Detecting Granular Time Series in Large Panels

C. Brownlees and *G. Mesters*

Universitat Pompeu Fabra and Barcelona GSE

Preliminary and Incomplete

May 1, 2017

Abstract

Large economic and financial panels often contain time series that influence the entire cross-section. We name such series *granular*. In this paper we introduce a panel data model that allows to formalize the notion of granular time series. We then propose a methodology, which is inspired by the network literature in statistics and econometrics, to detect the set of granulars when such set is unknown. The influence of the i-th series in the panel is measured by the norm of the i-th column of the inverse covariance matrix. We show that a detection procedure based on the column norms allows to consistently select granular series when the cross-section and time series dimensions are sufficiently large. Moreover, the asymptotic distribution of the column norms is derived in order to construct confidence intervals and carry out hypothesis testing. Importantly, we show that the methodology allows to consistently detect granulars also when the series in the panel are influenced by common factors. A simulation study shows that the proposed procedures perform satisfactorily in finite samples. We illustrate the methodology with applications in macroeconomics and finance.

JEL classification: C32, C43.

Keywords: granularity, network models, factor models, large-dimensional models.

1 Introduction

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

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

We begin by introducing a model for a panel of time series that formalizes the notion of granularity used in this paper. We assume that the panel is partitioned into a (finite) set of series labeled as granular and a remaining set of non-granular series. The granular series coincide with their respective idiosyncratic shocks, which we call granular shocks. Each non-granular series is modeled as a linear combination of the granular shocks and an idiosyncratic non-granular shock. We work under the assumption that the researcher does not observe whether a given series belongs to the set of granulars.

Our granular detection methodology is based on the properties of the inverse covariance

[1]Our work is complementary to the large macroeconomic literature that uses input-output tables, or other criteria such as firm size, to determine whether a certain series is granular, see among others Gabaix (2011), Foerster, Sarte and Watson (2011), Acemoglu et al. (2012), Di Giovanni and Levchenko (2012), Carvalho and Gabaix (2013), Di Giovanni, Levchenko and Mejean (2014), Bernard, Jensen, Redding and Schott (2016) and Gaubert and Itskhoki (2016). Instead, we develop methodology that classifies a series as granular on the basis of the properties of the covariance matrix of the panel.

matrix of the panel, hereafter concentration matrix, and is motivated by the literature on graphical and network models in statistics (see, among others, Lauritzen (1996), Meinshausen and Bühlmann (2006), Peng, Wang, Zhou and Zhu (2009), Pourahmadi (2013, Chapter 5)). As it is well known in this literature, the concentration matrix embeds the partial dependence structure of the panel: Series i and j are partially uncorrelated given the remaining ones if and only if the (i, j) element of the concentration matrix is zero. This motivates us to use as a natural measure of the influence of series i in the panel the norm of the i-th column (or row) of the concentration matrix: Series that influence the entire cross-section ought to have a larger column norm (or row) than those series that influence a small part of the panel only.

We develop a granular detection methodology based on the column norms of the concentration matrix. We show that under appropriate identification assumptions the column norms of the population concentration matrix allow to detect the granular series in the panel. More precisely, we establish that the column norms corresponding to the granular series are larger than the non-granular ones. This implies that ranking series in the panel according to the value of the column norm ranks the granular series higher than the non-granular ones. Next we show that the ratio among subsequent ordered column norms is maximized when comparing the column norms of the last granular with the first non-granular series. This implies that we can identify the number of granular series as the index that maximizes the sequential column norm ratio. This criterion is analogous to the eigenvalue ratio criterion proposed by Ahn and Horenstein (2013) for the selection of the number of factors.

In large panels of time series common factors typically explain a large proportion of total variability. See for instance the research by Foerster et al. (2011), Long and Plosser (1987) and Forni and Reichlin (1998), for earlier work on the trade-off between idiosyncratic and aggregate shocks in macroeconomics. To this extent, we consider an extension of the model in which the series in the panel are additionally influenced by a set of common factors. We show that, under appropriate assumptions, the column norms of the concentration matrix maintain their detection properties in this setting. Importantly, the results imply that in order to detect granulars the researcher does not need to know the number of common factors.

We operationalize our identification results by estimating the column norms of the concentration matrix using the sample covariance matrix of the panel. We show that the column norms of the inverse sample covariance matrix consistently estimate their population analog when the cross-sectional dimension and the number of time series observations are large. This result allows us to establish that our column norm estimator leads to consistent ranking and selection of the granular series. Our estimation results rely on distributional and dependence assumptions made in the literature on large dimensional covariance estimation, see for example Fan, Liao and Mincheva (2011).

We derive the asymptotic distribution of the column norms. For this we consider methods used in the literature on many weak instruments, see for example Hausman, Newey, Woutersen, Chao and Swanson (2012). This allows us to construct confidence intervals for the column norms of the concentration matrix. Moreover, we propose a test for the null-hypothesis of equality between two column norms that can be used to assess whether two series exert the same influence on the panel. See Pesaran and Yamagata (2012) and Fan, Liao and Yao (2015) for other recent examples of related high dimensional tests.

Finally, we address the estimation of the granular panel model. In order estimate the model with common factors, knowledge of the number of factors is required. To this extent we also show that when the number of factors is unknown appropriate modifications of existing methods for factor selection introduced in the literature (e.g. Onatski (2010)) can be applied. We then cast the model in state space form and estimate the model parameters using standard Kalman filter techniques, see Durbin and Koopman (2012). The estimated granular model can be used, for instance, to study the economic importance of the granular shocks using impulse response analysis and variance decompositions.

A simulation study is carried out to assess the performance of our methodology in finite samples. In the study we simulate a granular model with common factors and then use our granular detection methodology to recover the granular series. Results show that the granular detection methodology procedures performs satisfactorily in finite samples when the strength of the granulars is sufficiently large. Further, we show that our methodology compares favorably to detection methods that are based on principal components analysis, see Stock and Watson (2002a), Bai and Ng (2006) and Parker and Sul (2016).

We apply our methodology in two empirical studies. First, we consider detecting granular series in a large panel of industrial production series that was previously considered in Foerster et al. (2011). 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. We introduce the empirical studies in more detail in the next section.

The remainder of this paper is organized as follows. Section 2 presents the granular panel model analyzed in this work and formalizes the granular detection problem. This section also discusses applications of this specification in macroeconomics and finance. Section 3 introduces our granular detection methodology and it establishes its large sample properties. Section 4 compares our methodology to alternative methods based on principal components analysis and maximum likelihood. Section 5 discusses how to carry out inference on the granular panel model. Section 6 carries out a simulation study to assess the finite sample performance of the proposed methodology. Section 7 presents the results of two empirical applications of our methodology. Concluding remarks follow in Section 8.

2 The granular detection problem

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

We assume that there are k (fixed) time series whose idiosyncratic shocks g_t influence the entire panel[2]. We label these time series as granular and the shocks g_t as granular shocks.

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

For simplicity and without loss of generality we assume that the granular series are the first k series in the panel. The other $n-k$ time series are the non-granular series whose idiosyncratic shocks are denoted by ϵ_t. All series in the panel are influenced by a set of r common shocks, or factors, f_t. The granular panel data model with common factors is defined as

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

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

In this paper we work under the assumption that the data is generated according to model (1) and that the researcher does not know (i) which series are granular and (ii) the number of granular series k.[3] Our objective is to introduce a methodology that allows to consistently recover this information from the data. We point out here that in our framework the researcher does not need to know the number of common factors r in order to detect granular series. However, in order to carry out inference on the model knowledge of r is required. Existing methods for factor selection introduced in the literature (e.g. Onatski (2010)) may be applied for this task.

It is important to clarify that while model (1) has a factor model representation the methodology that we introduce in this paper is different from the standard techniques that are adopted in the factor model literature, like maximum likelihood and principal components. Specifically, our detection strategy is based on the partial correlation properties of the panel and in Section 4 we compare our methodology to the more standard methods.

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

[3]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, see Bai and Li (2012, Section 4 and Table 1) for a detailed overview.

2.1 Granular sectors in industrial production

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

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

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

2.2 Granular institutions in the financial system

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

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

We can cast the problem of detecting systemic institutions as yet another instance of a granular detection problem. Following, Diebold and Yılmaz (2014) we may consider a panel of volatility measure for a set of large US financial institutions. We assume the panel is generated by model (1) and the methodology introduced in this work may be used detect granular/systemic institutions while controlling for market risk, and other economy wide sources of risk, through the common factors.

3 Methodology

3.1 Granular panel model

Before considering the complete granular panel specification of (1) we first outline our methodology for a simplified version of the model where there are no common factors. More precisely, we consider the model

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

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

We propose a granular detection strategy that is based on the properties of the concentration matrix of the panel. It is straightforward to check that the covariance matrix $\mathbf{\Sigma} = \text{Var}(y_t)$ and the concentration matrix $\mathbf{K} = \mathbf{\Sigma}^{-1}$ of the panel are given by

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

Assume, for simplicity, that the norms of the columns of the $\boldsymbol{\beta}$ matrix are larger than one and that $\mathbf{\Sigma}_\epsilon$ is the identity matrix. Then, it is straightforward to verify that the norms of the first k columns (or rows) of the concentration matrix are larger than the norms of the last $(n-k)$ columns (or rows). Thus, the set of granular series can be identified simply by checking which series are associated with the largest column (or row) norms of the concentration matrix.

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

$$\|\mathbf{K}_i\| \quad \text{for} \quad i = 1, \ldots, n, \tag{3}$$

where $\mathbf{K}_i$ denotes the i-th column of $\mathbf{K}$.[4][5]

Our detection strategy has a natural interpretation in terms of a partial correlation network model, e.g. Pourahmadi (2013, Chapter 5). The partial correlation network representation of the panel consists of a graph defined over n vertices where each series corresponds to a vertex and vertices i and j are connected by an edge if i and j are correlated given the remaining series in the panel. The concentration matrix embeds the partial dependence structure of the panel: Series i and j are partially uncorrelated if the (i, j) element of the concentration matrix $\mathbf{K}$ is zero.[6] Thus, heuristically, granular time series can be though of as

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

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

[6]More precisely, we have that the partial correlation between series i and j ρ^{ij} is related to the concen-

hubs in a partial correlation network representation of the panel and the granular detection parameter $\|\mathbf{K}_i\|$ can be though as a parameter proportional to the number of connections, or degree, of each vertex. The top panel of Figure 1 shows the partial correlation network representation of the panel when $n = 6$, $g = 1$ and $\mathbf{\Sigma}_\epsilon$ is a diagonal matrix.

