximize the explained variance. Our definition of granular series, as formulated in assumption 1, does not necessarily imply those series that explain the most variance.

Second, it follows that principal components based methods perform well when the granulars explain a significant portion of the variance. At the same time, when the granular shocks have low signal-to-noise ratios principal components methods have low detection power. A simple example of this is the following. Consider the baseline granular model (2) with $k = 1$ and the following parametrization

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

where ι_n is an $n \times 1$ vector of ones and δ and c are constants. For this model the condition number of $\boldsymbol{\Sigma}_\epsilon$ is given by $\kappa_\epsilon = 1$ and the identification assumption 1-(iv) is satisfied when $\delta > 1$. In contrast, the detection power of the principal components method depends on the value of c. When c is large the first principal component will not correlate with the granular series.[20] Moreover, estimators for the number of factors, such as those developed 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 also present in the model. When these explain a large portion of the variance the principal components method primarily detects these and the subsequent regressions, that aim to find which series correlate most with the estimated factors, will detect also those series that load on the common factors. However, notice that this difficulty also affects 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 to be sufficiently strong, and the factor loading on the granular series $\boldsymbol{\Lambda}_1$ not to be too large.[21]

Third, recovering the *number* of granular series is difficult using any principal components based method. To see this consider again the simple example above, but now for the case

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

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

where c is small relative to δ. Clearly for large n, T the R^2 from the regression of the first principal component on the first series will tend 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 difficult, because the same principal component can also be generated by multiple correlated granular series.

More specifically, 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 $\mathbf{\Sigma}$. The column norm statistics are able to distinguish between these two scenarios with ease.

These observations are verified in the Monte Carlo study in the next section. There we confirm that the detection power of the principal components based methods does not perform satisfactorily under weak factor settings and when there are additional common factors in the model. Also, and of utmost importance, in the empirical section we show that the documented rankings of granular series can be vastly different for principal components and column norm methods.

4.2 Likelihood based methods

Doz et al. (2012), Bai and Li (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 straightforward method for granular detection would be to consider the granular model for different orderings of the variables in y_t and to subsequently compare appropriate goodness-of-fit statistics across the different orderings.

To outline the practical difficulty with this approach notice that for a given k this would involve estimating $\binom{n}{k}$ possible models. For $k = 1$ or $k = 2$ this approach is quite feasible. But for $n = 100$ and $k = 3$ this already involves estimating 161700 different models making this a computationally prohibitive task, see also Elliott, Gargano and Timmermann (2013). Potentially smart search algorithms could be considered but we do not explore this route further. Also, determining the number of granular series is difficult using such approach as different combinations of granular series and common factors can lead to observationally 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).[22] 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 parameter consistency in Doz et al. (2012) and Bai and Li (2012) are violated and the limiting distribution is unknown. We conclude that detecting granular series in a computationally efficient manner for large panels using likelihood based methods is currently infeasible.

5 Simulation Study

We perform a simulation study to assess the finite sample performance of our proposed methodology. We evaluate the performance of the detection methods based on the granular statistic $\|\hat{\mathbf{K}}_i\|$ under different data generating processes. The outcome criteria that we are interested in are as follows. First, we evaluate the fraction of the true granular series that correspond to the k largest granular statistics and second we consider the frequency by which we correctly select the number of granular series. We compare the performance of our

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

granular statistics to other methods that are based on principal components analysis.

5.1 Simulation design

We generate data panels from the granular panel data model with common factors given in equation (6). We consider data panels with dimensions $n = 50, 100$ and $T = 200, 400$. The number of granular series that we include is equal to $k = 3, 5$ and the number of common factors that we include is equal to $r = 0, 3, 5$.

The granular shocks and common factors follow the vector autoregressive process

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

The variance matrix $\mathbf{\Sigma}_\eta = \text{Var}(\eta_{f,t}, \eta_{g,t})$ has ones on the main diagonal and correlation coefficient c_η on the off-diagonal elements. We note that c_η captures the contemporaneous correlation among the granular shocks and the common factor shocks. We vary its value by taking $c_\eta = 0, 0.5$. 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 $\epsilon_t = \mathbf{\Sigma}_\epsilon^{1/2} e_t$ such that $\text{Var}(\epsilon_t) = \mathbf{\Sigma}_\epsilon$. For the latter we consider (a) diagonal, (b) banded and (c) sparse structures. For the diagonal structure we have $\mathbf{\Sigma}_{\epsilon,i,j}^{1/2} \sim U(0.5, 1.5)$ for all $i = j$ and zero else. For the banded structure we have $\mathbf{\Sigma}_{\epsilon,i,j} \sim U(0.5, 1.5)$ 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 else. Finally, the sparse structure is given by $\mathbf{\Sigma}_{\epsilon,i,j}^{1/2} \sim U(0.5, 1.5)$ if $i = j$, if $i > j$ $\mathbf{\Sigma}_{\epsilon,i,j}^{1/2} \sim U(-0.5, 0.5)$ with probability $0.2/[\sqrt{n-k}\log(n-k)]$ and zero else, and if $i < j$ the value is zero. Notice that this implies that for each case $\mathbf{\Sigma}_{\epsilon}^{1/2}$ is lower triangular and $\mathbf{\Sigma}_{\epsilon} = \mathbf{\Sigma}_{\epsilon}^{1/2}\mathbf{\Sigma}_{\epsilon}^{1/2\prime}$. The banded structure is similar as in Stock and Watson (2002a) and Bai and Ng (2002) whereas the sparse structure is similar as considered in Fan et al. (2011).

The influence of the granular shocks is determined by β. We vary the variance of the granular loadings in order to change the magnitude of their effect. In particular, we have $\beta_{i,j} \sim NID(0, \sigma_b^2)$, 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. Small values of σ_b^2 reflect the scenario where the common factors explain more variance in the observations when compared to the granular shocks. Such settings are argued to be empirically relevant in for example Foerster et al. (2011).

In total we have six dimensions along which we vary the granular panel data model: (i) panel dimensions, (ii) number of granulars, (iii) number of factors, (iv) effect of the granulars, (v) correlation among the granulars and factors and (vi) specification of the non-granular shocks. For each possible combination across these six dimension we draw $S = 1000$ different data panels.

5.2 Granular detection results

We begin by studying the finite sample properties of our ranking methods. Corollaries 1 and 2, that are based on the consistency of the column norms, imply that we can correctly identify the granular series when n and T become large. For each simulated panel we rank the series in the panel according to the column norms of the concentration matrix, and then we select the number of granulars using the column ratio statistic given in equation (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).

We summarize the performance of the detection procedure by reporting the average proportion of correctly ranked granular series and the proportion of correctly selected number of granulars. Given the large number of simulations considered, we only discuss the case where the non-granular errors have the banded and sparse designs. These cases are the most relevant for empirical applications. We only report the cases where the correlation between the granular shocks and the common factors is fixed at $c_\eta = 0.5$. Changing this coefficient has no effect on the detection results based on the column norm statistic. Finally, we only show $r = 0, 5$ and leave out the intermediate case where $r = 3$.

We present the results for the average proportion of correctly ranked granular series in Table 1 and the proportions of correctly selected number of granulars in Table 2. The tables reveal 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 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 for reasonably connected granular series we should expect to detect them 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. This confirms the identification results derived in Lemmas 3 and 4.

Clearly, the key parameter of our simulation setting is the standard deviation of granular standard deviation coefficients σ_b^2. We investigate its interaction with the amount of cross-sectional correlation in the non-granular shocks. These correlations are captured by c_ϵ in our simulation design for banded errors. In Tables 1 and 2 this value was fixed at 0.2 and we now vary it between 0 and 0.9. The identification lemmas 1 and 3 suggests that the interaction between σ_β^2 and c_ϵ is crucial.

Figure 2 reports the plots of the proportion of correctly ranked granulars and the proportion of correct selection of the number of granulars as a function of the granular strength (as measured by σ_β) and the non-granular shocks dependence (as measured by c_ϵ). The plots show that when the degree of dependence among non-granular shocks is weak, our granular detection methodology performs satisfactorily even when the strength of the granulars is modest. On the other hand, when the degree of dependence among non-granular shocks is strong our procedure requires the strength of the granular shocks to be much larger to detect granulars with sufficiently high probability.

Overall, the simulation study conveys that the granular detection methodology procedures performs satisfactorily in finite samples provided that the strength of the granulars is sufficiently large.

5.2.1 Comparison to Factor Model Based Methods

In this section we compare the performance of our methods with granular identification procedures that are based on principal components analysis, see for example Stock and Watson (2002b), Bai and Ng (2006) and Parker and Sul (2016).

Here we consider a straightforward implementation of such methods: (i) estimate the number of factors using Bai and Ng (2002), (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 Table 3 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 theoretical discussion in Section 4 where we illustrated that the factor based methods become difficult for weak granular series, see also Onatski (2012).

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, see the discussion in Section 4. Finally, we comment that more refined principal components based procedures, such as Parker and Sul (2016), did not lead to different results. This follows as their methods rely on the same initial R^2 ranking.

6 Empirical Applications

6.1 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 $n = 138$ different sectors for a total of $T = 431$ time

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

periods.[24] The panel is standardized such that each series has mean zero and unit variance.

A preliminary factor analysis conveys evidence of factors in the panel. Using the IC2 criteria proposed in Bai and Ng (2002) we find that there is one common factor in the panel. This is confirmed by the estimators proposed in Onatski (2010) and Ahn and Horenstein (2013) (see also Foerster et al. (2011) who find one or two factors for a similar panel).

We apply our methodology to detect the set of granular time series. We investigate the granular set for different sampling periods and compare our results to principal components based methods. Finally, we adopt standard methods to study the economic importance of the granular series.

6.1.1 Granular detection results

In the top panel of Figure 3 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 of the granular rankings which we report in the top panel of Table 4. 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 bottom panel of Figure 3 we show the column norm ratios. The estimator $\hat{k}$, given in equation 11, indicates that there are two granular series in our panel.

In summary, our granular detection methodology identifies a model with two granular series: "Motor Vehicle Parts" and "Automobiles and Light Duty Motor Vehicles".

[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.

6.1.2 Time dependence in granular detection

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 over the two sub-samples in the middle and bottom panels of Table 4. 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.

6.1.3 Comparison to principal components based methods

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 Table 5 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

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

components have difficulty detecting these, see the discussion in Section 4.

6.1.4 Influence of the granular series

We study the economic importance of the granular series using impulse response analysis, see Stock and Watson (2016). We take the set of granular series as given. In order to carry out the analysis we require estimates for the parameters of model (6). We model the granular shocks and common factors by a vector autoregressive process of order one, see equation (13), and rely on likelihood based methods to estimate the model parameters.[26] Since the granular series are highly correlated we consider the model with only one granular series: "Motor Vehicle Parts".[27] The number of common factors r is set to one.[28]

After imposing the normalizing restrictions of assumption 2-(i) all model parameters are identified and we estimate the parameters by maximum likelihood where the likelihood is constructed from the output of the Kalman filter. Details are given in Durbin and Koopman (2012, Chapter 4 and 7). The impulse responses from a one standard deviation shock to the granular series are computed as outlined in Stock and Watson (2016, Section 5). We compute the responses of all non-granular series in the panel and compute standard errors using the parametric bootstrap described in Stock and Watson (2016, Section 5.1.3). The impulse responses are interpretable as a shock to the automobile, or transportation, industry while controlling for general economic conditions via the common factor. It is important to emphasize that this shock is a reduced form shock and does not carry a causal interpretation.

In the left panels of Figure 5 we show the impulse responses for several series which are chosen based on the magnitude of the loadings $\boldsymbol{\beta}$. In particular, we show the impulse

[26] Specification with more elaborate dynamics for the granular shocks and common factors did not lead to different results. This is because we do our analysis for the growth rates of industrial production on a monthly frequency which implies that the persistence in the shocks is low.

[27] The unconditional correlation between the granular series "Motor Vehicle Parts" and "Automobiles and Light Duty Motor Vehicles" is 0.857. The results for a model with "Automobiles and Light Duty Motor Vehicles" as granular series are similar and therefore not presented. Alternatively, one could model both granular series with a scalar shock $\eta_{g,t}$.

[28] The choice for $r = 1$ is motivated by our preliminary factor analysis mentioned above. In particular, the IC2 criteria of Bai and Ng (2002) and the eigenvalue ratio criteria ofAhn and Horenstein (2013) estimated $r = 1$. We comment that the model with $r = 2$ common factors gave nearly identical results for the granular impulse responses.

response for the series that correspond to the 14th largest β_i (top 10%), the median β_i and the 124th largest β_i (bottom 10%). We find that in general the responses are short lived, and that within half a year the influence on the growth rates has vanished. This is due to the fact that the growth rates of industrial production have little persistence. Nevertheless, we find that the impulse response that corresponds to the largest 14th largest β_i is significantly positive and remain so for 4 months ahead. In total 65 out of 136 non-granular series have a significant positive response for the initial time period. This increases to 111 series for one-period ahead after which it decays gradually. For comparison purposes, in the right panels we also plot the responses of the same series to a shock to the common factor. It is clear that the influence of the common factor is overall much larger. For most series the magnitude of the impulse response for the common factor is more than twice as large on impact.

