Dominant Sectors in the US: A Factor Model Analysis of Sectoral Industrial Production

Soroosh Soofi Siavash*

February 2015

Abstract

This paper considers the role of sector-specific growth shocks as major sources of cross-sector comovement in the business cycle. The dominant sectors are identified according to the number of factors detected in the residuals after treating the production growth in some candidate sectors as observed factors. We build on the properties of the principal component method and analytically show this approach can consistently identify the dominant units (sectors) in a large panel data set even if the assumption of strongly influential factors is moderately relaxed. We provide empirical evidence that growth in a few industrial sectors in the US provide suitable approximations for an unknown common factor. Using data on the intersectoral material input-outputs and the cross-sector capital flows, we find that the dominant sectors have an important role as suppliers of capital products to other sectors.

*Goethe University Frankfurt, ssoofi@hof.uni-frankfurt.de. I am grateful to Mehdi Hosseinkouchack and Sebastian Kripfganz for very helpful comments.

1 Introduction

Comovement across sectors in an economy is one of the striking features of the business cycle. This feature appears in large pairwise correlation between the sectoral growth rates in terms of output, value added, and employment. Shea (2002) finds an average correlation of 0.28 using annual growth rate of gross output for 126 industries in the US over the period 1960 to 1986. Even larger correlation is found using data on employment. Other studies report similar results such as Long and Plosser (1987), Foerster, Sarte, and Watson (2011), and Carvalho (2014).

One approach to explaining intersectoral comovement relies on factor model analysis, where it is assumed that a few common shocks account for a large portion of comovement across sectors. Comparison of sectoral versus aggregate shocks have been provided by Long and Plosser (1987), Shea (2002), and Foerster et al. (2011). A common finding is that the idiosyncratic shocks to sectors have important roles in explaining overall movements, where it is shown that around 50% of overall variability in the US is attributable to the sector-specific shocks in some periods. Another strand of literature considers propagation of shocks due to intersectoral complementarities (e.g. Long and Plosser (1983); Horvath (1998); Dupor (1999)). These studies create multisector general equilibrium models, where production technology of a sector (or a firm) is linked to other sectors through input-output interactions. Complementarity creates intersectoral linkages transmitting growth disturbances to downstream and upstream sectors. Sequential shocks propagation might lead to widespread comovement, and in turn to substantial aggregate variability. A common finding among these studies is that the extent of shock propagation depends on structure of the cross-sector linkages. The sectors which have a larger number of trading partners have more important roles in spreading aggregate and idiosyncratic disturbances throughout the model. A link between these two strands of literature is provided by Foerster et al. (2011). In this paper, we view sector-specific shocks as potential sources of overall comovement, and our goal is to identify the dominant sectors whose growth disturbances act as common macroeconomic shocks.

The concept of micro shocks acting as sources of macroeconomic variability was originally established by Jovanovic (1987), and Durlauf (1993). In the presence of intersectoral complementarity, the sectors sharing linkages tend to comove. If the optimal decisions of the firms is to form linkages with a particular firm A, which might be a producer of a broadly used capital product or a general purpose material, we would expect strong comovement among all the firms with the source of comovement being attributable to the shocks to firm A. Acemoglu, Carvalho, Ozdaglar, and Tahbaz-Salehi (2012) elaborate on the role of structure of the linkages as a transmission mechanism, and argue that micro shocks would have nontrivial contributions if they come from the firms having considerably more important roles as suppliers of inputs to others. The authors provide empirical evidence for the presence of such substantial asymmetries in favor of their argument.[1] Similar evidence is presented by Carvalho (2014).

We focus on identification of the dominant sectors. We do not impose any structure on the intersectoral interactions implied for example from the input-output linkages. Instead, we view the presence of a few unknown common factors as a starting point, and build our analysis on the statistical factor model analysis. From the literature on the principal component estimator, we know that number of aggregate shocks and the space spanned by them can be consistently estimated when the number of cross-section units (N) and time series observations (T) jointly go to infinity ($N,T \to \infty$).[2] To identify the dominant sectors,

[1] Gabaix (2011) focuses on size of the firms in an economy and develops the hypothesis that idiosyncratic shocks to large firms would have nontrivial contributions to aggregate fluctuations. The author provides empirical evidence that a weighted average of idiosyncratic shocks to the top US firms explain around one-third of fluctuations in the GDP growth.

[2] The properties of the principal component estimator are discussed in detail by Stock and Watson

we use the approach from Parker and Sul (2015) (henceforth, PS). In particular, production growth of a candidate sector is treated as an observed common factor. Including the candidate variables into the factor model, we can identify the dominant sectors according to number of factors estimated in the residuals. If a candidate sector is indeed dominant, the production growth in this sector covers the space of an underlying factor, which leads to a reduction in the number of factors in the residuals. In Section 2, we provide definition of a dominant sector and review the PS approach.

Dominance of a unit (sector) as defined by PS relates to factor analysis and the concept of unknown common factors or diffuse indexes which are usually meant to explain statistical properties of panel data sets. As opposed to a statistical factor model, a model of production network incorporates the observed input-output linkages across sectors and uses the network techniques to discuss the distinguishing properties of the sectors or firms. In Section 3, we build on this literature and provide a link between dominance of a sector and its network centrality. The outdegree centrality is considered which provides a measure of sectors centrality consistent with our factor analysis. Starting from the solution of a multisector general equilibrium model (Acemoglu et al. (2012), and Holly and Petrella (2012)), we obtain a factor model representation as a reduced form of the model which features the growth in the most central sectors acting as common factors.

In factor analysis and principal component estimation, it is usually assumed that the factors are common and potentially affect all the cross-sectional units in the data set. This feature together with the assumption that there is limited dependency left among the idiosyncratic errors ensure consistency of the estimator. In Section 4, we deviate from such a common factor structure and analytically assess consistency of the PS approach under different degree of factors influence.

(1998), Bai and Ng (2002), and Bai (2003), among others.

The contribution of this section is motivated by the fact that the principal component method becomes less reliable or even inconsistent as dominance of factors relative to the idiosyncratic error reduces (Boivin and Ng (2006), and Onatski (2012)). To weaken the assumption of commonness of factors, we closely follow Kapetanios and Marcellino (2010) who use an analytically appealing way to incorporate less-than-strong factors to track the implication of different degree of factor influence on the principal component method.[3] The analysis proceeds in two steps. In the first step, it is shown that the principal component method remains consistent for each time t under a mild deviation from the case of common factors. Under this setting, the second step provides the conditions required to consistently identify the dominant units using the PS approach. A Monte Carlo study is presented in Section 5 to assess small sample properties of the identification approach following the analytical analysis.

Section 6 presents our empirical work, where the dominant sectors in the US are identified using the data set from Foerster et al. (2011). The data set contains disaggregated data on sectoral industrial production in the period covering 1972 to 2007. Two common shocks are detected among production growth rates. Our findings are as follows. We provide evidence that one of these common shocks is attributable to the shocks arising in a few sectors, including the heavy machinery industries, while the other one appears external to the model. This data set excludes some service sectors like financial industry. The latter shock could be attributable to the shocks arising in these industries or another aggregate shock, such as monetary policy affecting demand for durable goods. In addition, we show the dominant sectors mostly have important role as suppliers of capital products to others. Adopting a network perspective, this result is implied from centrality analysis of two tables of input-output linkages; in-

[3] See Chudik, Pesaran, and Tosetti (2011) for an elaboration on factor influence and definitions of strong and weak factors.

tersectoral material purchases and cross-sector capital flows. Furthermore, we show the growth rate of the sectors whose higher centrality is implied from the table of capital flows tend to have more significant relations with the other sectoral growth rates in comparison to those implied from the table of intersectoral material input-outputs.

A brief note on notation. We use 'hat' to present the least squares estimates and 'tilde' to present the principal component estimates. We use K as a generic finite number which is independent of N and T, and $C_{NT} = min\{\sqrt{N}, \sqrt{T}\}$ is an important divergence rate as $N,T \to \infty$. $||A|| = \sqrt{tr(A'A)}$ is the Frobenius norm of the matrix A. Let $\rho_1(A) > \rho_2(A) > \cdots > \rho_r(A)$ denote absolute value of the first r eigenvalues of A in descending order with $\rho_1(A)$ being the spectral radius, and $||A||_2 = \sqrt{\rho_1(A'A)}$ denotes the spectral norm of A. $a_n = O(b_n)$ states that the sequence $\{a_n\}$ is at most of order b_n, and $x_n = O_p(y_n)$ states that the random variable x_n is at most of order y_n in probability. Convergence in probability is denoted by $\xrightarrow{p}$.

2 Factor Model and Review of the PS Approach

To elaborate on the PS approach to identification of dominant sectors, we begin with defining a static approximate factor model for sectoral growth rates. Let x_{it} denote the ith sector's production grow rate, for $i = 1,...,N$ and $t = 1,...,T$. We assume the underlying data generating process is a r-factor model

$$x_t = \Lambda f_t + e_t, \tag{1}$$

where $x_t = (x_{1t},...,x_{Nt})'$, f_t is a $r \times 1$ vector of unknown factors, Λ is the $N \times r$ matrix of factor loadings, and e_t contains the idiosyncratic errors. The ith row of x_t, Λf_t, and e_t give the process of the ith sector's production growth rate as

$x_{it} = \lambda_i' f_t + e_{it}$ with $\lambda_i' f_t$ being the common component of sector i. We also use the matrix form of the model

$$X = F\Lambda' + e, \tag{2}$$

where $X = (x_1, ..., x_T)'$ is the $T \times N$ matrix of all observations in our panel data set, $F = (f_1, ..., f_T)'$, and $e = (e_1, ..., e_T)'$. Here, f_t is unknown, and potentially attributable to growth disturbances of some sectors. Our goal is to identify these sectors.

Our model assumptions are as follows.