We impose a number of assumptions on the components of model (2) to establish the identification results.[7]

Assumption 1. *We assume that*

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

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

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

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

$$\|\boldsymbol{\beta}_i\| > \sqrt{1 + \kappa_\beta}\, \kappa_\epsilon$$

where κ_β *and* κ_ϵ *are, respectively, the condition number of the matrices* $\mathbf{D}_\beta$ *and* $\mathbf{\Sigma}_\epsilon$.

Assumption (i) ensures that none of the granular series are linear combinations of each other. Assumption (ii) implies that $\mathbf{\Sigma}_\epsilon$ is invertible. Assumption (iii) is standard for regression models, see for example White (2000, Chapter 2). Assumption (iv) characterizes the granular model. First, we require $\boldsymbol{\beta}'\boldsymbol{\beta}$ to be non-vanishing when n increases. Second, we require the norm of the columns of the granular loading matrix to be larger than a threshold that depends on the degree of collinearity among the non-granular shocks and the granular

tration matrix $\mathbf{K}$ through the relation

$$\rho^{ij} = -\frac{k_{ij}}{\sqrt{k_{ii}k_{jj}}} \, .$$

[7] The following notation is adopted. The k-th largest eigenvalue of an $N \times N$ matrix $\boldsymbol{B}$ is denoted as $\mu_k(\boldsymbol{B})$, $\boldsymbol{B} > 0$ indicates that $\boldsymbol{B}$ is positive definite and $\boldsymbol{B} \geq 0$ indicates that $\boldsymbol{B}$ is positive semi-definite. As a matrix norm we generally adopt the spectral norm that is given for a general $M \times N$ matrix $\boldsymbol{A}$ by $\|\boldsymbol{A}\| = \sqrt{\mu_1(\boldsymbol{A}'\boldsymbol{A})}$.

loadings. The threshold is such that, the larger the degree of collinearity the larger the column norms of the loading matrix. Intuitively, the granular loading matrix determines the strength of the granular series in model and assumption (*iv*) (and its variants to appear below) ensures that the granular series are strong enough to be detected. The assumption implies that the higher the collinearity among the non-granular shocks and the granular loadings, the larger the column norms of the granular loading matrix have to be for successful granular detection.

Assumption 1 is sufficient to rank the granular series higher than the non-granular ones when ordering series on the basis of the column norms of the concentration matrix of y_t. The following lemma establishes the population ranking result formally.

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

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

All proofs are collected in the appendix.

In order to select the number of granular time series in the panel we use a strategy inspired by eigenvalue ratio criterion proposed by Ahn and Horenstein (2013) for the selection of the number of factors. Let $\mathbf{K}_{(s)}$ denote the s–th largest column of the concentration matrix.[8] Consider the ratio between two subsequent ordered column norms, that is

$$\|\mathbf{K}_{(s)}\| \;/\; \|\mathbf{K}_{(s+1)}\| \;, \tag{4}$$

for $s = 1, ..., n-1$. Heuristically, the column norms are large for granular series and small otherwise. Thus, the ratio ought to be largest when comparing the last column norm corresponding to the granular series with the first column norm corresponding to the non-granular series. This suggests that the sequential column norm ratio ought to be maximized when s is equal to k. In order to identify the number of granulars using the sequential column norm

[8] Columns are ordered on the basis of their norm.

ratio we need strengthen assumption 1 (*iv*).

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

$$\begin{aligned} \|\boldsymbol{\beta}_i\| &> 2(1+\kappa_\beta)^{3/2}\kappa_\epsilon^2 \,, && (5) \\ \|\boldsymbol{\beta}_i\| &> 2\sqrt{1+\kappa_\beta}\kappa_\epsilon^2 \frac{\mu_{n-k}(\boldsymbol{\Sigma}_\epsilon)}{\mu_k(\boldsymbol{\Sigma}_g)} \,. && (6) \end{aligned}$$

We emphasize that in practical situations condition (5) is the one likely to be more binding. Given the stronger condition on the loading matrix we obtain the following lemma.

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*

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

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

Clearly, (4) is not the only function of the concentration matrix $\mathbf{K}$ that identifies k. In fact, several other functions of the elements of the concentration matrix can be used to identify the number of granular series. For instance, one could consider appropriate variants of selection criteria introduced in the factor model literature, see, among others, the criteria in Onatski (2009) and Cavicchioli, Forni, Lippi and Zaffaroni (2016).

We briefly compare our assumptions and identification results to the factor model literature. Two main differences can be noted in our setup. First, assumptions (*iv*)/(*iv**) reflect that the granular loadings are not orthogonal to each other. Second, we do not rely on assuming that the norm of the loadings matrix is proportional to n, see, among others, Bai and Ng (2002), Bai (2003), Ahn and Horenstein (2013) and Cavicchioli et al. (2016). Instead, we impose lower bounds on the norm of the columns of the granular loading matrix that are sufficient to carry out granular detection. Assuming that the norm of the granular loading matrix does not grow with n is dictated by empirical realism.[9] As we document in the empirical application, while there is indeed evidence of series that have an important influence over the entire panel these series typically do not explain an overall variance fraction that justifies that their importance is proportional to n. Last, we point out that assumptions $(iv)/(iv^*)$ are comparable to weak factor assumptions considered in Onatski (2009), Onatski (2010) and Onatski (2012). See also Pesaran (2012) and Chudik, Pesaran and Tosetti (2011) for more discussion on the distinction between weak and strong factors.

3.2 Granular panel model with common factors

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

$$\begin{aligned} y_{1:k,t} &= \boldsymbol{\Lambda}_1 f_t + g_t \\ y_{k+1:n,t} &= \boldsymbol{\Lambda}_2 f_t + \boldsymbol{\beta} g_t + \epsilon_t, \end{aligned} \tag{7}$$

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. All other components are the same as in the previous section. To identify the granular series in this setting we make the following additional assumptions.

Assumption 2. *We assume that*

(i) $\mathrm{E}(f_t) = 0$ *and* $\mathrm{E}(f_t f_t') = \boldsymbol{I}_r$.

(ii) $\mathrm{E}(f_t g_t') = \boldsymbol{\Sigma}_{fg}$ *for all* t *and* $\boldsymbol{\Sigma}_g - \boldsymbol{\Sigma}_{gf}\boldsymbol{\Sigma}_{fg} \succ 0$, *where* $\boldsymbol{\Sigma}_{gf} = \boldsymbol{\Sigma}_{fg}'$.

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

(iv) Let $\boldsymbol{\Lambda} = (\boldsymbol{\Lambda}_1', \boldsymbol{\Lambda}_2')'$ *and* $\boldsymbol{\Lambda}'\boldsymbol{\Lambda} \to \boldsymbol{D}_\lambda$ *as* $n \to \infty$, *with* $\mu_r(\boldsymbol{D}_\lambda) > 0$ *and* $\mu_1(\boldsymbol{D}_\lambda) < \infty$.

[9] If there would be strong granulars, explaining the common factors in factors models would not be so hard.

Assumption (i) is a standard normalization assumption for factor models, see for example Doz, Giannone and Reichlin (2012). Assumption (ii) allows for contemporaneous correlation between the factors and the granular shocks. The correlation is restricted by requiring the Schur complement $\mathbf{\Sigma}_g - \mathbf{\Sigma}_{gf}\mathbf{\Sigma}_{fg}$ to be positive definite. Assumptions $\mathrm{E}(f_t f_t') = \boldsymbol{I}_r$ and $\mathbf{\Sigma}_g - \mathbf{\Sigma}_{gf}\mathbf{\Sigma}_{fg} > 0$ together are the same as the requirement that the variance of f_t and g_t is positive definite, see Horn and Johnson (2013, pp. 25). Notice that this assumption is weaker when compared to the literature on network models with common factors which typically assumes no correlation between the common factors and the network. This allows, importantly, for feedback effects between common factors and granular series. Assumption (iii) imposes that the factors, similar as the granular shocks, are independent from the non-granular shocks. Assumption (iv) imposes the factors are non-vanishing when n increases. Overall, this set of additional assumptions for the factors corresponds to the weak factor model assumptions considered in Onatski (2009) and Onatski (2012).

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

Lemma 3. *Let y_t be generated by model* (7) *under assumptions 1 (i)–(iv) and 2 (i)–(iv). Then $\mathbf{K} = \mathbf{\Sigma}^{-1}$, where $\mathbf{\Sigma} = \mathrm{Var}(y_t)$, exists and we have for $n > N$ that*

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

Lemma 4. *Let y_t be generated by model* (7) *under assumptions 1 (i)–(iii) and (iv*) and 2 (i)–(iv). Then we have for for $n > N$ and $k > 0$ that*

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

It is important to emphasize that in our framework the presence of factors does not alter the detection properties of the column norms of the concentration matrix. It is also important to emphasize that in population in order to carry out granular detection it is not required to know the number of factors r.

3.3 Estimation

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

We need to impose appropriate dependence and distributional assumptions on y_t in order to establish the large sample properties of our estimator. Let $\mathcal{F}_{-\infty}^{0}$ and $\mathcal{F}_{s}^{\infty}$ denote the σ-algebras generated by $\{y_s : -\infty \leqslant s \leqslant 0\}$ and $\{y_s : t \leqslant s \leqslant \infty\}$, respectively. We define the α-mixing coefficients of the y_t process as

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

We make the following assumptions.

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

(i) $\{y_t\}$ is stationary and ergodic.

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

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

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

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

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

These are similar to the dependence and distributional assumptions made in Fan et al. (2011). In particular, assumption 3 (*ii*) states that y_t is strongly mixing and assumption 3 (*iii*) states that the marginal distributions of the components of y_t have generalized-exponential tails. The parameter γ defined in assumption 3 (*iv*) is a key quantity in this work and measures the degree of dependence and tail thickness of the data: The smaller the parameter the more dependent and thick tailed the data are.