Next, having established that the granular series has a positive influence over a non-negligible part of the panel we investigate which sectors are most influenced. For this we plot in Figure 6 the impulse responses per sector level where the specific impulse response is chosen based on the median estimate of β_i within each sector. We find that the granular influence is highest for various textile sectors, wood products, furniture and fabricated metal products. These sectors rely heavily on transportation for both receiving intermediate goods and delivering final products. Interestingly, food, mining and electric and gas utilities are all not influenced by the granular shocks. These more primary consumption goods are less influenced by shocks to the automobile industry.

We conclude that there is evidence for granular influence in US industrial production. The influence comes through the automobile, or transportation, industry and has significant effects on several industries that depend on it.

6.2 Granular detection in the Eurozone financial system

We study the presence of granular series in a CDS spreads panel of Eurozone financial institutions. More precisely, we consider a panel of daily growth rates of CDS spreads for a set of financial entities across 10 Eurozone countries from January 3rd, 2006 to December 31, 2013. The panel contains $n = 69$ institutions for $T = 2080$ time periods. Table 6 contains the list of companies in the panel. CDS spreads are a market based measure of default risk that have been used for systemic risk analysis in a number of papers, like Ang and Longstaff (2013) and Oh and Patton (2017). Under appropriate assumptions (cf Oh and Patton (2017)), changes in CDS spreads reflect changes in the underlying probability of default of an institution. The panel is standardized such that each series has mean zero and unit variance.

The results of a preliminary factor analysis reveal that the panel also has a factor structure. In particular, the IC2 criteria selects one factor for the panel. Principal component analysis shows that the first component explains 28.8% of the total variability whereas the second component explains 3.7%.

Analogously to the previous section, we apply our methodology to detect the granulars in the panel and then study granular influence using impulse response analysis. We also report some robustness checks regarding stability of the rankings over different subsamples and a comparison with principal components.

6.2.1 Granular detection results

We show the granular statistics in the top panel of Figure 4. The plot shows that the top two series in the rankings have a rather large value for the granular statistic. Interestingly, these are the two largest Spanish banks: Banco Santander and BBVA. It is important to recall that following the burst of the Spanish housing bubble and the beginning of the Great Financial Crisis the Spanish financial system has been in great distress which in turn spilled over other financial institutions in the Eurozone. The worsening of the distress following the European sovereign debt crisis forced the Spanish government to apply for a European rescue package to recapitalize its financial system. Besides, these two banks the top 10 is also dominated by large German and French financial institutions. In the bottom panel of Figure 4 we show the column ratio statistics. Their estimate for the number of granulars is equal to two.

6.2.2 Time dependence in granular detection

It is interesting to assess the stability of the rankings over different subsets. In particular, it is interesting to investigate granularity rankings before the beginning of the Great Financial Crisis of 2008–2009 as well as the European sovereign debt crisis of 2010–2012. We report these in the middle and bottom panels of Table 4. The rankings show that Banco Santander and BBVA are consistently ranked as top granular institutions. The number of selected granular institutions selected by the column ratio statistic varies depending on the sample. In the January 2006 to September 2008 sample, the list of selected granulars contains BBVA only, while in the January 2006 to April 2010 the selected granulars are Banco Santander, BBVA, Assicurazioni Generali and Allianz.

6.2.3 Comparison to principal components based methods

We also construct granularity rankings on the basis of principal components, following the procedure described in the previous section. We report the rankings over the full sample as well as two subsamples in Table 5. Interestingly, for this application our procedure and principal components deliver substantially close rankings. In particular, Banco Santander and BBVA are identified, respectively, as the first and third most granular institutions in all samples. More generally, the rank correlation between the two sets of rankings is above 0.90 in all samples considered. A possible explanation for the large overlap is that the granular series explain a relatively large proportion of variability in the panel, see the discussion in Section 4. We investigate this explanation further below.

6.2.4 Influence of the granular series

Last, we carry out an impulse response analysis to study the influence of granular institutions on the rest of the Eurozone financial system. In particular, in this application we focus on the impact of Santander. In order to construct impulse response functions, we estimate the granular panel data model (6) using the same steps outlined in the industrial production application. In particular, the granular panel specification we estimate considers one factor and one granular series (Santander).

In the left panels of Figure 7 we show the impulse responses for different series chosen on the basis of the magnitude of the loadings $\boldsymbol{\beta}$. Specifically, we report the impulse response for the series that correspond to the 7th largest β_i (top 10%), the median β_i and the 63rd largest β_i (bottom 10%). For comparison purposes, in the right panels of the figure we plot the responses of the same series to a shock to the common factor. We find that responses are short lived and that after a few days the influence of the granular shock becomes insignificant. This is to be expected given the weak persistence of the CDS growth rates. That said, the effect of the granular shock is significant for all series in the panel and its magnitude is comparable, yet weaker, to the one of the factor.

In Figure 8 we plot impulse responses per country where the specific impulse response is chosen based on the median estimate of β_i within each country. We find that the granular influence is highest for financial entities in the Eurozone periphery, in particular Ireland, Italy and Portugal. This is consistent with the fact that during the European Sovereign debt crisis Eurozone periphery financial institutions experienced strong spillover effects among them, in particular Spanish and Italian banks. We note that the impact of the granular shock on Greek banks is not particularly strong. However, it has to be noted that throughout the sample Greek CDSs have a large variability and weak correlation with all other series in the panel.

7 Conclusion

In this work we introduce a panel model in which the idiosyncratic shocks of a (finite) 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. Two empirical applications are used to showcase our methodology. In the first application, we analyze granualarity in US industrial production sectors. Results show that several sectors in the automobile industry, in particular "Motor Vehicle Parts" are selected as granular. In the second application we study granularity in a panel of CDS spreads of Eurozone financial institutions. In this case our methodology detect as granulars the two largest Spanish banks: Banco Santander and BBVA.

References

Acemoglu, D., Carvalho, V. M., Ozdaglar, A. and Tahbaz-Salehi, A.: 2012, The Network Origens of Aggregate Fluctuations, *Econometrica* **80**, 1977–2016.

Acemoglu, D., Ozdaglar, A. and Tahbaz-Salehi, A.: 2015, Systemic Risk and Stability in Financial Networks, *American Economic Review* **105**, 564–608.

Adrian, T. and Brunnermeier, M. K.: 2016, CoVaR, *American Economic Review* **106**, 1705–1741.

Ahn, S. C. and Horenstein, A. R.: 2013, Eigenvalue Ratio Test for the Number of Factors, *Econometrica* **80**, 1203–1227.

Ang, A. and Longstaff, F.: 2013, Systemic Sovereign Credit Risk: Lessons from the U.S. and Europe, *Journal of Monetary Economics* **60**, 493–510.

Ansley, C. F. and Kohn, R.: 1986, A note on reparameterizing a vector autoregressive moving average model to enforce stationarity, *Journal of Statistical Computation and Simulation* **24**, 99–106.

Atalay, E.: 2017, How Important Are Sectoral Shocks?, *American Economic Journal: Macroeconomics* . forthcoming.

Bai, J.: 2003, Inferential Theory for Factor Models of Large Dimensions, *Econometrica* **71**, 135–171.

Bai, J. and Li, K.: 2012, Statistical Analysis of Factor Models of High Dimension, *The Annals of Statistics* **40**, 436–465.

Bai, J. and Li, K.: 2016, Maximum Likelihood Estimation and Inference for Approximate Factor Models of High Dimension, *Review of Economics and Statistics* **98**, 298–309.

Bai, J., Li, K. and Lu, L.: 2016, Estimation and inference of FAVAR models, *Journal of Business and Economic Statistics* **34**, 620–641.

Bai, J. and Ng, S.: 2002, Determining the Number of Factors in Approximate Factor Models, *Econometrica* **70**, 191–221.

Bai, J. and Ng, S.: 2006, Evaluating latent and observed factors in macroeconomics and finance, *Journal of Econometrics* **131**, 507–537.

Bai, Z. and Yin, Y.: 1993, Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix, *Annals of Probability* **21**, 1275–1294.

Bernanke, B. S., Boivin, J. and Eliasz, P.: 2005, Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach, *Quarterly Journal of Economics* **120**, 387–422.

Bernard, A. B., Jensen, J. B., Redding, S. J. and Schott, P. K.: 2017, Global Firms, *Journal of Economic Literature* . forthcoming.

Billio, M., Getmansky, M., Lo, A. and Pellizzon, L.: 2012, Econometric measures of connectedness and systemic risk in the finance and insurance sectors, *Journal of Financial Economics* **104**, 535–559.

Brownlees, C. and Engle, R.: 2017, SRISK: A Conditional Capital Shortfall Measure of Systemic Risk, *Review of Financial Studies* pp. 48–79.

Carvalho, V. and Gabaix, X.: 2013, The Great Diversification and its Undoing, *American Economic Review* **103**, 1697–1727.

Cavicchioli, M., Forni, M., Lippi, M. and Zaffaroni, P.: 2016, Eigenvalue Ratio Estimators for the Number of Dynamic Factors. Working paper.

Chudik, A., Pesaran, M. H. and Tosetti, E.: 2011, Testing Weak Cross-Sectional Dependence in Large Panels, *The Econometrics Journal* **14**, 45–90.

Di Giovanni, J. and Levchenko, A. A.: 2012, International Trade and Aggregate Fluctuations in Granular Economies, *Journal of Political Economy* **120**, 1083–1132.

Di Giovanni, J., Levchenko, A. A. and Mejean, I.: 2014, Firms, Destinations and Aggregate Fluctuations, *Econometrica* **82**, 1303–1340.

Diebold, F. X. and Yılmaz, K.: 2014, On the network topology of variance decompositions: Measuring the connectedness of financial firms, *Journal of Econometrics* **182**, 119–134.

Doz, C., Giannone, D. and Reichlin, L.: 2012, A Quasi Maximum Likelihood Approach for Large Approximate Dynamic Factor Models, *Review of Economics and Statistics* **94**, 1014–1024.

Durbin, J. and Koopman, S. J.: 2012, *Time Series Analysis by State Space Methods; 2nd Edition*, Oxford University Press, Oxford.

Elliott, G., Gargano, A. and Timmermann, A.: 2013, Complete Subset Regressions, *Journal of Econometrics* **177**, 357–373.

Fan, J., Liao, Y. and Liu, H.: 2016, An overview of the estimation of large covariance and precision matrices, *Econometrics Journal* **19**, 1–32.

Fan, J., Liao, Y. and Mincheva, M.: 2011, High-Dimensional Covariance Matrix Estimation in Approximate Factor Models, *Annals of Statistics* **39**, 3320–3356.

Foerster, A. T., Sarte, P. D. G. and Watson, M. W.: 2011, Sectoral versus Aggregate Shocks: A Structural Factor Analaysis of Industrial Production, *Journal of Political Economy* **119**, 1–38.

Forni, M. and Reichlin, L.: 1998, Lets Get Real: A Dynamic Factor Analytical Approach to Disaggregated Business Cycle, *Review of Economic Studies* **65**, 453–474.

Gabaix, X.: 2011, The Granular Origens of Aggregate Fluctuations, *Econometrica* **79**, 733–772.

Gaubert, C. and Itskhoki, O.: 2016, Granular Comparative Advantage. Working paper.

Ghysels, E., Liuy, H. and Raymondz, S.: 2017, Granularity and Risk in Equity Markets, *Technical report*.

Hautsch, N., Schaumburg, J. and Schienle, M.: 2015, Financial network systemic risk contributions, *Review of Finance* p. 685738.

Heywood, H. B.: 1931, On Finite Sequences of Real Numbers, *Proceedings of the Royal Society A* **134**, 486–501.

Horn, R. A. and Johnson, C. R.: 2013, *Matrix Analysis - 2nd edition*, Cambridge University Press, Cambridge.

Jungbacker, B. and Koopman, S.: 2015, Likelihood-based Analysis for Dynamic Factor Analysis for Measurement and Forecasting, *Econometrics Journal* **18**, C1–C21.

Kolenikov, S. and Bollen, K. A.: 2012, Testing Negative Error Variances. Is a Heywood Case a Symptom of Misspecification?, *Sociological Methods and Research* **41**, 124–167.

Lauritzen, S. L.: 1996, *Graphical Models*, Oxford University Press, Oxford.

Long, J. and Plosser, C.: 1987, Sectoral vs. Aggregate Shocks In The Business Cycle, *American Economic Review* **77**, 333–336.

Meinshausen, N. and Bühlmann, P.: 2006, High dimensional graphs and variable selection with the lasso, *The Annals of Statistics* **34**, 1436–1462.