Assumption FM1 (Factors): *$E(||f_t||^4) \leq K < \infty$, and $F'F/T \xrightarrow{p} \Sigma_f$ where Σ_f is a full rank matrix.*

Assumption FM2 (Loadings): *λ_i is deterministic such that $||\lambda_i|| \leq K < \infty$ for all i, and $\Lambda'\Lambda/N \rightarrow \Sigma_\Lambda$ where Σ_Λ is a full rank matrix.*

Assumption FM3 (Weakly dependent idiosyncratic errors): *For all N and T;* (i) *$E(e_{it}) = 0$ and $E(|e_{it}|^8) \leq K < \infty$.* (ii) *Let $E(e_{it}e_{js}) = \tau_{ij,ts}$, $|\tau_{ij,ts}| \leq \tau_{ij}$ for all (t,s), and $|\tau_{ij,ts}| \leq \tau_{ts}$ for all (i,j); $\frac{1}{N}\sum_{i,j=1}^{N} \tau_{ij} \leq K$, $\frac{1}{T}\sum_{t,s=1}^{T} \tau_{ts} \leq K$, and $\frac{1}{NT}\sum_{i,j,t,s=1} |\tau_{ij,ts}| \leq K$.* (iii) *$E|N^{-1/2}\sum_{i=1}^{N}(e_{is}e_{it} - E(e_{is}e_{it}))|^4 \leq K$ for all (t,s).*

Assumption FM4 (Weak dependence between factors and idiosyncratic errors, and moment conditions): *(i) $E\left(N^{-1}\sum_{i=1}^{N} ||T^{-1/2}\sum_{t=1}^{T} f_t e_{it}||^2\right) \leq K$ for all i.* (ii) *$E||(NT)^{-1/2}\sum_{s=1}^{T}\sum_{i=1}^{N} f_s[e_{is}e_{it} - E(e_{is}e_{it})]||^2 \leq K$ for all t.* (iii) *$E||(NT)^{-1/2} \times \sum_{s=1}^{T}\sum_{i=1}^{N} f_s\lambda_i' e_{it}||^2 \leq K$.*

Assumptions FM1-4 are similar to the assumptions of Bai (2003). According to Assumptions FM1-2, there are r common factors. Assumption FM2 ensures that Λ is full column rank, and the number of nonzero elements in each of its columns is proportional to N as indicated by convergence of $\Lambda'\Lambda/N$ to a full rank matrix. Assumptions FM3 and FM4 respectively allow for limited dependence in e_{it} across i and t, and limited dependence between the factors and the idiosyncratic errors. These assumptions let the model deviate from a classical factor model which assumes f_t and e_{it} are i.i.d. to the extent that some moment conditions which hold under a classical factor model, still hold. More particularly, the model (2) has the sample covariance structure (divided additionally by N)

$$\frac{X'X}{NT} = \frac{\Lambda F'F\Lambda'}{NT} + \frac{\Lambda F'e + e'F\Lambda}{NT} + \frac{e'e}{NT}.$$

Under Assumptions FM1-2, $\Lambda F'F\Lambda'/NT$ has always r nonzero eigenvalues. In contrast, the eigenvalues of $(\Lambda F'e + e'F\Lambda)/NT$ and $e'e/NT$ would be diminishing at a suitable rate as $N,T \to \infty$. These results hold similarly to a classical factor model despite allowing for weak dependence in e and weak dependence between F and e. Thus, the contribution of the common component in variance of x_t substantially dominates that of idiosyncratic errors when N and T are large. These features let us estimate r and F using simple approaches.

We here use the principal component method. The method minimizes the average sum of squared errors

$$V(r) = \min_{F,\Lambda} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(x_{it} - \lambda_i' f_t\right)^2, \tag{3}$$

to estimate F and Λ. Using the normalization $F'F/T = I_r$ where I_r is a $r \times r$ identity matrix, the estimate of matrix of the factors, $\tilde{F}$, is $\sqrt{T}$ times the first r eigenvectors of the $T \times T$ matrix XX'/NT, and the estimate of matrix of the loadings is $\tilde{\Lambda} = X'\tilde{F}(\tilde{F}'\tilde{F})^{-1} = X'\tilde{F}/T$. If we treat $V(k)$ as a function of k for $k = 0, \ldots, r_{max}$ such that k is the number of principal components included to obtain residuals, we can estimate r using the IC criteria from Bai and Ng (2002). In particular, r can be estimated by minimizing the following criterion

$$IC2(k) = log(V(k)) + k\left(\frac{N+T}{NT}\right) log(min\{N,T\}),$$

with the penalty term given by the second term.[4] Let $\hat{r} = argmin_{0\leq k\leq r_{max}} IC2(k)$. Then, $\hat{r}$ equals r with a probability approaching one as $N,T \to \infty$.

To describe the factors which are attributable to the production growth rates of sectors, we need to define precisely a dominant sector. Following PS, an individual unit (sector) is dominant if its variable provides an approximation for a factor. This approximation can be defined in terms of an approximate dominant leader (ADL)

Definition of Approximate Dominant Leader (ADL): *Sector i is an ADL for the true factor l,* f_{lt}, *if and only if* $f_{lt} = x_{it} + o_{lit}$, *such that* $o_{lit} = \nu_{lit}/\sqrt{T}$ *with* $Var(\nu_{lit}) = \sigma_{li} \leq K$ *for all* (l,i).

In particular, x_i is an approximation for f_l when it is not asymptotically distinguishable from it when $T \to \infty$. Including the term o_{lit} allows x_{it} to deviate from f_{jt} for some periods.

Suppose f_t can be divided in terms of a $r_1 \times 1$ vector which are attributable to ADLs, and a $r_2 \times 1$ vector of the common factors having external sources, where $r = r_1 + r_2$. Notice according to the definition of an ADL, there might be more than one x_{it} for $i = 1,\ldots,N$ which are approximations for single factor l since each variable is allowed to deviate from the true factor in a limited way. We elaborate on this issue later. For now, suppose the first sector in the panel data set is the only ADL in the model which corresponds to f_{1t} such that $f_{1t} = x_{1t} + o_{11t}$,

[4] We here focus on IC2 which shows a good small sample performance.

and f_{-1t} is the $(r-1)\times 1$ vector of the other factors. Inserting x_{1t} in (1), we obtain

$$x_t = \Lambda_1 x_{1t} + \Lambda_2 f_{-1t} + error_t, \quad (4)$$

where Λ_1 and Λ_2 are the corresponding matrices of loadings, and

$$error_t = \Lambda_1 o_{1t} + e_t$$

contains the new idiosyncratic errors. Introduction of the approximation errors does not affect the distinguishable behavior of the first and the remaining eigenvalues of $X'X/NT$ for large N and T, which is ensured by their diminishing behavior as T increases, i.e. $o_{lit} = O_p(T^{-1/2})$.

The key to identification of ADLs is that f_t, properly scaled, is consistently estimated by $\tilde{f}_t$ at each time t as shown by Bai (2003). The estimation error of the principal component method for the unobserved f_t is asymptotically negligible. Nevertheless, f_t is identified from $\tilde{f}_t$ up to scale. This can be seen by noting that f_t and λ_i are not separately identified since $(\lambda_i' M^{-1} M f_t)$ creates an equivalent model for the same observations by any invertible matrix M.

To assess whether any x_{it} is an approximation for a true factor we use the PS approach. The approach looks into x_{it}s one at a time, and includes a single variable, say x_{1t}, together with a vector of $(r-1)$ principal components into the model (4). Retrieving the least squares residuals, whether x_{1t} is an ADL or not is implied according to the number of factors in the residuals. If the variable is an ADL, there is at least one $(r-1)$ combination of the principal components which can together with x_{1t}, spans the space of common factors. Thus, zero factors will be detected using IC2.

To summarize, the ADLs can be identified from the following steps

1. Estimate r and use the principal component method to obtain $\tilde{f}_t$.

2. For a given r, consider a set of m candidate sectors with the variable set $\{x_{jt} : j \in \mathscr{I}\}$ where $\mathscr{I}$ is the set of the sectors. Do the following steps for each j.

   - Consider r vectors containing the principal components $\tilde{f}_{-kt}$, for $k = 1,...,r$, where $\tilde{f}_{-kt}$ corresponds to $\tilde{f}_t$ excluding its kth element. For each combination $\{x_{jt}, \tilde{f}_{-kt}\}$, estimate the residuals in the following regression

$$x_t = \beta_{jk} x_{jt} + \Lambda^{\star}_{jk} \tilde{f}_{-kt} + e^{\star}_{jkt}, \tag{5}$$

     where β_{jk} and $\Lambda^{\star}_{jk}$ are the coefficient matrices.

   - Select x_{jt} as an ADL if the number of factors estimated on the residuals from at least one of these r regressions is zero.

Some remarks:

Notice in Step 2 of the algorithm, we need a set of candidate sectors as a starting point. In the absence of such candidates, we can use an R^2-criterion, where the variables having relatively larger explanatory power for the principal components, would be selected as potential dominant sectors. To check if x_{jt} is a good candidate, we can estimate the following regression for each principal component

$$\tilde{f}_{kt} = a_{kj} x_{jt} + b'_{kj} \tilde{f}_{-kt} + \eta^{\star}_{kjt}, \; for \, k = 1,...,r. \tag{6}$$

If x_{jt} is indeed a good candidate, $\hat{\eta}^{\star}_{kjt}$ would diminish as $N, T \to \infty$ resulting in a relatively higher R^2. Repeating this analysis for every variable in the data set, we obtain m candidate series by selecting $\bar{m}$ variables with the highest R^2 for each principal component, where $m = r \cdot \bar{m}$.[5]

Clustering of ADLs: The concept of ADLs allows production growth of a dominant sector to deviate from a true factor for some periods. This might lead to selection of more than one ADL corresponding to the same factor. To gain a better understanding of how dominant sectors relate to each other and in turn to a true factor, we can use a clustering technique. This technique looks at all combinations of the ADLs each consisting of r growth series, and check whether they can span the space of the factors. To elaborate on it, consider a two-factor model and a case of four ADLs (in our empirical work discussed in Section 6, this is the implied setting). Consider all pairs of the selected variables (there are six pairs considering all combinations of two out of four). Including these variables as observed factors, we can put the ADLs in one or two groups according to the following regression

$$x_t = C_l \begin{bmatrix} x_{l_1 t} \\ x_{l_2 t} \end{bmatrix} + \zeta_{il}, for\, l = 1, ..., 6,$$

where C_l is a $N \times 2$ matrix of coefficients, and l_1 and l_2 are indexes of the ADLs in the lth pair. If a factor is detected in the residuals of all six regressions, it is implied that all four ADLs belong to a single group since their production growth capture only the effect of one of the underlying factors. But if in some regressions we detect a factor and in others we detect zero factors, it can be implied that the paired sectors resulting in zero factors belong to two different groups. Using this technique, we conclude in our model that four ADLs can be put in a single group as there is always a factor is residuals.

In the statistical setting described so far, the ADLs are detected with a prob-

[5]PS suggest to set m equal to 10% of N so that there are enough candidates for the analysis.

ability approaching one as $N,T \to \infty$. The next section sheds more light on dominance of a sector building on the literature on production networks.

3 Dominant Sectors in Production Networks

Centrality is a key aspect of production networks. Thinking of an economy as a network comprising N sectors with the network matrix characterized by the corresponding input-output matrix, the central sectors are the most important suppliers of inputs to others, and lie at the center of propagation of technology disturbances. In this section, we turn to a stylized structural factor model, which provides a precise definition of the common and idiosyncratic shocks, and incorporates the cross-sector dependencies which are attributable to the input-output linkages. Our goal is to provide a link between dominance of a sector as defined before, and the concept of centrality from the network perspective. To do so, we start with a structural model, and then, describe the network structure under which a factor representation similar to (4) can be obtained.