These assumptions allow to apply a Bernstein-type inequality for mixing processes derived in Merlevede, Peligrad and Rio (2011), which is one of the main tools needed to establish the results of this section. Notice that we directly impose the assumptions on the observed series y_t instead of on f_t, g_t and ϵ_t separately. This is only for convenience and the results of this section may also be obtained by assuming that f_t, g_t and ϵ_t satisfy assumption 3. We establish the following result concerning the sample covariance matrix.

Theorem 1. *Let y_t be generated by model (7) under assumptions 2 and assumption 3. Suppose $n \to \infty$ and $T = O(n^{2/\gamma-1})$. Then, for any $\eta > 0$ there exists positive constants $C_1, \ldots, C_5$ such that*

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

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

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

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

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

Theorem 1 (*i*) establishes a Bai-Yin-type law (Bai and Yin (1993)) for the largest and smallest eigenvalues of the sample covariance matrix of strongly mixing data with generalized-exponential tails. The theorem states that the eigenvalues of the sample covariance matrix are bounded away from zero and infinity when n and T are large. Note that recent contributions in the factor modeling literature often circumvent this step and impose this as an additional assumption, see for example Ahn and Horenstein (2013) and Moon and Weidner (2015). The proof of the theorem follows the arguments laid out in Vershynin (2012), with appropriate modifications for the present setting. Let us emphasize that the theorem extends the results of Vershynin (2012) for a set of assumptions that are more convenient for economic and financial applications. Theorem 1 (i) facilitates the derivation of the subsequent parts (ii) to (iv).

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

$$\mathcal{E}_{\mathsf{R}} = \left\{ \|\hat{\mathbf{K}}_i\| > \|\hat{\mathbf{K}}_j\| \text{ for all } i = 1, ..., k \text{ and } j = k+1, ..., n \right\}, \tag{8}$$

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

Corollary 1. *Let y_t be generated by model* (2) *under assumptions 1 and 3. Consider the event $\mathcal{E}_{\mathsf{R}}$ defined in equation* (8). *Suppose that $n \to \infty$ and $T = O(n^{2/\gamma - 1})$. Then, for any $\eta > 0$*

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

In other words, the corollary shows that the granular statistic consistently ranks the granular series ahead of the non-granular ones. Second, in light of Lemma 2, it is natural to estimate the number of granular series by

$$\hat{k} = \arg \max_{s=1,...,n-1} \|\hat{\mathbf{K}}_{(s)}\| / \|\hat{\mathbf{K}}_{(s+1)}\| \ , \tag{9}$$

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

$$\mathcal{E}_{\mathsf{S}} = \left\{ \hat{k} = k \right\}, \tag{10}$$

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

Corollary 2. *Let y_t be generated by model* (2) *under assumptions 1 (i)-(iii) and (iv*) and 3. Consider the event $\mathcal{E}_{\mathsf{S}}$ defined in equation* (10)*. Suppose that $n \to \infty$ and $T = O(n^{2/\gamma-1})$. Then, for any $\eta > 0$*

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

A number of additional comments are in order. It may be possible to obtain better convergence rates for the granular statistic by employing some appropriate regularized covariance estimator. However, this typically requires making additional assumptions on the model. It is important to highlight that the results of this section can also be obtained by making assumptions similar to those in Stock and Watson (2002a), Bai and Ng (2002) and Doz et al. (2012). Such results rely on weaker distributional assumptions than the ones spelled out in assumption 3. However they rely on stronger dependence assumptions.

3.4 Asymptotic distribution

We derive the limiting distribution granular statistics $\|\hat{\mathbf{K}}_i\|$ in order to construct confidence intervals and to carry hypothesis testing. To this extent it is useful to derive the regression representation of the series in the panel. For each series $y_{i\,t}$ in the panel we have that

$$y_{i,t} = \boldsymbol{y}'_{-i,t}\gamma_i + u_{i,t} \ ,$$

where $\boldsymbol{y}_{-i,t} = (y_{1,t}, \ldots, y_{i-1,t}, y_{i+1,t}, \ldots, y_{n,t})'$ and $\mathrm{Var}(u_{i,t}) = \sigma^2_{u_i}$. This 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.4.[10]. Then, column norm can be written as

$$\|\mathbf{K}_i\| = \frac{\sqrt{1+\gamma_i'\gamma_i}}{\sigma^2_{u_i}} \ .$$

The sample column norm is given by

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

where $\hat{\gamma}_i = (\boldsymbol{Y}_{-i}'\boldsymbol{Y}_{-i})^{-1}\boldsymbol{Y}_{-i}'y_{i,.}$ with $y_{i,.} = (y_{i,1},\ldots,y_{i,T})'$, $\boldsymbol{Y}_{-i} = (y_{1,.},\ldots,y_{i-1,.},y_{i+1,.},\ldots,y_{n,.})$ and $\hat{u}_{i,.} = y_{i,.} - \boldsymbol{Y}_{-i}\hat{\gamma}_i$. We notice that $\|\hat{\mathbf{K}}_i\|$ is a function of $\hat{\gamma}_i'\hat{\gamma}_i$ and $\frac{1}{T}\hat{u}_{i,.}'\hat{u}_{i,.}$. We derive the joint limiting of these two quadratic forms and an application of the delta method subsequently gives the distribution of $\|\hat{\mathbf{K}}_i\|$. The joint distribution is derived using comparable techniques as developed in the literature on many weak instruments and many regressors, e.g. Hausman et al. (2012), Chao, Swanson, Hausman, Newey and Woutersen (2012), Chao, Hausman, Newey, Swanson and Woutersen (2014), Cattaneo, Jansson and Newey (2017a) and Cattaneo, Jansson and Newey (2017b). See also Kelejian and Prucha (2001) and Kelejian and Prucha (2010) for related approaches. To facilitate the derivation we make an additional technical assumption.

Assumption 4. *For each auxiliary regression representation,* $y_{i,.} = \boldsymbol{Y}_{-i}\gamma_i + u_{i,.}$ *the disturbances can be written as* $u_{i,.} = \boldsymbol{\Sigma}_{u_i}^{1/2}\zeta_{i,.}$, *with* $T \times T$ *lower triangular matrix* $\boldsymbol{\Sigma}_{u_i}^{1/2}$ *and* $\mathrm{Var}(u_{i,.}|\boldsymbol{Y}_{-i}) = \boldsymbol{\Sigma}_{u_i}^{1/2}\boldsymbol{\Sigma}_{u_i}^{1/2'} = \boldsymbol{\Sigma}_{u_i}$. *Further, we assume that there exist positive constants* $c_{u_i,u}$, $c_{u_i,l}$ *and* c_γ, *such that*

(i) $\limsup_{T\to\infty} \|\boldsymbol{\Sigma}_{u_i}^{1/2}\|_1 < c_{u_i,u}$ *and* $\limsup_{T\to\infty} \|\boldsymbol{\Sigma}_{u_i}^{-1/2}\|_1 < c_{u_i,l}$

(ii) $\{\zeta_{i,t}\}$ *is independent and identically distributed conditional on* $\boldsymbol{Y}_{-i}$.

(iii) $\limsup_{n\to\infty} \sum_{j\neq i} |\gamma_{i,j}| < c_\gamma$

[10]This also follows directly from the inverse formula for block matrices, see Magnus and Neudecker (2007, page 12).

Part (i) of this assumption restricts the correlation among the residuals. Part (ii) assumes the existence of an underlying independent process $\zeta_{i,t}$ and part (iii) ensures that the coefficients γ_i remain bounded. The assumptions are more relaxed when compared to other papers that deal with distributions in high dimensional settings, e.g. Pesaran and Yamagata (2012) and Fan et al. (2015), and the literature on many weak instruments, e.g. Chao and Swanson (2005) and Hausman et al. (2012). In these works the disturbances are typically considered independent, whereas we allow for correlation following the approach of Kelejian and Prucha (2010).

We obtain the following limiting distribution.

Theorem 2. *Let y_t be generated by model (2) under identification assumption 1 and sampling assumptions 3 and 4. We have when $n, T \to \infty$ with $n/T \to c \in (0,1)$ that*

$$\frac{\sqrt{T}(\|\hat{\mathbf{K}}_i\| - \delta_{K_i})}{\sigma_{K_i}} \xrightarrow{d} N(0,1), \quad \textit{where}$$

$$\delta_{K_i} = \sqrt{\frac{1}{\delta_{u_i}^2} + \frac{\delta_{\gamma_i}}{\delta_{u_i}^2}}, \qquad \sigma_{K_i}^2 = \left[\frac{1}{2\delta_{u_i}\sqrt{1+\delta_{\gamma_i}}}, -\frac{\sqrt{1+\delta_{\gamma_i}}}{\delta_{u_i}^2}\right] \mathbf{\Sigma}_{V_i} \left[\frac{1}{2\delta_{u_i}\sqrt{1+\delta_{\gamma_i}}}, -\frac{\sqrt{1+\delta_{\gamma_i}}}{\delta_{u_i}^2}\right]'$$

and $\delta_{\gamma_i} = \gamma_i'\gamma_i + \frac{1}{T}\sum_{t=1}^T \boldsymbol{A}_{i,1,tt}$, $\delta_{u_i} = \frac{1}{T}\sum_{t=1}^T \boldsymbol{A}_{i,2,tt}$ and $\mathbf{\Sigma}_{V_i}$ is 2×2 symmetric with elements