Merlevede, F., Peligrad, M. and Rio, E.: 2011, A bernstein type inequality and moderate deviations for weakly dependent sequences, *Probability Thoery and Related Fields* **151**, 435–474.

Mulally, A. R.: 2008, Examining the State of the Domestic Automobile Industry. Hearing, United States Senate Committee on Banking, Housing and Urban Affairs.

Oh, D. H. and Patton, A. J.: 2017, Time-varying systemic risk: Evidence from a dynamic copula model of cds spreads, *Journal of Business & Economic Statistics* . forthcoming.

Onatski, A.: 2009, Testing Hypotheses about the Number of Factors in Large Factor Models, *Econometrica* **77**, 1447–1479.

Onatski, A.: 2010, Determining the Number of Factors from Empirical Distribution of Eigenvalues, *Review of Economics and Statistics* **92**, 1004–1016.

Onatski, A.: 2012, Asymptotics of the Principal Components Estimator of Large Factor Models with Weakly Influential Factors, *Journal of Econometrics* **168**, 244–258.

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

Peng, J., Wang, P., Zhou, N. and Zhu, J.: 2009, Partial correlation estimation by joint sparse regression models, *Journal of the American Statistical Association* **104**, 735–746.

Pesaran, M. H.: 2012, Testing Weak Cross-Sectional Dependence in Large Panels, *Econometric Reviews* **34**, 1089–1117.

Pesaran, M. H. and 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. and Watson, M. W.: 2002a, Forecasting Using Principal Components From a Large Number of Predictors, *Journal of the American Statistical Association* **97**, 1167–1179.

Stock, J. H. and Watson, M. W.: 2002b, Macroeconomic Forecasting Using Diffusion Indexes, *Journal of Business and Economic Statistics* **220**, 147–162.

Stock, J. H. and Watson, M. W.: 2016, Factor Models and Structural Vector Autoregressions in Macroeconomics, *in* J. B. Taylor and H. Uhlig (eds), *Handbook of Macroeconomics*, Elsevier, Amsterdam, North Holland.

Vershynin, R.: 2012, Introduction to the non-asymptotic analysis of random matrices, *Compressed Sensing*, Cambridge University Press, Cambridge, pp. 210–268.

White, H.: 2000, *Asymptotic Theory for Econometrians - 2nd edition*, Academic Press, San Diego, California.

Table 1: Granular Ranking Probabilities

n	T	k	r	0.01	0.05	0.10	0.25	0.50	0.75	1.00
Banded error design										
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
Sparse error design										
50	200	3	0	0.201	0.809	0.968	0.999	1.000	1.000	1.000
100	200	3	0	0.225	0.937	0.995	1.000	1.000	1.000	1.000
50	400	3	0	0.214	0.931	0.998	1.000	1.000	1.000	1.000
100	400	3	0	0.312	0.992	1.000	1.000	1.000	1.000	1.000
50	200	5	0	0.269	0.816	0.956	0.998	1.000	1.000	1.000
100	200	5	0	0.293	0.931	0.996	1.000	1.000	1.000	1.000
50	400	5	0	0.284	0.917	0.993	1.000	1.000	1.000	1.000
100	400	5	0	0.373	0.989	1.000	1.000	1.000	1.000	1.000
50	200	3	5	0.138	0.695	0.872	0.966	0.982	0.982	0.986
100	200	3	5	0.177	0.859	0.977	0.996	0.997	0.998	0.998
50	400	3	5	0.147	0.832	0.954	0.989	0.994	0.996	0.996
100	400	3	5	0.232	0.967	0.997	1.000	1.000	1.000	1.000
50	200	5	5	0.214	0.733	0.891	0.972	0.987	0.988	0.991
100	200	5	5	0.241	0.881	0.971	0.997	0.999	0.999	0.999
50	400	5	5	0.222	0.839	0.953	0.991	0.996	0.997	0.997
100	400	5	5	0.311	0.963	0.997	0.999	1.000	1.000	1.000

The table reports the average proportion of correctly ranked granulars.

Table 2: Granular Selection Probabilities

n	T	k	r	0.01	0.05	0.10	0.25	0.50	0.75	1.00
						Banded error design				
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
						Sparse error design				
50	200	3	0	0.105	0.201	0.520	0.909	0.977	0.990	0.994
100	200	3	0	0.095	0.409	0.790	0.983	0.997	1.000	1.000
50	400	3	0	0.085	0.397	0.844	0.989	0.997	1.000	1.000
100	400	3	0	0.118	0.739	0.979	1.000	1.000	1.000	1.000
50	200	5	0	0.042	0.117	0.392	0.813	0.948	0.978	0.992
100	200	5	0	0.047	0.248	0.766	0.968	0.996	0.999	1.000
50	400	5	0	0.047	0.248	0.669	0.953	0.988	1.000	0.997
100	400	5	0	0.046	0.632	0.960	0.999	1.000	1.000	1.000
50	200	3	5	0.128	0.156	0.264	0.528	0.608	0.576	0.603
100	200	3	5	0.092	0.231	0.539	0.760	0.778	0.778	0.806
50	400	3	5	0.115	0.205	0.460	0.688	0.738	0.715	0.694
100	400	3	5	0.121	0.510	0.792	0.892	0.884	0.867	0.873
50	200	5	5	0.047	0.087	0.166	0.429	0.565	0.636	0.637
100	200	5	5	0.036	0.176	0.467	0.750	0.832	0.848	0.824
50	400	5	5	0.038	0.131	0.315	0.668	0.727	0.737	0.753
100	400	5	5	0.036	0.426	0.761	0.907	0.913	0.913	0.928

The table reports the average proportion of times the correct number of granulars is selected.

Table 3: Granular Ranking Probabilities: R_i^2 versus $\|\mathbf{K}_i\|$

n	T	k	r	0.01	0.05	0.10	0.25	0.50	0.75	1.00
Banded error design										
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
Sparse error design										
50	200	3	0	0.685	0.622	0.692	0.704	0.764	0.850	0.908
100	200	3	0	0.524	0.524	0.653	0.556	0.485	0.529	0.621
50	400	3	0	0.765	0.775	0.820	0.840	0.890	0.926	0.948
100	400	3	0	0.600	0.690	0.659	0.553	0.486	0.533	0.579
50	200	5	0	0.807	0.619	0.792	0.830	0.902	0.940	0.953
100	200	5	0	0.508	0.790	0.894	0.928	0.951	0.977	0.981
50	400	5	0	0.627	0.543	0.739	0.661	0.664	0.768	0.848
100	400	5	0	0.706	0.793	0.777	0.668	0.673	0.707	0.773
50	200	3	5	0.390	0.460	0.712	0.904	0.964	0.967	0.966
100	200	3	5	0.143	0.465	0.788	0.963	0.986	0.984	0.987
50	400	3	5	0.306	0.355	0.645	0.830	0.859	0.871	0.849
100	400	3	5	0.102	0.440	0.685	0.794	0.830	0.826	0.823
50	200	5	5	0.451	0.535	0.744	0.917	0.965	0.968	0.964
100	200	5	5	0.184	0.472	0.792	0.969	0.987	0.989	0.993
50	400	5	5	0.369	0.507	0.755	0.890	0.915	0.910	0.903
100	400	5	5	0.129	0.539	0.794	0.872	0.866	0.860	0.850

The table reports the ratio between the average proportion of correctly ranked granulars based on the R2 statistic and the the column norm statistic.

Table 4: Granular series US Industrial Production

1972 – 2007

Sector	$\|\hat{\mathbf{K}}_{(i)}\|$	$\frac{\|\hat{\mathbf{K}}_{(i)}\|}{\|\hat{\mathbf{K}}_{(i+1)}\|}$
Motor Vehicle Parts	11.284	1.161
Automobiles and Light Duty Motor Vehicles	9.718	1.603
Aluminum Extruded Products	6.062	1.004
Plastics Products	6.039	1.004
Miscellaneous Aluminum Materials	6.013	1.049
Motor Vehicle Bodies	5.732	1.076
Paper and Paperboard Mills	5.328	1.095
Household and Institutional Furniture and Kitchen Cabinets	4.867	1.067
Commercial and Service Industry Machines	4.561	1.016
Motor Homes	4.490	1.011

1972 – 1983

Sector	$\|\hat{\mathbf{K}}_{(i)}\|$	$\frac{\|\hat{\mathbf{K}}_{(i)}\|}{\|\hat{\mathbf{K}}_{(i+1)}\|}$
Motor Vehicle Parts	4546.027	1.150
Household and Institutional Furniture and Kitchen Cabinets	3953.440	1.340
Plastics Products	2949.915	1.083
Commercial and Service Industry Machines	2722.886	1.069
Automobiles and Light Duty Motor Vehicles	2547.430	1.004
Foundries	2538.000	1.057
Organic Chemicals	2401.269	1.170
Semiconductors and Other Electronic Components	2053.029	1.027
Farm Machinery and Equipment	1998.120	1.022
Animal Slaughtering and Meat Processing Ex Poultry	1954.887	1.036

1984 – 2007

Sector	$\|\hat{\mathbf{K}}_{(i)}\|$	$\frac{\|\hat{\mathbf{K}}_{(i)}\|}{\|\hat{\mathbf{K}}_{(i+1)}\|}$
Motor Vehicle Parts	26.253	1.039
Automobiles and Light Duty Motor Vehicles	25.261	1.356
Aluminum Extruded Products	18.626	1.014
Miscellaneous Aluminum Materials	18.365	1.183
Motor Vehicle Bodies	15.530	1.455
Truck Trailers	10.675	1.009
Carpet and Rug Mills	10.575	1.010
Paper and Paperboard Mills	10.474	1.064
Motor Homes	9.843	1.034
Concrete and Products	9.519	1.035

The table reports the ranking of granular series for US industrial production.

Table 5: GRANULAR SERIES US INDUSTRIAL PRODUCTION BASED ON R^2

1972 – 2007

Sector	R^2
Plastics Products	0.651
Household and Institutional Furniture and Kitchen Cabinets	0.520
Metal Valves Except Ball and Roller Bearings	0.462
Architectural and Structural Metal Products	0.448
Other Miscellaneous Manufacturing	0.441
Fabricated Metals Spring and Wire Products	0.422
Commercial and Service Industry Machines	0.405
Fabricated Metals Forging and Stamping	0.402
Coating Engraving Heat Treating and Allied Activities	0.355
Other Textile Product Mills	0.332

1972 – 1983

Sector	R^2
Plastics Products	0.733
Household and Institutional Furniture and Kitchen Cabinets	0.634
Metal Valves Except Ball and Roller Bearings	0.626
Architectural and Structural Metal Products	0.597
Fabricated Metals Spring and Wire Products	0.588
Other Miscellaneous Manufacturing	0.534
Other Textile Product Mills	0.486
Fabricated Metals Forging and Stamping	0.486
Plastics Materials and Resins	0.485
Communication and Energy Wires and Cables	0.481

1984 – 2007

Sector	R^2
Plastics Products	0.407
Commercial and Service Industry Machines	0.368
Architectural and Structural Metal Products	0.328
Other Miscellaneous Manufacturing	0.327
Household and Institutional Furniture and Kitchen Cabinets	0.301
Coating Engraving Heat Treating and Allied Activities	0.294
Fabricated Metals Forging and Stamping	0.265
Metalworking Machinery	0.232
Metal Valves Except Ball and Roller Bearings	0.217
Fabricated Metals Spring and Wire Products	0.215

The table reports the ranking of granular series for US industrial production.

Table 6: Eurozone Financial Institutions

Country	Abbrev.	Name	Country	Abbrev.	Name
Austria	WAG	Austria		INT	Intesa Sanpaolo
	ERS	Erste Bank Group		MIL	Banca Popolare di Milano
	RAI	Raiffeisen Bank International		MPS	Banca Monte dei Paschi di Siena
Belgium	FOR	Fortis / Ageas Holding		POP	Banco Popolare S.C.
	KBC	KBC Bank		UDP	Unione di Banche Italiane SCPA
France	DEX	Dexia Crédit Local		UNI	UniCredit
	AXA	AXA France	Ireland	ANG	Anglo Irish Bank
	BQE	Banque Fédérative du Crédit Mutuel		DEP	Depfa PLC
	PEU	Banque PSA Finance		GOV	Governor and Company of the Bank of Ireland
	BNP	BNP Paribas		NAT	Irish Nationwide Bank
	AGR	Crédit Agricole		ILP	Irish Life and Permanent
	CIC	Crédit Industriel et Comercial	Netherland	AEG	Aegon NNV
	SOC	Société Générale		NIB	NIBC Bank
	WEN	Wendel		RAB	Rabobank
	GEC	Cecine		SNS	SNS Bank
	KLE	Klepierre		VAN	F. van Lanschot Bankiers
	SOP	Sophia GE		ING	Ing Bank
Germany	ALL	Allianz		ABN	Abn Amro Bank
	DBA	Deutsche Bank		ROD	Rodamco Europe
	COM	Commerzbank		ACH	Achmea Holding
	DBZ	DZ Bank	Portugal	BCO	Banco Comercial Portugues
	MRV	Münchner Rückversicherung		CAI	Caixa Geral de Depositos
	NLB	Norddeutsche Landesbank		BPI	Banco Portugues de Investimiento
	HSH	HSH Nordbank		SAN	Espirito Santo Financial Group
	LBW	Landesbank Baden-Württemberg	Spain	PAS	Banco Pastor
	BLB	Bayerische Landesbank		SAB	Banco de Sabadell
	LBB	Landesbank Berlin		MED	Caja de Ajorros del Mediterraneo
	LHT	Landesbank Hessen - Thühringen		INT	Bankinter
	WLB	West LB / Portigon AG		CAV	Caixa de Ahorros de Valencia
Greece	NBG	National Bank of Greece		SAN	Banco Santander
	EFG	EFG Eurobank Ergasias		BBV	Banco Bilbao Vizcaya Argentaria
Italy	ASG	Assicurazioni Generali		CAB	Caja de Ahorros y Pensiones de Barcelona
	LAV	Banca Nazionale de Lavoro		MPM	Caja de Ahorros y Monde de Piedad de Madrid
	LEA	Banca Italease		POP	Banco Popular Espanol
	MED	Mediobanca			

The table reports the list of entity names, abbreviations and countries of the series in the CDS panel.

