Detecting Granular Time Series in Large Panels

C. Brownlees and *G. Mesters*

Universitat Pompeu Fabra and Barcelona GSE

Preliminary and Incomplete
November 30, 2016

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 develop methodology to detect and test the set of granulars in a panel when such set is unknown. The methods are based on the norms of the columns of the concentration matrix of the panel. We establish that the norms consistently identify a ranking and the number of granulars when the cross-section and time-series dimensions are sufficiently large. The distribution of the column norms is derived in order to conduct hypothesis tests. Importantly, we show that the methodology is unaffected when the series in the panel are influenced by common factors. A simulation study shows that the proposed detection and testing procedures perform satisfactorily in finite samples. We illustrate the methodology with applications in macroeconomics and finance.

JEL classification: C32; C43

Keywords: granularity, factor models, network models, hypothesis testing.

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 – can generate aggregate fluctuations. This is called by Gabaix (2011) the granular hypothesis. These models have been applied to study aggregate fluctuations in macroeconomics and financial stability in finance.

One of the main hurdles in applying 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 the determine the set of granular entities from the data. In this paper we tackle this challenge by (i) introducing a model that allows us to formalize the granular detection problem for a panel of time series and (ii) developing a methodology to detects the set of granular series from the data when such set is unknown.

We begin by introducing a model for a panel of time series to formalize 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 the their respective idiosyncratic shocks, which we call granular shocks. The non-granular series are a linear combination of the granular shocks and an idiosyncratic non-granular shock. The granular and non-granular shocks are assumed to be uncorrelated. We work under the assumption that the econometrician does not observe whether a given series belongs to the set of granulars or not and our objective is to introduce a methodology that allows to detect the set of granulars series from the data.

An important property of the model we introduce is that the inverse covariance matrix of the panel, hereafter concentration matrix, allows to recover the set of granulars. Under appropriate identification assumptions the norms of the columns of the concentration matrix that are associated with the granular series are larger when compared to to those of the non-granular series. Intuitively, this follows as the concentration matrix can be interpreted as a rescaled partial correlation network and since the granular series have more “connections” its column norms are larger, see Pourahmadi (2013, section 5.2). Indeed, we show that under appropriate assumptions on strength of the granular shocks that we can rank the series in the panel according to granularity based on the column norms. Further more, by comparing the ratios among the subsequent ordered column norms we can determine the number of granular series.

Given the identification results we operationalize the detection procedure by replacing the concentration matrix by its sample equivalent. We establish the consistency of the norms of the columns of the sample concentration matrix for panels that consist of stationary time series and where the non-granular idiosyncratic shocks exhibit weak cross-sectional and time series dependence. The assumptions are similar to those considered in the factor model literature, see for example Ahn and Horenstein (2013). Based on the consistency of the sample column norms we show that the implied ranking of the granular series is consistent and also that the number of granular series can be consistently estimated.

The consistency results for the detection method rely on the identification assumptions. We are interested in assessing the uncertainty surrounding these assumptions. For this purpose we introduce a hypothesis test for the null-hypothesis that a particular column norm is equal to another column norm. We derive the limiting distribution of the test statistic using similar methods as in Pesaran and Yamagata (2012) and Fan, Liao and Yao (2015). Using this test as a building block, we develop a supremum type test statistic that tests the correct number of granular series. Other relevant hypotheses tests are also considered.

Typically, common factors explain a large proportion of the variance of aggregate fluctuations. See for instance the research by Foerster, Sarte and Watson (2011), Long and Plosser (1987), Forni and Reichlin (1998), and Shea (2002) for earlier work on the trade-off between idiosyncratic and aggregate shocks. To incorporate this we consider an extension of the baseline model in which all the series in the panel are additionally influenced by a set of common factors. The number of common factors is unknown. We show that the presence of the factors does not break down the methodology, and we provide extensions of the identification, estimation and testing results established for the baseline case. The intuition behind this result is that while the correlation patterns that are generated by common factors and granular series are similar the partial correlation structure that they generate is vastly different. This implies that the granular series can still be detected using the concentration matrix. After the set of granular series is determined standard methods can be used to determine the number of common factors. Finally, the complete model with granular series and common factors can be studied.

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. In the experiment we considers a large possible number of different variants of the data generating process. Among other things, we study the performance of the procedure under different degrees of cross-sectional and time-series dependence. 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 for 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). Certain sectors in the US economy are suspected to have effects over other a large number of other sectors. An example is the automobile sector as argued in Acemoglu et al. (2012). Our methodology finds 6 granular sectors that include plastic products, semi-conductor and automobile sectors. Further the complete model has one common factor which explains the majority of the variance in the panel.

In the second study, we use our framework to detect granulars in a panel of volatility measures of large US financial firms during the 2007–2009 Great Financial Crisis. The application is close in spirit to the work of Billio, Getmansky, Lo and Pellizzon (2012) and

Diebold and Yılmaz (2014) who associate systemic risk to those institutions that have the highest degree of interdependence in the system. Our methodology identifies as granulars JP Morgan, Northern Trust and Bank of America and the top ten rankings of the most granular institutions in the panel contains most of the financial institutions classified by the Financial Stability Board as either globally or domestically systemic.

Our work is related to different strands of the literature. First, in several ways our work relates to the large literature on factor models, see Stock and Watson (2011) for a review. In particular, the set of granular series form observable factors in a factor model for the non-granular series. Further, since the interpretation of common factors is known to be difficult, several recent methods have proposed to associate estimated common factors with individual time series, see for example Parker and Sul (2016) and Siavash (2016) [1]. We discuss the differences between this approach and our approach both from a theoretical perspective and in the simulation study. This paper is also related to the literature on networks in econometrics and statistics, as in Meinshausen and Bühlmann (2006) Peng, Wang, Zhou and Zhu (2009) and Diebold and Yılmaz (2014). In particular, as in the statistical graphical literature, the key parameter of interest of our framework is the inverse covariance matrix of the panel, which encodes the partial dependence properties of the data.

The remainder of this paper is organized as follows. Section 2 introduces the granular detection problem and discusses its relevance for applications in macroeconomics and finance. Section 3 develops our methodology for our baseline granular panel model. Section 4 extends the methodology for a general model that also allows for common factors. Section 5 carries out a simulation study to assess the finite sample performance of the proposed techniques. Section 5.3 presents the results of two empirical applications of our methodology. Concluding remarks follow in Section 6.

[1] This approach proceeds by replacing estimated latent factors by observed series and then determining whether the observed series capture similar the variation. In contrast our methodology is able to detect granular series without requiring an estimate for the common factors nor the number of common factors. This avoids the strong assumptions that are typically required for consistent estimation in factor models, see Onatski (2012).

2 The granular detection problem

We have observations of variable $y_{i,t}$, that is associated with entity i and time t. Data is available for $i = 1, \ldots, n$ entities and T time periods. For each time period $t = 1, \ldots, T$ we collect the variables in the vector $y_t = (y_{1,t}, \ldots, y_{n,t})'$. We use the notation $y_{i:j,t}$ with $i < j$ to denote the $(j-i+1)$-dimensional time series associated with 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 shocks g_t as granular shocks. 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 given by ϵ_t. Finally, all series are influenced by a set of r common shocks, or factors, f_t.

The granular panel data model combines the different shocks as follows

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

where the loading matrices $\boldsymbol{\beta}$, $\boldsymbol{\Lambda}_1$ and $\boldsymbol{\Lambda}_2$ have dimensions $(n-k) \times k$, $k \times r$ and $(n-k) \times r$, respectively. Precise assumptions on the model are spelled out in the following sections. Let us point out here that the loading matrix $\boldsymbol{\beta}$ is key in this framework as it determines how the non-granular series are affected by the granular shocks.

In this paper we work under the assumption that the data is generated according to model (1) but that the econometrician does not know which series are granular, nor the number of granular series, nor the number of common factors in the panel[2]. Our objective is to introduce a methodology that allows us to consistently recover this information from the data.

The granular model of (1) allows us to formalize the notion that certain time-series have

[2]We notice that *if* the set of granular series is known, then model (1) is equivalent to a factor model with a specific identification restriction.

a pervasive influence over the entire panel. Detecting the set of granulars has important economic implications. Next, we discuss two leading examples that illustrate how our framework can be used to formalize different applications in macro and finance. We return to these examples in the empirical study.

2.1 Sectoral versus aggregate shocks in US 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 in 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 have been proposed.

First, some sector specific shocks, or idiosyncratic shocks, may affect many other sectors. Proposed mechanisms for this explanation include that certain large sectors may have a pervasive effect on the entire index, see Gabaix (2011), alternatively, Acemoglu et al. (2012) have argued that not the size of the sectors but its degree of connectedness determines whether a certain sector affects the index. Idiosyncratic shocks to sectors that have influence over many other sectors may be able to drive aggregate volatility.

Second, aggregate shocks may exist that influence many sectors at the same time, see Foerster et al. (2011). These shocks stand in contrast to the idiosyncratic shocks in the sense that there exact origin is unknown. Examples include, monetary policy shocks, exchange rate shocks and uncertainty shocks. It is possible that a mixture of both idiosyncratic and aggregate shocks determines aggregate volatility.

The model (1) can disentangle both explanations. When we define $y_{1:n,t}$ as the vector of sector specific industrial production outcomes, model (1) implies that these are determined by the k granular shocks g_t and the r aggregate shocks f_t. Both have influence over the entire panel. Our methodology determines which sectors are granular, how many sectors are granular and how many common factors exist. Finally, when the model is specified it enables a variance decomposition to determine which explanations bears more weight empirically.

2.2 Financial stability and systemic risk.

One of the lessons from the financial crisis is that the distress of few yet highly influential financial firms can impose significant negative externalities on the entire economy. The size of the shocks to the financial institutions together with how they are connected to each other has large implications for contagion and risk throughout the financial system, see Acemoglu et al. (2015). These ideas have motivated a large literature that aims at detecting and ranking institutions in the financial system according to their "systemicness".

A number of influential contributions have proposed to measure the systemic risk of an institution on the basis of network models like in Billio et al. (2012) and Diebold and Yılmaz (2014). These authors measure systemic risk based on the analysis of large panels of equity returns or volatility measures. Broadly speaking, these contributions measure how systemic an institution is on the basis of the number/magnitude of the spillovers effects of that institution on the rest of the panel. Despite the intuitive appeal of these proposals, these papers do not introduce a model that precisely defines when an institutions is indeed systemic, and, consequently, they do not establish the properties of their selection procedures.

In this work we cast the problem of detecting systemic institutions as an instance of a granular detection problem by assuming that the panel of target measures of interest is generated by model (1). The model allows for a selection of institutions to influence the entire panel while allowing for market risk, and other economy wide sources of risk, through the common factors.

3 Methodology

3.1 Model and Identification

Before considering the complete granular panel data model (1) we first outline our methodology for a simplified version of the model where there are no common factors. In particular, we consider a panel time series model with k granular series and $n - k$ non-granular series that is formulated as follows.

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

where $y_{1:k,t}$ denotes the outcomes of the k granular series, g_t is the $k \times 1$ vector of non-structural granular shocks, $y_{k+1:n,t}$ denotes the outcomes of the $n-k$ non-granular series, $\boldsymbol{\beta}$ is the $(n-k) \times k$ parameter matrix that captures the effect of the process g_t and ϵ_t is the $(n-k) \times 1$ idiosyncratic disturbance vector of the non-granular series.

The model for the non-granular series in (2) is a standard multivariate regression model with k low rank regressors. Estimation and inference for model (2) is standard when the researcher knows which series are the granular series. In our setting we apply estimation methods and hypothesis tests to determine which series are the granular series.

Without further assumptions for the components of model (2) determining which series are granular is impossible. Therefore we impose restrictions on g_t, $\boldsymbol{\beta}$ and ϵ_t that allow us to separate the granular series from the non-granular series.

Assumption 1. *For model* (2) *we assume that*

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

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

(iii) $\mathrm{E}(g_t \epsilon_{i,t}) = 0$ *for all* i, t

(iv) We have that $\boldsymbol{\beta}'\boldsymbol{\beta} \to \boldsymbol{D}$ *as* $n \to \infty$, *with* $\mu_k(\boldsymbol{D}) > 0$ *and* $\mu_1(\boldsymbol{D}) < \infty$. *Also, there exists an integer* $N > 0$ *such that for all* $n > N$ *the columns of* $\boldsymbol{\beta}$, *denoted by* $\boldsymbol{\beta}_i$ *for* $i = 1, \ldots, k$, *satisfy*

$$\min_{i=1,\ldots,k} \frac{\|\boldsymbol{\beta}'\boldsymbol{\beta}_i\|}{\|\boldsymbol{\beta}\|} > C_{\Sigma_\epsilon}$$

where C_{Σ_ϵ} *is the conditioning number of the matrix* $\boldsymbol{\Sigma}_\epsilon$.

Assumption (i) ensures that none of the granular series are linear combinations of each other. Assumption (ii) requires all eigenvalues of $\boldsymbol{\Sigma}_\epsilon$ to be positive, which is necessary as otherwise a linear combination of the non-granular series cannot be distinguished from the granular series. Assumption (iii) is standard for regression models, see for example White (2000, Chapter 2).

Assumption (iv) in a way defines the granular model[3]. The loading matrix $\boldsymbol{\beta}$ determines the "strength" of the granular series in model (2). We require $\boldsymbol{\beta}'\boldsymbol{\beta}$ to be non-vanishing when n increases. This is similar when compared to the factor model literature. However, as the granular shocks correspond directly to specific time series we can derive an exact condition for the relationship between $\boldsymbol{\beta}$ and the correlation among non-granular shocks $\boldsymbol{\Sigma}_\epsilon$ that will be sufficient for the detection of the granular series.

When compared to the factor model literature two main differences can be noted. First, assumption (iv) reflects that the granular loadings are not orthogonal to each other, if they where the condition (iv) would simplify to $\min_{i=1,\ldots,k} \|\boldsymbol{\beta}_i\| > C_{\Sigma_\epsilon}$. Assumption (iv) is stronger for correlated granular series. Second, we do not require the strength of the loading matrix to grow proportional to n, see also Onatski (2010) and Onatski (2012). This is in contrast to the work of Bai and Ng (2002) and Bai (2003) for example.

The assumptions 1 are sufficient to separate the granular series from the non-granular series. The statistics that we use to identify the granular series are the column norms of the inverse covariance matrix. In particular, for model (2) under assumption 1 the norms of the columns that corresponds to granular time series are larger when compared to those that of the non-granular series.

The following lemma establishes this formally.

Lemma 1. *Let y_t be generated by model (2). The granular detection statistics are defined as*

$$\|\boldsymbol{K}_i\|, \text{ where } \quad \boldsymbol{K} = (\boldsymbol{K}_1, \ldots, \boldsymbol{K}_n) \quad \text{and} \quad \boldsymbol{K} = \boldsymbol{\Sigma}_y^{-1}.$$

[3]The conditioning number in assumption (iv) can be relaxed to $\max_{j=1,\ldots,n-k} \frac{\|\boldsymbol{\Sigma}_{j,\epsilon}^{-1}\|}{\mu_{n-k}(\boldsymbol{\Sigma}_\epsilon^{-1})}$ for which holds $\max_{j=1,\ldots,n-k} \frac{\|\boldsymbol{\Sigma}_{j,\epsilon}^{-1}\|}{\mu_{n-k}(\boldsymbol{\Sigma}_\epsilon^{-1})} \leq C_{\Sigma_\epsilon}$.

Under assumption 1 (i)-(iv) $\boldsymbol{K}$ *exists and we have for* $n > N$ *that*

$$\|\boldsymbol{K}_i\| > \|\boldsymbol{K}_{k+j}\| \qquad \forall \quad i = 1, \dots, k, \quad \text{and} \quad j = 1, \dots, n-k,$$

where $\boldsymbol{K}_i$ *denotes the ith column of* $\boldsymbol{K} = \boldsymbol{\Sigma}_y^{-1}$ *with* $\boldsymbol{\Sigma}_y = \text{Var}(y_t)$.