$$\begin{aligned}
\Sigma_{V_i,11} &= \tfrac{2}{T}\sum_{t=1}^T\sum_{s=1}^T \boldsymbol{A}_{i,1,ts}^2 + \tfrac{1}{T}\sum_{t=1}^T \boldsymbol{b}_{i,t}^2 + \tfrac{1}{T}\sum_{t=1}^T \boldsymbol{A}_{i,1,tt}^2(\xi_{i,t}^{(4)} - 3) + \tfrac{2}{T}\sum_{t=1}^T \boldsymbol{A}_{i,1,tt}\boldsymbol{b}_{i,t}\xi_{i,t}^{(3)} \\
\Sigma_{V_i,12} &= \tfrac{2}{T}\sum_{t=1}^T\sum_{s=1}^T \boldsymbol{A}_{i,1,ts}\boldsymbol{A}_{i,2,ts} + \tfrac{1}{T}\sum_{t=1}^T \boldsymbol{A}_{i,1,tt}\boldsymbol{A}_{i,2,tt}(\xi_{i,t}^{(4)} - 3) + \tfrac{2}{T}\sum_{t=1}^T \boldsymbol{A}_{i,2,tt}\boldsymbol{b}_{i,t}\xi_{i,t}^{(3)} \\
\Sigma_{V_i,22} &= \tfrac{2}{T}\sum_{t=1}^T\sum_{s=1}^T \boldsymbol{A}_{i,2,ts}^2 + \tfrac{1}{T}\sum_{t=1}^T \boldsymbol{A}_{i,2,tt}^2(\xi_{i,t}^{(4)} - 3),
\end{aligned}$$

with $\xi_{i,t}^{(3)} = \mathrm{E}\left(\zeta_{i,t}^3|\boldsymbol{Y}_{-i}\right)$, $\xi_{i,t}^{(4)} = \mathrm{E}\left(\zeta_{i,t}^4|\boldsymbol{Y}_{-i}\right)$ and

$$\begin{aligned}
\boldsymbol{A}_{i,1} &= \tfrac{1}{T}\mathbf{\Sigma}_{u_i}^{1/2'}\boldsymbol{Y}_{-i}(\tfrac{1}{T}\boldsymbol{Y}_{-i}'\boldsymbol{Y}_{-i})^{-2}\boldsymbol{Y}_{-i}'\mathbf{\Sigma}_{u_i}^{1/2} \\
\boldsymbol{A}_{i,2} &= \mathbf{\Sigma}_{u_i}^{1/2'}(\boldsymbol{I}_T - \boldsymbol{Y}_{-i}(\boldsymbol{Y}_{-i}'\boldsymbol{Y}_{-i})^{-1}\boldsymbol{Y}_{-i}')\mathbf{\Sigma}_{u_i}^{1/2} \\
\boldsymbol{b}_i &= 2\mathbf{\Sigma}_{u_i}^{1/2'}\boldsymbol{Y}_{-i}\left(\tfrac{1}{T}\boldsymbol{Y}_{-i}'\boldsymbol{Y}_{-i}\right)^{-1}\gamma_i.
\end{aligned}$$

A number of comments are in order. First, under the null hypothesis $H_0 : \|\mathbf{K}_i\| = 0$ the distribution of the granular statistics simplifies as then $\gamma_i = 0$. Hence, one can modify Theorem 2 by setting $\boldsymbol{b}_i = 0$ which yields the distribution of $T(\|\hat{\mathbf{K}}_i\| - \delta_{K_i})/\sigma_{K_i}$ under this null-hypothesis. In this paper we are not explicitly interested in this hypothesis, but we note that when dropping the ith element from $\mathbf{K}_i$ this result can be used to determine whether series i is partial correlated to any other series.

Second, it is important to emphasize that the limiting distribution does not depend on the specific structure of the concentration matrix that is imposed by the granular model (2). The theorem is general in the sense that it applies for all settings where the population concentration matrix $\mathbf{K}$ has bounded eigenvalues and where the observations that are used to construct the sample concentration matrix satisfy the assumptions 3 and 4.

4 Comparison to other methods

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

4.1 Principal components based methods

Stock and Watson (2002b), Bai and Ng (2006) and Parker and Sul (2016) propose methods based on principal components analysis to give meaning to the otherwise hard to interpret estimates for the common factors. They estimate the factors in an approximate factor model using principal components and subsequently use regression analysis to find the series that correlate most with the factors. Upon first sight, it seems that these methods could be adopted to detect granular series as well and it is true that in some settings these methods will yield the same set of granular series when compared to detection based on the column norms. However, there are several important scenarios in which a principal components based method will not be able to detect the granular series.

First, a conceptual difference is that any principal components based method amounts to detecting the series that explain the most variance in the panel. This follows as the common factors from principal components are estimated to maximize the explained variance. Our definition of granular series, as formulated in assumption 1, does not necessarily imply those series that explain the most variance. This implies that in noisy data panels detection based on principal components will perform less well.

A simple example of this is the following. Consider the baseline granular model (2) with $k = 1$ and the following parametrization

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

where ι_{n-1} is a vector of ones of length $n-1$ and δ and c are constants. For this model the conditioning number of $\boldsymbol{\Sigma}_\epsilon$ is given by $\kappa_\epsilon = 1$ and the identification assumption 1-(iv) is satisfied when $\delta > \sqrt{2}$. In this setting the granular series will be detected using the column norm statistic whenever $\delta > \sqrt{2}$.

In contrast, the detection power of the principal components method depends on the value of c. When c is large the first principal component will not correlate with the granular series. Moreover, estimators for the number of factors, such as those developed in Bai and Ng (2002) and Ahn and Horenstein (2013) will not detect any factors.

This problem becomes more prominent when common factors f_t are also present in the model. When these explain a lot of variance the principal components method primarily detects these and the subsequent regressions, that aim to find which series correlate most with the estimated factors, will detect also those series that load on the common factors.

Second, recovering the *number* of granular series is difficult using a principal components based method. To see this consider again the simple example above, but now for the case where c is small relative to δ. Clearly for $n, T \to \infty$ the R^2 from the regression of the first principal component on the first series will tend to one. However, for small c the R^2's from the other regressions also become arbitrarily close to one. This makes selecting the number of granular series difficult, because the same principal component can also be generated by multiple correlated granular series. In particular, a model with two highly correlated granular shocks will imply that PCA will estimate one common factor and generate the exactly the same R^2's sequence. The column norm statistics are able to distinguish between these two scenarios with ease.

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

4.2 Maximum likelihood based methods

A relatively straightforward method for performing granular detection would be to estimate the granular model for different orderings of the variables in y_t and then comparing the marginal likelihoods, or some other statistic, of the different models. To outline the practical difficulty with this approach consider the simplified model without factors 2 for a given k. This would involve estimating $\binom{n}{k}$ possible models. For $k = 1$ or $k = 2$ this approach is quite feasible. But for $n = 100$ and $k = 3$ this already involves estimating 161700 different models making this a computationally prohibitive task, see also Elliott, Gargano and Timmermann (2013). When also allowing for common factors in the model the computational task becomes even harder as more factors increase the computational difficulty. Potentially smart search algorithms could be considered but we do not explore this route further.

Alternatively, it is possible to estimate a factor representation of the model without imposing identification restrictions on the loading matrix. For example

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

where $h_t = (f_t', g_t')'$ and fixing the variance of h_t to the unit matrix. The difficulty of this approach lies again in the second step when one needs to figure out which series correspond with the estimated factors the most. In particular, maximum likelihood will assign positive variance to the errors $\zeta_{i,t}$ that correspond to the granular series and the rotation of the factors is not fixed. This makes the subsequent regressions of the estimated factors on the individual series unreliable since the factors are estimated based on maximizing the likelihood of an essentially under-identified model which can yield a very different optimal rotation of the factors. The same difficulty for detecting the number of granular series as for principal components applies here as well.

5 Inference on the granular panel model

5.1 Determining the number of common factors

Once the granular series have been identified and tested, we complete the specification of model (7) 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.

5.2 Model evaluation and inference

A question of central interest is whether the granular series have economically meaningful impact. So far our methodology has been predominantly focused on statistical measures for detecting granular series. In this section we outline how to use knowledge concerning granular series – and the number of common factors – to asses the economic importance of the granular series. This requires inference on model (7) and a subsequent discussion on how to obtain relevant impulse responses and variance decompositions.

In general, we follow Bernanke, Boivin and Eliasz (2005) and conduct likelihood based inference on the complete model. Similar as in their factor augmented vector autoregressive model we assume that the granular shocks and the common factors are modeled by

$$\begin{bmatrix} f_t \\ g_t \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Phi}_{ff,1} & \boldsymbol{\Phi}_{fg,1} \\ \boldsymbol{\Phi}_{gf,1} & \boldsymbol{\Phi}_{gg,1} \end{bmatrix} \begin{bmatrix} f_{t-1} \\ g_{t-1} \end{bmatrix} + \cdots + \begin{bmatrix} \boldsymbol{\Phi}_{ff,p} & \boldsymbol{\Phi}_{fg,p} \\ \boldsymbol{\Phi}_{gf,p} & \boldsymbol{\Phi}_{gg,p} \end{bmatrix} \begin{bmatrix} f_{t-p} \\ g_{t-p} \end{bmatrix} + \begin{bmatrix} \eta_{f,t} \\ \eta_{g,t} \end{bmatrix}, \quad (11)$$

where the $\boldsymbol{\Phi}$'s are the autoregressive matrices and the η's are the disturbances for which we assume $(\eta_{f,t}, \eta_{g,t})' \sim iid(0, \boldsymbol{\Sigma}_\eta)$. Following assumption 2 we have that $\eta_{f,t}$ and $\eta_{g,t}$ can be arbitrarily correlated among each other and for identification purposes we restrict the variance of $\eta_{f,t}$ to unity, see also Doz et al. (2012).

The state equation (11) combined with the measurement equation (7) implies a linear state space model in the sense of Durbin and Koopman (2012). The parameters of the model can be summarized in the vector $\boldsymbol{\psi}$ which includes the loading matrices $\boldsymbol{\beta}$ and $\boldsymbol{\Lambda}$, and also the parameters of the state equation (11). For the non-granular shocks ϵ_t different assumptions can be made depending on their persistence and their correlation among each other. See Jungbacker and Koopman (2015) for different possibilities for the error terms within a likelihood based framework. Once the complete model is specified the parameters can be estimated by maximum likelihood for which the likelihood can be evaluated using the Kalman filter and the expectation-maximization algorithm, see Durbin and Koopman (2012, Chapter 4).