The key to capturing the production growth of the dominant sectors in the space spanned by the principal components is that there is a substantial heterogeneity among the dominant sectors and the others, which is reflected in the larger contribution of ADLs in overall variability of X. This feature can be simply incorporated into a structural factor model. Suppose a structural relationship between x_t, and a vector of technology innovations, z_t, is given as

$$x_t = \Gamma_N x_t + z_t, \tag{7}$$

where Γ_N is the $N \times N$ matrix of the cross-sector dependencies. This matrix contains exclusion restrictions for the equilibrium values of x_{it}, $i = 1,...,N$. In addition, z_t can be written as

$$z_t = B_N f_t^2 + \varepsilon_t, \tag{8}$$

where f_t^2 denotes the $r_2 \times 1$ vector of external common disturbances, and ε_t is a vector of sector-specific technology disturbances, which are independent across i with $E(\varepsilon_{it}^2) = \sigma_i^2$. The $N \times r_2$ coefficient matrix B_N captures how f_t^2 affect productivity in individual sectors. The model (7)-(8) is a general static relationship which admits the solutions of the multisector general equilibrium models of Acemoglu et al. (2012) and Holly and Petrella (2012), and furthermore, it brings in f_t^2 as common disturbances to productivity of sectors following Foerster et al. (2011).[6]

More particularly, these models characterize Γ_N as

$$\Gamma_N = A_N W_N, \tag{9}$$

where $A_N = diag(a_1, ..., a_N)$ is a diagonal matrix with $a_i \in (0,1)$ denoting a parameter of the ith sector production function, for $i = 1, ..., N$, and the $N \times N$ matrix W_N corresponds to the input-output matrix. The ijth element of W_N satisfies $w_{ij} \in [0,1]$ for all (i,j), and gives the share of the production of sector j in total intermediate inputs used by sector i. As a convention, the input shares are normalized so that they sum up to unity where $\sum_{j=1}^{N} w_{ij} = 1$ for all i.[7]

[6] Acemoglu et al. (2012) and Holly and Petrella (2012) consider a static variation of the general equilibrium model of Long and Plosser (1983). The solution of a model of this type can be generally cast is an autoregressive-moving-average in form of ARMA(1,1) depending on the timing of material input delivery and whether capital is included as discussed in detail by Foerster et al. (2011).

[7] The production function of sector i is assumed to be Cobb-Douglas with constant returns to scale

$$X_{it} = exp(z_{it})(\prod_{j=1}^{N} X_{ijt}^{w_{ij}})^{a_i} L_{it}^{1-a_i}, for\, i = 1, ..., N,$$

where X_{it} is the sector's production, X_{ijt} is the amount of the products of sector j used in the production of sector i, L_{it} is the amount of labor hired, and a_i denotes the share of total intermediate input, such that $a_i \in (0,1)$. Note $\sum_{j=1}^{N} w_{ij} = 1$ ensures that the function satisfies constant returns to scale. Holly and Petrella (2012) create their model such that their model features common factors, and ε_{it} is serially correlated and independent across i. Acemoglu et al. (2012) focus on a network assessment of the propagation of only ε_t abstracting from external common shocks, and obtain $A = aI_N$ assuming $a = a_1 = ... = a_N$ where

The matrix of cross-sector dependencies, Γ_N, specifies how the sector-specific technology disturbances, as well as, the common productivity components propagate across sectors. Looking at column i of Γ_N, every element amounts to the direct influence of production growth in sector i on another sector, and more interestingly for our analysis, the overall influence of an individual sector in the model can be measured by sum of all the elements in the column. Based on this matrix, we can define network centrality for sectors. We focus on the outdegree centrality which is simply defined as column-sums of Γ_N. Let $c_i^{out}(N) = \sum_{j=1}^{N} a_j w_{ji}$ denote the outdegree centrality of sector i. Another measure to capture centrality in production networks is the Katz-Bonacich centrality.[8] This measure relates to column-sums of R_N in a reduced form of the model (7) as

$$x_t = R_N z_t, \tag{10}$$

where $R_N = (I - \Gamma_N)^{-1}$. We here focus on the outdegree centrality based on Γ_N, as opposed to R_N which captures propagation of the productivity shocks. We do so, to remain consistent with the identification approach which looks at x_t rather than the unobserved z_t.

In the previous section, commonness of factors in (4) was defined according to the following condition in Assumption FM2

$$\frac{1}{N}\begin{bmatrix} \Lambda_1'\Lambda_1 & \Lambda_1'\Lambda_2 \\ \Lambda_2'\Lambda_1 & \Lambda_2'\Lambda_2 \end{bmatrix} \to \Sigma_\Lambda, \tag{11}$$

with Σ_Λ being a full rank matrix. In particular, this condition corresponds to the case that each column of Λ has potentially N nonzero elements to ensure the diagonal elements of $N^{-1}\Lambda'\Lambda$ remain nonzero for all N

I_N is an identity matrix.

[8] See Acemoglu, Ozdaglar, and Tahbaz-Salehi (2015).

$$N^{-1}\sum_{i=1}^{N}\lambda_{ij}^{2}\rightarrow\tau\,\text{with}\,\tau>0\,\text{for all}\,j.$$

To bring a similar structure into (7) and (8), we suppose some sectors have substantially larger outdegree centrality in comparison to the others, which is reflected in fat-tailed behavior of the the distribution $\{c_i^{out}(N), i=1,\ldots,N\}$. Let $\mathscr{I}$ be the index set of the r_1 most central sectors receiving the largest outdegree centrality, $\Gamma_{N,\mathscr{I}}$ be the $N\times r_1$ matrix containing the columns of Γ_N corresponding to these sectors, and $\Gamma_{N,-\mathscr{I}}$ be the $N\times N$ matrix whose $\mathscr{I}$ columns contain zeros and the rest of its columns are the same as those in Γ_N. To incorporate these cross-sector asymmetries, we make the following assumption on the coefficient matrices

Assumption SM (Coefficient matrices Γ and B): *For all N,* (i) *Let* $S_N=[\Gamma_{N,\mathscr{I}},B_N]$ *be the* $N\times r$ *matrix of deterministic coefficients with bounded elements.* S_N *satisfies* $S_N'S_N/N^{1-2\alpha}\rightarrow\Sigma_S$ *where* Σ_S *is a full rank matrix, and* $\alpha\in[0,1/2)$. (ii) *The spectral norm of* $\Gamma_{N,-\mathscr{I}}$ *satisfies* $||\Gamma_{N,-\mathscr{I}}||_2<1$.

The parameter α indicates how influential the factors are in the model. It generalizes the condition (11) which corresponds to $\alpha=0$, and allows the degree of influence to vary with α such that the smaller the value of α, the greater the influence of $x_{\mathscr{I},t}$ and f_t^2 on the sectoral growth rates. As $\alpha\in[0,1/2)$, the nonzero elements of S_N (in each column) increases with N, which distinguishes the central sectors from the others (peripheral sectors) whose limited dependencies are particularly modeled by $||\Gamma_{N,-\mathscr{I}}||_2<1$ in second part of the assumption.

When the economy comprises a large number of sectors, the distribution of the outdegree centrality becomes strongly fat-tailed since the centrality measure of the most central sectors is an increasing function of N, while the others always share a limited number of ties among themselves and the central sectors. Note S_N has potentially $N^{1-\alpha}$ nonzero elements in each of its columns according to Assumption SM. This structure accords with the findings of various studies in the network literature. It covers an extreme case of a star network, where there is a node at the center sharing ties with all others, while the others have neighborhood relationships only with this central node, that corresponds to $\alpha = 0$ and $\Gamma_{N,-\mathscr{I}}$ being diagonal. In a more general case, in the presence of the central sectors whose measures are $O(N^{1-\alpha})$ where $\alpha \in [0, 1/2)$, the fat-tailed behavior defined above accords with a scale free distribution (Pareto distribution) with a tail parameter lying in the interval $[1, 2)$.[9] The source of such network behavior can be attributed to an efficient outcome of a network formation model with distance-based utility functions, where the cost of forming a tie relative to its benefit is in an intermediate range.[10] Carvalho and Voigtländer (2014) use this type of cost-benefit decision making for production of new innovative products, where the producer of the new product searches for inputs among his current suppliers and the suppliers of his own suppliers. The implication of such evolving behavior is that the producers who already have important roles in supplying inputs are more prone to be selected as suppliers, and as they become closer in supply chains to an increasing number of producers, their chance to be selected by more producers grows even larger.

Consider the structural model (7)-(8) which is now written as

$$x_t = \Gamma_{N,\mathscr{I}} x_{\mathscr{I},t} + B_N f_t^2 + \Gamma_{N,-\mathscr{I}} x_t + \varepsilon_t.$$

Rearranging the terms and the multiplying both sides by $\bar{\Gamma}_N = (I_N - \Gamma_{N,-\mathscr{I}})^{-1}$, we obtain

[9]See Gabaix (2009) for a survey on the presence of Pareto distributions as a key feature of economic and finance networks, and some theoretical motivations for it.

[10]See Chapter 6 of Jackson (2010).

$$x_t = \bar{\Lambda}_{N,1} x_{\mathscr{I},t} + \bar{\Lambda}_{N,2} f_t^2 + u_t, \quad (12)$$

where $\bar{\Lambda}_{N,1} = \bar{\Gamma}_N \Gamma_{N,\mathscr{I}}$, and $\bar{\Lambda}_{N,2} = \bar{\Gamma}_N B_N$, and $u_t = \bar{\Gamma}_N \varepsilon_t$. The factor model representation (12) is obtained as a reduced form having $x_{\mathscr{I},t}$ initially separated from x_t. It is similar to the model of the previous section since on one hand, the error terms in u_t, whose cross-section dependence is governed by $\bar{\Gamma}_N$, are weakly dependent. This weak dependence is implied from Assumption SM.ii.[11] On the other hand, the particular arrangement of nonzero elements in the columns of S_N leads the $r \times 1$ vector $\xi_t = [x'_{\mathscr{I},t}, f_t^{2\prime}]'$ to have dominating role relative to u_t.