All proofs are collected in the appendix. The lemma implies a ranking of the n time series according to their explanatory effect on the other series. In this perspective the ranking is based on the rescaled sum of squared partial correlations between each series and all others.

When we strengthen assumption 1 (iv) we can identify the number of granular series. In particular, we assume

(iv) Let the columns of* $\boldsymbol{\beta}$ *be ordered such that* $\|\boldsymbol{\beta}'\boldsymbol{\beta}_1\| \geq \ldots \geq \|\boldsymbol{\beta}'\boldsymbol{\beta}_k\|$. *There exists an integer* $N^* > 0$ *such that for all* $n > N^*$

$$\frac{\|\boldsymbol{\beta}'\boldsymbol{\beta}_k\|}{\|\boldsymbol{\beta}\|} > A_l + C_{\Sigma_\epsilon}^2 \frac{\|\boldsymbol{\beta}'\boldsymbol{\beta}_l\|}{\|\boldsymbol{\beta}'\boldsymbol{\beta}_{l+1}\|}$$

for all $l = 1, \dots, k-1$ *where* $A_l = C_{\Sigma_\epsilon} \frac{\mu_1(\boldsymbol{\Sigma}_g^{-1})}{\mu_{n-k}(\boldsymbol{\Sigma}_\epsilon^{-1})\|\boldsymbol{\beta}'\boldsymbol{\beta}_{l+1}\|}$

We emphasize that A_l is typically negligible in practical situations. The second term $C_{\Sigma_\epsilon}^2 \frac{\|\boldsymbol{\beta}'\boldsymbol{\beta}_l\|}{\|\boldsymbol{\beta}'\boldsymbol{\beta}_{l+1}\|}$ typically dominates the bound. We can replace C_{Σ_ϵ} by $\max_{j=1,\dots,n-k} \frac{\|\boldsymbol{\Sigma}_{j,\epsilon}^{-1}\|}{\mu_{n-k}(\boldsymbol{\Sigma}_\epsilon^{-1})}$, which is a less restrictive assumption but is less easy to interpret. Given the stronger condition on the loading matrix we obtain the following lemma that identifies k.

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

$$k = \underset{s=1,\dots,n-1}{\arg\max} \|\boldsymbol{K}_s\| \;/\; \|\boldsymbol{K}_{s+1}\|$$

where $\boldsymbol{K}_i$ *denotes the ith column of* $\boldsymbol{K} = \boldsymbol{\Sigma}_y^{-1}$ *with* $\boldsymbol{\Sigma}_y = \text{Var}(y_t)$.

The stronger assumption (iv*) can now be understood intuitively. First, to ensure that the column ratios among the non-granular series ($s = k+1, \dots, n-1$) are smaller when compared to $\|\boldsymbol{K}_k\|^2 \ / \ \|\boldsymbol{K}_{k+1}\|^2$ we need to allow for less cross-sectional correlation among the idiosyncratic components. A sufficient condition is given by $\frac{\|\boldsymbol{\beta}'\boldsymbol{\beta}_k\|}{\|\boldsymbol{\beta}\|} > C^2_{\Sigma_\epsilon}$. Second, to ensure that the ratios among the columns of the granular series ($s = 1, \ldots, k-1$) are smaller than $\|\boldsymbol{K}_k\|^2 \ / \ \|\boldsymbol{K}_{k+1}\|^2$ we additionally need that the decrease in strength from $\boldsymbol{\beta}_l$ to $\boldsymbol{\beta}_{l+1}$ (for the ordered $\boldsymbol{\beta}$'s) is not too large. This explain the scaling by the ratio $\|\boldsymbol{\beta}'\boldsymbol{\beta}_l\|/\|\boldsymbol{\beta}'\boldsymbol{\beta}_{l+1}\|$ in assumption (iv*). A similar condition was also found for the eigenvalue ratio statics for factor models in the simulation study of Ahn and Horenstein (2013). Given the structure of our model we are able to derive this condition explicitly.

Lemmas 1 and 2 are sufficient for the identification of the granular series. Lemma 1 ranks the series in order of granularity and lemma 2 determines the number of granular series.

It is important to notice that Lemma 2 is not the only function of the column norms that identifies k. In fact, many functions of the elements of the precision matrix $\boldsymbol{K}$ can identify the number of granular series. An interesting alternative choice is given by

$$k = \arg\min_{s=1,\ldots,n} \sum_{j=1}^{s}\sum_{i=1}^{N}(\boldsymbol{K}^2_{i,j} - \bar{\boldsymbol{K}}_s)^2 + \sum_{j=s+1}^{n}\sum_{i=1}^{N}(\boldsymbol{K}^2_{i,j} - \bar{\boldsymbol{K}}_{ns})^2$$

where $\bar{\boldsymbol{K}}_s = (Ns)^{-1}\sum_{j=1}^{s}\sum_{i=1}^{N}\boldsymbol{K}^2_{i,j}$ and $\bar{\boldsymbol{K}}_{ns} = (N(N-s))^{-1}\sum_{j=s+1}^{n}\sum_{i=1}^{N}\boldsymbol{K}^2_{i,j}$. This has the analogy of a break estimator.

3.2 Estimation

Next, we estimate the granular statistics $\|\boldsymbol{K}_i\|$, for $i = 1, \ldots, n$, using the sample of T observations for the vectors y_t. Let $\hat{\boldsymbol{\Sigma}} = T^{-1}\sum_{t=1}^{T} y_t y_t'$ and we define $\hat{\boldsymbol{K}} = \hat{\boldsymbol{\Sigma}}^{-1}$ as the sample concentration matrix. The sample equivalent of $\boldsymbol{K}$ allows us to operationalize the identification lemmas from the previous section. We need to impose conditions on the dependence properties and moments of g_t and ϵ_t such that the sample column norms converge to their population counterparts. For this several strategies are possible and we adopt similar assumptions when compared to the literature on approximate factor models, see Stock and Watson (2002) Bai and Ng (2002), Bai (2003) and **?**.

Assumption 2. *For model* (2) *we assume that for each* i, j *there exists a positive constants* $c_0 < c_1 < \infty$

(i) $\mathrm{E}\left(\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_t g_t' - \mathbf{\Sigma}_g\|^2\right) \leq c_1$.

(ii) $\mathrm{E}\left(\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_t \epsilon_{i,t}\|^2\right) \leq c_1$

(iii) $\mathrm{E}\left((\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \epsilon_{i,t}\epsilon_{j,t} - \mathbf{\Sigma}_{\epsilon,ij})^2\right) \leq c_1$

(iv) $0 < c_0 \leq \liminf_{n,T\to\infty} \mu_{n-k}(\boldsymbol{E}\boldsymbol{E}'/T) \leq \limsup_{n,T\to\infty} \mu_1(\boldsymbol{E}\boldsymbol{E}'/T) \leq c_1$ *where* $\boldsymbol{E} = (\epsilon_1, \ldots, \epsilon_T)$.

The assumptions are identical to those in Doz, Giannone and Reichlin (2012). They allow for a great deal of heterogeneity and dependence in the processes for g_t and $\epsilon_{i,t}$. Based on these assumptions the following theorem establishes the consistency of the column norms.

Theorem 1. *Let* y_t *be generated by model* (2) *under assumption 2 (i)-(ii) and 1 (i)-(iv) then for each* $i \in \{1, \ldots, n\}$ *we have that*

$$\|\hat{\boldsymbol{K}}_i\| - \|\boldsymbol{K}_i\| = \mathcal{O}_p\left(\frac{n}{T}\right)$$

where $\hat{\boldsymbol{K}}_i$ *denotes the ith column of* $\hat{\boldsymbol{K}}$ *with* $\hat{\boldsymbol{K}} = \left(T^{-1}\sum_{t=1}^{T} y_t y_t'\right)^{-1}$ *and* $\boldsymbol{K}_i$ *is the ith column of* $\boldsymbol{K} = \mathbf{\Sigma}_y^{-1}$ *with* $\mathbf{\Sigma}_y = \mathrm{Var}(y_t)$.

We use Theorem 1 to establish the following corollaries.

Corollary 1. *Let* y_t *be generated by model* (2) *under assumption 2 (i)-(ii) and 1 (i)-(iv). Consider the event* $\mathcal{E}$ *defined as*

$$\mathcal{E} = \left\{\|\hat{\boldsymbol{K}}_i\| > \|\hat{\boldsymbol{K}}_j\| \;\; \textit{for all} \;\; i = 1, ..., k \; \textit{and} \; j = k+1, ..., n\right\} .$$

Then,

$$\Pr(\mathcal{E}) = 1 - \mathcal{O}\left(\frac{n^2}{T}\right) .$$

Corollary 2. *Let y_t be generated by model* (2) *under assumption 2 (i)-(ii) and 1 (i)-(iv*). Let $\hat{k}$ be defined as*

$$\hat{k} = \arg \max_{s=1,\ldots,n-1} \|\hat{\boldsymbol{K}}_s\| / \|\hat{\boldsymbol{K}}_{s+1}\|.$$

Then,

$$\Pr(\hat{k} = k) = 1 - \mathcal{O}\left(\frac{n^2}{T}\right) .$$

Remark:

Ideally, when choosing an appropriate estimate for $\boldsymbol{K} = \mathrm{Var}(y_t)^{-1}$ we would like to exploit the structure of our model to obtain a fast convergence rate, see Pourahmadi (2013). In particular, for model (2) if the granular series were known we could adopt the covariance estimator for approximate factor models that was proposed in Fan, Liao and Mincheva (2011). However, since the granular series are unknown we need to rely on a more standard estimator for the concentration matrix. In future work we aim to adopt a regularized estimator for estimating the concentration matrix.

3.3 Asymptotic distribution

To asses the uncertainty around the estimated granular statistics $\|\hat{\boldsymbol{K}}_i\|$ we derive their limiting distribution. To facilitate the derivation of the limiting distribution we notice that the statistics can be written as

$$\|\hat{\boldsymbol{K}}_i\| = \frac{\sqrt{1 + \hat{\gamma}_i' \hat{\gamma}_i}}{\frac{1}{T} \hat{u}_{i,\cdot}' \hat{u}_{i,\cdot}},$$

where $\hat{\gamma}_i = (\boldsymbol{Y}_{-i}' \boldsymbol{Y}_{-i})^{-1} \boldsymbol{Y}_{-i}' y_{i,\cdot}$ with $y_{i,\cdot} = (y_{i,1}, \ldots, y_{i,T})'$, $\boldsymbol{Y}_i = (y_{1,\cdot}, \ldots, y_{i-1,\cdot}, y_{i+1,\cdot}, \ldots, y_{n,\cdot})$ and $\hat{u}_{i,\cdot} = y_{i,\cdot} - \boldsymbol{Y}_{-i} \hat{\gamma}_i,$. This representation follows from the regression representation for the columns of the inverse covariance matrix, see Pourahmadi (2013, Section 5.2) and the discussion in the Appendix A.1.

We notice that $\|\hat{\boldsymbol{K}}_i\|$ is a function of $\hat{\gamma}_i' \hat{\gamma}_i$ and $\frac{1}{T} \hat{u}_{i,\cdot}' \hat{u}_{i,\cdot}$. We derive the joint limiting of these two quadratic forms using the central limit theorem that is proposed in Kelejian and Prucha (2010). An application of the delta method then gives the distribution of $\|\hat{\boldsymbol{K}}_i\|$. The result is summarized in the following theorem.

3.4 Testing granular hypotheses

We are interested in testing hypotheses that concern the number of granular series. Many of the hypotheses of interest are composites of the simple null-hypothesis $H_0: \; \|\boldsymbol{K}_i\| = \|\boldsymbol{K}_j\|$, which asserts that the granular statistics for series i and j are equal. We first propose a test statistic for this hypothesis and derive its limiting distribution. Then we show that several composite hypotheses of interest can be based on the simple test.

3.5 Discussion

Next, we aim to justify the use of the norms of the columns of the concentration matrix for detecting granular series.

4 Granular panel data with common factors

4.1 General model and identification

Next, we extend the granular panel data model (2) to allows for the presence of common factors. The model is given by

$$\begin{aligned} y_{1:k,t} &= \boldsymbol{\Lambda}_1 f_t + g_t \\ y_{k+1:n,t} &= \boldsymbol{\Lambda}_2 f_t + \boldsymbol{\beta} g_t + \epsilon_t, \end{aligned} \tag{3}$$

where f_t is the $r \times 1$ vector of common dynamic factors and $\boldsymbol{\Lambda}_1$ and $\boldsymbol{\Lambda}_2$ are the $k \times r$ and $(N-k) \times r$ loadings matrices. We again work under the condition that the researcher does not know which series are the granular series. To compare, suppose that we would known which series are the granular series and adopt a vector autoregressive model for the g_t and f_t, the model would be similar to the factor augmented VAR that was proposed in Bernanke, Boivin and Eliasz (2005) see also Bai, Li and Lu (2016). Further, we also assume that the number of common factors r is unknown.

To identify the granular series and separate them from the common factors we make the following assumptions .

Assumption 3. *For model* (3) *we assume that*

(i) The assumptions 1 hold.

(ii) $\mathrm{E}(f_t) = 0$ *and* $\mathrm{E}(f_t f_t') = \boldsymbol{\Sigma}_f$.

(iii) $\mathrm{E}(f_t g_t') = 0$ *for all* t.

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

(v) Let $\boldsymbol{\Lambda} = (\boldsymbol{\Lambda}_1', \boldsymbol{\Lambda}_2')'$ *where* $\boldsymbol{\Lambda}_2'\boldsymbol{\Lambda}_2 = \boldsymbol{D}_{2,nT}$ *and* $\boldsymbol{\Lambda}'\boldsymbol{\Lambda} = \boldsymbol{D}_{nT}$ *and* $D_{2,nT}$ *and* D_{nT} *are diagonal with positive diagonal elements.*

(vi) $\mu_i(\boldsymbol{S}) \neq 1$ *for all* $i = 1, \ldots, n$ *where* $\boldsymbol{S} = \boldsymbol{K}\boldsymbol{\Lambda}(\boldsymbol{I}_r + \boldsymbol{\Lambda}'\boldsymbol{K}\boldsymbol{\Lambda})^{-1}\boldsymbol{\Lambda}'$.

Given assumptions 1, the following lemma identifies the granular series.

Lemma 3. *Let* y_t *be generated by model* (3) *under assumption 3 (i)-(vi). Then we have for large* n *that*

$$\|\boldsymbol{M}_i\| > \|\boldsymbol{M}_{k+j}\| \qquad \forall \quad i = 1, \ldots, k, \quad \textit{and} \quad j = 1, \ldots, n-k,$$

where $\boldsymbol{M}_i$ *denotes the* i*th column of* $\boldsymbol{M} = \boldsymbol{\Sigma}_y^{-1}$ *with* $\boldsymbol{\Sigma}_y = \mathrm{Var}(y_t)$.

The following lemma identifies the number of granular series.

Lemma 4. *Let* y_t *be generated by model* (3) *under assumptions 1 (i)-(vi) and the additional assumption (iv*). Then we have for large* n *that*

$$k = \underset{s=1,\ldots,n-1}{\arg\max} \, \|\boldsymbol{M}_s\| \; / \; \|\boldsymbol{M}_{s+1}\|$$

where $\boldsymbol{M}_i$ *denotes the* i*th column of* $\boldsymbol{M} = \boldsymbol{\Sigma}_y^{-1}$ *with* $\boldsymbol{\Sigma}_y = \mathrm{Var}(y_t)$.

4.2 Granular detection and testing

Assumption 4. *For model* (3) *we assume that*

(i) f_t *is a stationary ergodic vector sequence, such that* $T^{-1}\sum_{t=1}^{T} f_t f_t' \xrightarrow{p} \boldsymbol{\Sigma}_f$.