Typically, we estimate a restricted version of what is believed to be the true model as often the variance of ϵ_t is restricted to be diagonal. For parameter consistency in such settings we appeal to the quasi-maximum likelihood theories that were developed in Doz et al. (2012), Bai and Li (2012) and Bai and Li (2015). Given the parameter estimates we can compute impulse responses and variance decompositions similar as in Bernanke et al. (2005).

6 Simulation Study

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

6.1 Simulation design

We generate data panels from the granular panel data model with common factors given in equation (7). We consider data panels with dimensions $n = 50, 100$ and $T = 200, 400$. The number of granular series that we include is equal to $k = 3, 5$ and the number of common factors that we include is equal to $r = 0, 3, 5$.

The granular shocks and common factors follow the vector autoregressive process given in equation (11). We consider a lag lenght of $p = 1$. The variance matrix $\mathbf{\Sigma}_\eta$ has ones on the main diagonal and correlation coefficient c_η on the off-diagonal elements. We note that c_η captures the contemporaneous correlation among the granular shocks and the common factor shocks. We vary its value by taking $c_\eta = 0, 0.5$. The elements for the diagonal of $\mathbf{\Phi}_1 = [\mathbf{\Phi}_{ff,1}, \mathbf{\Phi}_{fg,1} : \mathbf{\Phi}_{gf,1}, \mathbf{\Phi}_{gg,1}]$ 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.

We generate the non-granular idiosyncratic shocks from

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

with $\mathbf{\Gamma}$ diagonal with elements $\mathbf{\Gamma}_{ii} \sim U(0.5, 0.95)$, where $U(0.5, 0.95)$ indicates the uniform distribution over the range (0.5,0.95). This ensures that e_t follows a stationary vector autoregressive process with variance $\boldsymbol{I}_{n-k}$. From this we generate $\epsilon_t = \mathbf{\Sigma}_\epsilon^{1/2} e_t$ such that $\text{Var}(\epsilon_t) = \mathbf{\Sigma}_\epsilon$. For the latter we consider (a) diagonal, (b) banded and (c) sparse structures. For the diagonal structure we have $\mathbf{\Sigma}_{\epsilon,i,j}^{1/2} \sim U(0.5, 1.5)$ for all $i = j$ and zero else. For the banded structure we have $\mathbf{\Sigma}_{\epsilon,i,j}^{1/2} \sim U(0.5, 1.5)$ if $i = j$, $\mathbf{\Sigma}_{\epsilon,i,j}^{1/2} = c_\epsilon$ with $c_\epsilon = 0.2$ if $i > j$ and $i - j < 10$ and zero else. Finally, the sparse structure is given by $\mathbf{\Sigma}_{\epsilon,i,j}^{1/2} \sim U(0.5, 1.5)$ if $i = j$, if $i > j$ $\mathbf{\Sigma}_{\epsilon,i,j}^{1/2} \sim U(-0.5, 0.5)$ with probability $0.2/[\sqrt{n-k}\log(n-k)]$ and zero else, and if $i < j$ the value is zero. Notice that this implies that for each case $\mathbf{\Sigma}_\epsilon^{1/2}$ is lower triangular and $\mathbf{\Sigma}_\epsilon = \mathbf{\Sigma}_\epsilon^{1/2}\mathbf{\Sigma}_\epsilon^{1/2\prime}$. The banded structure is similar as in Stock and Watson (2002a) and Bai and Ng (2002) whereas the sparse structure is similar as considered in Fan et al. (2011).

The 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 $\sigma_b^2 = 0.01, 0.05, 0.1, 0.25, 0.5, 0.75, 1$. The loadings of the common factors are drawn from a standard normal distribution. For small values of σ_b^2 this reflect the situation where the common factors explain more variance in the observations when compared to the granular shocks. Such settings are argued to be empirically relevant in for example Foerster et al. (2011).

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

6.2 Granular detection results

We begin by studying the finite sample properties of our ranking methods. Corollaries 1 and 2, that are based on the consistency of the column norms, imply that we can correctly identify the granular series when n and T become large. For each simulated panel we rank the series in the panel according to the column norms of the concentration matrix, and then we select the number of granulars using the column ratio statistic 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 simulations considered, we only discuss the case where the non-granular errors have the banded and sparse designs. These cases are the most relevant for empirical applications. Only, we only show the case where the correlation between the granular shocks and the common factors is fixed at $c_\eta = 0.5$. Changing this coefficient has no effect on the detection results for the column norm.

We present the results for average proportion of correctly ranked granular series in Table 1 and the proportions of correctly selected number of granulars in Table 2. The tables reveal some interesting patterns.

First, the key parameter for which the outcomes fluctuate the most is the magnitude of the granular loadings as captured by the standard deviation coefficients σ_b^2. When this variance is close to zero this implied that the granular loadings are close to zero and by result ranking the granulars correctly becomes more challenging. When the variance increases the percentage of correctly ranked granulars increases rapidly. Notice that when $\sigma_b^2 = 0.1$, which still implies that the coefficients are on average local-to-zero for $n = 100$ the detection rate is nearly perfect. Hence for reasonably connected granular series we should expect to detect them easily. For estimating the correct number of granular series a similar pattern is detected. However, as obtaining the correct number of granulars requires a stronger identification condition, see Lemma 2, we see that the percentages are overall lower for this statistic.

Second, the panel dimensions imply that larger panels n, T improve the detection results. Both increases in n and T improve the ranking and the estimation of the number of granular series. We remark here that the case where $n \approx T$, which is not covered by our theory, produces less good results.

Third, differences between $k = 3$ and $k = 5$ granular series are small. Also, the performance of the methodology is only mildly affected by the numbers of factors in the specification. Only, the selection of the number of granular series suffers slightly when including additional common factors. This confirms the identification results derived in Lemmas 3 and 4.

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

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

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

6.2.1 Comparison to Factor Model Based Methods

In this section we compare the performance of our methods with granular identification procedures that are based on principal components analysis, see for example Stock and Watson (2002b), Bai and Ng (2006) and Parker and Sul (2016). The latter methods are based on regressing the individual series on the estimated factors from principal components analysis. First, we consider a straightforward implementation of such method: (i) estimate the number of factors using Bai and Ng (2002), (ii) regress each time series on the common factors and (iii) rank according to the R^2 of this regression. For estimating the number of factors we use the IC2 criteria from Bai and Ng (2002) which gave slightly better results when compared to the eigenvalue ratio estimator from Ahn and Horenstein (2013). Ranking based on the R^2 are commonly reported in the factor model literature (e.g. Stock and Watson (2002b) and Foerster et al. (2011)) and we emphasize that they are not designed for granular detection but for interpreting the common factors. Nevertheless, it is interesting to compare our methodology to this procedure.

In Table 3 we show the ratios between the percentage of correctly ranked granulars based on the R^2's and the column norm statistic. We find that in all cases the column norm statistic determines a better ranking when compared to the R^2's. On average – across all specifications – we find that the column norm ranking method performs 25% better. The differences are large for small values of σ_b^2 and tend to zero when the influence of the granular series becomes larger. This is in line with the theoretical discussion in Section 4 where we illustrated that the factor based methods become difficult for weak granular series, see also Onatski (2012).

Second, a potentially more advanced method for detecting granular series based on pca is proposed in Parker and Sul (2016). As previously discussed, it is important to emphasize that the Parker and Sul (2016) algorithm is designed for a different problem in comparison to the one considered here. Nevertheless, the two approaches have some similarities and it is interesting to analyze the performance of the Parker and Sul (2016) algorithm in the framework of this paper. To this extent, we apply the Parker and Sul (2016) algorithm to detect granular series. As far as the implementation of the Parker and Sul (2016) procedure is concerned, we follow exactly the algorithm described in their paper.

7 Empirical Applications

7.1 Granular series in US industrial production

We study the presence of granular series in US industrial production, see also Forni and Reichlin (1998), Foerster et al. (2011), Pesaran and Yang (2016), Siavash (2016) and Atalay (2017). We consider a panel of sector specific industrial production time series for $n = 138$ different sectors in the US economy ranging from 1972 until 2007[11]. Each time series concerns monthly industrial production growth rates and in total we have $T = 431$ time periods. The panel is standardized such that each series has mean zero and unit variance. The questions that we aim to answer are: (a) are there granular sectors in US industrial production?, (b) do the granular sectors significantly predict aggregate industrial production?, (c) how much variance do the granular sectors explain relative to the common factors? To answer these questions we apply our granular detection methods.

In Figure 3 we show the ordered column norms $\|\hat{\mathbf{K}}_i\|$ of the concentration matrix. In panel (a) we show all norms together with their 95% confidence bounds. The bounds are based on the asymptotic approximation given in Theorem 2 and computed using the variance estimator given in Lemma **??**. To improve visibility the largest twenty column norms are also shown in Panel (c). We find that there are two series that are clearly distinct from the others: Motor Vehicle Parts and Automobiles and Light Duty Motor Vehicles. Both sectors fall within the automobile industry which was signaled as a potentially granular industry during the financial crises by Alan R. Mulally, the chief executive of Ford, see Mulally (2008) and the discussion in Acemoglu et al. (2012).

The importance of the automobile industry is further amplified in Table 8 where we provide the details for the series corresponding to the largest ten column norms. We find that in the top ten; four series are directly related to the automobile industry. Other potentiually granular sectors that we find are related to aluminum, plastics and paper products. We emphasize that after the first six or seven sectors the differences in the column norms become

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

very small.

In panels (b) and (d) of Figure 3 we show the column norm ratios. The estimator $\hat{k}$, given in equation 10, indicates that there are two granular series in our panel. We conduct a hypothesis test to verify that the second and third column norms are significantly distinct, see section **??**. The null hypothesis of equal column norms is rejected with for a test with level $\alpha = 0.05$. A second hypothesis test for the difference between the seventh and eighth largest column norms could not be rejected.

Next, we determine the number of common factors that remain after taking into account the granular series. Using the IC2 criteria proposed in Bai and Ng (2002) we find that there is one common factors left in the panel. This is confirmed by the common factor estimators proposed in Onatski (2010) and Ahn and Horenstein (2013), see also Foerster et al. (2011) who find one or two factors for a similar panel.

In summary, our granular detection method and specification tests indicate that a model with two granular series (Motor Vehicle Parts and Automobiles and Light Duty Motor Vehicles) and one common factor agrees with the data.