Table 7: Granular series Eurozone Financials CDS

January 2006 – December 2013

Entity	$\|\hat{\mathbf{K}}_{(i)}\|$	$\frac{\|\hat{\mathbf{K}}_{(i)}\|}{\|\hat{\mathbf{K}}_{(i+1)}\|}$
SAN (es)	6.176	1.066
BBV (es)	5.796	1.224
ASG (it)	4.737	1.004
ALL (de)	4.716	1.065
MRV (de)	4.430	1.056
BNP (fr)	4.197	1.004
SOC (fr)	4.180	1.005
AXA (fr)	4.158	1.011
AGR (fr)	4.113	1.004
INT (it)	4.098	1.059

January 2006 – September 2008

Entity	$\|\hat{\mathbf{K}}_{(i)}\|$	$\frac{\|\hat{\mathbf{K}}_{(i)}\|}{\|\hat{\mathbf{K}}_{(i+1)}\|}$
SAN (es)	6.767	1.127
ASG (it)	6.007	1.017
BBV (es)	5.908	1.017
ALL (de)	5.811	1.121
MRV (de)	5.183	1.094
COM (de)	4.739	1.019
AXA (fr)	4.652	1.051
AEG (nl)	4.428	1.014
AGR (fr)	4.365	1.001
DBA (de)	4.359	1.020

January 2006 – April 2010

Entity	$\|\hat{\mathbf{K}}_{(i)}\|$	$\frac{\|\hat{\mathbf{K}}_{(i)}\|}{\|\hat{\mathbf{K}}_{(i+1)}\|}$
SAN (es)	7.319	1.056
BBV (es)	6.934	1.143
ASG (it)	6.069	1.007
ALL (de)	6.028	1.282
AXA (fr)	4.702	1.008
AGR (fr)	4.665	1.015
SAN (pt)	4.596	1.020
INT (it)	4.505	1.002
MRV (de)	4.498	1.025
BCO (pt)	4.387	1.057

The table reports the ranking of granular series for the Eurozone Financial CDS panel.

Table 8: Granular series Eurozone Financials CDS based on R^2

January 2006 – December 2013

Sector	R^2
SAN (es)	0.725
AGR (fr)	0.706
BBV (es)	0.705
ASG (it)	0.697
SOC (fr)	0.693
INT (it)	0.691
MPS (it)	0.691
BNP (fr)	0.689
ALL (de)	0.673
DBA (de)	0.672

January 2006 – September 2008

Sector	R^2
SAN (es)	0.701
MRV (de)	0.693
BBV (es)	0.692
ASG (it)	0.692
AGR (fr)	0.688
SOC (fr)	0.668
DBA (de)	0.667
ALL (de)	0.667
POP (it)	0.647
COM (de)	0.645

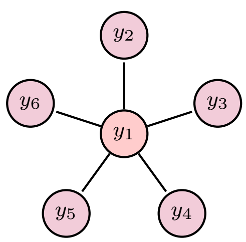
January 2006 – April 2010

Sector	R^2
SAN (es)	0.726
AGR (fr)	0.726
BBV (es)	0.721
ASG (it)	0.689
MPS (it)	0.686
DBA (de)	0.676
SOC (fr)	0.676
INT (it)	0.676
ALL (de)	0.665
COM (de)	0.659

The table reports the ranking of granular series for the Eurozone Financial CDS panel.

Figure 1: PARTIAL CORRELATION NETWORK REPRESENTATION

$$
\begin{aligned}
y_{1\,t} &= g_t & g_t &\sim D(0,1) \\
y_{2:6,t} &= \beta g_t + \epsilon_t & \epsilon_t &\sim D(0,\mathbf{I}_5)
\end{aligned}
$$

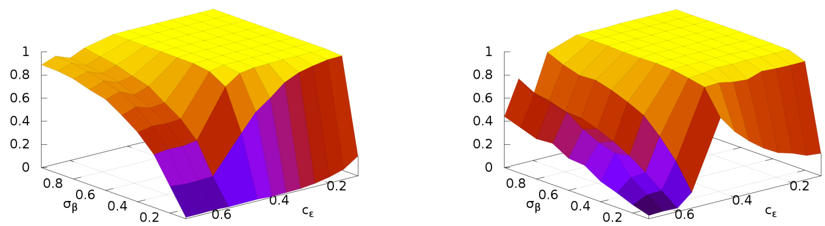


$$
\mathbf{K} = \begin{bmatrix}
1+\beta'\beta & -\beta_1 & -\beta_2 & -\beta_3 & -\beta_4 & -\beta_5 \\
-\beta_1 & 1 & 0 & 0 & 0 & 0 \\
-\beta_2 & 0 & 1 & 0 & 0 & 0 \\
-\beta_3 & 0 & 0 & 1 & 0 & 0 \\
-\beta_4 & 0 & 0 & 0 & 1 & 0 \\
-\beta_5 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
$$

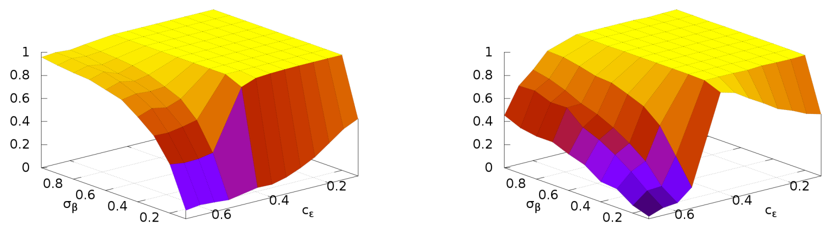
Partial Correlation Network

Concentration Matrix

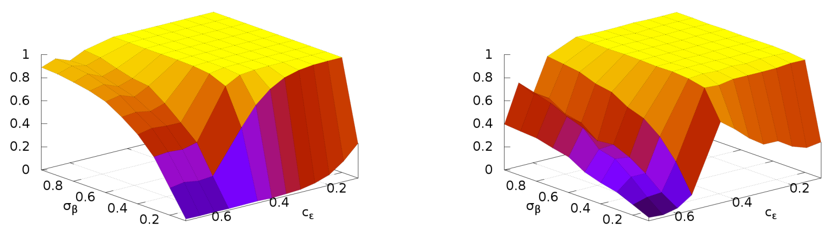
Figure 2: Granular Ranking and Selection Probabilities