Theorem 2. *Let* y_t *be generated by model* (3) *under assumption 4 (i)-(iii) and 3 (i)-(iv) then for all* $i = 1, \dots, n$ *we have that*

$$\|\hat{\boldsymbol{M}}_i\| - \|\boldsymbol{M}_i\| = \mathcal{O}_p\left(\frac{n}{T}\right)$$

where $\hat{\boldsymbol{M}}_i$ *denotes the ith column of* $\hat{\boldsymbol{M}}$ *with* $\hat{\boldsymbol{M}} = \left(T^{-1}\sum_{t=1}^{T} y_t y_t'\right)^{-1}$ *and* $\boldsymbol{M}_i$ *is the ith column of* $\boldsymbol{M} = \boldsymbol{\Sigma}_y^{-1}$ *with* $\boldsymbol{\Sigma}_y = \mathrm{Var}(y_t)$.

4.3 Determining the number of common factors

Once the granular series have been identified and tested, we complete the specification of model (3) by determining the number of common factors. Several estimators and hypothesis tests have been previously proposed, see Bai and Ng (2002), Onatski (2009), Onatski (2010) and Ahn and Horenstein (2013) for examples. We build on these estimators.

4.4 Model estimation

Once the granular time series and the number of common factors are known, we can estimate the parameters of the granular panel data model. A restricted version of model (3) can be written in state space form and the parameters can be estimated using maximum likelihood. The likelihood can be evaluated using the Kalman filter and the expectation-maximization algorithm, see Durbin and Koopman (2012).

Despite the fact that we estimate a restricted version of the true model, quasi-maximum likelihood theories that were developed in Bai and Li (2012) and Bai and Li (2015) can be adopted to show consistency of the parameters. In a second step the Kalman filter and smoothing recursions can be adopted to obtain estimates for the common factors f_t and the granular shocks g_t. The consistency and asymptotic normality for the estimates of f_t and g_t is discussed in Doz et al. (2012), Bai and Li (2012) and Bai and Li (2015).

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 $\|\hat{\boldsymbol{K}}_i\|$ under different data generating processes. The outcome criteria that we are interested in are as follows. First, we study the percentage of correctly selected granular series. That is the fraction of the number of true granular series that correspond to the k largest granular statistics. Second, we asses frequency by which we correctly select the number of granular series. Third and finally, we evaluate the size and power properties of the hypothesis tests. This includes an evaluation of the asymptotic distribution of the granular statistic.

5.1 Simulation design

We generate data panels from the granular panel data model with common factors given in equation (3). 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 $r = 0, 5$.

The granular shocks and common factors, summarized in $h_t = (f_t', g_t')'$, follow a vector autoregressive process of order one.

$$h_t = \boldsymbol{\Phi} h_{t-1} + \eta_t \qquad \eta_t \sim NID(0, \boldsymbol{\Sigma}_\eta),$$

where $\boldsymbol{\Sigma}_\eta = \text{diag}(\boldsymbol{\Sigma}_{11,\eta}, \boldsymbol{\Sigma}_{22,\eta})$ with $\boldsymbol{\Sigma}_{11,\eta} = \boldsymbol{I}_r$ and $\boldsymbol{\Sigma}_{22,\eta}$ is $k \times k$ with ones on the main diagonal and correlation coefficient c_η in the off-diagonal elements. We note that c_η captures the contemporaneous correlation between the granular shocks. We vary its value by taking

$c_\eta = 0.0, 0.9$. The common factors shocks are kept orthogonal with unit variance. The elements for the diagonal of $\mathbf{\Phi}$ 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 h_t admits a stationary vector autoregressive process. The strength of the granular shocks is determined by $\boldsymbol{\beta}$. 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 $\sigma_b^2 = 0.1, 0.25, 0.5, 0.75, 1$. The loadings of the common factors are drawn from a standard normal distribution. Experiments that varied the strength of the common factors had no influence on the granular detection methods, nor the hypothesis tests.

For the idiosyncratic, non-granular, shocks $\epsilon_{i,t}$ we consider four different specifications that allow for different types of correlation across i and t. We have the following specification for $\epsilon_{i,t}$

$$\epsilon_{i,t} = \sqrt{\frac{1-\rho^2}{1+2Jc_\epsilon^2}}\xi_{it},$$

$\xi_{i,t} = \rho\xi_{i,t-1} + v_{i,t} + \sum_{h=\max(i-J,1)}^{i-1} c_\epsilon v_{h,t} + \sum_{h=i+1}^{\min(i+J,n)} c_\epsilon v_{h,t}$ where $v_{i,t} \sim NID(0, 1)$. This specification for the error terms is similar when compared to Bai and Ng (2002), Onatski (2010) and Ahn and Horenstein (2013) and it allows for serial and cross-sectional correlation. The first scaling term on the right hand side ensures that $\text{Var}(\epsilon_{i,t}) = 1$ for most series[4] and thus when we vary σ_b^2 we are changing the relative importance of the granular series. We consider four scenarios for the parameters that are required for drawing $\epsilon_{i,t}$: (a) iid series ($\rho = c_\epsilon = J = 0$), (b) serially correlated series ($\rho = 0.5$ and $c_\epsilon = J = 0$), (c) cross-sectional correlated series ($\rho = 0$, $c_\epsilon = 0.5$ and $J = 10$) and (d) serially and cross-sectional correlated series ($\rho = 0.5$, $c_\epsilon = 0.5$ and $J = 10$). These four cases are analogous to the ones considered in Ahn and Horenstein (2013).

In total we have six dimensions along which we vary the granular panel data model: 1. panel dimensions, 2. number of granular, 3. number of factors, 4. effect of the granular, 5.

[4]In particular, the variance will be one for all series $J + 1 \leq i \leq N - J$.

correlation among the granulars and 6. specification of the non-granular shocks. For each possible combination across these six dimension we draw $S = 1000$ different data panels. For each panel we rank, select and test the granular series.

5.2 Ranking 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 inverse sample covariance matrix, and then we select the number of granulars using the column ratio statistic of corollary 2. 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 cases considered, we provide more detailed simulation results for a number of benchmark scenarios and provided summaries for the remaining ones.

We present the results of our benchmark scenario in Table **??**. For our benchmark scenario the number of granulars k is 5, the number of factors r is 4 and the standard deviation of the granulars coefficients b is 0.4. The table reports the average proportion of correctly ranked granulars and average number of selected granulars as a function of T and N and for the four different dependence settings considered. In the iid case (setting (a)), the granular detection methodology is able to perfectly recover the set of granulars for all possible combinations of N and T. Interestingly, allowing for serial correlation (setting (b)) only mildly worsens the performance of the methodology. Once cross-sectional dependence is present (setting (c)), the performance of the granular detection procedure declines when the panel dimensions are not sufficiently large. In particular, the granular selection procedure has the tendency of selecting a larger number of granular of time series. Finally, when the granualar shocks are both serially and cross-sectional correlated, the performance of the granular detection methodology is roughly the same one of setting (c) with slightly worse statistics.

Inspection of the whole set of replications (not reported in the text) shows some interesting patterns. First, the performance of the methodology is only mildly affected by the numbers of factors in the specification. The simulation results show that the performance statistics are negatively correlated with the number of factors in the specification but the decreases is mild. As far as the number of granular shocks in the model the estimation results do not show a clear pattern in the performance statistics and the number of granulars considered in the model.

A key parameter of our simulation setting is the standard deviation of granular standard deviation coefficients b. In order to investigate its role in Figure **??** and Figure **??** we show the performance statistics of the granular ranking and selection procedures as a function of b. In particular, the picture shows ranking and selection statistics when the number of factors r is 3, the number of granulars k is 5, the cross-sectional dimension N is 50 for the four dependence settings considered in the study and different sample sizes T. As far as the ranking statistics are concerned, the picture shows that the proportion of correctly ranked statistics rapidly grows as a function b. When cross-sectional dependence is present in the model larger values of b are required to obtain consistent ranking. As far as the selection statistic are concerned, the picture shows that as b grows the selection procedure selects on average the correct number of granulars. When the magnitude of b is small, the selection procedure has the tendency to choose a larger number of granulars on average. Again, when cross sectional is present larger values of b are required for the granular selection procedure to perform satisfactorily.

In figure 1 we report the plot of the probability of correct ranking and (b) the average selected number of granulars relative to the correct number of granulars as a function of the standard deviation of the granular coefficients β and the coefficient c_ϵ which determines the correlation of the the non-granualr coefficients. These simulation have been obtained in the case a DGP with $r = 1$ factors, $k = 5$ granulars and panel dimensions N and T equal to, respectively, 50 and 400.

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.

5.3 Testing results

sectionEmpirical Applications

5.4 Granular Detection in the Financial System

In this application we focus on detecting granulars in a panel of volatility measures of large US financial institutions. The application is close in spirit to the work of, among others, Billio et al. (2012) and Diebold and Yılmaz (2014).

We consider a panel of large US financial firms during the 2007-2009 Great Financial Crisis. The list of companies is in Table 3. The sample roughly matches the same companies used in other studies (see Brownlees and Engle (2016) and Acharya, Pedersen, Philippon and Richardson (2016)). It is important to stress that we only consider firms that have been trading throughout the sample, which implies that a number of institutions such as Lehman Brothers, Bear Stearns, Freddie Mac and Fannie Mae are not included in our analysis. The sample period spans March 1st 2007 to March 1st 2009. Following, Diebold and Yılmaz (2014) we measure volatility using the high-low range Parkinson (1980)

$$\tilde{\sigma}^2_{i,t} = 0.361 \left(p^{\mathsf{high}}_{i,t} - p^{\mathsf{low}}_{i,t} \right) ,$$

where $p^{\mathsf{high}}_{i,t}$ and $p^{\mathsf{low}}_{i,t}$ denote respectively the max and the min log price of stock i on day t.[5] As it is customary, we analyse the log of the high-low range rather than its level. The final panel dimension is $T = 503$ and $N = 88$.

[5] It is important to acknowledge that more precise estimators based on high frequency data could also have been employed (see, *inter alia*, Andersen, Bollerslev, Diebold and Labys (2003), Barndorff-Nielsen, Hansen, Lunde and Shephard (2008)). The high-low range however has been documented to perform well relative to more advanced alternatives (Alizadeh, Brandt and Diebold (2002)).

The volatility panel exhibits the typical stylized facts documented in the literature. Table 4 reports summary statistics on the series in the panel. The series have positive skewness, excess kurtosis and a strong degree of persistence. There is strong evidence of a single factor structure: A principal component analysis reveals that the 1st principal component explains roughly 76% of the overall variation in the panel. The first principal component can be associated with the overall degree of volatility in the market, in fact, the correlation with the high-low range of the S&P 500 is 88%. The second principal component on the other hand explains less than 4% of the overall variation in the panel. These results are in-line with the evidence documented in Luciani and Veredas (2015) and Barigozzi and Hallin (2015).

We apply the granular detection methodology described in the previous sections to the volatility panel. Table 5 reports the granular rankings of the top twenty firms as well as the value of the concentration matrix column ratio statistic. It is natural to think of the granular financial institutions in the volatility panel as systemic. To this extent, the table also flags the institutions that are classified as either Globally Systemic (G-SIB) or Domestically Systemic (D-SIB) by the Financial Stability Board. Inspection of the full set of results reveals that larger firms are typically ranked higher: The rank correlation between $\|\hat{\boldsymbol{K}}_i\|$ and firm size is 0.39. The top ten includes a number of financial institutions that have been indeed deeply involved with the financial crisis and its unwinding, that is Bank of America, JPMorgan and Wells Fargo. These are also firms classified by the Financial Stability Board as G-SIBs. The top ten also contains several D-SIBs like Northern Trust, Comerica. The $\hat{k}$ ratio statistic however indicates that only a small number of firms in the panel is granulars: Bank of America, JP Morgan and Northern Trust.

We explore more thoroughly the relation between our granular statistic and the set of SIFIs identified by the FSB. To this extent we define a SIFI binary response indicator s_i as a n-dimensional vector of dummies, the i-th element of which is one if institutions i is either a D-SIB or a G-SIB and zero otherwise. We then model the SIFI indicator using the following logit regression model

$$\text{logit}(p_i) = c_0 + c_1 \|\hat{\boldsymbol{K}}_i\| + c_2 \text{vol}_i + c_3 \text{siz}_i + c_4 \text{lvg}_i \,,$$

where vol_i denotes the average volatility, siz_i denotes the size i and lvg_i the leverage of firm i. We report in Table 6 the estimation results of the logit regression under different sets of restrictions. The estimation results show that the granular ranking statistic and size contribute significantly to probability of being a SIFI whereas volatility and leverage are not significant. Also the magnitude of the psuedo-R^2 shows that the contribution of the granular statistic is sizeable.

6 Conclusion

In this work we introduce a panel model in which the idiosyncratic shocks of a (finite) subset of time-series influences 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 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. We apply our framework to study systemic risk in finance using a panel of volatility measures during the financial crisis (Diebold and Yılmaz (2014)). The methodology delivers economically meaningful rankings of the most systemic institutions in the panel and identifies in particular JP Morgan, Northern Trust and Bank of America as the granular institutions in the panel.

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.

Acharya, V., Pedersen, L., Philippon, T. and Richardson, M.: 2016, Measuring Systemic Risk, *Review of Financial Studies* p. forthcoming.

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

Alizadeh, S., Brandt, M. W. and Diebold, F. X.: 2002, Range-based estimation of stochastic volatility models, *The Journal of Finance* **57**, 1047–1091.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P.: 2003, Modeling and forecasting realized volatility, *Econometrica* **71**, 579–625.

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.

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.: 2015, Maximum Likelihood Estimation and Inference for Approximate Factor Models of High Dimension, *Review of Economics and Statistics* . forthcoming.

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

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

Barigozzi, M. and Hallin, M.: 2015, Generalized dynamic factor models and volatilities: recovering the market volatility shocks, *The Econometrics Journal* .

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. and Shephard, N.: 2008, Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise, *Econometrica* **76**, 1481–1536.

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.

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.: 2016, SRISK: A Conditional Capital Shortfall Measure of Systemic Risk, *Review of Financial Studies* p. forthcoming.

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.

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

Fan, J., Liao, Y. and Yao, J.: 2015, Power enhancement in high-dimensional cross-sectional tests, *Econometrica* **83**, 1497–1541.

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.

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

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

Kelejian, H. K. and Prucha, I. R.: 2010, Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances, *Journal of Econometrics* **157**, 53–67.

Luciani, M. and Veredas, D.: 2015, Estimating and forecasting large panels of volatilities with approximate dynamic factor models, *Journal of Forecasting* **34**, 163–176.

Magnus, J. R. and Neudecker, H.: 2007, *Matrix Differential Calculus with Applications in Statistics and Econometrics (Third Edition)*, Wiley, Chichester.

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

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.

Parkinson, M.: 1980, The Extreme Value Method for Estimating the Variance of the Rate of Return, *The Journal of Business* **53**, 61–65.

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. and Yamagata, T.: 2012, Testing CAPM with a Large Number of Assets. IZA discussion paper No. 6469.

Pourahmadi, M.: 2013, *High Dimensional Covariance Estimation*, John Wiley & Sons, Hoboken, New Jersey.

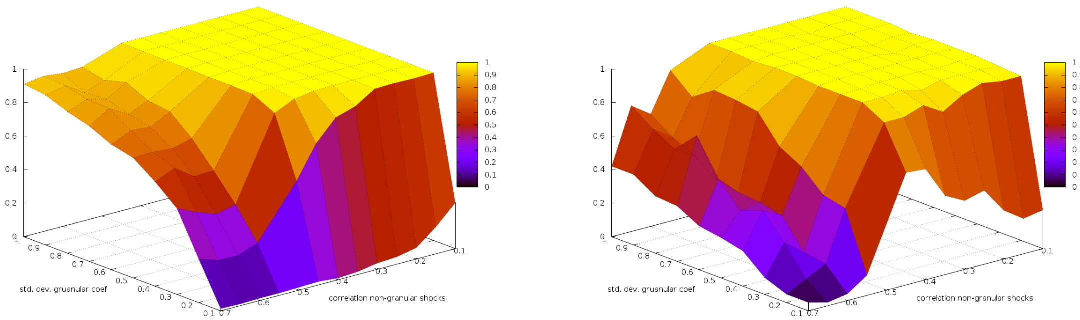
Siavash, S. S.: 2016, Dominant Sectors in the US: A Factor Model Analysis of Sectoral Industrial Production. Working paper.