7.1.1 Time dependence in granular detection

Next, we consider the stability of the granular detection method for different sampling periods. In particular, we follow Foerster et al. (2011) and split the sample into two different periods, 1972-1983 and 1984-2007, and repeat the previous analysis. We note that for the 1972-1983 sampling period our methodology is not very reliable as $T = 143$ for this period. This makes n and T very close and this is not covered by our theory.

In Figure 4 we show the largest ten column norm statistics for both sampling periods. For the 1972-1983 period we find no significant granular series. The two granular series from the full sample analysis remain in the top five of series, see Table 8, but the standard errors are very large and we consider the estimates in general unreliable.

For the 1984-2007 sampling period we find a similar ranking as for the full sample. In particular, nine of the top ten series are also in the top twenty for the full sample. The top five series is practically unchanged. The estimator for k now indicates that there are five granular series in the model. This confirms the finding in Foerster et al. (2011) who find that idiosyncratic shocks have become more important in recent years.

7.1.2 Comparison to other methods

We compare our granular detection method to methods based on principal components. Based on the simulation section we only consider a ranking based on the R^2's of the regression of the ith series on the principal components. A similar ranking is also presented in Foerster et al. (2011)[12] and we follow their construction by using two principal components.

In Table 9 we show the selected granular series that result from the R^2 ranking. We find a quite different set of granular series.

7.2 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 4. 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)

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

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

where $p_{i,t}^{\text{high}}$ and $p_{i,t}^{\text{low}}$ denote respectively the max and the min log price of stock i on day t.[13] 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 $N = 88$.

The volatility panel exhibits the typical stylized facts documented in the literature. Table 5 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 6 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 $\|\hat{\mathbf{K}}_i\|$ 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.

[13] 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)).

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

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

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 7 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.

8 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.

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

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

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

Amemiya, T.: 1985, *Advanced Econometrics*, Harvard University Press, Cambridge, Massachusetts.

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

Anderson, T. W. and Gupta, S. D.: 1963, Some Inequalities on Characteristic Roots of Matrices, *Biometrika* **50**, 522–524.

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

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

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

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

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

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

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

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

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.

Bernard, A. B., Jensen, J. B., Redding, S. J. and Schott, P. K.: 2016, Global Firms. CEP Discussion Paper No 1420.

Billingsley, P.: 1986, *Probability and Measure*, Wiley, Chichester.

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.

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

Cattaneo, M. D., Jansson, M. and Newey, W.: 2017a, Alternative asymptotics and the partially linear model with many regressors, *Econometric Theory* p. forthcoming.

Cattaneo, M. D., Jansson, M. and Newey, W.: 2017b, Inference in linear regression models with many covariates and heteroskedasticity. Working paper.

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

Chao, J. C., Hausman, J. A., Newey, W. K., Swanson, N. R. and Woutersen, T.: 2014, Testing overidentifying restrictions with many instruments and heteroskedasticity, *Journal of Econometrics* **178**, 15–21.

Chao, J. C. and Swanson, N. R.: 2005, Consistent Estimation with a Large Number of Weak Instruments, *Econometrica* **73**, 1673–1692.

Chao, J. C., Swanson, N. R., Hausman, J. A., Newey, W. K. and Woutersen, T.: 2012, Asymptotic Distribution of JIVE in a Heteroskedastic IV Regression with Many Instruments, *Econometric Theory* **28**, 42–86.

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

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

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

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

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

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

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

Fan, J., Liao, Y. and 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.

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

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

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

Hall, P. and Heyde, C. C.: 1980, *Martingale Limit Theory and its Application*, Academic Press, New York.

Hausman, J. A., Newey, W. K., Woutersen, T., Chao, J. C. and Swanson, N. R.: 2012, Instrumental variable estimation with heteroskedasticity and many instruments, *Quantitative Economics* **3**, 211–255.

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

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

Kelejian, H. K. and Prucha, I. R.: 2001, On the asymptotic distribution of the Moran I test statistic with applications, *Journal of Econometrics* **104**, 219–257.

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.

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

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

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.

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

Moon, H. R. and Weidner, M.: 2015, Dynamic Linear Panel Regression Models with Interactive Fixed Effects, *Econometric Theory* . forthcoming.

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

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

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

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

Pesaran, M. H. and Yang, C. F.: 2016, Econometric analysis of production networks with dominant units. Institute for New Economic Thinking: Working Paper No. 16-25.

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

Rao, C. R.: 1973, *Linear Statistical Inference and Its Applications (Second Ed.)*, Wiley, New York.

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

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

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

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

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

Table 1: Granular Ranking Probabilities

n	T	k	r	0.01	0.05	0.10	0.25	0.50	0.75	1.00
					Banded error design					
50	200	3	0	0.140	0.800	0.967	1.000	1.000	1.000	1.000
100	200	3	0	0.140	0.926	0.996	1.000	1.000	1.000	1.000
50	400	3	0	0.207	0.941	0.998	1.000	1.000	1.000	1.000
100	400	3	0	0.266	0.992	1.000	1.000	1.000	1.000	1.000
50	200	5	0	0.206	0.813	0.954	0.997	1.000	1.000	1.000
100	200	5	0	0.252	0.932	0.997	1.000	1.000	1.000	1.000
50	400	5	0	0.187	0.905	0.992	1.000	1.000	1.000	1.000
100	400	5	0	0.334	0.991	1.000	1.000	1.000	1.000	1.000
50	200	3	5	0.103	0.676	0.876	0.969	0.983	0.985	0.989
100	200	3	5	0.148	0.865	0.972	0.998	0.998	0.998	0.999
50	400	3	5	0.104	0.800	0.953	0.989	0.992	0.994	0.994
100	400	3	5	0.184	0.967	0.996	0.999	1.000	0.999	1.000
50	200	5	5	0.156	0.721	0.884	0.970	0.988	0.988	0.993
100	200	5	5	0.216	0.878	0.976	0.998	0.999	0.998	0.999
50	400	5	5	0.150	0.822	0.950	0.990	0.995	0.997	0.998
100	400	5	5	0.273	0.966	0.997	1.000	0.999	0.999	1.000
					Sparse error design					
50	200	3	0	0.201	0.809	0.968	0.999	1.000	1.000	1.000
100	200	3	0	0.225	0.937	0.995	1.000	1.000	1.000	1.000
50	400	3	0	0.214	0.931	0.998	1.000	1.000	1.000	1.000
100	400	3	0	0.312	0.992	1.000	1.000	1.000	1.000	1.000
50	200	5	0	0.269	0.816	0.956	0.998	1.000	1.000	1.000
100	200	5	0	0.293	0.931	0.996	1.000	1.000	1.000	1.000
50	400	5	0	0.284	0.917	0.993	1.000	1.000	1.000	1.000
100	400	5	0	0.373	0.989	1.000	1.000	1.000	1.000	1.000
50	200	3	5	0.138	0.695	0.872	0.966	0.982	0.982	0.986
100	200	3	5	0.177	0.859	0.977	0.996	0.997	0.998	0.998
50	400	3	5	0.147	0.832	0.954	0.989	0.994	0.996	0.996
100	400	3	5	0.232	0.967	0.997	1.000	1.000	1.000	1.000
50	200	5	5	0.214	0.733	0.891	0.972	0.987	0.988	0.991
100	200	5	5	0.241	0.881	0.971	0.997	0.999	0.999	0.999
50	400	5	5	0.222	0.839	0.953	0.991	0.996	0.997	0.997
100	400	5	5	0.311	0.963	0.997	0.999	1.000	1.000	1.000

The table reports the average proportion of correctly ranked granulars.

Table 2: Granular Selection Probabilities

n	T	k	r	0.01	0.05	0.10	0.25	0.50	0.75	1.00
Banded error design										
50	200	3	0	0.098	0.190	0.520	0.883	0.979	0.992	0.994
100	200	3	0	0.104	0.377	0.773	0.969	0.995	0.999	1.000
50	400	3	0	0.093	0.404	0.833	0.988	0.998	1.000	1.000
100	400	3	0	0.090	0.732	0.974	0.999	1.000	1.000	1.000
50	200	5	0	0.044	0.101	0.357	0.832	0.959	0.969	0.986
100	200	5	0	0.048	0.265	0.782	0.981	0.995	0.998	0.999
50	400	5	0	0.042	0.207	0.661	0.963	0.993	0.997	0.999
100	400	5	0	0.047	0.652	0.941	0.998	1.000	1.000	1.000
50	200	3	5	0.117	0.136	0.310	0.482	0.594	0.587	0.615
100	200	3	5	0.115	0.273	0.556	0.771	0.789	0.781	0.795
50	400	3	5	0.098	0.188	0.418	0.656	0.704	0.708	0.723
100	400	3	5	0.135	0.525	0.785	0.882	0.884	0.875	0.878
50	200	5	5	0.036	0.068	0.141	0.428	0.575	0.626	0.641
100	200	5	5	0.042	0.159	0.507	0.784	0.825	0.838	0.852
50	400	5	5	0.044	0.105	0.296	0.608	0.708	0.746	0.759
100	400	5	5	0.047	0.425	0.787	0.902	0.916	0.922	0.917
Sparse error design										
50	200	3	0	0.105	0.201	0.520	0.909	0.977	0.990	0.994
100	200	3	0	0.095	0.409	0.790	0.983	0.997	1.000	1.000
50	400	3	0	0.085	0.397	0.844	0.989	0.997	1.000	1.000
100	400	3	0	0.118	0.739	0.979	1.000	1.000	1.000	1.000
50	200	5	0	0.042	0.117	0.392	0.813	0.948	0.978	0.992
100	200	5	0	0.047	0.248	0.766	0.968	0.996	0.999	1.000
50	400	5	0	0.047	0.248	0.669	0.953	0.988	1.000	0.997
100	400	5	0	0.046	0.632	0.960	0.999	1.000	1.000	1.000
50	200	3	5	0.128	0.156	0.264	0.528	0.608	0.576	0.603
100	200	3	5	0.092	0.231	0.539	0.760	0.778	0.778	0.806
50	400	3	5	0.115	0.205	0.460	0.688	0.738	0.715	0.694
100	400	3	5	0.121	0.510	0.792	0.892	0.884	0.867	0.873
50	200	5	5	0.047	0.087	0.166	0.429	0.565	0.636	0.637
100	200	5	5	0.036	0.176	0.467	0.750	0.832	0.848	0.824
50	400	5	5	0.038	0.131	0.315	0.668	0.727	0.737	0.753
100	400	5	5	0.036	0.426	0.761	0.907	0.913	0.913	0.928