Our discussion elaborates on three aspects of our analysis. First, given that the model (7)-(8) is the true model, it is expected that the dominant sectors whose production growth act as major sources of covariation, match the central sectors whose greater outdegree centrality is implied from the corresponding input-output table. Second, this section refers to a stylized general equilibrium framework focusing on the material input-output linkages. There might be other types of intersectoral linkages such as cross-sector capital flows (Foerster et al. (2011) and Bouakez et al. (2014)), or more probably there exists a combination of them. The PS approach to identification of dominant sectors is based on the statistical factor model, which is in general less restrictive than structural model

[11]This assumption implies convergence of the power series $\bar{\Gamma}_N = I_N + \Gamma_{N,-\mathscr{I}} + \Gamma^2_{N,-\mathscr{I}} + \cdots$ for all N. Notice $||\Gamma_{N,-\mathscr{I}}||_2 < 1$ is stronger than $\rho_1(\Gamma_{N,-\mathscr{I}}) < 1$ which is made in vector autoregressive (VAR) models for covariance-stationary time series. This ensures boundedness of variance of u_t in the reduced form model (12) as $N \to \infty$. To see this, consider

$$\Sigma_u = E(\bar{\Gamma}_N \varepsilon_t \varepsilon_t' \bar{\Gamma}_N') \leq \bar{\sigma}^2 \bar{\Gamma}_N \bar{\Gamma}_N',$$

where $\bar{\sigma}^2 = max_i\{\sigma_i^2\}$. Note Σ_u has bounded eigenvalues given that

$$\rho_1(\bar{\Gamma}_N \bar{\Gamma}_N') = (||\bar{\Gamma}_N||_2)^2 \leq K < \infty,$$

since $||\Gamma_{N,-\mathscr{I}}||_2 < 1$. One can come up with examples such that $\rho_1(\Gamma_{N,-\mathscr{I}}) < 1$ and $||\Gamma_{N,-\mathscr{I}}||_2 \geq 1$ where variance of u_{it} for different i is increasing in N. See the example in Section 3 from Chudik and Pesaran (2011) in the context of infinite dimensional VARs. This way, separating $\Gamma_{N,\mathscr{I}}$ initially from Γ_N and then obtaining the reduced form of the model, we can incorporate extreme asymmetries in terms of the outdegree centrality which otherwise would not be allowed under the setting from Acemoglu et al. (2012) when $N \to \infty$.

described in this section in the sense that it does not make any particular assumption on sources of the linkages. Thus, it can be used at first. Next, some measures of relative importance of the sectors in terms of supplying material inputs, supplying capital products, or sectors size can be used to assess if there are common features among the identified sectors. Third, centrality as discussed in this section is a more general concept than dominance of a sector from the previous section, because centrality of a sector varies with $\alpha \in [0, 1/2)$, while only the case of $\alpha = 0$ was covered in the discussion on the identification approach. The next section brings the parameter α into the factor model in a consistent way, and assesses performance of the principal component method in capturing the ADLs under different values of α.

4 Identification in Less-Than-Strong Factor Models

We now consider the factor model of Section 2, and suppose the matrix of loading has a similar structure to the coefficient matrix S_N as defined by Assumption SM. More particularly, the influence of factors varies with the parameter α, such that a larger value of α indicates that a smaller part of covariability is attributable to the common component. This section intends to analytically answer two questions. Does the principal component method consistently capture the ADLs at each time t when α deviates from 0, and how far above 0 it can become in order that the estimator remains consistent? What is the range of α for which the PS approach can consistently identify the ADLs? Theorems 1 and 2 in this section answer these questions, respectively. The proofs are provided in the appendix.

We suppose Λ is the matrix of loading which satisfies Assumption FM2. To deviate from the case of $\alpha = 0$, we consider the following data generating process

$$x_t = \Lambda_N f_t + e_t, \tag{13}$$

for $t = 1, ..., T$, with the assumption that

$$\Lambda_N' \Lambda_N / N^{1-2\alpha} \to \Sigma_{\Lambda_N},$$

where Σ_{Λ_N} is a full rank matrix with $\alpha \in [0, 1/2)$. Chudik et al. (2011) define strong and semi-strong factors according to column-sums of $\Lambda_N = (\lambda_{N,il})$. Let $s_l = \sum_{i=1}^{N} |\lambda_{N,il}|$, for $l = 1, ..., r$, where $s_l = O(N^{1-\alpha})$. Then, f_t is a vector of strong factors if $\alpha = 0$, and the factors are semi-strong given that $\alpha \in (0, 1/2)$.[12] Kapetanios and Marcellino (2010) consider a less-than-strong factor model by assuming that Λ_N in (13) satisfies

$$\Lambda_N = N^{-\alpha} \Lambda. \tag{14}$$

This builds on the literature on instrumental variable estimator in the presence of so many weak instruments where an instrument appears to be weak as sample size goes to infinity, and it provides a mathematically tractable way to assess properties of the principal component method for different values of α. We here closely follow Kapetanios and Marcellino (2010) and adopt the equation (14).

The denominator N^{α} makes the contribution of the common component in $X'X/NT$ a negative function of α. Lemma A.1 in the appendix shows

$$\rho_1(X'X/NT) = O_p(N^{-2\alpha}),$$

if $\alpha \in [0, 1/2)$ and $N^{2\alpha}/T \to K < \infty$. This term is driven by $\Lambda_N f_t$ which maintains its dominating role relative to e_t when $\alpha \in [0, 1/2)$. As α increases, this

[12]Notice Chudik et al. (2011) define these rates in terms of 1-α which makes for example a strong factor correspond to $1 - \alpha = 1$.

dominance diminishes, and when α goes out of this range and equals 1/2, the relative dominance goes away. From the literature on factor analysis, we know the factor estimates and performance of the factor number estimators are affected as factors becomes less dominating relative to the idiosyncratic errors. Under this setting, it is not clear if the principal component method remains consistent for all values $\alpha \in [0, 1/2)$. Kapetanios and Marcellino (2010) establish that the average squared difference between the factors estimates and the scaled true factors diminishes if $\alpha \in [0, 1/4)$ and $N^{4\alpha}/T \to 0$. Onatski (2012) shows the principal component method is inconsistent considering the case of $\alpha = 1/2$.[13]

For the purpose of identifying the ADLs, we lay down the sufficient conditions required for consistent estimation of factors at each time t. Before doing so, we assume the number of factors, r, is known. Though, this assumption seems reasonable under $\alpha = 0$, it is expected that performance of IC2 deteriorates similarly to the principal component method for larger values of α. We here use IC2, and confine our assessment of this factor number estimator to our simulation studies where we show IC2 performs well in small samples for the range of α for which the principal component method turns out to be performing well.

The following theorem and corollary establish the time-t convergence of the principal component method for different values of α.

Theorem 1. *Let $\Lambda_N = N^{-\alpha}\Lambda$. Under Assumptions FM1-4, $\tilde{f}_t$ converges to $H'f_t$ as $N, T \to \infty$ given that $\alpha \in [0, 1/4)$ and $N^{4\alpha}/T \to 0$, such that the convergence rate is given as*

$$\tilde{f}_t - H'f_t = O_p(\frac{N^{4\alpha}}{C_{NT}^2}) + O_p(N^{\alpha-1/2}),$$

[13]Note consistency of factor estimates depend also on ratio of N and T as $N, T \to \infty$. The conditions mentioned here such as $N^{2\alpha}/T \to K < \infty$ for the case of $\alpha \in [0, 1/2)$, and $N^{4\alpha}/T \to 0$ for the case of $\alpha \in [0, 1/4)$ are ensured by the general condition; $N/T \to K < \infty$. This condition which is not stringent is directly assumed by Onatski (2012).

with $H = (\Lambda_N'\Lambda_N/N)\left(F'\tilde{F}/T\right)V_{NT}^{-1}$ *being a rotation matrix,* V_{NT} *is a diagonal matrix containing the first r eigenvalues of* XX'/NT*, and* $C_{NT}^2 = min\{N,T\}$*.*

Corollary 1. *Under the conditions of Theorem 1,*
(i) if $N^{1/2+3\alpha}/C_{NT}^2 \to 0$*, we have*

$$N^{1/2-\alpha}\left(\tilde{f}_t - H'f_t\right) = O_p(1),$$

(ii) if $N^{1/2+3\alpha}/C_{NT}^2 \to \tau > 0$*, we have*

$$\frac{C_{NT}^2}{N^{4\alpha}}\left(\tilde{f}_t - H'f_t\right) = O_p(1).$$

The theorem follows Theorem 1 from Bai (2003). It implies $\tilde{f}_t$ is a consistent estimator given that $N^{4\alpha}/C_{NT}^2 \to 0$. This condition holds under $\alpha \in [0,1/4)$ and $N^{4\alpha}/T \to 0$. Corollary 1 puts these results in terms of the Bai's theorem focusing only on the convergence rates. Bai shows

$$N^{1/2}\left(\tilde{f}_t - H'f_t\right) = O_p(1),$$

given that $N^{1/2}/T \to 0$, but as $N^{1/2}/T \to \tau > 0$, he shows

$$T\left(\tilde{f}_t - H'f_t\right) = O_p(1).$$

Under $\alpha > 0$, there are slower convergence rates such that a nonzero α leads the rates to decline from $1/N^{1/2}$ to $1/N^{1/2-\alpha}$ in (*i*), and from $1/T$ to $N^{4\alpha}/min\{N,T\}$ in (*ii*). It appears that as factors become less influential, the principal component estimator converges to the scaled f_t at a slower rate. We need $\alpha \in [0,1/4)$ to maintain consistency.

Finally, it comes to identification of the ADLs using the PS approach. To present the results, we consider a general relationship between a candidate variable, x_{jt}, and a true factor, f_{lt}, as follows

$$f_{lt} = x_{jt} + \nu_{ljt}/\sqrt{T} + \delta_{lj}\xi_{jt}, \tag{15}$$

where ν_{ljt} and ξ_{jt} are two random variables with finite variances. In this way, x_{jt} is an ADL if $\delta_{lj} = 0$. To identify dominant sectors, we follow the steps described in Section 2. Recall that for each candidate sector, the regression model (5) is estimated r times for the combinations of the principal components and the candidate sector's production growth, and next, whether the sector is a dominant sector is implied according to the number of factors in the residuals $\hat{e}^{\star}_{jkt}$, for $k = 1, ..., r$, obtained from these regressions. Let $\hat{\#}(\hat{e}^{\star}_{jkt})$ denote the corresponding number factors estimated. The following theorem summarizes the results.

Theorem 2. *Let* $\Lambda_N = N^{-\alpha}\Lambda$. *Given that* $\alpha \in [0, 1/6]$ *and* $N/T \to K < \infty$, *and under Assumptions FM1-4*
(i) if $\delta_{lj} = 0$, *we have*

$$\underset{N,T\to\infty}{Prob}\left[\hat{\#}(\hat{e}^{\star}_{j1t}) = 0, \text{or}, \hat{\#}(\hat{e}^{\star}_{j2t}) = 0, \text{or}, ...\hat{\#}(\hat{e}^{\star}_{jkt}) = 0\right] = 1,$$

(ii) if $\delta_{lj} \neq 0$, *we have*

$$\underset{N,T\to\infty}{Prob}\left[\hat{\#}(\hat{e}^{\star}_{j1t}) = 0, \text{or}, \hat{\#}(\hat{e}^{\star}_{j2t}) = 0, \text{or}, ...\hat{\#}(\hat{e}^{\star}_{jkt}) = 0\right] = 0.$$

Theorem 2 provides the sufficient conditions required to consistently identify the ADLs, and extends the results of PS for the case of α greater than zero. Putting the results from this section in terms of centrality in production networks, it is implied that the most central sectors would be detected using a large panel data set of sectoral production growth rates when their outdegree centrality are $O(N^{1-\alpha})$ for $\alpha \in [0,1/6]$. In the next section, we conduct Monte Carlo experiments to assess small sample properties of this identification approach, and we show under relatively less constraining conditions it still performs well in detecting the ADLs.