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

Stock, J. H. and Watson, M. W.: 2011, Dynamic Factor Models, *in* M. P. Clements and D. F. Hendry (eds), *Oxford Handbook of Economic Forecasting*, Oxford University Press, Oxford.

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

Figure 1: Granular Detection Probabilities



(a) (b)

The figure display (a) the probability of correct ranking and (b) probability of correct selection of the number of granulars as a function of the standard deviation of the granular coefficients β and the coefficient c_ϵ which determines the correlation of the non-granular shocks.

Table 1: Granular Ranking

N	T	k	r	c_η	0.01	0.05	0.10	0.25	0.50	0.75	1.00
50	200	3	0	0	0.006	0.829	0.991	1.000	1.000	1.000	1.000
100	200	3	0	0	0.103	0.996	1.000	1.000	1.000	1.000	1.000
50	400	3	0	0	0.006	0.879	0.998	1.000	1.000	1.000	1.000
100	400	3	0	0	0.136	0.999	1.000	1.000	1.000	1.000	1.000
50	200	5	0	0	0.009	0.774	0.982	1.000	1.000	1.000	1.000
100	200	5	0	0	0.135	0.997	1.000	1.000	1.000	1.000	1.000
50	400	5	0	0	0.008	0.826	0.993	1.000	1.000	1.000	1.000
100	400	5	0	0	0.180	1.000	1.000	1.000	1.000	1.000	1.000
50	200	3	5	0	0.004	0.838	0.994	1.000	1.000	1.000	1.000
100	200	3	5	0	0.105	0.998	1	1.000	1.000	1.000	1.000
50	400	3	5	0	0.005	0.880	0.998	1.000	1.000	1.000	1.000
100	400	3	5	0	0.137	0.999	1.000	1.000	1.000	1.000	1.000
50	200	5	5	0	0.009	0.787	0.981	1.000	1.000	1.000	1.000
100	200	5	5	0	0.135	0.996	1.000	1.000	1.000	1.000	1.000
50	400	5	5	0	0.009	0.831	0.991	1.000	1.000	1.000	1.000
100	400	5	5	0	0.176	0.999	1.000	1.000	1.000	1.000	1.000
50	200	3	0	0.9	0.009	0.831	0.991	1.000	1.000	1.000	1.000
100	200	3	0	0.9	0.109	0.998	1	1.000	1.000	1.000	1.000
50	400	3	0	0.9	0.008	0.890	0.999	1.000	1.000	1.000	1.000
100	400	3	0	0.9	0.157	1.000	1.000	1.000	1.000	1.000	1.000
50	200	5	0	0.9	0.013	0.791	0.982	1.000	1.000	1.000	1.000
100	200	5	0	0.9	0.151	0.997	1.000	1.000	1.000	1.000	1.000
50	400	5	0	0.9	0.011	0.844	0.991	1.000	1.000	1.000	1.000
100	400	5	0	0.9	0.205	0.999	1.000	1.000	1.000	1.000	1.000
50	200	3	5	0.9	0.008	0.831	0.992	1.000	1.000	1.000	1.000
100	200	3	5	0.9	0.103	0.997	1.000	1.000	1.000	1.000	1.000
50	400	3	5	0.9	0.006	0.881	0.998	1.000	1.000	1.000	1.000
100	400	3	5	0.9	0.148	1.000	1.000	1.000	1.000	1.000	1.000
50	200	5	5	0.9	0.007	0.789	0.982	1.000	1.000	1.000	1.000
100	200	5	5	0.9	0.131	0.996	1.000	1.000	1.000	1.000	1.000
50	400	5	5	0.9	0.009	0.831	0.993	1.000	1.000	1.000	1.000
100	400	5	5	0.9	0.186	0.999	1.000	1.000	1.000	1.000	1.000

The table shows simulation results for the granular ranking method. We report the average proportion of correctly detected granulars.

Table 2: Granular Selection

N	T	k	r	c_η	0.01	0.05	0.10	0.25	0.50	0.75	1.00
50	200	3	0	0	0.105	0.170	0.689	0.953	0.993	0.995	1.000
100	200	3	0	0	0.114	0.790	0.971	0.999	1.000	1.000	1.000
50	400	3	0	0	0.132	0.229	0.759	0.972	0.995	0.998	1.000
100	400	3	0	0	0.123	0.925	0.996	1.000	1.000	1.000	1.000
50	200	5	0	0	0.059	0.049	0.462	0.913	0.99	0.996	0.999
100	200	5	0	0	0.034	0.703	0.967	0.999	1.000	1.000	1.000
50	400	5	0	0	0.047	0.066	0.578	0.956	0.99	0.998	1.000
100	400	5	0	0	0.046	0.896	0.998	1.000	1.000	1.000	1.000
50	200	3	5	0	0.130	0.204	0.671	0.969	0.994	0.993	0.997
100	200	3	5	0	0.121	0.764	0.979	1.000	1.000	1.000	1.000
50	400	3	5	0	0.131	0.248	0.736	0.964	0.995	0.999	0.999
100	400	3	5	0	0.122	0.912	1.000	1.000	1.000	1.000	1.000
50	200	5	5	0	0.054	0.053	0.484	0.933	0.99	0.996	1.000
100	200	5	5	0	0.047	0.689	0.976	1.000	1.000	1.000	1.000
50	400	5	5	0	0.057	0.064	0.593	0.942	0.992	0.998	0.999
100	400	5	5	0	0.051	0.882	0.996	1.000	1.000	1.000	1.000
50	200	3	0	0.9	0.113	0.199	0.700	0.961	0.986	0.997	0.999
100	200	3	0	0.9	0.108	0.798	0.978	1.000	1.000	1.000	1.000
50	400	3	0	0.9	0.127	0.250	0.810	0.974	0.997	1.000	1.000
100	400	3	0	0.9	0.153	0.925	0.997	1.000	1.000	1.000	1.000
50	200	5	0	0.9	0.052	0.056	0.508	0.919	0.989	0.991	0.999
100	200	5	0	0.9	0.047	0.692	0.974	1.000	1.000	1.000	1.000
50	400	5	0	0.9	0.054	0.076	0.613	0.943	0.994	0.999	0.997
100	400	5	0	0.9	0.042	0.916	0.999	1.000	1.000	1.000	1.000
50	200	3	5	0.9	0.134	0.182	0.688	0.952	0.993	0.997	0.998
100	200	3	5	0.9	0.140	0.765	0.974	1.000	1.000	1.000	1.000
50	400	3	5	0.9	0.127	0.239	0.755	0.973	0.994	0.998	0.999
100	400	3	5	0.9	0.136	0.912	0.997	1.000	1.000	1.000	1.000
50	200	5	5	0.9	0.058	0.058	0.439	0.917	0.986	0.994	0.998
100	200	5	5	0.9	0.050	0.673	0.969	1.000	1.000	1.000	1.000
50	400	5	5	0.9	0.054	0.050	0.612	0.944	0.995	0.998	1.000
100	400	5	5	0.9	0.052	0.899	0.994	1.000	1.000	1.000	1.000

The table reports the simulation results for the granular selection method. We report the average proportion of times that we select the correct number of granulars.

Table 3: Descriptive Statististics

Ticker	Company Name	Ticker	Company name
ANAT	American National Insurance Co	GNW	Genworth Financial
AMP	Signature Bank/New York NY	HRB	HR Block
AFG	America Financial Group	HBHC	Hancock Holding Co
AIG	American International Group	THG	The Hanover Insurance Group
AINV	Apollo Investment	HIG	Hartford Financial Services Group/The
ASB	Associated Banc-Corp	HBAN	Huntington Bancshares/OH
AIZ	Assurant	ISBC	Investors Bancorp
BOH	Bank of Hawaii	JNS	Janus Capital Group
ALL	Allstate	JLL	Jones Lang LaSalle
AMG	Affiliated Managers Group.	KMPR	Kemper
AXP	American Express Co	KEY	KeyCorp
BAC	Bank of America	LM	Legg Mason
BK	The Bank of New York Mellon.	LNC	Lincoln National
BOKF	BOK Financial	MBI	MBIA
BRO	Brown and Brown	MCY	Mercury General
CFFN	Capitol Federal Financial	NYCB	New York Community Bancorp
C	Citigroup.	NTRS	Northern Trust
COF	Capital One Financial	ORI	Old Republic International
GS	Goldman Sachs Group	PBCT	People s United Financial
JPM	JPMorgan	PNC	PNC Financial
MET	MetLife.	PFG	Principal Financial Group
MS	Morgan Stanley	PRA	ProAssurance
SPG	Simon Property Group	PB	Prosperity Bancshares
USB	U.S. Bancorp	PRU	Prudential Financial
WFC	Wells Fargo	RJF	Raymond James Financial
CSH	Cash America International	RF	Regions Financial
CBG	CBRE Group	SBNY	Signature Bank/New York NY
CNA	CNA Financial	SLM	SLM
CNO	CNO Financial Group	STT	State Street
CNS	Cohen and Steers	SF	Stifel Financial
CMA	Comerica	STI	SunTrust Banks
CBSH	Commerce Bancshares/MO	SIVB	SVB Financial Group
CACC	Credit Acceptance	SNV	Synovus Financial
CFR	Cullen/Frost Bankers	TCB	TCF Financial
ETFC	E*TRADE Financial	TMK	Torchmark
EWBC	East West Bancorp	UMBF	UMB Financial
EV	Eaton Vance	UNM	Unum Group
ERIE	Erie Indemnity Co	VLY	Valley National Bancorp
EZPW	Ezcorp	WDR	Waddell and Reed Financial
FII	Federated Investors	WAFD	Washington Federal
FCNCA	First Citizens BancShares/NC	WBS	Webster Financial
FHN	First Horizon National	WTM	White Mountains Insurance Group Ltd
FCE-A	Forest City Realty Trust	WRB	WR Berkley
FULT	Fulton Financial	ZION	Zions Bancorporation

The table reports the list of tickers and company names of the financial panel.

Table 4: Descriptive Statististics

	Mean	Std Dev	Skew	Kurt	ACF(1)	ACF(22)
$q_{0.25}$	-3.131	0.507	0.352	2.717	0.725	0.417
Median	-3.003	0.574	0.508	2.955	0.771	0.478
$q_{0.75}$	-2.916	0.633	0.746	3.301	0.816	0.527

The table reports the 1st quartile, median and 3rd quartile

Table 5: Granular Rankings

Rank	Granulars	K-Ratio	G-SIB	D-SIB
1	JPMorgan	3.951	✓	
2	Northern Trust	2.148		✓
3	Bank of America	16.894	✓	
4	Commerce Bancshares/MO	14.520		
5	Comerica	0.265		✓
6	Allstate	2.673		
7	Torchmark	2.296		
8	Wells Fargo	0.498	✓	
9	U.S. Bancorp	0.202		✓
10	Bank of Hawaii	3.195		
11	Cullen/Frost Bankers	0.335		
12	Associated Banc-Corp	1.008		
13	American Express Co	14.521		✓
14	Goldman Sachs Group	2.460	✓	
15	Prosperity Bancshares	1.900		
16	MetLife.	0.740		✓
17	Valley National Bancorp	5.926		
18	UMB Financial	1.609		
19	Citigroup.	0.184	✓	
20	Regions Financial	3.420		✓

The table reports

Table 6: SIFI Prediction

$\|\boldsymbol{K}_i\|$	0.441^{***} (0.103)				1.072^{***} (0.057)
vol_i		0.868 (1.734)			2.628^{***} (1.045)
siz_i			0.431^{***} (0.093)		0.763^{***} (0.286)
lvg_i				0.008 (0.007)	-0.033 (0.023)
$\tilde{R}^2$	0.354	0.003	0.405	0.014	0.745

The table reports

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 $\boldsymbol{B}$ the k-th largest eigenvalue of $\boldsymbol{B}$ is denoted as $\mu_k(\boldsymbol{B})$. For an $M \times N$ matrix $\boldsymbol{A}$ the k-th largest singular value of $\boldsymbol{A}$ is denoted as $\sigma_k(\boldsymbol{A})$. As a matrix norm we generally adopt the spectral norm is given by $\|\boldsymbol{A}\|_2 = \sqrt{\mu_1(\boldsymbol{A}'\boldsymbol{A})}$. We drop the index when no confusion can arise and write $\|\boldsymbol{A}\|_2 = \|\boldsymbol{A}\|$. The frobenius norm is given by $\|\boldsymbol{A}\|_F = \sqrt{\sum_{i=1}^N \sum_{j=1}^N a_{i,j}^2} = \sqrt{\text{Trace}(\boldsymbol{A}'\boldsymbol{A})}$. We have $\|\boldsymbol{A}\|_2 \leq \|\boldsymbol{A}\|_F \leq \text{rank}(\boldsymbol{A})\|\boldsymbol{A}\|_2$. For a square matrix $\boldsymbol{B}$ we let $\boldsymbol{B} \succ 0$ indicate that $\boldsymbol{B}$ 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.

Proof of Lemma 1. Under assumptions 1 (i), (ii) and (iii) the population variance of the observations $\boldsymbol{\Sigma}_y = \text{Var}(y_t)$ and its inverse $\mathbf{K} = \boldsymbol{\Sigma}_y^{-1}$ are given by

$$\boldsymbol{\Sigma}_y = \begin{bmatrix} \boldsymbol{\Sigma}_g & \boldsymbol{\Sigma}_g\boldsymbol{\beta}' \\ \boldsymbol{\beta}\boldsymbol{\Sigma}_g & \boldsymbol{\beta}\boldsymbol{\Sigma}_g\boldsymbol{\beta}' + \boldsymbol{\Sigma}_\epsilon \end{bmatrix} \qquad \mathbf{K} = \begin{bmatrix} \boldsymbol{\Sigma}_g^{-1} + \boldsymbol{\beta}'\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta} & -\boldsymbol{\beta}'\boldsymbol{\Sigma}_\epsilon^{-1} \\ -\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta} & \boldsymbol{\Sigma}_\epsilon^{-1} \end{bmatrix}, \tag{4}$$

where the inverse follows from the inverse formula for block matrices, see Magnus and Neudecker (2007, page 12).

First, we show that assumptions 1 ensure that $\boldsymbol{K}$ exist. A lower bound on the minimum eigenvalue of $\boldsymbol{\Sigma}_y$ can be established by finding an upper bound on the largest eigenvalue of its inverse $\boldsymbol{K}$. Note that

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

Thus,

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

For the first term we have

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

When combining with the second term and doing the inverse we have

$$\begin{aligned} \mu_n(\boldsymbol{\Sigma}_y) &= \|\boldsymbol{K}\|^{-1} \\ &\geq [(1 + \|\boldsymbol{\beta}\|)^2 \max\{\mu_1(\boldsymbol{\Sigma}_g^{-1}), \mu_1(\boldsymbol{\Sigma}_\epsilon^{-1})\}]^{-1} \\ &= (1 + \|\boldsymbol{\beta}\|)^{-2} \min\{\mu_n(\boldsymbol{\Sigma}_g), \mu_n(\boldsymbol{\Sigma}_\epsilon)\} \end{aligned}$$

We notice assumption 1 (iv) implies that for $n \to \infty$ we have $\|\boldsymbol{\beta}\| = \sqrt{\mu_1(\boldsymbol{\beta}'\boldsymbol{\beta})} \to \sqrt{\mu_1(\boldsymbol{D})} < \infty$. This implies that as long as the minimum limiting eigenvalues of $\mu_n(\boldsymbol{\Sigma}_g)$ and $\mu_n(\boldsymbol{\Sigma}_\epsilon)$

remain positive the smallest eigenvalue of $\mathbf{\Sigma}_y$ will remain to be bounded away from zero.