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

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

n	T	k	r	0.01	0.05	0.10	0.25	0.50	0.75	1.00
Banded error design										
50	200	3	0	0.062	0.582	0.835	0.977	0.993	0.995	0.995
100	200	3	0	0.000	0.561	0.860	0.991	0.997	0.999	0.999
50	400	3	0	0.056	0.701	0.949	0.999	1.000	1.000	1.000
100	400	3	0	0.006	0.855	0.993	1.000	1.000	1.000	1.000
50	200	5	0	0.044	0.636	0.843	0.954	0.977	0.979	0.976
100	200	5	0	0.036	0.655	0.899	0.992	0.997	0.998	0.999
50	400	5	0	0.000	0.618	0.874	0.981	0.992	0.992	0.994
100	400	5	0	0.002	0.754	0.966	0.999	1.000	1.000	1.000
50	200	3	5	0.419	0.331	0.544	0.830	0.935	0.952	0.954
100	200	3	5	0.153	0.218	0.591	0.938	0.991	0.999	0.996
50	400	3	5	0.365	0.299	0.554	0.876	0.963	0.980	0.979
100	400	3	5	0.136	0.201	0.655	0.983	0.999	1.000	1.000
50	200	5	5	0.446	0.347	0.502	0.736	0.859	0.884	0.891
100	200	5	5	0.207	0.250	0.536	0.904	0.983	0.991	0.992
50	400	5	5	0.461	0.312	0.492	0.762	0.899	0.936	0.942
100	400	5	5	0.160	0.242	0.604	0.967	0.998	1.000	0.999
Sparse error design										
50	200	3	0	0.685	0.622	0.692	0.704	0.764	0.850	0.908
100	200	3	0	0.524	0.524	0.653	0.556	0.485	0.529	0.621
50	400	3	0	0.765	0.775	0.820	0.840	0.890	0.926	0.948
100	400	3	0	0.600	0.690	0.659	0.553	0.486	0.533	0.579
50	200	5	0	0.807	0.619	0.792	0.830	0.902	0.940	0.953
100	200	5	0	0.508	0.790	0.894	0.928	0.951	0.977	0.981
50	400	5	0	0.627	0.543	0.739	0.661	0.664	0.768	0.848
100	400	5	0	0.706	0.793	0.777	0.668	0.673	0.707	0.773
50	200	3	5	0.390	0.460	0.712	0.904	0.964	0.967	0.966
100	200	3	5	0.143	0.465	0.788	0.963	0.986	0.984	0.987
50	400	3	5	0.306	0.355	0.645	0.830	0.859	0.871	0.849
100	400	3	5	0.102	0.440	0.685	0.794	0.830	0.826	0.823
50	200	5	5	0.451	0.535	0.744	0.917	0.965	0.968	0.964
100	200	5	5	0.184	0.472	0.792	0.969	0.987	0.989	0.993
50	400	5	5	0.369	0.507	0.755	0.890	0.915	0.910	0.903
100	400	5	5	0.129	0.539	0.794	0.872	0.866	0.860	0.850

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

Table 4: 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 5: 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 6: 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

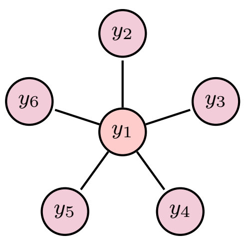
Table 7: SIFI Prediction

$\|\boldsymbol{K}_i\|$	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

Figure 1: Partial Correlation Network Representation

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

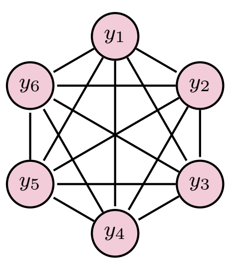


Partial Correlation Network

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

Concentration Matrix

$$
y_t = \Lambda f_t + \epsilon_t \quad \epsilon_t \sim D(0,\mathbf{I}_6)
$$

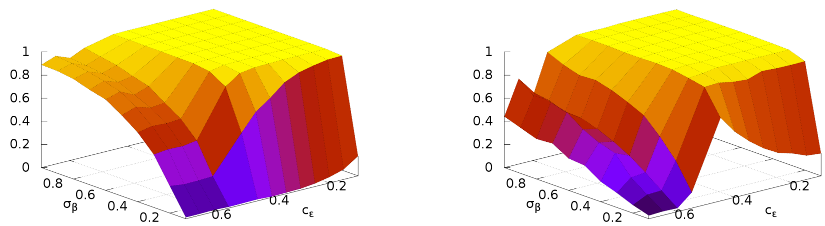


Partial Correlation Network

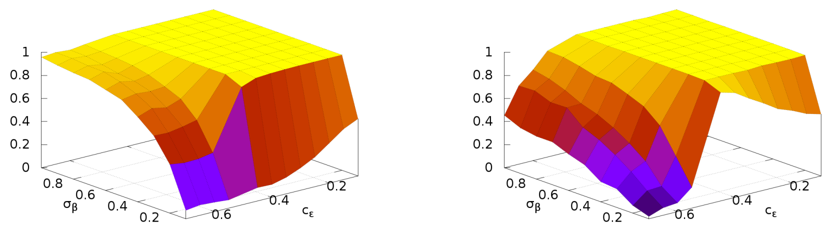
$$
\mathbf{K} = -\frac{1}{1+\Lambda'\Lambda}\begin{bmatrix}
\Lambda_1^2-1 & \dots & \dots & \dots & \dots & \Lambda_1\Lambda_6 \\
\Lambda_2\Lambda_1 & \Lambda_2^2-1 & \dots & \dots & \dots & \Lambda_2\Lambda_6 \\
\Lambda_3\Lambda_1 & \Lambda_3\Lambda_2 & \Lambda_3^2-1 & \dots & \dots & \Lambda_3\Lambda_6 \\
\Lambda_4\Lambda_1 & \Lambda_4\Lambda_2 & \Lambda_4\Lambda_3 & \Lambda_4^2-1 & \dots & \Lambda_4\Lambda_6 \\
\Lambda_5\Lambda_1 & \Lambda_5\Lambda_2 & \Lambda_5\Lambda_3 & \Lambda_5\Lambda_4 & \Lambda_5^2-1 & \Lambda_5\Lambda_6 \\
\Lambda_6\Lambda_1 & \Lambda_6\Lambda_2 & \Lambda_6\Lambda_3 & \Lambda_6\Lambda_4 & \Lambda_6\Lambda_5 & \Lambda_6^2-1
\end{bmatrix}
$$

Concentration Matrix

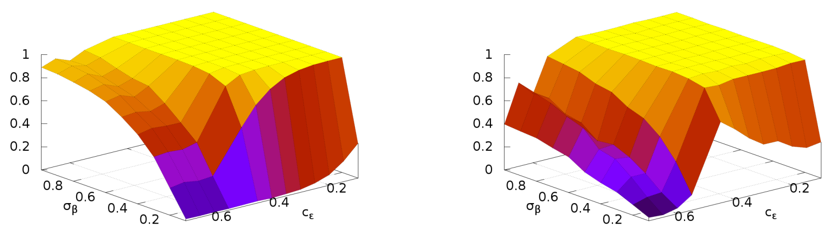
Figure 2: Granular Ranking and Selection Probabilities