5 Monte Carlo Study

In the simulation study, we consider the following data generating process

$$x_{it} = \lambda_{N,i1} f_{1t} + \lambda_{N,i2} f_{2t} + \sqrt{\theta} e_{it}, \text{ for } i = 1, ..., N, \ and \ t = 1, ..., T,$$

where the parameter θ governs the-signal-to-noise ratio. The factors are generated such that they are correlated. In particular,

$$\begin{bmatrix} f_{1t} \\ f_{2t} \end{bmatrix} = U \begin{bmatrix} w_{1t} \\ w_{2t} \end{bmatrix},$$

where U is obtained from the Cholesky decomposition of $\Omega = Var(f_t)$ with $\Omega = [2, 0.2; 0.2, 1]$, and

$$w_{it} = \rho_i w_{it-1} + \varepsilon_{it}, \text{ for } i = 1, 2,$$

with $\varepsilon_{it} \overset{i.i.d}{\sim} N(0, 1-\rho_i^2)$.

Elements of the loading matrix $\Lambda_N = (\lambda_{N,ij})$ are drawn from a normal distribution according to

$$\lambda_{N,ij} \overset{i.i.d}{\sim} N(0, \sigma^2_{N,\alpha}),$$

with $\sigma_{N,\alpha} = 1/N^{\alpha}$. Once the loadings are generated randomly, the model is more general in comparison to the case of deterministic loadings which was assumed before. The analytical results would not be affected anyway if $\lambda_{N,ij}$ has finite fourth moment and it is independent from the factors and the idiosyncratic errors, as it is generated here.

The error term i follows

$$e_{it} = \phi e_{it-1} + v_{it} + \beta \sum_{j \neq i, j=-J}^{J} v_{i+j,t}, \tag{16}$$

with $v_{it} \overset{i.i.d}{\sim} N(0, (1-\phi)/(1+2J\beta^2))$. It is serially correlated and also cross-sectionally dependent on the J-upstream and downstream neighbors.

Two sets of experiments are conducted. In Experiment 1, we estimate number of factors under different values of α using IC2. This sheds light on performance of the estimator under different degree of factors influence. In Experiment 2, we turn to identification of ADLs. Two cases are considered. In the first case, there are two ADLs, which are generated according to the following process and are replaced by the first two units in the panel data set such that

$$x_{jt} = f_{jt} + \nu_{jt}/\sqrt{T}, \text{ for } j = 1, 2, \tag{17}$$

with $\nu_{jt} \overset{i.i.d}{\sim} N(0,1)$ for $j = 1, 2$. A case of four ADLs is also considered, where the first four variables in the data set provide approximations for the underlying factors as follows

$$x_{jt} = \begin{cases} f_{1t} + \nu_{j1t}/\sqrt{T}, \text{ for } j = 1, 2, \\ f_{2t} + \nu_{j2t}/\sqrt{T}, \text{ for } j = 3, 4, \end{cases} \tag{18}$$

where $\nu_{jlt} \overset{i.i.d}{\sim} N(0,1)$ for all (j,l). The latter case incorporates more than one ADL for each factor.

This is a general setting which closely follows the simulation study design of PS, and in addition, it incorporates different values of α to adjust the factors influence in the model. We consider $\{\alpha = 0, 1/6, 0.25, 0.3, 0.4, 0.5\}$. The values of α covers the whole range of strong to relatively weak factors. The case of $\alpha = 0$ is considered by PS to assess performance of the identification approach. The authors also consider $\alpha = 0.5$ to assess performance of the factor number estimator. In our simulation study, we extend these analyses considering different intermediate α. Among these values, 1/6 and 0.25 are relevant for the asymptotic properties explained by Theorems 1-2. Furthermore, 0.3 and 0.4 are included to provide a general picture for the intermediate values in $[0, 0.5]$. In the primary simulation studies, a more detailed incrementing of the intermediate values was used. The main findings remain the same regarding performance of the factor number estimator and the PS approach in certain ranges, also that their performance deteriorates for α outside of these ranges as it approaches 0.5. To save space, the results only for these values are reported.

Other parameters values are set as follows. For serial correlation in w_t and the error terms, we consider $\rho_i = 0.5$ for $i = 1, 2$, and $\phi = 0.5$. We assume each units has eight neighbors with $J = 4$, and $\beta = 0.1$. The signal-to-noise ratio is set equal to one, $\theta = 1$. To generate the panel data sets, we consider $\{N, T = 50, 100, 200, 500\}$. The data set is initially generated for $T + 100$ observations, and $N + 20$ units. Next, the first 100 observations are dropped from the beginning of the series, and we keep N units from the middle of cross-section units. The Monte Carlo results are obtained according to 2000 replications.

Experiment 1. Estimation of Number of Factors

Table (1) reports the frequency of estimating two factors in the left panel and the average number of factors estimated in the right panel under different α. IC2 performs well for relatively small values of α. For $\alpha \leq 1/6$, this result holds

			Frequency of Estimating Two Factors						Avg. Number of Factors Estimated					
		α	0	1/6	0.25	0.3	0.4	0.5	0	1/6	0.25	0.3	0.4	0.5
N	T													
50	50		0.63	0.85	0.53	0.21	0.01	0.00	2.51	2.08	1.60	1.10	0.25	0.04
50	100		0.46	0.86	0.83	0.51	0.03	0.00	2.80	2.14	1.95	1.57	0.51	0.05
50	200		0.35	0.89	0.93	0.72	0.03	0.00	3.03	2.11	2.03	1.78	0.74	0.04
50	500		0.29	0.96	0.99	0.88	0.03	0.00	3.24	2.05	2.00	1.89	0.85	0.03
100	50		0.67	0.91	0.61	0.22	0.01	0.00	2.41	2.08	1.69	1.11	0.18	0.03
100	100		0.96	1.00	0.81	0.23	0.00	0.00	2.04	2.00	1.81	1.20	0.10	0.00
100	200		0.95	1.00	0.99	0.66	0.00	0.00	2.05	2.00	1.99	1.66	0.37	0.00
100	500		0.93	1.00	1.00	0.95	0.00	0.00	2.07	2.00	2.00	1.95	0.76	0.00
200	50		0.81	0.96	0.58	0.12	0.00	0.00	2.21	2.03	1.60	0.93	0.05	0.01
200	100		0.99	1.00	0.88	0.18	0.00	0.00	2.01	2.00	1.88	1.16	0.03	0.00
200	200		1.00	1.00	0.99	0.39	0.00	0.00	2.00	2.00	1.99	1.39	0.03	0.00
200	500		1.00	1.00	1.00	0.97	0.00	0.00	2.00	2.00	2.00	1.97	0.44	0.00
500	50		0.97	0.99	0.27	0.01	0.00	0.00	2.03	1.99	1.21	0.44	0.00	0.00
500	100		1.00	1.00	0.70	0.02	0.00	0.00	2.00	2.00	1.70	0.87	0.00	0.00
500	200		1.00	1.00	1.00	0.16	0.00	0.00	2.00	2.00	2.00	1.16	0.00	0.00
500	500		1.00	1.00	1.00	0.89	0.00	0.00	2.00	2.00	2.00	1.89	0.00	0.00

Table 1: The frequency of detecting two factors and the average number of factors estimated are reported. IC2 is used.

specially for $N,T \geq 100$. When $\alpha = 0.25$, we observe that the frequency of detecting both of the factors approaches one, but it happens for relatively larger N and T, such as $N \geq 100$ and $T \geq 200$. The overall performance of IC2 is better under $\alpha = 1/6$ relative to the strong case of $\alpha = 0$. This is attributable to tendency of this criterion to overestimate the number of factors, $r = 2$, when N and T are relatively small, while $\hat{r}^{IC2}$ tends to decline with an increase in α. This offsets the tendency of IC2 to overestimate and leads to more frequent detection of exactly two factors for $N,T = 50$ under $\alpha = 1/6$. Furthermore, we observe a substantial reduction in the frequencies as α becomes larger. As the factors become less dominating relative to the idiosyncratic errors, $\hat{r}^{IC2}$ tends to become smaller which leads in turn to the lower frequencies. This result accords with the results of Kapetanios and Marcellino (2010) and also our analytical findings that factors can be consistently estimated for relatively small values of α.

Experiment 2. Identification of ADLs

To assess small sample properties of the identification approach, we initially assume that the ADLs are known. Considering the case of two ADLs, it means that we start the identification procedure by taking the first two variables in the data set as candidate variables. We then look into the frequency that the PS approach correctly detects them as dominant units. The left panel in Table (2) presents the frequency of detecting x_{1t}. Our results indicate a good small sample performance for $\alpha \leq 1/6$, since the ADL is identified with a frequency close to one for almost all combinations of N and T. This result accords with the sufficient conditions we obtained in our analytical findings in Theorem 2. For $\alpha = 0.25$, the frequencies slightly reduce, but they recover as N and T increase. Furthermore, very similar to the pattern we previously observed for estimation of number of factors, our results show a poor performance of the PS approach in identifying the dominant units when the data is generated using larger values of α.

In most of applications, the potential dominant units are not known, and we need to select a set of candidate variables first. We develop our simulation study incorporating four ADLs according to (18). In this case, we consider a different generating process for the idiosyncratic errors of the nondominant units. In particular, for $i = 5, \ldots, N$, e_{it} is now generated according to

$$e_{it} \overset{i.i.d}{\sim} N(0, \sigma_i^2), \tag{19}$$

where $\sigma_i^2 = \frac{1}{T}\sum_t C_{it}^2$ with $C_{it} = \lambda_{N,i1} f_{1t} + \lambda_{N,i2} f_{2t}$. Under the previous data generating process for disturbances, there is a chance that a nondominant unit acts like an ADL, but this possibility would be ruled out under (19).