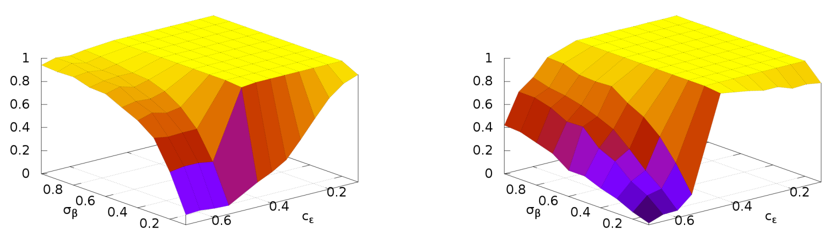
$T = 200 \quad N = 50$



$T = 200 \quad N = 100$



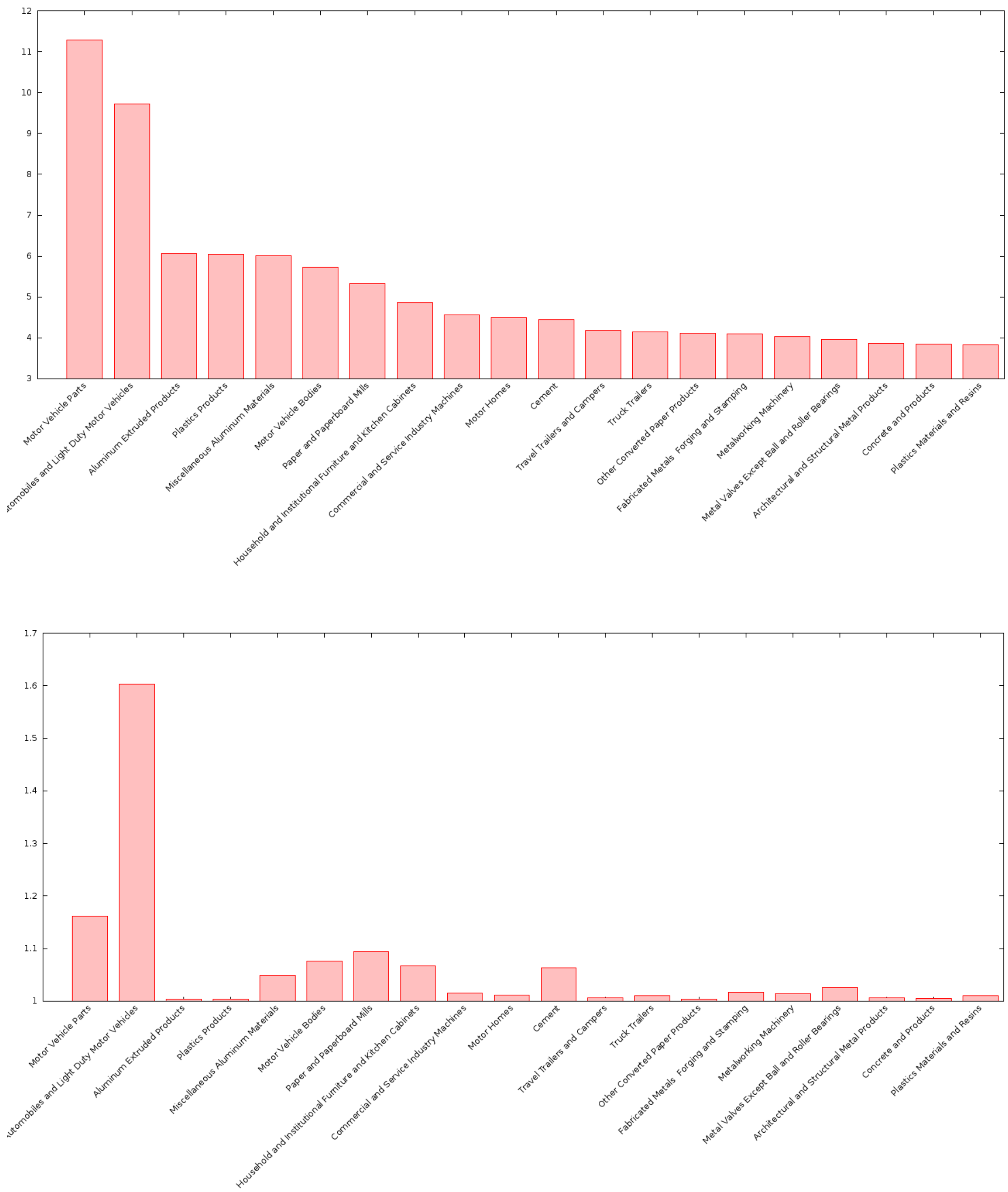
$T = 400 \quad N = 50$



$T = 400 \quad N = 100$

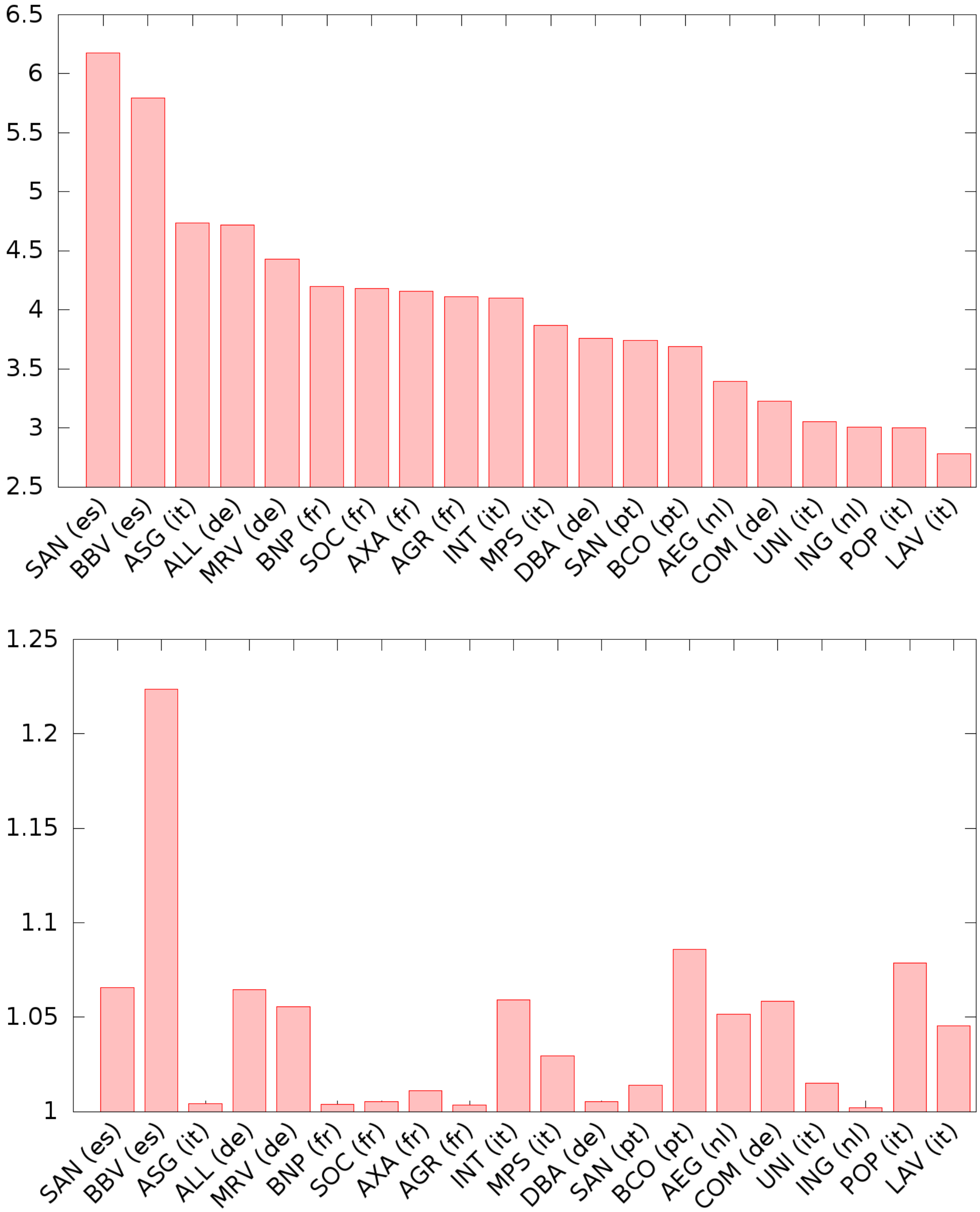
The figure shows the proportion of correctly ranked granulars (left panel) and of correctly selecting the number of granulars (right panel) as a function of the standard deviation of the granular loadings σ_b and the coefficient controlling the degree of dependence of the idiosyncratic shocks c_ϵ.

Figure 3: Granular detection results for the IP series



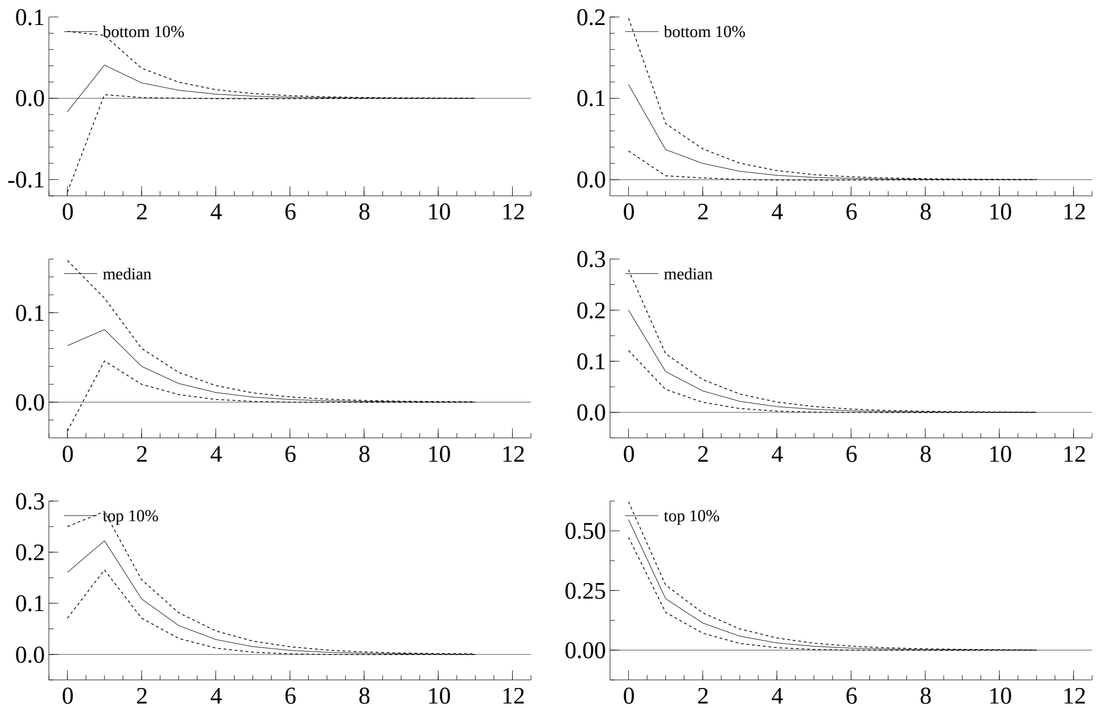
The figure shows the ordered column norms (top panel) and the column norm ratios (bottom panel) for a panel of $n = 138$ industrial production monthly growth rates from 1972 until 2007 ($T = 431$).

Figure 4: Granular detection results for the CDS spreads series



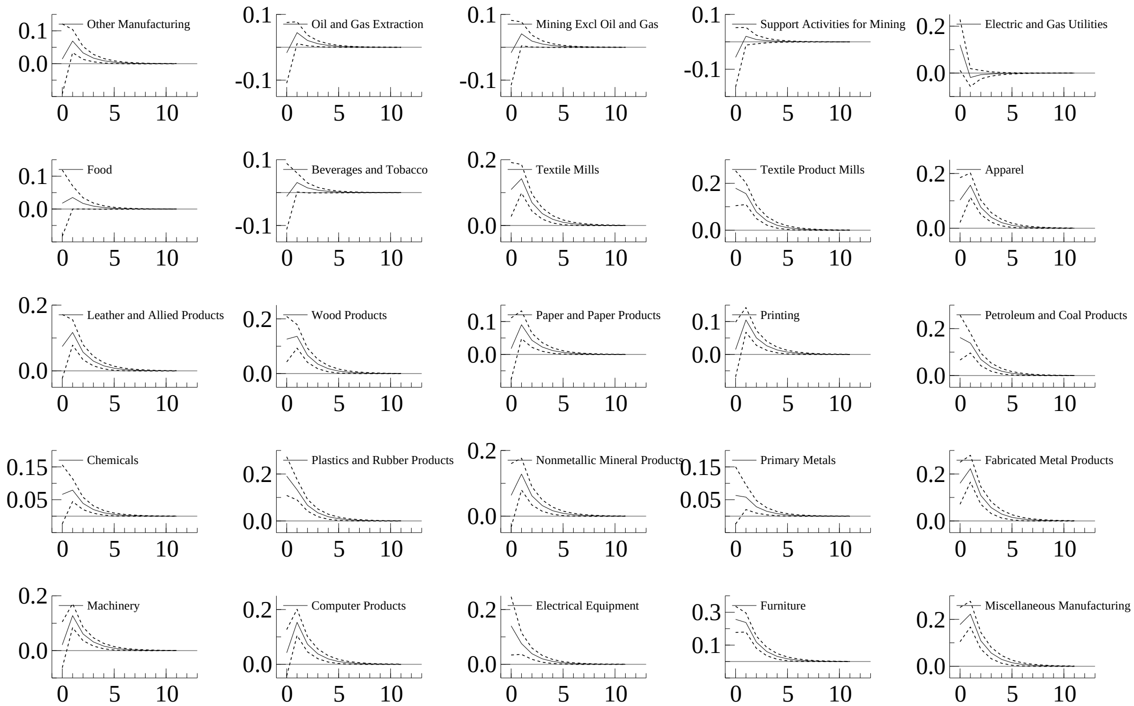
The figure shows the ordered column norms (top panel) and the column norm ratios (bottom panel) for a panel of daily growth rates of CDS spreads for a set of $n = 69$ financial entities across 10 Eurozone countries from January 3rd, 2006 to December 31, 2013 ($T = 2080$).

Figure 5: IMPULSE RESPONSES INDUSTRIAL PRODUCTION



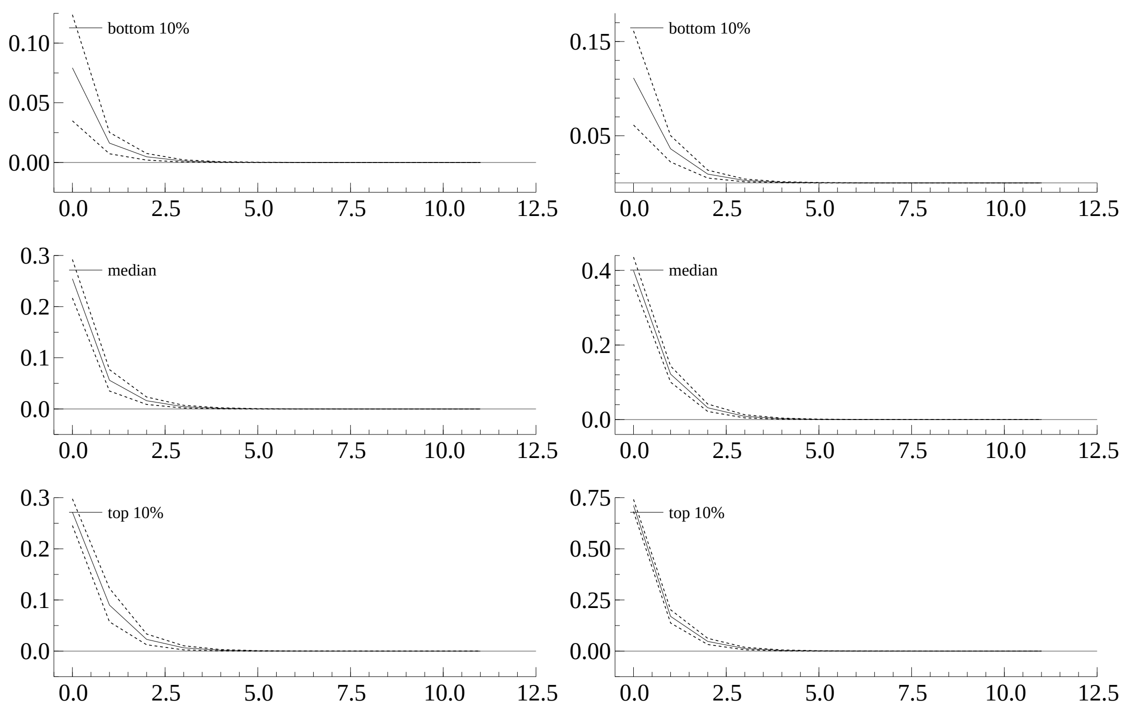
We show the response of a 1-standard deviation shock to the granular series (left column) and common factor (right column) on the 14th largest (top 10%), the median and the 124th largest (bottom 10%) non-granular industrial production series, where the ranking is based on the loadings $\boldsymbol{\beta}$.