Next, we show that 1 (iv) is sufficient for showing that $\|\boldsymbol{K}_i\|^2 = \|\boldsymbol{K}e_{n,i}\|^2 \geq \|\boldsymbol{K}e_{n,k+j}\|^2 = \|\boldsymbol{K}_{k+j}\|^2$ for each $i \in \{1, ..., k\}$ and $j \in \{1, ..., n-k\}$, where $\boldsymbol{K}$ is defined above in equation (4).

We note that

$$\begin{aligned} \|\mathbf{K}e_{n,i}\|^2 &= \|(\mathbf{\Sigma}_g^{-1} + \boldsymbol{\beta}'\mathbf{\Sigma}_\epsilon^{-1}\boldsymbol{\beta})e_{k,i}\|^2 + \|\mathbf{\Sigma}_\epsilon^{-1}\boldsymbol{\beta}e_{k,i}\|^2 \\ \|\mathbf{K}e_{n-k,j}\|^2 &= \|\boldsymbol{\beta}'\mathbf{\Sigma}_\epsilon^{-1}e_{n-k,j}\|^2 + \|\mathbf{\Sigma}_\epsilon^{-1}e_{n-k,j}\|^2. \end{aligned}$$

To find the conditions for $\|\mathbf{K}e_{n,i}\|^2 \geq \|\mathbf{K}e_{n,k+j}\|^2$ we study

(a) $\|(\mathbf{\Sigma}_g^{-1} + \boldsymbol{\beta}'\mathbf{\Sigma}_\epsilon^{-1}\boldsymbol{\beta})e_{k,i}\|^2 \geq \|\boldsymbol{\beta}'\mathbf{\Sigma}_\epsilon^{-1}e_{n-k,j}\|^2$

(b) $\|\mathbf{\Sigma}_\epsilon^{-1}\boldsymbol{\beta}e_{k,i}\|^2 \geq \|\mathbf{\Sigma}_\epsilon^{-1}e_{n-k,j}\|^2$

Starting with condition (b), we have that

$$\begin{aligned} \|\mathbf{\Sigma}_\epsilon^{-1}\boldsymbol{\beta}e_{k,i}\|^2 &= \|\mathbf{\Sigma}_\epsilon^{-1}\boldsymbol{\beta}_i\|^2 && \text{definition } e_{k,i} \\ &\geq \mu_{n-k}^2(\mathbf{\Sigma}_\epsilon^{-1})\|\boldsymbol{\beta}_i\|^2 && \mathbf{\Sigma}_\epsilon \succ 0 \\ \|\mathbf{\Sigma}_\epsilon^{-1}e_{n-k,j}\|^2 &= \|\mathbf{\Sigma}_{j,\epsilon}^{-1}\|^2 && \text{change notation} \end{aligned}$$

such that we may conclude that (b) holds if

$$\min_{i=1,\ldots,k} \|\boldsymbol{\beta}_i\| > \max_{j=1,\ldots,n-k} \frac{\|\mathbf{\Sigma}_{j,\epsilon}^{-1}\|}{\mu_{n-k}(\mathbf{\Sigma}_\epsilon^{-1})}$$

Now, we consider (a) for which we have

$$\begin{aligned} \|(\mathbf{\Sigma}_g^{-1} + \boldsymbol{\beta}'\mathbf{\Sigma}_\epsilon^{-1}\boldsymbol{\beta})e_{k,i}\|^2 &\geq \|\boldsymbol{\beta}'\mathbf{\Sigma}_\epsilon^{-1}\boldsymbol{\beta}_i\|^2 && \mathbf{\Sigma}_g \succ 0 \\ &\geq \mu_{n-k}^2(\mathbf{\Sigma}_\epsilon^{-1})\|\boldsymbol{\beta}'\boldsymbol{\beta}_i\|^2 && \Sigma_\epsilon \succ 0 \\ \|\boldsymbol{\beta}'\mathbf{\Sigma}_\epsilon^{-1}e_{n-k,j}\|^2 &\leq \|\boldsymbol{\beta}\|^2\|\mathbf{\Sigma}_{j,\epsilon}^{-1}\|^2 && \text{Cauchy-Schwartz} \end{aligned}$$

So we obtain

$$\min_{i=1,\ldots,k} \frac{\|\boldsymbol{\beta}'\boldsymbol{\beta}_i\|}{\|\boldsymbol{\beta}\|} > \max_{j=1,\ldots,n-k} \frac{\|\mathbf{\Sigma}_{j,\epsilon}^{-1}\|}{\mu_{n-k}(\mathbf{\Sigma}_\epsilon^{-1})}.$$

Since, $\min_{i=1,\ldots,k} \|\boldsymbol{\beta}'\boldsymbol{\beta}_i\|/\|\boldsymbol{\beta}\| \leq \min_{i=1,\ldots,k} \|\boldsymbol{\beta}_i\|$, we find that condition (b) is implied by condition (a). Further, for all $j = 1, \ldots, n-k$ we have

$$\frac{\|\mathbf{\Sigma}_{j,\epsilon}^{-1}\|}{\mu_{n-k}(\mathbf{\Sigma}_\epsilon^{-1})} \leq \frac{\mu_1(\mathbf{\Sigma}_\epsilon^{-1})}{\mu_{n-k}(\mathbf{\Sigma}_\epsilon^{-1})} = \frac{\mu_n^{-1}(\mathbf{\Sigma}_\epsilon)}{\mu_1^{-1}(\mathbf{\Sigma}_\epsilon)} = \frac{\mu_1(\mathbf{\Sigma}_\epsilon)}{\mu_n(\mathbf{\Sigma}_\epsilon)} = C_{\Sigma_\epsilon}$$

where C_{Σ_ϵ} is the conditioning number of the matrix Σ_ϵ. This completes the proof. □

Proof of Lemma 2. Assume, without loss of generality that the columns of $\boldsymbol{K}$ are ordered in decreasing order by their squared column 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)-(iii), see (4). We need that

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

First, we consider the case where $j-k = s = k+1,\dots,n$. We have

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

which is 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. First, $\frac{a}{c} > \frac{e}{g}$ is given by

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

for which we have from lemma 1

$$\frac{\|(\boldsymbol{\Sigma}_g^{-1} + \boldsymbol{\beta}'\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta})e_{k,k}\|^2}{\|\boldsymbol{\beta}'\boldsymbol{\Sigma}_\epsilon^{-1}e_{n-k,1}\|^2} \geq \frac{\mu_{n-k}^2(\boldsymbol{\Sigma}_\epsilon^{-1})\|\boldsymbol{\beta}'\boldsymbol{\beta}_k\|}{\mu_1^2(\boldsymbol{\Sigma}_\epsilon^{-1})\|\boldsymbol{\beta}\|^2}$$

and

$$\frac{\|\boldsymbol{\beta}'\boldsymbol{\Sigma}_\epsilon^{-1}e_{n-k,j}\|^2}{\|\boldsymbol{\beta}'\boldsymbol{\Sigma}_\epsilon^{-1}e_{n-k,j+1}\|^2} \leq \frac{\mu_1^2(\boldsymbol{\Sigma}_\epsilon^{-1})}{\mu_{n-k}^2(\boldsymbol{\Sigma}_\epsilon^{-1})}$$

which implies the condition

$$\frac{\|\boldsymbol{\beta}'\boldsymbol{\beta}_k\|}{\|\boldsymbol{\beta}\|} > C_{\Sigma_\epsilon}^2.$$

where C_{Σ_ϵ} is the conditioning number of the matrix C_{Σ_ϵ}. Notice that we have imposed that column k corresponds to the weakest granular. Next, $\frac{a}{c} > \frac{f}{h}$ is equal to

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

The bound for the right hand side is given by

$$\frac{\|\boldsymbol{\Sigma}_\epsilon^{-1}e_{n-k,j}\|^2}{\|\boldsymbol{\Sigma}_\epsilon^{-1}e_{n-k,j+1}\|^2} \leq C_{\Sigma_\epsilon}^2.$$

When combined with the bound for $\frac{a}{c}$ derived above we have the condition

$$\frac{\|\boldsymbol{\beta}'\boldsymbol{\beta}_k\|}{\|\boldsymbol{\beta}\|} > C_{\Sigma_\epsilon}^2$$

Next, $\frac{b}{d} > \frac{e}{g}$ is given by

$$\frac{\|\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta}e_{k,k}\|^2}{\|\boldsymbol{\Sigma}_\epsilon^{-1}e_{n-k,1}\|^2} > \frac{\|\boldsymbol{\beta}'\boldsymbol{\Sigma}_\epsilon^{-1}e_{n-k,j}\|^2}{\|\boldsymbol{\beta}'\boldsymbol{\Sigma}_\epsilon^{-1}e_{n-k,j+1}\|^2}$$

where we use

$$\frac{\|\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta} e_{k,k}\|^2}{\|\boldsymbol{\Sigma}_\epsilon^{-1} e_{n-k,1}\|^2} \geq \frac{\mu_{n-k}^2(\boldsymbol{\Sigma}_\epsilon^{-1})\|\boldsymbol{\beta}_k\|^2}{\mu_1^2(\boldsymbol{\Sigma}_\epsilon^{-1})}$$

and the bound for $\frac{e}{g}$ was given above such that we have

$$\|\boldsymbol{\beta}_k\| > C_{\Sigma_\epsilon}^2$$

Finally, for the last part $\frac{b}{d} > \frac{f}{h}$ we have

$$\frac{\|\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta} e_{k,k}\|^2}{\|\boldsymbol{\Sigma}_\epsilon^{-1} e_{n-k,1}\|^2} \geq \frac{\|\boldsymbol{\Sigma}_\epsilon^{-1} e_{n-k,j}\|^2}{\|\boldsymbol{\Sigma}_\epsilon^{-1} e_{n-k,j+1}\|^2}$$

which by the bounds derived above implies the condition

$$\|\boldsymbol{\beta}_k\| \geq C_{\Sigma_\epsilon}^2.$$

Next, we consider the case where $i = s = 1, \dots, k-1$. We have

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

which is of the form $\frac{a'+b'}{c'+d'} > \frac{e'+f'}{g'+h'}$ that is implied by $\frac{a'}{c'} > \frac{e'}{g'}$, $\frac{a'}{c'} > \frac{f'}{h'}$, $\frac{b'}{d'} > \frac{e'}{g'}$ and $\frac{b'}{d'} > \frac{f'}{h'}$. Starting with $\frac{a'}{c'} > \frac{e'}{g'}$ that is given by

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

after taking the square root on both sides we have for the left hand side

$$\frac{\|(\boldsymbol{\Sigma}_g^{-1} + \boldsymbol{\beta}'\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta})e_{k,k}\|}{\|\boldsymbol{\beta}'\boldsymbol{\Sigma}_\epsilon^{-1} e_{n-k,1}\|} \geq \frac{\mu_{n-k}(\boldsymbol{\Sigma}_\epsilon^{-1})\|\boldsymbol{\beta}'\boldsymbol{\beta}_k\|}{\mu_1(\boldsymbol{\Sigma}_\epsilon^{-1})\|\boldsymbol{\beta}\|}$$

and we have for the right hand side

$$\frac{\|(\boldsymbol{\Sigma}_g^{-1} + \boldsymbol{\beta}'\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta})e_{k,i}\|}{\|(\boldsymbol{\Sigma}_g^{-1} + \boldsymbol{\beta}'\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta})e_{k,i+1}\|} \leq \frac{\mu_1(\boldsymbol{\Sigma}_g^{-1}) + \mu_1(\boldsymbol{\Sigma}_\epsilon^{-1})\|\boldsymbol{\beta}'\boldsymbol{\beta}_i\|}{\mu_{n-k}(\boldsymbol{\Sigma}_\epsilon^{-1})\|\boldsymbol{\beta}'\boldsymbol{\beta}_{i+1}\|}$$

Such that we have the sufficient inequality

$$\frac{\mu_{n-k}(\boldsymbol{\Sigma}_\epsilon^{-1})\|\boldsymbol{\beta}'\boldsymbol{\beta}_k\|}{\mu_1(\boldsymbol{\Sigma}_\epsilon^{-1})\|\boldsymbol{\beta}\|} > \frac{\mu_1(\boldsymbol{\Sigma}_g^{-1}) + \mu_1(\boldsymbol{\Sigma}_\epsilon^{-1})\|\boldsymbol{\beta}'\boldsymbol{\beta}_i\|}{\mu_{n-k}(\boldsymbol{\Sigma}_\epsilon^{-1})\|\boldsymbol{\beta}'\boldsymbol{\beta}_{i+1}\|}$$

which can be rewritten as

$$\frac{\|\boldsymbol{\beta}'\boldsymbol{\beta}_k\|}{\|\boldsymbol{\beta}\|} > \frac{\mu_1(\boldsymbol{\Sigma}_\epsilon^{-1})\mu_1(\boldsymbol{\Sigma}_g^{-1}) + \mu_1^2(\boldsymbol{\Sigma}_\epsilon^{-1})\|\boldsymbol{\beta}'\boldsymbol{\beta}_i\|}{\mu_{n-k}^2(\boldsymbol{\Sigma}_\epsilon^{-1})\|\boldsymbol{\beta}'\boldsymbol{\beta}_{i+1}\|}$$

Next, we consider $\frac{a'}{c'} > \frac{f'}{h'}$ which is given by

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

For the right hand side we have

$$\frac{\|\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta}e_{k,i}\|^2}{\|\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta}e_{k,i+1}\|^2} \leq \frac{\mu_1^2(\boldsymbol{\Sigma}_\epsilon^{-1})\|\boldsymbol{\beta}_i\|^2}{\mu_n^2(\boldsymbol{\Sigma}_\epsilon^{-1})\|\boldsymbol{\beta}_{i+1}\|^2}$$

and combined with the bound for the left hand side that was derived above we have that

$$\frac{\|\boldsymbol{\beta}'\boldsymbol{\beta}_k\|}{\|\boldsymbol{\beta}\|} > C_{\Sigma_\epsilon}^2 \frac{\|\boldsymbol{\beta}_i\|}{\|\boldsymbol{\beta}_{i+1}\|}$$

Next, for $\frac{b'}{d'} > \frac{e'}{g'}$ we have

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

and the bound for the left hand side is given by

$$\frac{\|\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta}e_{k,k}\|^2}{\|\boldsymbol{\Sigma}_\epsilon^{-1}e_{n-k,1}\|^2} \geq \frac{\mu_n^2(\boldsymbol{\Sigma}_\epsilon^{-1})\|\boldsymbol{\beta}_k\|^2}{\mu_1^2(\boldsymbol{\Sigma}_\epsilon^{-1})}$$

and when combined with the bound derived above we have

$$\|\boldsymbol{\beta}_k\| > \frac{\mu_1(\boldsymbol{\Sigma}_\epsilon^{-1})\mu_1(\boldsymbol{\Sigma}_g^{-1}) + \mu_1^2(\boldsymbol{\Sigma}_\epsilon^{-1})\|\boldsymbol{\beta}'\boldsymbol{\beta}_i\|}{\mu_{n-k}^2(\boldsymbol{\Sigma}_\epsilon^{-1})\|\boldsymbol{\beta}'\boldsymbol{\beta}_{i+1}\|}.$$

Finally for the last condition $\frac{b'}{d'} > \frac{f'}{h'}$ we have

$$\frac{\|\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta}e_{k,k}\|^2}{\|\boldsymbol{\Sigma}_\epsilon^{-1}e_{n-k,1}\|^2} > \frac{\|\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta}e_{k,i}\|^2}{\|\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta}e_{k,i+1}\|^2}$$

and the condition becomes

$$\|\boldsymbol{\beta}_k\| > C_{\Sigma_\epsilon}^2 \frac{\|\boldsymbol{\beta}_i\|}{\|\boldsymbol{\beta}_{i+1}\|}.$$

Summarizing, all conditions 8 that we derived are implied by

$$\frac{\|\boldsymbol{\beta}'\boldsymbol{\beta}_k\|}{\|\boldsymbol{\beta}\|} > \frac{\mu_1(\boldsymbol{\Sigma}_\epsilon^{-1})\mu_1(\boldsymbol{\Sigma}_g^{-1}) + \mu_1^2(\boldsymbol{\Sigma}_\epsilon^{-1})\|\boldsymbol{\beta}'\boldsymbol{\beta}_i\|}{\mu_{n-k}^2(\boldsymbol{\Sigma}_\epsilon^{-1})\|\boldsymbol{\beta}'\boldsymbol{\beta}_{i+1}\|}$$

which is the condition given in (iv*). □

Proof of Lemma 3. For model (3) under assumptions 3 (i), (ii), (iii) and (iv), the variance $\mathbf{\Omega} = \text{Var}(y_t)$ and its inverse $\mathbf{M} = \mathbf{\Omega}^{-1}$ are given by

$$\mathbf{\Omega} = \mathbf{\Sigma} + \mathbf{\Lambda}\mathbf{\Lambda}' \qquad \mathbf{M} = \boldsymbol{K} - \boldsymbol{K}\mathbf{\Lambda}(\boldsymbol{I}_r + \mathbf{\Lambda}'\boldsymbol{K}\mathbf{\Lambda})^{-1}\mathbf{\Lambda}'\boldsymbol{K},$$

where $\boldsymbol{\Sigma}$ and $\boldsymbol{K}$ are given in (4). We show that under (i), (v) and (vi) $\|\boldsymbol{M}e_{n,i}\|^2 \geq \|\boldsymbol{M}e_{n,k+j}\|^2$ for all $i = 1, \ldots, k$ and $j = 1, \ldots, n-k$.