Ten candidate variables are selected initially using the R^2 values obtained from regression (6), and then, the frequency of identifying all four ADLs is computed. The right panel in Table (2) reports the results. It is shown that all

		Frequency of Correctly Identifying x_{1t} (Case of Two Known ADLs)						Frequency of Identifying All Four ADLs (Case of Four Unknown ADLs)					
	α	0	1/6	0.25	0.3	0.4	0.5	0	1/6	0.25	0.3	0.4	0.5
N	T												
50	50	0.93	0.93	0.72	0.61	0.20	0.02	0.99	0.96	0.57	0.22	0.01	0.00
50	100	0.91	0.96	0.87	0.66	0.39	0.03	0.99	1.00	0.89	0.55	0.03	0.00
50	200	0.84	0.96	0.95	0.76	0.53	0.03	1.00	1.00	0.98	0.75	0.04	0.00
50	500	0.77	0.98	0.99	0.89	0.56	0.03	1.00	1.00	1.00	0.89	0.03	0.00
100	50	0.97	0.98	0.80	0.66	0.15	0.02	1.00	1.00	0.65	0.23	0.01	0.00
100	100	0.99	1.00	0.86	0.67	0.10	0.00	1.00	1.00	0.81	0.23	0.00	0.00
100	200	0.99	1.00	0.99	0.73	0.36	0.00	1.00	1.00	0.99	0.66	0.00	0.00
100	500	0.99	1.00	1.00	0.96	0.72	0.00	1.00	1.00	1.00	0.95	0.00	0.00
200	50	0.99	1.00	0.85	0.69	0.05	0.01	1.00	1.00	0.59	0.12	0.00	0.00
200	100	1.00	1.00	0.92	0.77	0.03	0.00	1.00	1.00	0.88	0.18	0.00	0.00
200	200	1.00	1.00	0.99	0.66	0.03	0.00	1.00	1.00	0.99	0.39	0.00	0.00
200	500	1.00	1.00	1.00	0.97	0.44	0.00	1.00	1.00	1.00	0.97	0.00	0.00
500	50	1.00	0.99	0.82	0.42	0.00	0.00	1.00	0.99	0.27	0.01	0.00	0.00
500	100	1.00	1.00	0.89	0.82	0.00	0.00	1.00	1.00	0.70	0.02	0.00	0.00
500	200	1.00	1.00	1.00	0.79	0.00	0.00	1.00	1.00	1.00	0.16	0.00	0.00
500	500	1.00	1.00	1.00	0.90	0.00	0.00	1.00	1.00	1.00	0.89	0.00	0.00

Table 2: The frequency of correctly identifying the ADLs is reported. First, x_{1t} and x_{2t} are incorporated as ADLs into the data set, and it is known that the first two units are potentially the dominant units. The left panel reports the frequency of x_{1t} being correctly detected. In the right panel, four ADLs are incorporated into the data set which are assumed to be unknown. The right panel reports the frequency of detecting all four ADLs, where initially the R^2-criterion is used to select the candidate variables, and the PS approach is used then to identify the ADLs among the candidate variables.

four ADLs are identified with a frequency close to one for $\alpha \leq 1/6$ for almost all combinations of N and T. For other values of α, the results are similar to our previous findings. It is worth mentioning a feature in the results corresponding to $\alpha = 0.25$, where the frequencies appear to be very different under the cases of large N and small T and the other way around. For example for N=50 and T=500, the ADLs are detected with a frequency equal to one, but in contrast for N=500 and T=50, the frequency reduces to 0.27. This goes in line with the second condition in Theorem 2, which confines divergence rate of N relative to T and suggests that, in a case of small N, a large T would help identify the ADLs as α increases and goes slightly above 1/6.

	α	0	1/6	0.25	0.3
N	*T*				
50	50	0.41	0.21	0.48	0.66
50	100	0.54	0.14	0.17	0.48
50	200	0.65	0.11	0.07	0.28
50	500	0.71	0.05	0.01	0.12
100	50	0.34	0.10	0.38	0.66
100	100	0.04	0.00	0.19	0.74
100	200	0.05	0.00	0.01	0.34
100	500	0.07	0.00	0.00	0.05
200	50	0.19	0.04	0.42	0.69
200	100	0.01	0.00	0.12	0.80
200	200	0.00	0.00	0.01	0.61
200	500	0.00	0.00	0.00	0.03
500	50	0.03	0.01	0.68	0.43
500	100	0.00	0.00	0.30	0.84
500	200	0.00	0.00	0.00	0.84
500	500	0.00	0.00	0.00	0.11

Table 3: The frequency that the nondominant units are falsely identified as ADLs is reported. Notice in this table in contrast to the previous tables, a larger frequency implies worse performance of the PS approach since the nondominant units are detected more frequently.

Last but not the least, we elaborate on the frequency that the nondominant units are falsely identified as ADLs. To do so, we consider the previous case of four ADLs and we compute how frequently any units other than first four units are selected. Table (3) reports the results for α up to 0.3 (we exclude α equal to 0.4 and 0.5, because IC2 hardly detects any factor in the first place, that in turn results in smaller frequencies for false identification). It is observed that for a small N and T, the nondominant units are selected frequently, but the performance of the PS approach recovers as N and T increase and the frequencies of false identification reduce to zero.

In sum, our results indicate good small sample performance of the PS approach for rather small values of α. Putting these results in terms of network centrality as discussed in Section 3, it means when T is large, in the model for example with $N = 200$, a dominant sector would be identified if productivity disturbances in this sector spread to at least $200^{1-1/6} \approx 83$ other sectors through the input-output linkages.

6 Dominant Sectors in the US

In this section, we identify dominant sectors in the US. We use the data set provided by Foerster et al. (2011). The data set contains monthly industrial production series according to different levels of sectoral disaggregation. It also provides the corresponding tables of intersectoral material purchases and intersectoral capital flows, where the authors distinguish between the material and capital inputs according to the period they are used in production. A product is viewed as a capital product if it is used for a period longer than one year. Our analysis is based on growth rate of quarterly industrial productions, where quarterly data are computed as average over the monthly values. We focus on three levels of disaggregation for the sectors (L3-5). This gives the combinations T=143 covering the period 1972Q2-2007Q4 and N=88, 117, and 138 respectively for L3-5.

We start with estimation of number of factors. Using IC2, we detect two factors among industrial productions at all three levels of disaggregation. IC1 provides exactly the same results. The results of the factor number estimator is robust to use of different expanding windows with ending dates matched to the last sixteen periods in the sample period. This result goes in line with the factor model analysis from Foerster et al. (2011), who show two common shocks account for a large portion of variability in aggregate industrial production. Taking the presence of two aggregate shocks in the model as a starting point, we now turn to identification of the ADLs. We first select ten candidate sectors according to the R^2-criterion (here, we select five variables for each principal component). Next, IC2 is used to determine the number of factors in the residuals obtained by regressing the sectoral growth rates on the rates of the potential dominant sectors

		L3 ($N = 88$)	L4 ($N = 117$)	L5 ($N = 138$)
1.	Commercial and Service Industrial Machinery/			
	Other General Purpose Machinery	1	1	1
2.	Metalworking Machinery	1	1	1
3.	Electrical Equipment	1	1	1
4.	Other Wood Products (under L3)	1		
	Millwork (under L4-5)		0.38	0.81

Table 4: The dominant sectors and the frequency that they are selected in different expanding windows are presented according to three levels of sectoral disaggregation (L3-5). There are 16 windows in total. Notice that the first three sectors are classified on the same name under L3-5, but ’millwork’ shows up only under L4-5. It is a subsector of ’other wood products’ under L3.

as described in Section 2. Entire of this section, we conduct our analysis on all three levels of disaggregation for the purpose of robustness check. This means identifying a sector as a dominant sector according to the fourth level disaggregation (L4), we view that sector as an ADL if one of its subsectors using the data for L5 and its supersector in L3 are also identified as dominant sectors. Furthermore, we check how sensitive the results are to different expanding windows.

Four sectors are identified as ADLs, which are presented in Table 4. The sectors ’commercial and services industrial machinery and other general purpose machinery’ (’commercial and services machinery’ for short), and ’metalworking machinery’, as well as, ’electrical equipment’ appear to be dominating in all expanding windows, while ’millwork’ (classified as ’other wood products’ under L3) is selected with a frequency lower than one (the frequencies are 0.38 and 0.81 using the growth rates corresponding to L4 and L5, respectively). Even after filtering out the first principal component (by regressing the rates on the first principal component and repeating the analysis on the retrieved residuals), we find that the first two sectors (’commercial and services machinery’, and ’metalworking machinery’) appear to be dominating in all windows.[14] The latter

[14]Here, three sectors are selected. The additional sector is ’machine shops; turned products; and screws, nuts, and bolts’.

analysis probably does not filter out the effects of one true factor, but a combination of both. However, it elaborates on dominance of two heavy machinery industries, which maintain their role as ADLs even after removing the first common component.

The dominant sectors identified according to estimation of number of factors might deviate from the true factors in some periods. This in turn might lead to identifying more than one dominant sector for a true factor. In a two-factor model, we can put the dominant sectors in one or two groups according to whether they correspond to the same factor by means of the clustering technique discussed in Section 2. In particular, we consider all possible pairs of the selected sectors. The sectors in a pair are put in different groups if they capture effects of the both underlying factors. When we conduct this clustering technique, we find that all four sectors presented in Table (4) belong to a single group. It holds irrespective of choice of the expanding windows. These results suggest one of the factors can be attributed to the idiosyncratic production growth in the identified sectors, while the other factor might have external sources. Note the data set excludes agriculture, and public, financial and service sectors. Such external sources could be attributed to shocks arising in the excluded sectors or other factors affecting the sectoral growth rates such as monetary policy.

Our analysis so far focused on the statistical properties of the production growth rates. We now look at whether there are common features among the dominant sectors. Following our discussion in Chapter 3 on dominance and centrality of a sector in a production network, we particularly consider two types of intersectoral interactions to assess the role of a sector as a supplier of products to others, namely; (a) intermediate materials, and (b) capital products. The material input-output table is provided for 1977 and 1998, and the table of capital flows is provided for 1998. To compare centrality of a sector implied from both of these tables, we use the tables for 1998 which is also closer to the middle of the sample period. We refer to them as W^{IO} and W^{CF}, respectively. We also consider; (c) size of the sectors. For comparison of size of the sectors we use their aggregation weights. The weight for sector i is computed as $s_i = \frac{1}{T}\sum_t s_{it}$, with s_{it} being its weight at time t.

To measure centrality, we use the outdegree and Katz-Bonacich centrality measures.[15] Looking at the centrality scores obtained from W^{IO} and W^{CF}, we learn that dominance of the sectors implied from our statistical analysis accords more with the patterns of the intersectoral capital flows. This is reflected in relatively higher ranking of the ADLs according to W^{CF}. Irrespective of the level of disaggregation, we observe that the sectors 'commercial and services machinery' and 'metalworking machinery' are ranked 1st and 2nd according to the outdegree centrality using the capital flows table (2nd and 1st according to the Katz-Bonacich centrality), while they show very moderate roles in terms of supplying materials to others (based on W^{IO}, 'commercial and services machinery' is ranked 43rd (41st), and 'metalworking machinery' is ranked 51st (42nd) according to the outdegree (Katz-Bonacich) centrality using L4). A similar pattern holds for 'electrical equipment' and 'millwork'. We also look at size of these sectors relative to others. It turns out that the largest sector among these four is 'commercial and services machinery' (ranked 11th), and coming after it, there appear 'electrical equipment' (36th), 'metalworking machinery' (38th), and 'millwork' (78th) out of 117 sectors. Overall, it seems that the results regarding the dominant sectors are more in line with the sectoral centrality implied from W^{CF}, rather than those implied from W^{IO} and relative size of the sectors.