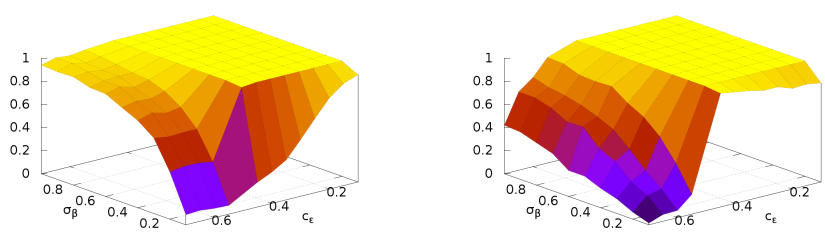
$T = 200 \quad N = 50$



$T = 200 \quad N = 100$



$T = 400 \quad N = 50$



$T = 400 \quad N = 100$

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

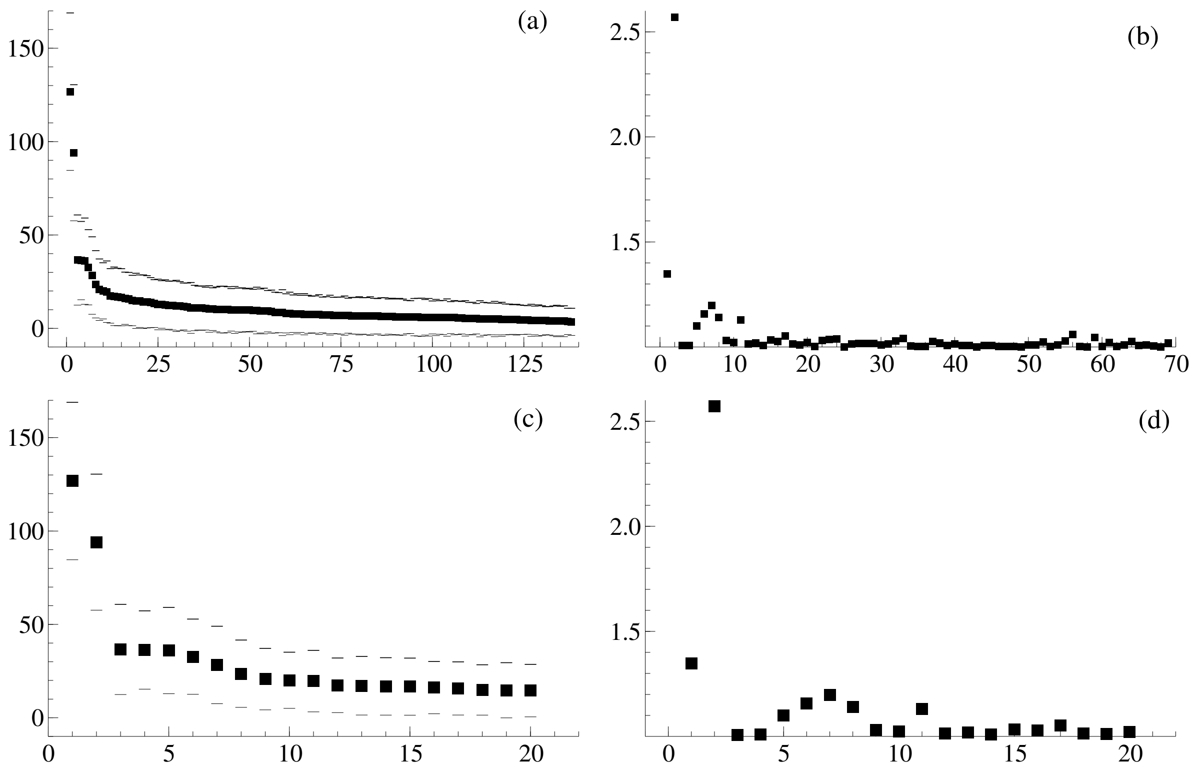


Figure 3: Granular detection results for industrial production series

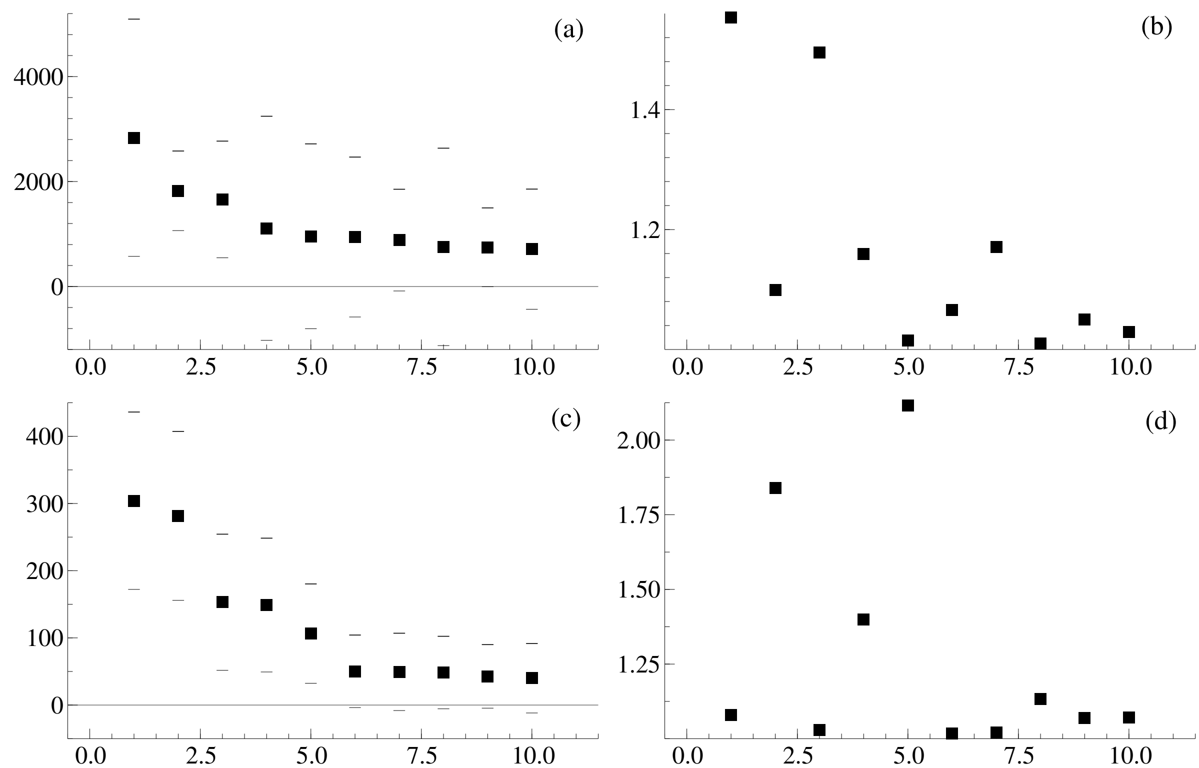


Figure 4: Granular detection results for industrial production series for different sampling periods.

Table 8: Granular series US Industrial Production

Sampling period 1972-2007

Sector	$\|\hat{\mathbf{K}}_{(i)}\|$	95% l	95% u	$\frac{\|\hat{\mathbf{K}}_{(i)}}{\hat{\mathbf{K}}_{(i+1)}\|}$
Motor Vehicle Parts	126.76	84.63	168.90	1.35
Automobiles and Light Duty Motor Vehicles	94.02	57.59	130.45	2.57
Aluminum Extruded Products	36.58	12.45	60.71	1.01
Plastics Products	36.30	15.32	57.27	1.01
Miscellaneous Aluminum Materials	35.99	12.89	59.08	1.10
Motor Vehicle Bodies	32.70	12.56	52.84	1.16
Paper and Paperboard Mills	28.26	7.52	49.01	1.20
Household and Institutional Furniture and Kitchen Cabinets	23.59	5.57	41.61	1.14
Commercial and Service Industry Machines	20.70	4.22	37.18	1.03
Motor Homes	20.07	5.04	35.10	1.02

Sampling period 1972-1983

Sector	$\|\hat{\mathbf{K}}_{(i)}\|$	95% l	95% u	$\frac{\|\hat{\mathbf{K}}_{(i)}}{\hat{\mathbf{K}}_{(i+1)}\|}$
Motor Vehicle Parts	2834.50	574.82	5094.20	1.55
Household and Institutional Furniture and Kitchen Cabinets	1824.41	1066.51	2582.31	1.10
Plastics Products	1659.92	548.32	2771.44	1.50
Organic Chemicals	1110.20	-1024.23	3244.52	1.16
Automobiles and Light Duty Motor Vehicles	957.29	-803.40	2718.02	1.02
Commercial and Service Industry Machines	942.99	-580.02	2466.01	1.07
Animal Slaughtering and Meat Processing Ex Poultry	884.35	-84.73	1853.41	1.17
Other Textile Product Mills	755.53	-1124.24	2635.33	1.01
Semiconductors and Other Electronic Components	747.73	-3.63	1499.14	1.05
Foundries	711.83	-433.10	1856.80	1.03

Sampling period 1984-2007

Sector	$\|\hat{\mathbf{K}}_{(i)}\|$	95% l	95% u	$\frac{\|\hat{\mathbf{K}}_{(i)}}{\hat{\mathbf{K}}_{(i+1)}\|}$
Motor Vehicle Parts	304.19	172.18	436.21	1.08
Automobiles and Light Duty Motor Vehicles	281.65	155.94	407.36	1.84
Aluminum Extruded Products	153.12	51.79	254.45	1.03
Miscellaneous Aluminum Materials	148.86	49.31	248.42	1.40
Motor Vehicle Bodies	106.44	32.53	180.35	2.12
Truck Trailers	50.32	-3.60	104.24	1.02
Carpet and Rug Mills	49.44	-8.12	106.99	1.02
Paper and Paperboard Mills	48.43	-5.55	102.41	1.13
Motor Homes	42.76	-4.50	90.013	1.07
Concrete and Products	40.00	-11.65	91.649	1.07

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

Table 9: Granular series US Industrial Production based on R^2

Sampling period 1972-2007

Sector	R^2
Plastics Products*	0.651
Household and Institutional Furniture and Kitchen Cabinets*	0.520
Metal Valves Except Ball and Roller Bearings	0.476
Architectural and Structural Metal Products	0.448
Other Miscellaneous Manufacturing	0.441
Sawmills and Wood Preservation	0.429
Reconstituted Wood Products	0.423
Fabricated Metals: Forging and Stamping	0.423
Fabricated Metals: Spring and Wire Products	0.422
Commercial and Service Industry Mach/Other Gen Purpose Mach	0.406

Sampling period 1972-1983

Sector	R^2
Plastics Products*	0.736
Household and Institutional Furniture and Kitchen Cabinets*	0.667
Metal Valves Except Ball and Roller Bearings	0.631
Architectural and Structural Metal Products	0.597
Other Electrical Equipment	0.596
Fabricated Metals: Spring and Wire Products	0.589
Fabricated Metals: Forging and Stamping	0.566
Commercial and Service Industry Mach/Other Gen Purpose Mach	0.548
Foundries	0.538
Other Miscellaneous Manufacturing	0.536

Sampling period 1984-2007

Sector	R^2
Carpet and Rug Mills	0.495
Reconstituted Wood Products	0.476
Breweries	0.443
Plastics Products*	0.439
Sawmills and Wood Preservation	0.398
Commercial and Service Industry Mach/Other Gen Purpose Mach	0.387
Metalworking Machinery	0.360
Other Miscellaneous Manufacturing	0.355
Architectural and Structural Metal Products	0.353
Boiler, Tank, and Shipping Containers	0.350

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

Table 10: State Space form Granular Series

	Autoregressive matrix $\mathbf{\Phi}$				Covariance matrix $\mathbf{\Sigma}_\eta$		
	$f_{1,t-1}$	$g_{1,t-1}$	$g_{2,t-1}$		$f_{1,t}$	$g_{1,t}$	$g_{2,t}$
$f_{1,t}$	0.173 $_{0.246}$	0.215 $_{0.105}$	0.169 $_{0.264}$	$f_{1,t}$	1.000	-	-
$g_{1,t}$	0.749 $_{0.162}$	0.137 $_{0.070}$	0.568 $_{0.175}$	$g_{1,t}$	-0.514 $_{0.020}$	0.432 $_{0.024}$	-
$g_{2,t}$	-0.066 $_{0.218}$	-0.022 $_{0.094}$	-0.134 $_{0.235}$	$g_{2,t}$	-0.867 $_{0.009}$	0.427 $_{0.017}$	0.782 $_{0.015}$

The table reports the matrices of the state equation.