We define $\boldsymbol{K}_s = \boldsymbol{K}e_{n,s}$ for $s = 1, \ldots, n$ as the sth column of $\boldsymbol{K}$. Further, let

$$\boldsymbol{S} = \boldsymbol{K}\boldsymbol{\Lambda}(\boldsymbol{I}_r + \boldsymbol{\Lambda}'\boldsymbol{K}\boldsymbol{\Lambda})^{-1}\boldsymbol{\Lambda}'$$

such that we have for all $s = 1, \ldots, n$

$$\begin{array}{rcll} \|\boldsymbol{M}e_{n,s}\|^2 & = & \|\boldsymbol{K}_s - \boldsymbol{S}\boldsymbol{K}_s\|^2 & \text{definition} \\ & = & (\boldsymbol{K}_s - \boldsymbol{S}\boldsymbol{K}_s)'(\boldsymbol{K}_s - \boldsymbol{S}\boldsymbol{K}_s) & \text{definition} \\ & = & \boldsymbol{K}_s'\boldsymbol{K}_s - \boldsymbol{K}_s'\boldsymbol{S}'\boldsymbol{K}_s - \boldsymbol{K}_s'\boldsymbol{S}\boldsymbol{K}_s + \boldsymbol{K}_s'\boldsymbol{S}'\boldsymbol{S}\boldsymbol{K}_s & \text{work out product} \\ & = & \boldsymbol{K}_s'(\boldsymbol{I}_n - \boldsymbol{S}' - \boldsymbol{S} + \boldsymbol{S}'\boldsymbol{S})\boldsymbol{K}_s & \text{regroup} \end{array}$$

We notice that $\boldsymbol{I}_n - \boldsymbol{S}' - \boldsymbol{S} + \boldsymbol{S}'\boldsymbol{S}$ is symmetric. We obtain the following inequality for $\|\boldsymbol{M}e_{n,i}\|^2 \geq \|\boldsymbol{M}e_{n,k+j}\|^2$

$$\boldsymbol{K}_i'(\boldsymbol{I}_n - \boldsymbol{S}' - \boldsymbol{S} + \boldsymbol{S}'\boldsymbol{S})\boldsymbol{K}_i \geq \boldsymbol{K}_{k+j}'(\boldsymbol{I}_n - \boldsymbol{S}' - \boldsymbol{S} + \boldsymbol{S}'\boldsymbol{S})\boldsymbol{K}_{k+j}$$

We rewrite

$$\boldsymbol{K}_i'(\boldsymbol{I}_n - \boldsymbol{S}' - \boldsymbol{S} + \boldsymbol{S}'\boldsymbol{S})\boldsymbol{K}_i - \boldsymbol{K}_{k+j}'(\boldsymbol{I}_n - \boldsymbol{S}' - \boldsymbol{S} + \boldsymbol{S}'\boldsymbol{S})\boldsymbol{K}_{k+j} = (\boldsymbol{K}_i - \boldsymbol{K}_{k+j})'(\boldsymbol{I}_n - \boldsymbol{S}' - \boldsymbol{S} + \boldsymbol{S}'\boldsymbol{S})(\boldsymbol{K}_i + \boldsymbol{K}_{k+j}) \geq 0$$

Now we have that

$$\begin{array}{rcl} (\boldsymbol{K}_i - \boldsymbol{K}_{k+j})'(\boldsymbol{I}_n - \boldsymbol{S}' - \boldsymbol{S} + \boldsymbol{S}'\boldsymbol{S})(\boldsymbol{K}_i + \boldsymbol{K}_{k+j}) & \geq & \mu_n(\boldsymbol{I}_n - \boldsymbol{S}' - \boldsymbol{S} + \boldsymbol{S}'\boldsymbol{S})(\boldsymbol{K}_i - \boldsymbol{K}_{k+j})'(\boldsymbol{K}_i + \boldsymbol{K}_{k+j}) \\ & = & \mu_n(\boldsymbol{I}_n - \boldsymbol{S}' - \boldsymbol{S} + \boldsymbol{S}'\boldsymbol{S})(\boldsymbol{K}_i'\boldsymbol{K}_i - \boldsymbol{K}_{k+j}'\boldsymbol{K}_{k+j}) \end{array}$$

which is greater then zero if $\mu_n(\boldsymbol{I}_n - \boldsymbol{S}' - \boldsymbol{S} + \boldsymbol{S}'\boldsymbol{S}) \geq 0$ and $\boldsymbol{K}_i'\boldsymbol{K}_i - \boldsymbol{K}_{k+j}'\boldsymbol{K}_{k+j} \geq 0$. The conditions for $\boldsymbol{K}_i'\boldsymbol{K}_i - \boldsymbol{K}_{k+j}'\boldsymbol{K}_{k+j} \geq 0$ are established in lemma 1. We note that

$$\boldsymbol{I}_n - \boldsymbol{S}' - \boldsymbol{S} + \boldsymbol{S}'\boldsymbol{S} = (\boldsymbol{I}_n - \boldsymbol{S})'(\boldsymbol{I}_n - \boldsymbol{S}) \succeq 0$$

which implies that $\mu_n(\boldsymbol{I}_n - \boldsymbol{S}' - \boldsymbol{S} + \boldsymbol{S}'\boldsymbol{S}) \geq 0$. We get an equality whenever $\boldsymbol{I}_n - \boldsymbol{S}' - \boldsymbol{S} + \boldsymbol{S}'\boldsymbol{S}$ has an eigenvalue that is zero. *We should look into this case more carefully..*

□

Proof of Lemma 4. Under the given assumptions Lemma 2 shows that

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

Further, from Lemma 3 we have that for all $i = 1, \ldots, n$

$$\|\boldsymbol{M}e_{n,i}\|^2 = \boldsymbol{K}_i'(\boldsymbol{I}_n - \boldsymbol{S}' - \boldsymbol{S} + \boldsymbol{S}'\boldsymbol{S})\boldsymbol{K}_i = \boldsymbol{K}_i'\boldsymbol{X}\boldsymbol{K}_i$$

where we have defined $\boldsymbol{X} = \boldsymbol{I}_n - \boldsymbol{S}' - \boldsymbol{S} + \boldsymbol{S}'\boldsymbol{S}$. We need to proof that for all $s = 1, \ldots, k-1, k+1, \ldots, n-1$ we have

$$\frac{\|\boldsymbol{M}e_{n,k}\|^2}{\|\boldsymbol{M}e_{n,k+1}\|^2} > \frac{\|\boldsymbol{M}e_{n,s}\|^2}{\|\boldsymbol{M}e_{n,s+1}\|^2}$$

This can be written as

$$\frac{\boldsymbol{K}_k'\boldsymbol{X}\boldsymbol{K}_k}{\boldsymbol{K}_{k+1}'\boldsymbol{X}\boldsymbol{K}_{k+1}} > \frac{\boldsymbol{K}_s'\boldsymbol{X}\boldsymbol{K}_s}{\boldsymbol{K}_{s+1}'\boldsymbol{X}\boldsymbol{K}_{s+1}}.$$

Next, we define for all $i = 1, \ldots, n$

$$\tilde{\boldsymbol{K}}_i = \frac{\boldsymbol{K}_i}{(\boldsymbol{K}_{i+1}'\boldsymbol{X}\boldsymbol{K}_{i+1})^{1/2}}$$

such that we may write

$$\frac{\boldsymbol{K}_i'\boldsymbol{X}\boldsymbol{K}_i}{\boldsymbol{K}_{i+1}'\boldsymbol{X}\boldsymbol{K}_{i+1}} = \tilde{\boldsymbol{K}}_i'\boldsymbol{X}\tilde{\boldsymbol{K}}_i.$$

Now the inequality becomes

$$\tilde{\boldsymbol{K}}_k'\boldsymbol{X}\tilde{\boldsymbol{K}}_k > \tilde{\boldsymbol{K}}_s'\boldsymbol{X}\tilde{\boldsymbol{K}}_s$$

which can be rewritten as

$$\begin{aligned} \tilde{\boldsymbol{K}}_k'\boldsymbol{X}\tilde{\boldsymbol{K}}_k - \tilde{\boldsymbol{K}}_s'\boldsymbol{X}\tilde{\boldsymbol{K}}_s &= (\tilde{\boldsymbol{K}}_k - \tilde{\boldsymbol{K}}_s)'\boldsymbol{X}(\tilde{\boldsymbol{K}}_k - \tilde{\boldsymbol{K}}_s) \\ &\geq \mu_n(\boldsymbol{X})(\tilde{\boldsymbol{K}}_k - \tilde{\boldsymbol{K}}_s)'(\tilde{\boldsymbol{K}}_k - \tilde{\boldsymbol{K}}_s) \\ &= \mu_n(\boldsymbol{X})(\tilde{\boldsymbol{K}}_k'\tilde{\boldsymbol{K}}_k - \tilde{\boldsymbol{K}}_s'\tilde{\boldsymbol{K}}_s) > 0. \end{aligned}$$

The last inequality is true if $\mu_n(\boldsymbol{X}) > 0$ and $\tilde{\boldsymbol{K}}_k'\tilde{\boldsymbol{K}}_k - \tilde{\boldsymbol{K}}_s'\tilde{\boldsymbol{K}}_s > 0$. The first requirement is met by assumption (vi). We continue with the second which can be written as

$$\frac{\boldsymbol{K}_k'\boldsymbol{K}_k}{\boldsymbol{K}_{k+1}'\boldsymbol{X}\boldsymbol{K}_{k+1}} - \frac{\boldsymbol{K}_s'\boldsymbol{K}_s}{\boldsymbol{K}_{s+1}'\boldsymbol{X}\boldsymbol{K}_{s+1}} > 0$$

Now we define

$$\check{\boldsymbol{K}}_{i+1} = \frac{\boldsymbol{K}_{i+1}}{(\boldsymbol{K}_i'\boldsymbol{K}_i)^{1/2}}$$

such that we can write the inequality as

$$\frac{1}{\check{\boldsymbol{K}}_{k+1}'\boldsymbol{X}\check{\boldsymbol{K}}_{k+1}} - \frac{1}{\check{\boldsymbol{K}}_{s+1}'\boldsymbol{X}\check{\boldsymbol{K}}_{s+1}} > 0$$

which becomes

$$\frac{\check{\boldsymbol{K}}_{s+1}'\boldsymbol{X}\check{\boldsymbol{K}}_{s+1} - \check{\boldsymbol{K}}_{k+1}'\boldsymbol{X}\check{\boldsymbol{K}}_{k+1}}{\check{\boldsymbol{K}}_{k+1}'\boldsymbol{X}\check{\boldsymbol{K}}_{k+1}\check{\boldsymbol{K}}_{s+1}'\boldsymbol{X}\check{\boldsymbol{K}}_{s+1}} > 0$$

This is greater than zero when the numerator is greater than zero, which implies

$$\check{\boldsymbol{K}}_{s+1}'\boldsymbol{X}\check{\boldsymbol{K}}_{s+1} - \check{\boldsymbol{K}}_{k+1}'\boldsymbol{X}\check{\boldsymbol{K}}_{k+1} \geq \mu_n(\boldsymbol{X})(\check{\boldsymbol{K}}_{s+1}'\check{\boldsymbol{K}}_{s+1} - \check{\boldsymbol{K}}_{k+1}'\check{\boldsymbol{K}}_{k+1}) > 0$$

which is greater than zero if $\mu_n(\boldsymbol{X}) > 0$ (see above) and $\check{\boldsymbol{K}}_{s+1}'\check{\boldsymbol{K}}_{s+1} - \check{\boldsymbol{K}}_{k+1}'\check{\boldsymbol{K}}_{k+1} > 0$. The latter requirement can be written as

$$\frac{\boldsymbol{K}_{s+1}'\boldsymbol{K}_{s+1}}{\boldsymbol{K}_s'\boldsymbol{K}_s} > \frac{\boldsymbol{K}_{k+1}'\boldsymbol{K}_{k+1}}{\boldsymbol{K}_k'\boldsymbol{K}_k}$$

Finally, when taking the inverse on both sides we have that

$$\frac{\boldsymbol{K}_s'\boldsymbol{K}_s}{\boldsymbol{K}_{s+1}'\boldsymbol{K}_{s+1}} < \frac{\boldsymbol{K}_k'\boldsymbol{K}_k}{\boldsymbol{K}_{k+1}'\boldsymbol{K}_{k+1}}$$

which was proven in Lemma 2. □

Proof of Theorem 1. From the inverse for block matrices, see Magnus and Neudecker (2007, page 12), it follows that the columns of the estimated precision matrix $\hat{\boldsymbol{K}}$ can be written as

$$\hat{\boldsymbol{K}}_i = -\hat{\boldsymbol{K}}_{ii}\hat{\boldsymbol{\Gamma}}_i \qquad \hat{\boldsymbol{\Gamma}}_i = (\hat{\gamma}_{i,1}, \ldots, \hat{\gamma}_{i,i-1}, 1, \hat{\gamma}_{i,i+1}, \ldots, \hat{\gamma}_{i,n})'$$

and $\hat{\gamma}_i = (\hat{\gamma}_{i,1}, \ldots, \hat{\gamma}_{i,i-1}, \hat{\gamma}_{i,i+1}, \ldots, \hat{\gamma}_{i,n})'$ is given by

$$\hat{\gamma}_i = \left(\frac{1}{T}\boldsymbol{Y}_{-i}'\boldsymbol{Y}_{-i}\right)^{-1}\frac{1}{T}\boldsymbol{Y}_{-i}'y_{i,.}$$

for $i = 1, \ldots, n$. Using the representation discussed in Section A.1 we obtain the error form $\hat{\gamma}_i - \gamma_i = \left(\frac{1}{T}\boldsymbol{Y}_{-i}'\boldsymbol{Y}_{-i}\right)^{-1}\frac{1}{T}\boldsymbol{Y}_{-i}'u_{i,.}$, where $u_{i,.} = y_{i,.} - \boldsymbol{Y}_{-i}\gamma_i$ is the population prediction error.