To further assess the relationship between the network centrality and the effect of a sector's growth rate on the growth rates of the other sectors, we conduct additional regression analysis. In particular, we regress all the rates on the rate of

[15]Considering the interaction matrix $W = (w_{ij})$, the Katz-Bonacich centrality for j is measured as $c_j^{KB} = \lambda \sum_{i=1}^{N} w_{ij} c_i^{KB} + \eta$ with η being a base centrality which is assumed to be equal among all nodes, and λ is a parameter. We use 1 and 0.5 for η and λ, respectively.

a sector whose higher centrality is implied from W^{IO} and W^{CF}, where we use the outdegree centrality to rank the sectors. In addition, in each regression we include a constant and a cross-section average of all the rates which accounts for a common factor in the model. Looking at the number of times the coefficient of a sector's growth rate appears to be significant, we obtain a simple benchmark on how significantly its production growth affects the others. We retrieve ten most central sectors according to each table and report the frequency of their coefficients being significant according to 95% confidence level in Table (5).[16] We here report the results for 117 sectors corresponding to L4. The results are similar based on the other levels of disaggregation and also to the case that the Katz-Bonacich centrality is used for ranking.

Two immediate features emerge from this table. First, the dominant sectors which appear among the most important suppliers of capital products (presented in the upper part of the table) have the largest number of significant coefficients. These sectors appear to be the most effective sectors having a large number of significant coefficients even after controlling for a cross-section average in the regression models. The 'commercial and service machinery' and 'metalworking machinery' have significant coefficients 25% and 24.14% of the times, respectively. This number is 21.55% for 'electrical equipment'. Furthermore, the results suggest that growth in production of the major capital products suppliers tend to significantly move with a larger number of other sectors than those having important roles as suppliers of input materials. When we repeat this analysis for 30 most central sectors, we obtain a mean value of 9.48 (7.76) and a median value of 7.33 (4.74) for number of the significant coefficients using W^{CF} (W^{IO}). Foerster et al. (2011) consider a class of the multisector general equi-

[16]To avoid the multiple testing problem, the critical values are adjusted based on the Holm-Bonferroni correction. Particularly, suppose $t_{(1)},...,t_{(N)}$ are the t-ratios in descending order with their corresponding adjusted critical value denoted by $p^a_{(1)},...,p^a_{(N)}$ for the significance level a. The critical values are computed as $p^a_{(i)} = \Phi^{-1}(1-\frac{a}{2(N-i)})$, where Φ^{-1} is inverse of the normal cumulative distribution function.

Rank	Sector	Fraction of the Times the Coefficient Being Significant (%)
Capital Flow Table (W^{CF})		
1	Commercial and Service Industry Mach/Other Gen Purpose Mach *	25.00
2	Metalworking Machinery *	24.14
3	Industrial Machinery	10.34
4	Computer and Peripheral Equipment	5.17
5	Navigational/Measuring/Electromedical/Control Instruments	6.03
6	Automobiles and Light Duty Motor Vehicles	8.62
7	Construction Machinery	4.31
8	Office And Other Furniture	3.45
9	Electrical Equipment *	21.55
10	Support Activities for Mining	10.34
Input-Output Table (W^{IO})		
1	Iron and Steel Products	7.76
2	Semiconductors and Other Electronic Components	6.03
3	Plastics Products	8.62
4	Electric Power Generation, Transmission and Distribution	3.45
5	Organic Chemicals	1.72
6	Paper and Paperboard Mills	5.17
7	Oil and Gas Extraction	2.59
8	Paperboard Containers	12.93
9	Resins and Synthetic Rubber	9.48
10	Sawmills and Wood Preservation	14.66
76 based on W^{CF} 83 based on W^{IO}	Millwork*	19.83

Table 5: The central sectors and the fraction of the times their growth rates have significant coefficients are presented. For example, a frequency of 7.76 for 'iron and steel products' says 7.76% of times the coefficient of this sector's growth rate is significant in a regression model having the rate of another sector as a dependent variable. The results are according to the 95% confidence interval.

librium models whose solution can be written as $x_t = \Theta x_{t-1} + R_0\varepsilon_t + R_1\varepsilon_{t-1}$, with Θ, R_0, and R_1 depending on the tables of cross-sector linkages, as well as, other underlying structural parameters of the model. Mostly these models abstract from capital or assume full depreciation of sector-specific capital within a period. The authors incorporate the cross-sector capital flows as another interaction channel. Looking at the cross-sector comovements and aggregate variability implied from the structural models, they show that incorporation of both intersectoral material and capital flows help better explain the key features observed in the data. Though, we do not impose the identifying restrictions related to the input-output tables, our results go in line with their results. In particular, our results highlight the important role of the major capital good producers in the US production network (specially the heavy machinery industries) whose production disturbances broadly spread to others, though they might moderately contribute in the supply chains of material inputs (they are not among the major energy industries nor the major general purpose material producers like the iron, and still producers).

7 Conclusions

This paper considers the possibility that the sector-specific shocks to production growth act as sources of aggregate variability. Our analysis is based on factor model analysis of sectoral production growth rates, where we identify the dominant sectors whose growth disturbances act as common macroeconomic shocks. For the purpose of identification, we treat production growth rates of some candidate sectors as observed factors. Implication about whether a sectoral growth rate is an unknown factor is based on the number of factors estimated in the residuals. Asymptotic properties of this identification approach are discussed in the literature under a strong factor model. We consider a model containing less-that-strong factors, and analytically show the approach identifies the dominant sectors given that they are influential on relatively large portions of other sectors in the model.

We provide evidence that production growth of a few sectors in the US provides approximations for a common factor. Using the input-output tables in terms of the intersectoral material purchases and capital flows, we show that the sectors identified as dominant sectors have important roles as suppliers of capital products to others. We further show that a more central role in terms of supplying capital products explains comovement among sectoral growth rates more significantly in comparison to a more central role as supplier of material inputs. These results highlight importance of the capital flows across-sectors, next to the material input-outputs, in analyzing intersectoral complementarity.

References

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

Acemoglu, D., A. Ozdaglar, and A. Tahbaz-Salehi (2015). Networks, shocks, and systemic risk. Technical report, National Bureau of Economic Research.

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

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

Boivin, J. and S. Ng (2006). Are more data always better for factor analysis? *Journal of Econometrics 132*(1), 169 – 194. Common Features.

Bouakez, H., E. Cardia, and F. Ruge-Murcia (2014). Sectoral price rigidity and aggregate dynamics. *European Economic Review 65*, 1 – 22.

Carvalho, V. M. (2014). From micro to macro via production networks. *Journal of Economic Perspectives 28*(4), 23–48.

Carvalho, V. M. and N. Voigtländer (2014, March). Input diffusion and the evolution of production networks. Working Paper 20025, National Bureau of Economic Research.

Chudik, A. and M. H. Pesaran (2011). Infinite-dimensional {VARs} and factor models. *Journal of Econometrics 163*(1), 4 – 22. Factor Structures in Panel and Multivariate Time Series Data.

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

Dupor, B. (1999). Aggregation and irrelevance in multi-sector models. *Journal of Monetary Economics 43*(2), 391 – 409.

Durlauf, S. N. (1993). Nonergodic economic growth. *The Review of Economic Studies 60*(2), 349–366.

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

Gabaix, X. (2009). Power Laws in Economics and Finance. *Annual Review of Economics 1*(1), 255–294.

Gabaix, X. (2011). The granular origins of aggregate fluctuations. *Econometrica 79*(3), 733–772.

Holly, S. and I. Petrella (2012). Factor demand linkages, technology shocks, and the business cycle. *Review of Economics and Statistics 94*(4), 948–963.

Horvath, M. (1998). Cyclicality and sectoral linkages: Aggregate fluctuations from independent sectoral shocks. *Review of Economic Dynamics 1*(4), 781 – 808.

Jackson, M. O. (2010). *Social and economic networks*. Princeton University Press.

Jovanovic, B. (1987). Micro shocks and aggregate risk. *The Quarterly Journal of Economics 102*(2), 395–409.

Kapetanios, G. and M. Marcellino (2010). Factor-gmm estimation with large sets of possibly weak instruments. *Computational Statistics and Data Analysis 54*(11), 2655 – 2675. The Fifth Special Issue on Computational Econometrics.

Long, John B., J. and C. I. Plosser (1983). Real business cycles. *Journal of Political Economy 91*(1), pp. 39–69.

Long, J. B. and C. I. Plosser (1987). Sectoral vs. aggregate shocks in the business cycle. *The American Economic Review 77*(2), 333–336.

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

Parker, J. and D. Sul (2015). Identification of unknown common factors: Leaders and followers. *Journal of Business and Economic Statistics* (forthcoming).

Shea, J. (2002). Complementarities and comovements. *Journal of Money, Credit and Banking 34*(2), 412–433.

Stock, J. H. and M. W. Watson (1998). Diffusion indexes. Technical report, National Bureau of Economic Research.

Appendix

We first describe the limiting behavior V_{NT}. This in turn gives the limiting behavior of the rotation matrix H.

Lemma A.1. *Let $\Lambda_N = N^{-\alpha}\Lambda$. Under Assumptions FM1-4, given that $\alpha \in [0,1/2)$ and $N^{2\alpha}/T \to K < \infty$, we have $||V_{NT}|| = O_p(N^{-2\alpha})$ and $||H|| = O_p(1)$.*

Proof: We consider $X'X/NT$ since $\rho_i(XX'/NT) = \rho_i(X'X/NT)$ for $i = 1,...,r$. We have

$$||V_{NT}|| = \sqrt{\sum_{i=1}^{r} \rho_i^2(X'X/NT)} \leq \sqrt{r}\rho_1(X'X/NT).$$

Thus, it suffices to look at $\rho_1(X'X/NT)$. Consider the following equation

$$\frac{X'X}{NT} = \frac{\Lambda_N F'F\Lambda_N'}{NT} + \frac{\Lambda_N F'e + e'F\Lambda_N}{NT} + \frac{e'e}{NT}.$$

The Weyl's eigenvalue inequality for Hermitian matrices implies

$$\rho_1(\frac{X'X}{NT}) \leq \rho_1(\frac{\Lambda_N F'F\Lambda_N'}{NT}) + \rho_1(\frac{\Lambda_N F'e + e'F\Lambda_N}{NT}) + \rho_1(\frac{e'e}{NT}).$$

Bai and Ng (2002) show $\rho_1(e'e/NT) = O_p(C_{NT}^{-2})$. Consider the first term

$$\frac{1}{NT}\rho_1(\Lambda_N F'F\Lambda_N') = \frac{1}{NT}\left(||F\Lambda_N'||_2\right)^2 \leq \frac{1}{NT}\left(||F||_2 \cdot ||\Lambda_N||_2\right)^2$$

$$= \rho_1(\frac{F'F}{T})\rho_1(\frac{\Lambda_N'\Lambda_N}{N}).$$

This shows the first term is $O_p(N^{-2\alpha})$ according to Assumptions FM1-2, since $\rho_1(\Lambda_N'\Lambda_N/N) = \rho_1(\Lambda'\Lambda/N^{1-2\alpha}) = O_p(N^{-2\alpha})$.