Figure 6: Grouped impulse responses industrial production



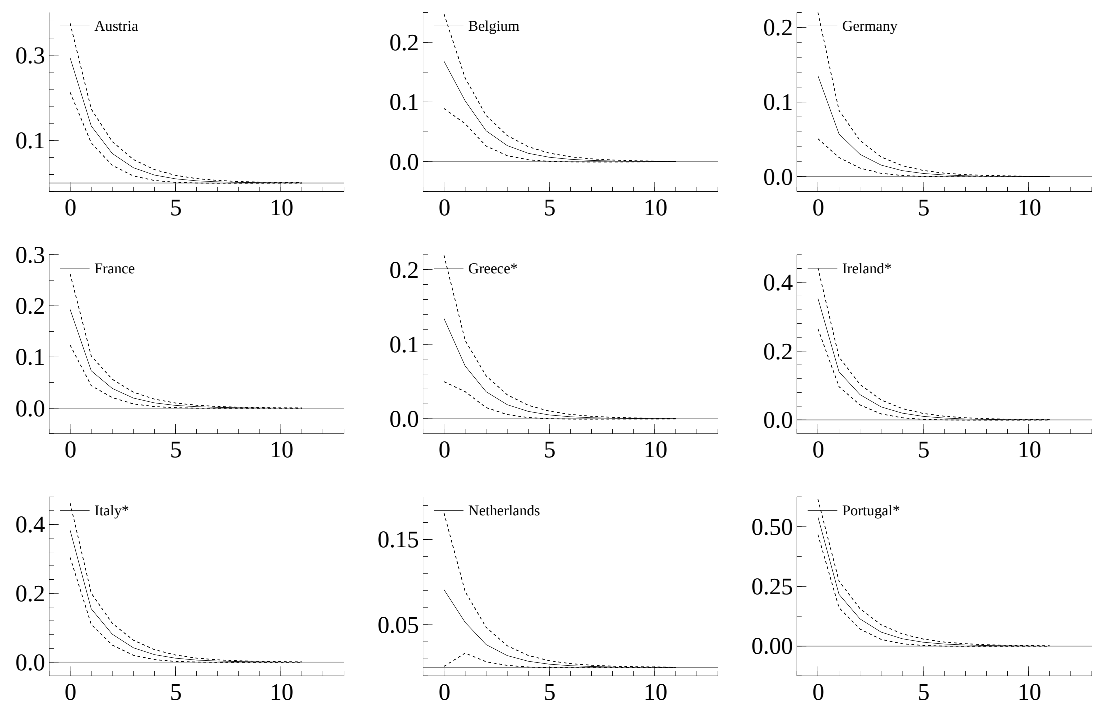
Impulse responses industrial production. We show the response of a 1-standard deviation shock to the granular series on the industrial production series that corresponds to the within sector median according to the loadings $\boldsymbol{\beta}$.

Figure 7: Impulse responses European CDS spreads



We show the response of a 1-standard deviation shock to the granular series (left column) and common factor (right column) on the 7th largest (top 10%), the median and the 64rd largest (bottom 10%) non-granular CDS growth rates, where the ranking is based on the loadings $\boldsymbol{\beta}$.

Figure 8: Grouped impulse responses European CDS spreads



We show the response of a 1-standard deviation shock to the granular series on the CDS growth rates that corresponds to the within country median according to the loadings $\boldsymbol{\beta}$.

A Proofs

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 k-th largest eigenvalue of $\mathbf{B}$ is denoted as $\mu_k(\mathbf{B})$. For an $M \times N$ matrix $\mathbf{A}$ the k-th 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 $\mathbf{B} > 0$ 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.

A.1 Identification Results

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

Proposition 1. *Let $\mathbf{A}$ be an $M \times N$ matrix and let $\mathbf{A}_i$ be the i-th 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

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

(ii) The first inequality follows from the fact that

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

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) $\|(\mathbf{A}^{-1} + \boldsymbol{d}'\mathbf{B}^{-1}\boldsymbol{d})e_{k,i}\|^2 \geqslant \mu_{n-k}^2(\mathbf{B}^{-1})\mu_k(\boldsymbol{d}'\boldsymbol{d})\|\boldsymbol{d}_i\|^2$

(ii) $\|(\mathbf{A}^{-1} + \boldsymbol{d}'\mathbf{B}^{-1}\boldsymbol{d})e_{k,i}\|^2 \leqslant 2\mu_1^2(\mathbf{A}^{-1}) + 2\mu_1^2(\mathbf{B}^{-1})\mu_1(\boldsymbol{d}'\boldsymbol{d})\|\boldsymbol{d}_i\|^2$

(iii) $\|\mathbf{B}^{-1}\boldsymbol{d}e_{k,i}\|^2 \geqslant \mu_{n-k}^2(\mathbf{B}^{-1})\|\boldsymbol{d}_i\|^2$

(iv) $\|\mathbf{B}^{-1}\boldsymbol{d}e_{k,i}\|^2 \leqslant \mu_1^2(\mathbf{B}^{-1})\|\boldsymbol{d}_i\|^2$

(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 assumptions 1-(i), 1-(ii) and 1-(iii) we can write $\boldsymbol{K}$, as given in equation (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

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

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 $i = 1, ..., k$ and $j = k + 1, ..., n$, which implies the claim of the lemma. Note that equation (3) implies that for $i = 1, \ldots, k$ 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 assumptions 1-(i), 1-(ii)

and 1-(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 $a,\dots,h > 0$. 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

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

and for (20) we have

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

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

Proof of Lemma 3. It is convenient to write the model in the factor representation (see Equation (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 $\mathbf{L}_2 = [\boldsymbol{\beta}\ \boldsymbol{\Lambda}_2]$ 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 Assumption 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 $\mathbf{X} = \mathbf{L}_2\mathbf{L}_2' + \mathbf{\Sigma}_\epsilon - \mathbf{L}_2\mathbf{L}_1'(\mathbf{L}_1\mathbf{L}_1')^{-1}\mathbf{L}_1\mathbf{L}_2'$. 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 $\mathbf{X} = \mathbf{L}_2\mathbf{M}_{L_1}\mathbf{L}_2' + \mathbf{\Sigma}_\epsilon = \hat{\mathbf{U}}\hat{\mathbf{U}}' + \mathbf{\Sigma}_\epsilon$ 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

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

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

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

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

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

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

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

The bound of interest becomes

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

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

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

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

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

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

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

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[29]

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

[29]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.

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

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

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

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

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

A.2 Estimation Results

The following proposition is similar to Lemma A.2 in Fan et al. (2011).

Proposition 3. *Let X_1 and X_2 be two random variables. Let γ, C be positive constants such that for any $s > 0$*

$$\Pr(|X_i| > s) \leqslant \exp(1 - (s/C)^\gamma),$$

for $i = 1, 2$. Then there exists a positive constant C' such that for any $s > 0$

$$\Pr(|X_1X_2| > s) \leqslant \exp(1 - (s/C')^{\gamma/2}) \ .$$

Proof. For any $s > 0$ we have that

$$\begin{aligned}\Pr(|X_1X_2| > s) &\leqslant \Pr(|X_1| > s^{1/2}) + \Pr(|X_2| > s^{1/2}) \leqslant 2\exp(1 - (s^{1/2}/C)^\gamma) \\ &= \exp(1 + \log 2 - (s/C^2)^{\gamma/2}) \ .\end{aligned}$$

Let $C' = (1 + \log 2)^{2/\gamma}C^2$. When $s > C'$ we have that

$$\log 2 - (s/C^2)^{\gamma/2} < -(s/C')^{\gamma/2} \ .$$

To see this note that, wlog, for $s = C'(1+\delta)^{2/\gamma}$ for any $\delta > 0$ we have that

$$\log 2 - (s/C^2)^{\gamma/2} + (s/C')^{\gamma/2} = \log 2 - (1 + \log 2)(1 + \delta) + (1 + \delta) = -\log 2 \cdot \delta \ .$$

This implies that when $s > C'$

$$\Pr(|X_1X_2| > s) \leqslant \exp(1 + \log 2 - (s/C^2)^{\gamma/2}) < \exp(1 - (s/C')^{\gamma/2}) \ . \tag{21}$$

When $s \leqslant C'$ we have that

$$\Pr(|X_1X_2| > s) \leqslant 1 \leqslant \exp(1 - (s/C')^{\gamma/2}) \ . \tag{22}$$

The inequalities in Equations (21) and (22) imply the claim of the proposition. □

The following proposition is in Merlevede et al. (2011) and is reported here for completeness.

Theorem 2. *Let X_t be a n-dimensional stationary and ergodic time series process with zero mean satisfying:*

(i) The $\{X_t\}$ process is α-mixing. There exists positive constants γ_1 and C_1 such that for all positive integers t we have that the α mixing coefficients satisfy

$$\alpha(t) \leqslant \exp(-C_1 t^{-\gamma_1}) \ .$$

(ii) There exists positive constants γ_2 and C_2 such that for any $s > 0$ and $i = 1, ..., n$

$$\Pr(|X_{i\,t}| > s) \leqslant \exp(1 - (s/C_2)^{\gamma_2}) \ .$$

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

Then, there exists positive constants C_3, C_4, C_5, C_6 and C_7 that only depend on C_1, C_2, γ_1 and γ_2 such that for T large enough

$$\mathrm{P}\left(\left|\frac{1}{T}\sum_{t=1}^{T} X_t\right| > s\right) \leqslant T\exp\left(-\frac{(Ts)^{\gamma}}{C_3}\right)+\exp\left(-\frac{(Ts)^2}{C_4(1+C_5T)}\right)+\exp\left(-\frac{(Ts)^2}{C_6T}\exp\left(\frac{(Ts)^{\gamma(1-\gamma)}}{C_7(\log(Ts))^{\gamma}}\right)\right) \ .$$

Proof. See Theorem 1 in Merlevede et al. (2011). □

The following three propositions are in Vershynin (2012) and are reported here for completeness.

Proposition 4. *Consider a matrix* $\mathbf{B}$ *that satisfies*

$$\|\mathbf{B}'\mathbf{B} - \mathbf{I}\| \leqslant \max(\delta, \delta^2) \tag{23}$$

for some $\delta > 0$. Then

$$1 - \delta \leqslant \mu_n(\mathbf{B}) \leqslant \mu_1(\mathbf{B}) \leqslant 1 + \delta \ . \tag{24}$$

Conversely, if $\mathbf{B}$ *satisfies* (24) *for some $\delta > 0$ then* $\|\mathbf{B}'\mathbf{B} - \mathbf{I}\| \leqslant 3\max(\delta, \delta^2)$.

Proof of Proposition 4. See Lemma 5.36 in Vershynin (2012). □

Proposition 5. *Let* $\mathbf{A}$ *be symmetric $n \times n$ matrix and let $\mathcal{N}_\varepsilon$ be an ε-net of S^{n-1} for some $\varepsilon \in (0,1)$. Then*

$$\|\mathbf{A}\| = \sup_{x \in S^{n-1}} |x'\mathbf{A}x| \leqslant (1-2\varepsilon)^{-1} \sup_{x \in \mathcal{N}_\varepsilon} |x'\mathbf{A}x| \ .$$

Proof of Proposition 5. See Lemma 5.4 in Vershynin (2012). □

Proposition 6. *Let S^{n-1} denoted the unit Euclidean sphere equipped with the Euclidean metric and let $\mathcal{N}(S^{n-1}, \varepsilon)$ denote an ε-net of S^{n-1}. Then, for every $\varepsilon > 0$ we have that $\mathcal{N}(S^{n-1}, \varepsilon) \leqslant \left(1 + \frac{2}{\varepsilon}\right)^n$.*

Proof of Proposition 6. See Lemma 5.2 in Vershynin (2012). □

Proof of Theorem 1. Part (i). We begin by noting that

$$\hat{\mathbf{\Sigma}}_y = \mathbf{\Sigma}_y^{1/2} \left(\frac{1}{T} \mathbf{V}'\mathbf{V} \right) \mathbf{\Sigma}_y^{1/2} ,$$

where $\mathbf{V}$ is an $T \times n$ matrix with t-th row defined as $\mathbf{\Sigma}_y^{-1/2} y_t$. It follows from an application of Ostrowsky's Theorem (Horn and Johnson (2013, Theorem 4.5.9)) to the last equation that

$$\mu_n(\mathbf{\Sigma}_y)\mu_n \left(\frac{1}{T} \mathbf{V}'\mathbf{V} \right) \leqslant \hat{\mathbf{\Sigma}}_y \leqslant \mu_1(\mathbf{\Sigma}_y)\mu_1 \left(\frac{1}{T} \mathbf{V}'\mathbf{V} \right) .$$