We have from the triangle inequality that

$$\begin{aligned} \|\hat{\boldsymbol{K}}_i\| - \|\boldsymbol{K}_i\| &= \boldsymbol{K}_{ii}\|\boldsymbol{\Gamma}_i\| - \hat{\boldsymbol{K}}_{ii}\|\hat{\boldsymbol{\Gamma}}_i\| \\ &\leq \left|\boldsymbol{K}_{ii} - \hat{\boldsymbol{K}}_{ii}\right|\|\boldsymbol{\Gamma}_i\| + \hat{\boldsymbol{K}}_{ii}\|\hat{\boldsymbol{\Gamma}}_i - \boldsymbol{\Gamma}_i\| \\ &= \left|\boldsymbol{K}_{ii} - \hat{\boldsymbol{K}}_{ii}\right|\|\boldsymbol{\Gamma}_i\| + \hat{\boldsymbol{K}}_{ii}\|\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i\| \end{aligned}$$

First, we consider

$$\begin{aligned} \|\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i\| &= \left\|\left(\tfrac{1}{T}\boldsymbol{Y}_{-i}'\boldsymbol{Y}_{-i}\right)^{-1}\tfrac{1}{T}\boldsymbol{Y}_{-i}'u_{i,.}\right\| \\ &\leq \left\|\left(\tfrac{1}{T}\boldsymbol{Y}_{-i}'\boldsymbol{Y}_{-i}\right)^{-1}\right\|\left\|\tfrac{1}{T}\boldsymbol{Y}_{-i}'u_{i,.}\right\| \end{aligned}$$

For the first term, Lemma 6 (i) and (ii) show that $0 < \lim_{n,T\to\infty}\|\frac{1}{T}\boldsymbol{Y}'\boldsymbol{Y}\| < \infty$, which implies that $0 < \lim_{n,T\to\infty}\|\frac{1}{T}\boldsymbol{Y}_{-i}'\boldsymbol{Y}_{-i}\| < \infty$. Hence, $\left\|\left(\frac{1}{T}\boldsymbol{Y}_{-i}'\boldsymbol{Y}_{-i}\right)^{-1}\right\| = \mathcal{O}_p(1)$. Next, we consider $\left\|\frac{1}{T}\boldsymbol{Y}_{-i}'u_{i,.}\right\|$. We can write

$$\left\|\frac{1}{T}\boldsymbol{Y}_{-i}'u_{i,.}\right\| = \sqrt{\frac{n}{T}}\sqrt{\frac{1}{n}\sum_{j\neq i}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}y_{j,t}u_{i,t}\right)^2}.$$

Hence to show that $\left\|\frac{1}{T}\boldsymbol{Y}_{-i}'u_{i,.}\right\| = \mathcal{O}_p(\sqrt{n/T})$ we need that $\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}y_{j,t}u_{i,t}\right)^2 = \mathcal{O}_p(1)$.

Using $u_{i,t} = y_{i,t} - \sum_{l\neq i} \gamma_{i,l} y_{l,t}$

$$\begin{aligned}
\mathrm{E}\left(\left(\tfrac{1}{\sqrt{T}}\sum_{t=1}^{T} y_{j,t}u_{i,t}\right)^2\right) &= \mathrm{E}\left(\left(\tfrac{1}{\sqrt{T}}\sum_{t=1}^{T} y_{j,t}y_{i,t} - \tfrac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{l\neq i}\gamma_{i,l}y_{j,t}y_{l,t}\right)^2\right) \\
&\leq 2\left\{\mathrm{E}\left(\left(\tfrac{1}{\sqrt{T}}\sum_{t=1}^{T} y_{j,t}y_{i,t} - \tfrac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathrm{E}(y_{j,t}y_{i,t})\right)^2\right)\right. \\
&\quad \left. +\mathrm{E}\left(\left(\tfrac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathrm{E}(y_{j,t}y_{i,t}) - \tfrac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{l\neq i}\gamma_{i,l}y_{j,t}y_{l,t}\right)^2\right)\right\}
\end{aligned}$$

For the first term we have

$$\begin{aligned}
&\sup_{i,j}\mathrm{E}\left(\left(\tfrac{1}{\sqrt{T}}\sum_{t=1}^{T} y_{j,t}y_{i,t} - \tfrac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathrm{E}(y_{j,t}y_{i,t})\right)^2\right) \\
&= \sup_{i,j}\mathrm{E}\left(\left(\beta_j'\left(\tfrac{1}{\sqrt{T}}\sum_{t=1}^{T} g_t g_t' - \boldsymbol{\Sigma}_g\right)\beta_i + \beta_j'\tfrac{1}{\sqrt{T}}\sum_{t=1}^{T} g_t\epsilon_{i,t} + \right.\right. \\
&\left.\left.\beta_i'\tfrac{1}{\sqrt{T}}\sum_{t=1}^{T} g_t\epsilon_{j,t} + \tfrac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\epsilon_{i,t}\epsilon_{j,t} - \mathrm{E}(\epsilon_{i,t}\epsilon_{j,t})\right)\right)^2\right) \\
&\leq \sup_{i,j} 4\left\{\|\beta_j\|\|\beta_i\|\mathrm{E}\left(\|\tfrac{1}{\sqrt{T}}\sum_{t=1}^{T} g_t g_t' - \boldsymbol{\Sigma}_g\|^2\right) + \|\beta_j\|\mathrm{E}\left(\|\tfrac{1}{\sqrt{T}}\sum_{t=1}^{T} g_t\epsilon_{i,t}\|^2\right)\right. \\
&\left. +\|\beta_i\|\mathrm{E}\left(\|\tfrac{1}{\sqrt{T}}\sum_{t=1}^{T} g_t\epsilon_{j,t}\|^2\right) + \mathrm{E}\left(\left(\tfrac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\epsilon_{i,t}\epsilon_{j,t} - \mathrm{E}(\epsilon_{i,t}\epsilon_{j,t})\right)\right)^2\right)\right\} \\
&\leq \|\beta_j\|\|\beta_i\|c_1 + \|\beta_j\|c_1 + \|\beta_i\|c_1 + c_1 = c_2 < \infty
\end{aligned}$$

where the final inequality follows from assumption 2 (i)-(iii). For the second term we notice that $\mathrm{E}(y_{j,t}y_{i,t}) = \mathrm{E}[\mathrm{E}(y_{i,t}|Y_{-i,t})y_{j,t}] = \mathrm{E}(\sum_{l\neq i}\gamma_{i,l}y_{l,t}y_{j,t}) = \sum_{l\neq i}\gamma_{i,l}\mathrm{E}(y_{l,t}y_{j,t})$, as $y_{i,t} = \sum_{l\neq i}\gamma_{i,l}y_{l,t} + u_{i,t}$. We have

$$\begin{aligned}
&\mathrm{E}\left(\left(\tfrac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{l\neq i}\gamma_{i,l}y_{j,t}y_{l,t} - \tfrac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathrm{E}(y_{j,t}y_{i,t})\right)^2\right) \\
&= \mathrm{E}\left(\left(\sum_{l\neq i}\gamma_{i,l}\tfrac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(y_{j,t}y_{l,t} - \mathrm{E}(y_{j,t}y_{l,t})\right)\right)^2\right) \\
&\leq \left(\sum_{l\neq i}\mathrm{E}\left(\left(\gamma_{i,l}\tfrac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(y_{j,t}y_{l,t} - \mathrm{E}(y_{j,t}y_{l,t})\right)\right)^2\right)^{1/2}\right)^2 \\
&\leq \left(\sum_{l\neq i}\left(\gamma_{i,l}^2\sup_{l,j}\mathrm{E}\left(\left(\tfrac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(y_{j,t}y_{l,t} - \mathrm{E}(y_{j,t}y_{l,t})\right)\right)^2\right)\right)^{1/2}\right)^2 \\
&\leq \left(\sum_{l\neq i}(\gamma_{i,l}^2 c_2)^{1/2}\right)^2 \leq c_2\left(\sum_{l\neq i}|\gamma_{i,l}|\right)^2 < \infty
\end{aligned}$$

where the first inequality follows from Minkowski's inequality and the third from part one. Finally, notice that $\sum_{l\neq i}|\gamma_{i,l}| < \infty$ as $\|\boldsymbol{K}_i\|_1 \leq \infty$ (THIS I NEED TO ADD TO ASSUMPTIONS).

Next, we consider $\left|\boldsymbol{K}_{ii} - \hat{\boldsymbol{K}}_{ii}\right| \leq |\boldsymbol{K}_{ii}||\boldsymbol{K}_{ii}^{-1} - \hat{\boldsymbol{K}}_{ii}^{-1}||\hat{\boldsymbol{K}}_{ii}|$. We have that $\boldsymbol{K}_{ii}^{-1} = u_{i,\cdot}'u_{i,\cdot}/T$ and $\hat{\boldsymbol{K}}_{ii}^{-1} = \hat{u}_{i,\cdot}'\hat{u}_{i,\cdot}/T$, where $\hat{u}_{i,\cdot} = y_{i,\cdot} - \boldsymbol{Y}_{-i}\hat{\gamma}_i$.

and from the inverse for block matrices

$$\begin{aligned}
\hat{\boldsymbol{K}}_{ii}^{-1} &= \tfrac{1}{T} y_{i,.}' y_{i,.} - \tfrac{1}{T} y_{i,.}' \boldsymbol{Y}_{-i} \left(\tfrac{1}{T}\boldsymbol{Y}_{-i}'\boldsymbol{Y}_{-i}\right)^{-1} \tfrac{1}{T}\boldsymbol{Y}_{-i}' y_{i,.} \\
&= \tfrac{1}{T} y_{i,.}' \boldsymbol{M}_{Y_{-i}} y_{i,.} \\
&= \tfrac{1}{T} u_{i,.}' \boldsymbol{M}_{Y_{-i}} u_{i,.} \\
&= \tfrac{1}{T} \hat{u}_{i,.}' \hat{u}_{i,.},
\end{aligned}$$

where $\hat{u}_{i,.} = y_{i,.} - \boldsymbol{Y}_{-i}\hat{\gamma}_i$. We have that

$$\begin{aligned}
\hat{u}_{i,.} &= y_{i,.} - \boldsymbol{Y}_{-i}\hat{\gamma}_i \\
&= u_{i,.} - \boldsymbol{Y}_{-i}(\gamma_i - \hat{\gamma}_i)
\end{aligned}$$

and

$$\frac{1}{T}\hat{u}_{i,.}'\hat{u}_{i,.} = \frac{1}{T}u_{i,.}'u_{i,.} - 2\frac{1}{T}u_{i,.}'\boldsymbol{Y}_{-i}(\hat{\gamma}_i - \gamma_i) + (\hat{\gamma}_i - \gamma_i)'\frac{1}{T}\boldsymbol{Y}_{-i}'\boldsymbol{Y}_{-i}(\hat{\gamma}_i - \gamma_i)$$

Hence, if the second and third term converge to zero we have that $\frac{1}{T}\hat{u}_{i,.}'\hat{u}_{i,.} \xrightarrow{p} \frac{1}{T}u_{i,.}'u_{i,.}$ for $n, T \to \infty$. We have using the results derived above

$$\begin{aligned}
\tfrac{1}{T}u_{i,.}'\boldsymbol{Y}_{-i}(\hat{\gamma}_i - \gamma_i) &\leq \|\tfrac{1}{T}u_{i,.}'\boldsymbol{Y}_{-i}\|\|\hat{\gamma}_i - \gamma_i\| \\
&= \mathcal{O}_p\left(\tfrac{n}{T}\right)
\end{aligned}$$

The final term

$$\begin{aligned}
(\hat{\gamma}_i - \gamma_i)'\tfrac{1}{T}\boldsymbol{Y}_{-i}'\boldsymbol{Y}_{-i}(\hat{\gamma}_i - \gamma_i) &\leq \|(\hat{\gamma}_i - \gamma_i)\|\|\tfrac{1}{T}\boldsymbol{Y}_{-i}'\boldsymbol{Y}_{-i}\|\|(\hat{\gamma}_i - \gamma_i)\| \\
&= \mathcal{O}_p\left(\tfrac{n}{T}\right)
\end{aligned}$$

where $\mu_1(\frac{1}{T}\boldsymbol{Y}_{-i}'\boldsymbol{Y}_{-i}) \leq \mu_1(\frac{1}{T}\boldsymbol{Y}'\boldsymbol{Y}) = \mathcal{O}_p(1)$, see Lemma 6 (i). Combining we have that

$$\begin{aligned}
\left|\boldsymbol{K}_{ii} - \hat{\boldsymbol{K}}_{ii}\right| &\leq \tfrac{1}{\hat{\boldsymbol{K}}_{ii}}\tfrac{1}{\boldsymbol{K}_{ii}}\left|\tfrac{1}{\boldsymbol{K}_{ii}} - \tfrac{1}{\hat{\boldsymbol{K}}_{ii}}\right| \\
&= \mathcal{O}_p\left(\tfrac{n}{T}\right)
\end{aligned}$$

Summarizing the proof

$$\begin{aligned}
\|\hat{\boldsymbol{K}}_i\| - \|\boldsymbol{K}_i\| &= \left|\boldsymbol{K}_{ii} - \hat{\boldsymbol{K}}_{ii}\right| \|\boldsymbol{\Gamma}_i\| + \hat{\boldsymbol{K}}_{ii}\|\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i\| \\
&= \mathcal{O}_p\left(\tfrac{n}{T}\right)
\end{aligned}$$

□

Proof of Corollary 1. The proof consist of showing that the probability of inconsistent selection of the granular series converges to zero as n and T grow large at the appropriate rates.

Note that

$$\begin{aligned}
\Pr(\mathcal{E}^c) &= \Pr\left(\min_{i=1,\ldots,k} \|\hat{\boldsymbol{K}}_i\| < \max_{j=k+1,\ldots,n} \|\hat{\boldsymbol{K}}_j\|\right) \\
&= \Pr\left(\min_{i=1,\ldots,k} \|\hat{\boldsymbol{K}}_i\| - \max_{j=k+1,\ldots,n} \|\hat{\boldsymbol{K}}_j\| < 0\right) \\
&\leq (n-k)\Pr\left(\min_{i=1,\ldots,k} \|\hat{\boldsymbol{K}}_i\| - \|\hat{\boldsymbol{K}}_j\| < 0\right) \\
&\leq (n-k)\,k\Pr\left(\|\hat{\boldsymbol{K}}_i\| - \|\hat{\boldsymbol{K}}_j\| < 0\right) \\
&= (n-k)\,k\Pr\left(\|\boldsymbol{K}_i\| - \|\boldsymbol{K}_j\| + (\|\hat{\boldsymbol{K}}_i\| - \|\boldsymbol{K}_i\|) - (\|\hat{\boldsymbol{K}}_j\| - \|\boldsymbol{K}_j\|) < 0\right) \\
&\leq (n-k)\,k\Pr\left(\|\boldsymbol{K}_i\| - \|\boldsymbol{K}_j\| - \left|\|\hat{\boldsymbol{K}}_i\| - \|\boldsymbol{K}_i\|\right| - \left|\|\hat{\boldsymbol{K}}_j\| - \|\boldsymbol{K}_j\|\right| < 0\right) \\
&= (n-k)\,k\Pr\left(\left|\|\hat{\boldsymbol{K}}_i\| - \|\boldsymbol{K}_i\|\right| + \left|\|\hat{\boldsymbol{K}}_j\| - \|\boldsymbol{K}_j\|\right| > \|\boldsymbol{K}_i\| - \|\boldsymbol{K}_j\|\right) \\
&\leq (n-k)\,k\,\frac{\mathrm{E}\left(\left|\|\hat{\boldsymbol{K}}_i\| - \|\boldsymbol{K}_i\|\right| + \left|\|\hat{\boldsymbol{K}}_j\| - \|\boldsymbol{K}_j\|\right|\right)}{\|\boldsymbol{K}_i\| - \|\boldsymbol{K}_j\|} \\
&= \mathcal{O}\left(\frac{n^2}{T}\right) .
\end{aligned}$$