We first show that if $n \to \infty$ and $T = O(n^{2/\gamma - 1})$ for any $\eta > 0$ there exists positive constants C_1 and C_2 such that

$$1 - C_1 \left(\sqrt{\frac{n}{T}} + \sqrt{\frac{\eta \log n}{T}} \right) \leqslant \mu_n \left(\frac{\mathbf{V}'\mathbf{V}}{T} \right) \leqslant \mu_1 \left(\frac{\mathbf{V}'\mathbf{V}}{T} \right) \leqslant 1 + C_2 \left(\sqrt{\frac{n}{T}} + \sqrt{\frac{\eta \log n}{T}} \right) , \tag{25}$$

with probability at least $1 - O(n^{-\eta})$. It follows from Proposition 4 that to show (25) it suffices to show that

$$\left\| \frac{1}{T} \mathbf{V}'\mathbf{V} - \mathbf{I} \right\| \leqslant \varepsilon = C^* \left(\sqrt{\frac{n}{T}} + \sqrt{\frac{\eta \log n}{T}} \right) ,$$

with probability at least $1 - O(n^{-\eta})$. It follows from Proposition 5 that we can bound the norm in the last expression using a $\frac{1}{4}$-net $\mathcal{N}$ of the unit sphere S^{n-1}, that is

$$\left\| \frac{1}{T} \mathbf{V}'\mathbf{V} - \mathbf{I} \right\| \leqslant 2 \max_{x \in \mathcal{N}} \left| \frac{1}{N} \| \mathbf{V}x \|_2^2 - 1 \right|$$

where $\mathcal{N}$. Then to complete the proof it suffices to show that

$$\max_{x \in \mathcal{N}} \left| \frac{1}{N} \| \mathbf{V}x \|_2^2 - 1 \right| \leqslant \frac{\varepsilon}{2}$$

with the prescribed probability. We begin by noting that, for a given x in S^{n-1}, we have

$$\frac{1}{T} \| \mathbf{V}x \|_2^2 - 1 = \frac{1}{T} \sum_{t=1}^{T} (V_t' x)^2 - 1 = \frac{1}{T} \sum_{t=1}^{T} \left((V_t' x)^2 - 1 \right) = \frac{1}{T} \sum_{t=1}^{T} Z_t ,$$

where $Z_t = (V_t' x)^2 - 1$. Notice that by construction Z_t has an expected value of zero. Moreover, it follows from Assumptions 3 and Proposition 3 that Z_t satisfies the Assumptions of the Bernstein-type inequality of Merlevede et al. (2011) (Theorem 2) and we have that there exists positive constants only depending on C_1, C_2, γ_1 and γ_3 of Assumption 3 labeled

C_3, C_4, C_5, C_6 and C_7 such that for T large enough

$$
\begin{aligned}
\mathrm{P}\left(\left|\frac{1}{T}\|\mathbf{V}x\|_2^2 - 1\right| > \frac{\varepsilon}{2}\right) &= \mathrm{P}\left(\left|\frac{1}{T}\sum_{t=1}^{T} Z_t\right| > \frac{\varepsilon}{2}\right) \\
&\leqslant T\exp\left(-\frac{(T\varepsilon)^\gamma}{C_3 2^\gamma}\right) \\
&\quad + \exp\left(-\frac{T^2\varepsilon^2}{4C_4(1+C_5T)}\right) \\
&\quad + \exp\left(-\frac{T^2\varepsilon^2}{4C_6T}\exp\left(\frac{(T\varepsilon)^{\gamma(1-\gamma)}}{2^{\gamma(1-\gamma)}C_7(\log(T\varepsilon))^\gamma}\right)\right) \; . \qquad (26)
\end{aligned}
$$

where $\gamma^{-1} = \gamma_1^{-1} + 2\gamma_2^{-1}$ and it is such that $\gamma < 1$. The first term on the right hand side of (26) can be bounded using Jensen's inequality by

$$
\begin{aligned}
T\exp\left(-\frac{(T\varepsilon)^\gamma}{C_3 2^\gamma}\right) &= T\exp\left(-\frac{(C^*)^\gamma(\sqrt{Tn} + \sqrt{T\eta\log n})^\gamma}{C_3 2^\gamma}\right) \\
&\leqslant T\exp\left(-\frac{(C^*)^\gamma}{C_3 2^\gamma}(Tn + T\eta\log n)^{\gamma/2}\right) \; .
\end{aligned}
$$

For $n \to \infty$ and $T = O(n^{2/\gamma - 1})$ there exists a positive constant C such that

$$
\frac{(Tn + T\eta\log n)^{\gamma/2}}{(n + \eta\log n)} = \frac{(Tn)^{\gamma/2}}{n}\frac{\left(1 + \frac{\eta\log n}{n}\right)^{\gamma/2}}{\left(1 + \frac{\eta\log n}{n}\right)} = O(1)\left(1 + \frac{\eta\log n}{n}\right)^{\gamma/2 - 1} < C \; .
$$

It follows that there exists a positive constant C' such that

$$
T\exp\left(-\frac{(T\varepsilon)^\gamma}{C_3 2^\gamma}\right) \leqslant \exp\left(-C'(n + \eta\log n)\right) \; . \qquad (27)
$$

The second term on the right hand side of (26) can be bounded by

$$
\begin{aligned}
\exp\left(-\frac{T^2\varepsilon^2}{4C_4(1+C_5T)}\right) &= \exp\left(-\frac{T^2(C^*)^2(\sqrt{n/T} + \sqrt{\eta\log n/T})^2}{4C_4(1+C_5T)}\right) \\
&\leqslant \exp\left(-\frac{T^2(C^*)^2(n/T + \eta\log n/T)}{4C_4(1+C_5T)}\right) \\
&= \exp\left(-\frac{(C^*)^2}{4C_4/T + 4C_4C_5}(n + \eta\log n)\right)
\end{aligned}
$$

where second inequality follows from the fact that $a^2 + b^2 \leqslant (a+b)^2$ if $a, b \geqslant 0$. For $n \to \infty$ and $T = O(n^{2/\gamma - 1})$ there exists a positive constant C'' such that

$$
\exp\left(-\frac{T^2\varepsilon^2}{4C_4(1+C_5T)}\right) \leqslant \exp\left(-C''(n + \eta\log n)\right) \; . \qquad (28)
$$

Using steps similar to the ones used to establish (27) and (28) it is straightforward to see that for $n \to \infty$ and $T = O(n^{2/\gamma - 1})$ the third term on the right hand side of (26) can be bounded by

$$\exp\left(-\frac{T^2\varepsilon^2}{4C_6 T}\exp\left(\frac{(T\varepsilon)^{\gamma(1-\gamma)}}{2^{\gamma(1-\gamma)}C_7(\log(T\varepsilon))^{\gamma}}\right)\right) \leqslant \exp\left(-C''''(n+\eta\log n)\right) . \tag{29}$$

where C'''' is a suitable positive constant. From (27) (28) and (29) we get that there exists a positive constant C such that for $n \to \infty$ and $T = O(n^{2/\gamma-1})$

$$\mathrm{P}\left(\left|\frac{1}{T}\|\mathbf{V}x\|_2^2 - 1\right| > \frac{\varepsilon}{2}\right) \leqslant 3\exp\left(-C(n+\eta\log n)\right).$$

It follows from Proposition 6 that we can choose the $\mathcal{N}$ net so that it has a cardinality $|\mathcal{N}| \leqslant 9^n$. Using the union bound we get that

$$\mathrm{P}\left(\max_{x\in\mathcal{N}}\left|\frac{1}{T}\|\mathbf{V}x\|_2^2 - 1\right| > \frac{\varepsilon}{2}\right) \leqslant 9^n \cdot 3\exp\left(-C(n+\eta\log n)\right) \leqslant 2\exp(-C\eta\log n) = O(n^{-\eta}) ,$$

which implies the claim.

Part (ii). It follows Proposition 4 and Part (i).

Part (iii). We begin by noting that

$$\|\hat{\mathbf{K}} - \mathbf{K}\| = \|\hat{\mathbf{K}}(\mathbf{\Sigma}_y - \hat{\mathbf{\Sigma}}_y)\mathbf{K}\| \leqslant \|\hat{\mathbf{K}}\|\|\mathbf{\Sigma}_y - \hat{\mathbf{\Sigma}}_y\|\|\mathbf{K}\| = \mu_n(\hat{\mathbf{\Sigma}}_y)^{-1}\|\mathbf{\Sigma}_y - \hat{\mathbf{\Sigma}}_y\|\mu_n(\mathbf{\Sigma}_y)^{-1} .$$

It follows from Part (i) and (ii) that the events

$$\left\{\|\mathbf{\Sigma}_y - \hat{\mathbf{\Sigma}}_y\| \leqslant C_3\sqrt{n/T}\right\} \text{ and } \left\{\mu_n(\hat{\mathbf{\Sigma}}_y) \geqslant \mu_n(\mathbf{\Sigma}_y) + C_1\sqrt{n/T}\right\}$$

are true at least with probability $1 - O(n^{-\eta})$. It follows from Fréchet inequality that with at least probability $1 - O(n^{-\eta})$ we have that

$$\|\hat{\mathbf{K}} - \mathbf{K}\| \leqslant \left(\mu_n(\mathbf{\Sigma}_y) + C_1\sqrt{n/T}\right)^{-1} \cdot C_3\sqrt{n/T} \cdot \mu_n(\mathbf{\Sigma}_y)^{-1} = C_4\sqrt{n/T} .$$

Part (iv). Note that

$$\begin{aligned} \|\hat{\mathbf{K}}_i\| - \|\mathbf{K}_i\| &= \|\hat{\mathbf{K}}_i + \mathbf{K}_i - \mathbf{K}_i\| - \|\mathbf{K}_i\| \leqslant \|\mathbf{K}_i\| + \|\hat{\mathbf{K}}_i - \mathbf{K}_i\| - \|\mathbf{K}_i\| = \|(\hat{\mathbf{K}} - \mathbf{K})e_i\| \\ &\leqslant \|\hat{\mathbf{K}} - \mathbf{K}\| . \end{aligned}$$

The claim then follows from Part(iii).

□

Proof of Corollary 1. The proof consist of showing that the probability of inconsistent selection of the granular series converges to zero under the assumptions of the corollary. First, note that by Theorem 1 (i) we have that that for $n \to \infty$ and $T = O(n^{2/\gamma-1})$ we have that

$\hat{\mathbf{K}}$ exists with high probability. Next, note that

$$
\begin{aligned}
\mathrm{P}(\mathcal{E}_R^c) &= \mathrm{P}\left(\min_{i=1,\dots,k}\|\hat{\mathbf{K}}_i\| < \max_{j=k+1,\dots,n}\|\hat{\mathbf{K}}_j\|\right) \\
&= \mathrm{P}\left(\max_{j=k+1,\dots,n}\|\hat{\mathbf{K}}_j\| + \max_{i=1,\dots,k}(-\|\hat{\mathbf{K}}_i\|) > 0\right) \\
&\leqslant (n-k)\,k \max_{\substack{j=k+1,\dots,n \\ i=1,\dots,k}} \mathrm{P}\left(\|\hat{\mathbf{K}}_j\| - \|\hat{\mathbf{K}}_i\| > 0\right) \\
&= (n-k)\,k \max_{\substack{j=k+1,\dots,n \\ i=1,\dots,k}} \mathrm{P}\left(\|\mathbf{K}_j\| - \|\mathbf{K}_i\| + (\|\hat{\mathbf{K}}_j\| - \|\mathbf{K}_j\|) - (\|\hat{\mathbf{K}}_i\| - \|\mathbf{K}_i\|) > 0\right) \\
&\leqslant (n-k)\,k \max_{\substack{j=k+1,\dots,n \\ i=1,\dots,k}} \mathrm{P}\left(\|\mathbf{K}_j\| - \|\mathbf{K}_i\| + \left|\|\hat{\mathbf{K}}_j\| - \|\mathbf{K}_j\|\right| + \left|\|\hat{\mathbf{K}}_i\| - \|\mathbf{K}_i\|\right| > 0\right) \\
&= (n-k)\,k \max_{\substack{j=k+1,\dots,n \\ i=1,\dots,k}} \mathrm{P}\left(\left|\|\hat{\mathbf{K}}_i\| - \|\mathbf{K}_i\|\right| + \left|\|\hat{\mathbf{K}}_j\| - \|\mathbf{K}_j\|\right| > \|\mathbf{K}_i\| - \|\mathbf{K}_j\|\right) \\
&\leqslant (n-k)\,k \max_{\substack{j=k+1,\dots,n \\ i=1,\dots,k}} \max_{l=i,j} \mathrm{P}\left(\left|\|\hat{\mathbf{K}}_l\| - \|\mathbf{K}_l\|\right| > \frac{\|\mathbf{K}_i\| - \|\mathbf{K}_j\|}{2}\right) .
\end{aligned}
$$

Note that by Theorem 1 (iv) for any $\eta' > 0$ there exists positive a positive constant C such that

$$
\mathrm{P}\left(\left|\|\hat{\mathbf{K}}_l\| - \|\mathbf{K}_l\|\right| > \frac{\|\mathbf{K}_i\| - \|\mathbf{K}_j\|}{2}\right) \leqslant 2\mathrm{P}\left(\|\hat{\mathbf{K}}_l\| - \|\mathbf{K}_l\| > C\sqrt{\frac{n}{T}}\right) = O(n^{-\eta'}) ,
$$

note that this follows from Lemma 1 that establishes $\|\mathbf{K}_i\| - \|\mathbf{K}_j\| > 0$. Thus, using the union bound we can bound the probability of misclassification by

$$
\mathrm{P}(\mathcal{E}_R^c) = nO(n^{-\eta'}) = O(n^{-\eta'+1}).
$$

The claim of the corollary then follows by choosing $\eta = \eta' - 1$. □

Proposition 7. *Let y_t be generated by model* (6) *under assumptions 2 and 3. Suppose $n \to \infty$ and $T = O(n^{2/\gamma - 1})$. Then, for any $\eta > 0$ there exists positive constant C such that for any $i = 1, \dots, n-1$*

$$
\frac{\|\hat{\mathbf{K}}_{(i)}\|}{\|\hat{\mathbf{K}}_{(i+1)}\|} - \frac{\|\mathbf{K}_{(i)}\|}{\|\mathbf{K}_{(i+1)}\|} \geqslant C\sqrt{\frac{n}{T}} .
$$

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

Proof of Proposition 7. For any $i = 1, \dots, n-1$ we have that

$$
\frac{\|\hat{\mathbf{K}}_{(i)}\|}{\|\hat{\mathbf{K}}_{(i+1)}\|} - \frac{\|\mathbf{K}_{(i)}\|}{\|\mathbf{K}_{(i+1)}\|} = \frac{\|\mathbf{K}_{(i)}\| + \|\hat{\mathbf{K}}_{(i)}\| - \|\mathbf{K}_{(i)}\|}{\|\mathbf{K}_{(i+1)}\| + \|\hat{\mathbf{K}}_{(i+1)}\| - \|\mathbf{K}_{(i+1)}\|} - \frac{\|\mathbf{K}_{(i)}\|}{\|\mathbf{K}_{(i+1)}\|} .
$$

Recall that $C_1 < \mu_n(\mathbf{K}) \leqslant \|\mathbf{K}_{(i+1)}\| \leqslant \mu_1(\mathbf{K}) < C_2$ where C_1 and C_2 are positive constants.

Condition on the event $\mathcal{E}_1 = \{\|\hat{\mathbf{K}}_{(i+1)}\| - \|\mathbf{K}_{(i+1)}\| \leqslant 1/2\mu_n(\mathbf{K})\}$. It follows from Theorem 1 that for $n \to \infty$, $T = O(n^{2/\gamma-1})$ and any $\eta > 0$ we have that $\mathrm{P}(\mathcal{E}_1) \geqslant 1 - O(T^{-\eta})$. This allows us to write

$$\frac{\|\hat{\mathbf{K}}_{(i)}\|}{\|\hat{\mathbf{K}}_{(i+1)}\|} - \frac{\|\mathbf{K}_{(i)}\|}{\|\mathbf{K}_{(i+1)}\|} \leqslant \frac{\|\mathbf{K}_{(i)}\| + \|\hat{\mathbf{K}}_{(i)}\| - \|\mathbf{K}_{(i)}\|}{\|\mathbf{K}_{(i+1)}\| - \left|\|\hat{\mathbf{K}}_{(i+1)}\| - \|\mathbf{K}_{(i+1)}\|\right|} - \frac{\|\mathbf{K}_{(i)}\|}{\|\mathbf{K}_{(i+1)}\|} .$$

Using the fact that for $0 < x < a/2$ we have that $1/(a-x) \leqslant 1/a + (2/a^2)x$ and after some computations we get that there are positive constants C_1, C_2 and C_3 such that

$$\begin{aligned}\frac{\|\hat{\mathbf{K}}_{(i)}\|}{\|\hat{\mathbf{K}}_{(i+1)}\|} - \frac{\|\mathbf{K}_{(i)}\|}{\|\mathbf{K}_{(i+1)}\|} &\leqslant C_1 \left|\|\hat{\mathbf{K}}_{(i)}\| - \|\mathbf{K}_{(i)}\|\right| + C_2 \left|\|\hat{\mathbf{K}}_{(i+1)}\| - \|\mathbf{K}_{(i+1)}\|\right| \\ &\quad + C_3 \left|\|\hat{\mathbf{K}}_{(i)}\| - \|\mathbf{K}_{(i)}\|\right| \left|\|\hat{\mathbf{K}}_{(i+1)}\| - \|\mathbf{K}_{(i+1)}\|\right| ,\end{aligned}$$

Condition on the events $\mathcal{E}_2 = \left\{\|\hat{\mathbf{K}}_{(i)}\| - \|\mathbf{K}_{(i)}\| \leqslant C_4\sqrt{n/T}\right\}$ and $\mathcal{E}_3 = \left\{\|\hat{\mathbf{K}}_{(i+1)}\| - \|\mathbf{K}_{(i+1)}\| \leqslant C_5\sqrt{n/T}\right\}$. It follows from Theorem 1 that for $n \to \infty$, $T = O(n^{2/\gamma-1})$ and any $\eta > 0$ we can choose constants C_4 and C_5 such that $\mathrm{P}(\mathcal{E}_i) \geqslant 1 - O(T^{-\eta})$ for $i = 2, 3$. Then we have that for n and T sufficiently large there exisits a positive constant C_6 such that

$$\frac{\|\hat{\mathbf{K}}_{(i)}\|}{\|\hat{\mathbf{K}}_{(i+1)}\|} - \frac{\|\mathbf{K}_{(i)}\|}{\|\mathbf{K}_{(i+1)}\|} \leqslant C_6\sqrt{\frac{n}{T}}.$$

Last, note that

$$\mathrm{P}(\mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3) \geqslant 1 - O(T^{-\eta}) ,$$

which implies the statement of the proposition. □

Proof of Corollary 2. The proof consists in showing that the probability of inconsistent selection of k converges to zero under the assumptions of the corollary. We note that the event of inconsistent selection can be written as

$$\mathcal{E}_S^c = \{\hat{k} \neq k\} = \bigcup_{\substack{j=1,\ldots,n-1 \\ j\neq k}} \left\{\frac{\|\hat{\mathbf{K}}_{(k)}\|}{\|\hat{\mathbf{K}}_{(k+1)}\|} < \frac{\|\hat{\mathbf{K}}_{(j)}\|}{\|\hat{\mathbf{K}}_{(j+1)}\|}\right\} .$$

We begin by analyzing the probability of the event $\mathcal{E}_S^c$ conditional on the event $\mathcal{E}_R$, i.e. that the granular series are correctly ordered. It follows from corollary 1 that for $n \to \infty$, $T = O(n^{2/\gamma-1})$ and any $\eta' > 0$ we have that $\mathrm{P}(\mathcal{E}_R) \geqslant 1 - O(T^{-\eta'})$. Using the union bound, we get that

$$\begin{aligned}\mathrm{P}(\mathcal{E}_S^c | \mathcal{E}_R) &\leqslant (n-2) \max_{\substack{j=1,\ldots,n-1 \\ j\neq k}} \mathrm{P}\left(\frac{\|\hat{\mathbf{K}}_{(k)}\|}{\|\hat{\mathbf{K}}_{(k+1)}\|} < \frac{\|\hat{\mathbf{K}}_{(j)}\|}{\|\hat{\mathbf{K}}_{(j+1)}\|}\right) \\ &= (n-2) \max_{\substack{j=1,\ldots,n-1 \\ j\neq k}} \mathrm{P}\left(\frac{\|\hat{\mathbf{K}}_{(j)}\|}{\|\hat{\mathbf{K}}_{(j+1)}\|} - \frac{\|\hat{\mathbf{K}}_{(k)}\|}{\|\hat{\mathbf{K}}_{(k+1)}\|} > 0\right) .\end{aligned}$$

Next we note that for any $j = 1, \ldots, n-k$ with $j \neq k$ we have

$$\begin{aligned}
\frac{\|\hat{\mathbf{K}}_{(j)}\|}{\|\hat{\mathbf{K}}_{(j+1)}\|} - \frac{\|\hat{\mathbf{K}}_{(k)}\|}{\|\hat{\mathbf{K}}_{(k+1)}\|} &= \frac{\|\mathbf{K}_{(j)}\|}{\|\mathbf{K}_{(j+1)}\|} - \frac{\|\mathbf{K}_{(k)}\|}{\|\mathbf{K}_{(k+1)}\|} + \left(\frac{\|\hat{\mathbf{K}}_{(j)}\|}{\|\hat{\mathbf{K}}_{(j+1)}\|} - \frac{\|\mathbf{K}_{(j)}\|}{\|\mathbf{K}_{(j+1)}\|}\right) - \left(\frac{\|\hat{\mathbf{K}}_{(k)}\|}{\|\hat{\mathbf{K}}_{(k+1)}\|} - \frac{\|\mathbf{K}_{(k)}\|}{\|\mathbf{K}_{(k+1)}\|}\right) \\
&\leqslant \frac{\|\mathbf{K}_{(j)}\|}{\|\mathbf{K}_{(j+1)}\|} - \frac{\|\mathbf{K}_{(k)}\|}{\|\mathbf{K}_{(k+1)}\|} + \left|\frac{\|\hat{\mathbf{K}}_{(j)}\|}{\|\hat{\mathbf{K}}_{(j+1)}\|} - \frac{\|\mathbf{K}_{(j)}\|}{\|\mathbf{K}_{(j+1)}\|}\right| + \left|\frac{\|\hat{\mathbf{K}}_{(k)}\|}{\|\hat{\mathbf{K}}_{(k+1)}\|} - \frac{\|\mathbf{K}_{(k)}\|}{\|\mathbf{K}_{(k+1)}\|}\right| \quad (30)
\end{aligned}$$

Note that here $\mathbf{K}_{(k)}$ denotes the population analog of $\hat{\mathbf{K}}_{(k)}$ (that is $\mathbf{K}_{(k)}$ is based on the same ordering on the columns used to rank $\hat{\mathbf{K}}_{(k)}$). Using the inequality in (30) we get that

$$\mathrm{P}\left(\frac{\|\hat{\mathbf{K}}_{(j)}\|}{\|\hat{\mathbf{K}}_{(j+1)}\|} - \frac{\|\hat{\mathbf{K}}_{(k)}\|}{\|\hat{\mathbf{K}}_{(k+1)}\|} > 0\right) \leqslant 2 \max_{l=k,j} \mathrm{P}\left(\left|\frac{\|\hat{\mathbf{K}}_{(l)}\|}{\|\hat{\mathbf{K}}_{(l+1)}\|} - \frac{\|\mathbf{K}_{(l)}\|}{\|\mathbf{K}_{(l+1)}\|}\right| \geqslant \frac{1}{2}\left(\frac{\|\mathbf{K}_{(k)}\|}{\|\mathbf{K}_{(k+1)}\|} - \frac{\|\mathbf{K}_{(j)}\|}{\|\mathbf{K}_{(j+1)}\|}\right)\right) . \quad (31)$$

where we have used the facts (i) for random variables X_1, X_2 with $P(X_1 \leqslant X_2) = 1$ we have that $P(X_1 > 0) \leqslant P(X_2 > 0)$; (ii) for positive random variables X_1, X_2 and constant C we have that $\mathrm{P}(X_1 + X_2 > C) \leqslant 2 \max_{i=1,2} \mathrm{P}(X_i > 1/2C)$.

We now focus on showing that the probability on the right hand side of (31) is small when n and T are large. First, note that lemma 4 implies that for any $j = 1, ..., n-1$ with $j \neq k$ we have that there exists a positive constant C such that

$$\frac{\|\mathbf{K}_{(k)}\|}{\|\mathbf{K}_{(k+1)}\|} - \frac{\|\mathbf{K}_{(j)}\|}{\|\mathbf{K}_{(j+1)}\|} > C.$$

By applying proposition 7 we have that for $n \to \infty$, $T = O(n^{2/\gamma - 1})$ and any $\eta' > 0$ there exists a positive constant C such that

$$\begin{aligned}
\mathrm{P}\left(\left|\frac{\|\hat{\mathbf{K}}_{l}\|}{\|\hat{\mathbf{K}}_{l+1}\|} - \frac{\|\mathbf{K}_{l}\|}{\|\mathbf{K}_{l+1}\|}\right| \geqslant \frac{1}{2}\left(\frac{\|\mathbf{K}_{k}\|}{\|\mathbf{K}_{k+1}\|} - \frac{\|\mathbf{K}_{j}\|}{\|\mathbf{K}_{j+1}\|}\right)\right) &\leqslant \mathrm{P}\left(\frac{\|\hat{\mathbf{K}}_{l}\|}{\|\hat{\mathbf{K}}_{l+1}\|} - \frac{\|\mathbf{K}_{l}\|}{\|\mathbf{K}_{l+1}\|} \geqslant C\sqrt{\frac{n}{T}}\right) \\
&\leqslant O(T^{-\eta'}) .
\end{aligned}$$

Thus, an upper bound on the probability of incorret selection is given by

$$\mathrm{P}(\mathcal{E}_S^c | \mathcal{E}_R) \leqslant n O(T^{-\eta'}) \leqslant O(T^{-\eta'+1}).$$

Finally, the unconditional probability of correct selection is bounded by

$$\mathrm{P}(\mathcal{E}_S) \geqslant \mathrm{P}(\mathcal{E}_S, \mathcal{E}_R) = \mathrm{P}(\mathcal{E}_S|\mathcal{E}_R)\mathrm{P}(\mathcal{E}_R) \geqslant (1 - O(T^{-\eta'+1}))(1 - O(T^{-\eta'})) \geqslant 1 - O(T^{-\eta'+1}) .$$

The claim of the corollary then follows by choosing $\eta = \eta' - 1$.

□