Note that the last step of the proof follows from Theorem 1 and Lemma 1. □

Proof of Corollary 2. The proof consist of showing that the probability of inconsistent selection of k converges to zero as n and T grow large at the appropriate rates. Note that

$$\begin{aligned}
\Pr(\hat{k} \neq k) &= \Pr\left(\frac{\|\hat{\boldsymbol{K}}_k\|}{\|\hat{\boldsymbol{K}}_{k+1}\|} < \max_{j=k+1,\ldots,n-1} \frac{\|\hat{\boldsymbol{K}}_j\|}{\|\hat{\boldsymbol{K}}_{j+1}\|}\right) \\
&\leq (n-k)\Pr\left(\frac{\|\hat{\boldsymbol{K}}_k\|}{\|\hat{\boldsymbol{K}}_{k+1}\|} - \frac{\|\hat{\boldsymbol{K}}_j\|}{\|\hat{\boldsymbol{K}}_{j+1}\|} < 0\right) \\
&= (n-k)\Pr\left(\frac{\|\boldsymbol{K}_k\|}{\|\boldsymbol{K}_{k+1}\|} - \frac{\|\boldsymbol{K}_j\|}{\|\boldsymbol{K}_{j+1}\|} + \left(\frac{\|\hat{\boldsymbol{K}}_k\|}{\|\hat{\boldsymbol{K}}_{k+1}\|} - \frac{\|\boldsymbol{K}_k\|}{\|\boldsymbol{K}_{k+1}\|}\right) - \left(\frac{\|\hat{\boldsymbol{K}}_j\|}{\|\hat{\boldsymbol{K}}_{j+1}\|} - \frac{\|\boldsymbol{K}_j\|}{\|\boldsymbol{K}_{j+1}\|}\right) < 0\right) .
\end{aligned}$$

Note that as n and T grow large such that $n/T \to 0$ we have that

$$\begin{aligned}
\frac{\|\hat{\boldsymbol{K}}_k\|}{\|\hat{\boldsymbol{K}}_{k+1}\|} - \frac{\|\boldsymbol{K}_k\|}{\|\boldsymbol{K}_{k+1}\|} &= \frac{\|\boldsymbol{K}_k\|}{\|\hat{\boldsymbol{K}}_{k+1}\|} + \frac{\|\hat{\boldsymbol{K}}_k\| - \|\boldsymbol{K}_k\|}{\|\hat{\boldsymbol{K}}_{k+1}\|} - \frac{\|\boldsymbol{K}_k\|}{\|\boldsymbol{K}_{k+1}\|} \\
&= \frac{\|\boldsymbol{K}_k\|}{\|\boldsymbol{K}_{k+1}\|} + \frac{\|\boldsymbol{K}_k\|}{\|\boldsymbol{K}_{k+1}\|^2} O_p(n/T) + \frac{1}{\|\hat{\boldsymbol{K}}_{k+1}\|} O_p(n/T) - \frac{\|\boldsymbol{K}_k\|}{\|\boldsymbol{K}_{k+1}\|} \\
&= O_p(n/T) .
\end{aligned}$$

This allows us to write

$$\begin{aligned}
\Pr(\hat{k} \neq k) &\leq n \Pr\left(\frac{\|\boldsymbol{K}_k\|}{\|\boldsymbol{K}_{k+1}\|} - \frac{\|\boldsymbol{K}_j\|}{\|\boldsymbol{K}_{j+1}\|} + O_p(n/T) < 0\right) \\
&= n \Pr\left(O_p(n/T) > \frac{\|\boldsymbol{K}_k\|}{\|\boldsymbol{K}_{k+1}\|} - \frac{\|\boldsymbol{K}_j\|}{\|\boldsymbol{K}_{j+1}\|}\right) \\
&\leq O\left(\frac{n^2}{T}\right)
\end{aligned}$$

where the last step follows from the fact that $\frac{\|\boldsymbol{K}_k\|}{\|\boldsymbol{K}_{k+1}\|} - \frac{\|\boldsymbol{K}_j\|}{\|\boldsymbol{K}_{j+1}\|}$ is positive by Lemma 2 and an application of Markov's inequality. □

Proof of Theorem 2. □

A.1 Alternative regression representation for the granular panel data model

To facilitate the proof of 1 we discuss a regression representation for the precision matrix $\boldsymbol{K}$. A more detailed discussion for general precision matrices can be found in Pourahmadi (2013, Section 5.2). Let $\boldsymbol{K}$ be defined as in (4) and let

$$\boldsymbol{D} = \text{diag}(\boldsymbol{K}_{11}, \ldots, \boldsymbol{K}_{nn}).$$

We define the linear transformation $u_t = \boldsymbol{D}^{-1}\boldsymbol{K}y_t$, where y_t is given by (2). After reordering, each row can be written as

$$y_{i,t} = \sum_{j \neq i} y_{j,t}\gamma_{i,j} + u_{i,t},$$

where $\gamma_{i,j} = -\boldsymbol{K}_{ij}/\boldsymbol{K}_{ii}$ and it follows that $\text{E}(u_{i,t}) = 0$, $\text{Var}(u_{i,t}) = \frac{1}{\boldsymbol{K}_{ii}}$ and $\text{Cov}(u_{i,t}, u_{j,t}) = \frac{\boldsymbol{K}_{ij}}{\boldsymbol{K}_{ii}\boldsymbol{K}_{jj}}$. Also, $u_{i,t}$ is uncorrelated with $y_{j,t}$ for all $j \neq i$. The vector representation for series i is given by

$$y_{i,.} = \boldsymbol{Y}_{-i}\gamma_i + u_{i,.} \tag{5}$$

where $y_{i,.} = (y_{i,1}, \ldots, y_{i,T})'$, $\boldsymbol{Y}_{-i} = (y_{1,.}, \ldots, y_{i-1,.}, y_{i+1,.}, \ldots, y_{n,.})$ and $\gamma_i = (\gamma_{i,1}, \ldots, \gamma_{i,i-1}, \gamma_{i,i+1}, \ldots, \gamma_{i,n})'$. For convenience we also define $\boldsymbol{\Gamma}_i = (\gamma_{i,1}, \ldots, \gamma_{i,i-1}, 1, \gamma_{i,i+1}, \ldots, \gamma_{i,n})'$ such that $\boldsymbol{K}_i = -\boldsymbol{K}_{ii}\boldsymbol{\Gamma}_i$.

A.2 Additional Lemmas

Lemma 5. *Let y_t be generated by model* (2) *under assumption 2 (i)-(iii) then*

$$\text{E}(|y_{i,t}|^4) < \infty \tag{6}$$

Proof: To show that $\text{E}(|y_{i,t}|^4) < \infty$ we first consider the case where i is a granular series. We have

$$\text{E}(|y_{i,t}|^4) = \text{E}(|g_{i,t}|^4) < \infty,$$

by assumption 2 (i). Next, for ith non-granular, let $\beta_{i,.}$ denote the ith row of $\boldsymbol{\beta}$ such that

$$\begin{aligned}
\mathrm{E}(|y_{i,t}|^4) &= \mathrm{E}(|\beta_{i,.}g_t + \epsilon_{i,t}|^4) \\
&\leq 8\,[\mathrm{E}(|\beta_{i,.}g_t|^4) + \mathrm{E}(|\epsilon_{i,t}|^4)] \\
&\leq 2^{3(k+1)} \textstyle\sum_{j=1}^{k} \mathrm{E}(|\beta_{i,j}g_{j,t}|^4) + 8\mathrm{E}(|\epsilon_{i,t}|^4) \\
&= 2^{3(k+1)} \textstyle\sum_{j=1}^{k} \beta_{i,j}^4 \mathrm{E}(|g_{j,t}|^4) + 8\mathrm{E}(|\epsilon_{i,t}|^4) < \infty
\end{aligned}$$

since k is fixed the result follows from 2 (i), (ii) and 1 (iii).

Lemma 6. *Let y_t be generated by model* (2) *under assumption 2 (i)-(ii) then*

(i) $\mu_1(\hat{\boldsymbol{\Sigma}}_y) = \mathcal{O}_p(1)$

(ii) $\mu_n(\hat{\boldsymbol{\Sigma}}_y) = \kappa + o_p(1)$

for some $\kappa > 0$ where $\hat{\boldsymbol{\Sigma}}_y = \frac{1}{T}\sum_{t=1}^{T} y_t y_t'$.

Proof. Proof of Lemma 6 For simplicity define

$$y_t = \boldsymbol{B}g_t + \tilde{\epsilon}_t$$

where $\boldsymbol{B} = (I_k, \boldsymbol{\beta}')'$ and $\tilde{\epsilon}_t = (0, \epsilon_t')'$. For part (i) we have

$$\begin{aligned}
\|\hat{\boldsymbol{\Sigma}}_y\| &= \|\tfrac{1}{T}\textstyle\sum_{t=1}^{T} y_t y_t'\| \\
&\leq \|\boldsymbol{B}\tfrac{1}{T}\textstyle\sum_{t=1}^{T} g_t g_t' \boldsymbol{B}'\| + 2\|\boldsymbol{B}\tfrac{1}{T}\textstyle\sum_{t=1}^{T} g_t \tilde{\epsilon}_t'\| + \|\tfrac{1}{T}\textstyle\sum_{t=1}^{T} \tilde{\epsilon}_t\tilde{\epsilon}_t'\| \\
&\leq \|\boldsymbol{B}\|^2\|\tfrac{1}{T}\textstyle\sum_{t=1}^{T} g_t g_t'\| + 2\|\boldsymbol{B}\|\|\tfrac{1}{T}\textstyle\sum_{t=1}^{T} g_t \tilde{\epsilon}_t'\| + \|\tfrac{1}{T}\textstyle\sum_{t=1}^{T} \tilde{\epsilon}_t\tilde{\epsilon}_t'\|
\end{aligned}$$

Now, $\lim_{n\to\infty}\|\boldsymbol{B}\|^2 = k + \lim_{n\to\infty}\|\boldsymbol{\beta}\|^2 = k + \mu_1(\boldsymbol{D}) < \infty$ by assumption 1 (iv). Next,

$$\|\frac{1}{T}\sum_{t=1}^{T} g_t g_t'\| \xrightarrow{p} \|\boldsymbol{\Sigma}_g\| < \infty$$

by assumption 2 (i) and hence, the first term is $\mathcal{O}_p(1)$. The last term gives

$$\begin{aligned}
\|\tfrac{1}{T}\textstyle\sum_{t=1}^{T} \tilde{\epsilon}_t\tilde{\epsilon}_t'\| &= \|\tfrac{1}{T}\textstyle\sum_{t=1}^{T} \epsilon_t\epsilon_t'\| \\
&= \|\boldsymbol{E}\boldsymbol{E}'/T\| = \mathcal{O}_p(1)
\end{aligned}$$

by assumption 2 (ii). For the middle term we have

$$\|\frac{1}{T}\sum_{t=1}^{T} g_t\tilde{\epsilon}_t'\| = \|\frac{1}{T}\sum_{t=1}^{T} g_t g_t'\|^{1/2}\|\frac{1}{T}\sum_{t=1}^{T} \epsilon_t\epsilon_t'\|^{1/2} = \mathcal{O}_p(1)$$

which follows from the preceding arguments.

For part (ii) we begin by noting that the model can be equivalently formulated as

$$y_t = \boldsymbol{M}\nu_t\,,$$

where

$$\boldsymbol{M} = \begin{bmatrix} \boldsymbol{I}_k & 0 \\ \boldsymbol{\beta} & \boldsymbol{I}_{n-k} \end{bmatrix} \text{ and } \nu_t = \begin{bmatrix} g_t \\ \epsilon_t \end{bmatrix}.$$

The sample covariance matrix of the data can then be expressed as

$$\hat{\boldsymbol{\Sigma}}_y = \frac{1}{T}\sum_{t=1}^{T} y_t y_t' = \boldsymbol{M}\,\frac{1}{T}\sum_{t=1}^{T} \nu_t \nu_t' \,\boldsymbol{M}' \ .$$

Since $\boldsymbol{M}$ is nonsingular and $\frac{1}{T}\sum_{t=1}^{T}\nu_t\nu_t'$ is symmetric, it follows from Ostrowski's theorem Horn and Johnson (2013, Theorem 4.5.9) that the minimum eigenvalue of $\hat{\boldsymbol{\Sigma}}_y$ satisfies

$$\mu_n(\hat{\boldsymbol{\Sigma}}_y) \ \ \geq \ \ \mu_n\left(\frac{1}{T}\sum_{t=1}^{T}\nu_t\nu_t'\right)\mu_n\left(\boldsymbol{M}\boldsymbol{M}'\right) \ .$$

It follows that $\hat{\boldsymbol{\Sigma}}_y$ is positive definite provided that $\frac{1}{T}\sum_{t=1}^{T}\nu_t\nu_t'$ and $\boldsymbol{M}\boldsymbol{M}'$ are positive definite. Consider first

$$\boldsymbol{M}\boldsymbol{M}' = \left[\begin{array}{cc} \boldsymbol{I}_k & \boldsymbol{\beta}' \\ \boldsymbol{\beta} & \boldsymbol{I}_{n-k} + \boldsymbol{\beta}\boldsymbol{\beta}' \end{array}\right] .$$

Since $\boldsymbol{I}_k$ and the Schur complement of $\boldsymbol{I}_k$ in $\boldsymbol{M}\boldsymbol{M}'$ which is $\boldsymbol{I}_{n-k}$ are positive definite it follows that $\boldsymbol{M}\boldsymbol{M}'$ is also positive definite for any n.

Next, it remains to check that $\frac{1}{T}\sum_{t=1}^{T}\nu_t\nu_t'$ is positive definite. We can do this by showing that $\hat{\boldsymbol{\Sigma}}_{gg}$ and the Schur complement of $\hat{\boldsymbol{\Sigma}}_{gg}$ in $\frac{1}{T}\sum_{t=1}^{T}\nu_t\nu_t'$ are positive definite. It follows from assumption 1(i) and 2(ii) that $\hat{\boldsymbol{\Sigma}}_{gg}$ is positive definite. Consider next the Schur complement of $\hat{\boldsymbol{\Sigma}}_{gg} = \frac{1}{T}\sum_{t=1}^{T} g_t g_t'$ in $\frac{1}{T}\sum_{t=1}^{T}\nu_t\nu_t'$. We have that

$$\begin{aligned}
\mu_n(\hat{\boldsymbol{\Sigma}}_{gg} - \hat{\boldsymbol{\Sigma}}_{\epsilon g}\hat{\boldsymbol{\Sigma}}_{gg}^{-1}\hat{\boldsymbol{\Sigma}}_{g\epsilon}) \ &\geq \ \mu_n(\hat{\boldsymbol{\Sigma}}_{gg}) - \mu_1(\hat{\boldsymbol{\Sigma}}_{\epsilon g}\hat{\boldsymbol{\Sigma}}_{gg}^{-1}\hat{\boldsymbol{\Sigma}}_{g\epsilon}) \\
&\geq \ \mu_n(\hat{\boldsymbol{\Sigma}}_{gg}) - \mu_1(\hat{\boldsymbol{\Sigma}}_{gg}^{-1})\mu_1(\hat{\boldsymbol{\Sigma}}_{\epsilon g}\hat{\boldsymbol{\Sigma}}_{g\epsilon}) \\
&= \ \mu_n(\hat{\boldsymbol{\Sigma}}_{gg}) - \mu_n(\hat{\boldsymbol{\Sigma}}_{gg})^{-1}\|\hat{\boldsymbol{\Sigma}}_{g\epsilon}\|^2 \\
&\geq \ \mu_n(\hat{\boldsymbol{\Sigma}}_{gg}) - \mu_n(\hat{\boldsymbol{\Sigma}}_{gg})^{-1}k^2\|\hat{\boldsymbol{\Sigma}}_{g\epsilon}\|_F^2 \\
&= \ \mu_n(\hat{\boldsymbol{\Sigma}}_{gg}) - \mu_n(\hat{\boldsymbol{\Sigma}}_{gg})^{-1}k^2 O\left(\frac{n}{T}\right) \ .
\end{aligned}$$

□