We now turn to the second term. Using the spectral radius of the square matrix in the second term as a lower bound for the Frobenius norm, we have

$$\rho_1(\frac{\Lambda_N F'e + e'F\Lambda_N}{NT}) \leq ||\frac{\Lambda_N F'e + e'F\Lambda_N}{NT}||$$

$$\leq ||\frac{\Lambda_N F'e}{NT}|| + ||\frac{e'F\Lambda_N}{NT}||.$$

For $||\Lambda_N F'e/NT||$, we have

$$||\frac{\Lambda_N F'e}{NT}|| = \frac{1}{\sqrt{N}T}\sqrt{tr\left[e'F(\frac{\Lambda_N'\Lambda_N}{N})F'e\right]} = O_p(N^{-\alpha}T^{-\frac{1}{2}})\sqrt{\frac{1}{NT}tr(e'FF'e)}.$$

where $tr(e'FF'e)/NT = ||\sum_{t=1}^{T} f_t e_t'||^2/NT = N^{-1}\sum_{i=1}^{N}||T^{-1/2}\sum_{t=1}^{T} f_t e_{it}||^2 = O_p(1)$ according to Assumption FM4-i. Thus, we have $||\Lambda_N F'e/NT|| = O_p(N^{-\alpha}T^{-\frac{1}{2}})$. The same result holds for$||e'F\Lambda_N/NT||$. These results show

$$||V_{NT}|| = O_p(N^{-2\alpha}) + O_p(N^{-\alpha}T^{-\frac{1}{2}}) + O_p(C_{NT}^{-2}).$$

Note the term $O_p(N^{-2\alpha})$ would dominate if $\alpha \in [0,1/2)$ *and* $N^{2\alpha}/T \to K < \infty$. Furthermore, consider H as follows

$$||H|| = ||(\Lambda_N'\Lambda_N/N)(F'\tilde{F}/T)V_{NT}^{-1}||$$

$$\leq ||(\Lambda_N'\Lambda_N/N)|| \cdot ||(F'F/T)||^{\frac{1}{2}} \cdot ||(\tilde{F}'\tilde{F}/T)||^{\frac{1}{2}} \cdot ||V_{NT}^{-1}||,$$

Assumptions FM1-2, together with $\Lambda_N = \Lambda/N^{\alpha}$ and the normalization $\tilde{F}'\tilde{F}/T =$

I_r imply

$$||H|| = O_p(N^{-2\alpha})||V_{NT}^{-1}|| = O_p(1).$$

This completes the proof. *Q.E.D.*

Lemma A.1. shows the contribution of the common component in overall variability of X would dominate that from the idiosyncratic shocks under $\alpha \in [0,1/2)$ and $N^{2\alpha}/T \to K < \infty$.

Kapetanios and Marcellino (2010) showed consistency of the principal component method for $\hat{F} = X\tilde{\Lambda}_N/N^{1-2\alpha}$, where $\tilde{\Lambda}_N$ is the principal component estimate of loadings. The following lemma deducts the same result for $\tilde{F}$.

Lemma A.2. *Let $\Lambda_N = N^{-\alpha}\Lambda$. Under Assumptions FM1-4, we have*

$$\frac{C_{NT}^2}{N^{4\alpha}}\left(\frac{1}{T}\sum_{t=1}^{T}||\tilde{f}_t - H'f_t||^2\right) = O_p(1). \quad (20)$$

given that $\alpha \in [0,1/2)$ and $N^{2\alpha}/T \to K < \infty$.

Proof: Substituting for $\tilde{\Lambda}_N = X'\tilde{F}/T$ in the formula for $\hat{F}$, we obtain $\hat{F} = XX'\tilde{F}/N^{1-2\alpha}T$. Also from the principal component analysis we know $XX'\tilde{F}/NT = \tilde{F}V_{NT}$. This implies $\tilde{F} = N^{-2\alpha}\hat{F}V_{NT}^{-1}$. Thus, we can write

$$\tilde{f}_t - H'f_t = \left(N^{-2\alpha}V_{NT}^{-1}\right)\left[\hat{f}_t - (\frac{\tilde{F}'F}{T})(\frac{\Lambda_N'\Lambda_N}{N^{1-2\alpha}})f_t\right].$$

The lemma follows Lemma A.1, together with Theorem 4 from Kapetanios and Marcellino (2010), which says

$$\frac{C_{NT}^2}{N^{4\alpha}}\left(\frac{1}{T}\sum_{t=1}^{T}||\hat{f}_t - (\frac{\tilde{F}'F}{T})(\frac{\Lambda_N'\Lambda_N}{N^{1-2\alpha}})f_t||^2\right) = O_p(1).$$

Q.E.D.

We can now provide the proof for Theorem 1, which closely follows Lemma A.2 and Theorem 1 from Bai (2003).

Proof of Theorem 1. We start with the following identity

$$\tilde{f}_t - H'f_t = V_{NT}^{-1}\left(\frac{1}{T}\sum_{s=1}^{T}\tilde{f}_s\gamma_N(s,t) + \frac{1}{T}\sum_{s=1}^{T}\tilde{f}_s\zeta_{st} + \frac{1}{T}\sum_{s=1}^{T}\tilde{f}_s\eta_{st} + \frac{1}{T}\sum_{s=1}^{T}\tilde{f}_s\xi_{st}\right) \tag{21}$$

where

$\gamma_N(s,t) = E(e_s'e_t/N)$,

$\zeta_{st} = e_s'e_t/N - \gamma_N(s,t)$,

$\eta_{st} = f_s'\Lambda_N'e_t/N$,

$\xi_{st} = f_t'\Lambda_N'e_s/N$.

Equation (21) can be written as

$$\tilde{f}_t - H'f_t = V_{NT}^{-1}\left[\left(\frac{1}{T}\sum_{s=1}^{T}(\tilde{f}_s - H'f_s)\gamma_N(s,t) + \frac{H'}{T}\sum_{s=1}^{T}f_s\gamma_N(s,t)\right)\right.$$

$$+\left(\frac{1}{T}\sum_{s=1}^{T}(\tilde{f}_s - H'f_s)\zeta_{st} + \frac{H'}{T}\sum_{s=1}^{T}f_s\zeta_{st}\right) + \left(\frac{1}{T}\sum_{s=1}^{T}(\tilde{f}_s - H'f_s)\eta_{st} + \frac{H'}{T}\sum_{s=1}^{T}f_s\eta_{st}\right)$$

$$\left.+\left(\frac{1}{T}\sum_{s=1}^{T}(\tilde{f}_s - H'f_s)\xi_{st} + \frac{H'}{T}\sum_{s=1}^{T}f_s\xi_{st}\right)\right]$$

$$= V_{NT}^{-1}[I + II + III + IIII + V + VI + VII + VIII].$$

Notice the main difference here in comparison to the setting from Bai (2003) is that $1/T\sum_{s=1}^{T}||\tilde{f}_s - H'f_s||^2$ is $O_p(N^{4\alpha}C_{NT}^{-2})$ instead of $O_p(C_{NT}^{-2})$. In addition, η_{st} and ξ_{st} contain $\Lambda_N = \Lambda/N^{\alpha}$ as opposed to Λ. From Lemma A.1, we know $||H|| = O_p(1)$, which implies the limiting behavior of *II* and *IIII* do not change here relative to those presented by Bai (2003). Thus, we have $II = O_p(1/T)$ and $IIII = O_p(1/\sqrt{NT})$.

Now, we continue by looking into *I*. We have

$$||\frac{1}{T}\sum_{s=1}^{T}(\tilde{f}_s - H'f_s)\gamma_N(s,t)|| \leq \left(\frac{1}{T}\sum_{s=1}^{T}||\tilde{f}_s - H'f_s||^2\right)^{\frac{1}{2}}\left[\frac{1}{\sqrt{T}}\left(\sum_{s=1}^{T}\gamma_N^2(s,t)\right)^{\frac{1}{2}}\right].$$

Assumption FM3-ii implies that $\left(\sum_{s=1}^{T}\gamma_N^2(s,t)\right)^{\frac{1}{2}}$ is bounded from above. Hence, $I = O_p(T^{-1/2}N^{2\alpha}C_{NT}^{-1})$ following Lemma A.2. For *III*, in a similar way we can write

$$||\frac{1}{T}\sum_{s=1}^{T}(\tilde{f}_s - H'f_s)\zeta_{st}|| \leq \left(\frac{1}{T}\sum_{s=1}^{T}||\tilde{f}_s - H'f_s||^2\right)^{\frac{1}{2}}\left(\frac{1}{T}\sum_{s=1}^{T}\zeta_{st}^2\right)^{\frac{1}{2}}$$

$$= O_p(N^{2\alpha}C_{NT}^{-1})\left(\frac{1}{NT}\sum_{s=1}^{T}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(e_{is}e_{it} - E(e_{is}e_{it}))\right]^2\right)^{\frac{1}{2}}.$$

We have $N^{-1/2}\sum_{i=1}^{N}(e_{is}e_{it} - E(e_{is}e_{it})) = O_p(1)$ from Assumption FM3-iii, and in turn it is implied that $III = O_p(N^{2\alpha-1/2}C_{NT}^{-1})$.

Looking at *VI* and *VIII*, we see their limiting behavior changes by a factor $N^{-\alpha}$ in comparison to the case that Λ is the loading matrix. Using the rates obtained by Bai (2003) which are $O_p(N^{-1/2})$ and $O_p(N^{-1/2}T^{-1/2})$ respectively, it follows that $VI = O_p(N^{-\alpha-1/2})$ and $VIII = O_p(N^{-\alpha-1/2}T^{-1/2})$.

We now consider V and VII. Bai (2003) shows they both are $O_p(N^{-1/2}C_{NT}^{-1})$. Following our argument about the changes in the rates due to the different convergence rate for $\tilde{F}$ ($N^{2\alpha}C_{NT}^{-1}$ replaces C_{NT}^{-1}), as well as, use of Λ/N^{α} which brings in a factor of $N^{-\alpha}$, we imply that $V\&VII = O_p(N^{-1/2}(N^{2\alpha}C_{NT}^{-1})N^{-\alpha}) = O_P(N^{\alpha-1/2}C_{NT}^{-1})$.

Using $V_{NT}^{-1} = O_p(N^{2\alpha})$ from Lemma A.1, we finally obtain

$$\tilde{f}_t - H'f_t = O_p(\frac{N^{4\alpha}}{C_{NT}^2}) + O_p(N^{\alpha-1/2}). \tag{22}$$

which completes the proof. *Q.E.D.*

The points (i) and (ii) in Corollary 1 are just implied from (22). Having the results from Theorem 1, we can finally provide the proof for Theorem 2.

Proof of Theorem 2. Consider the least squares residuals using $\tilde{f}_t$

$$\hat{e}_{it} = e_{it} - \left(\hat{\lambda}'_{N,i} - \lambda'_{N,i}H'^{-1}\right)\tilde{f}_t - \lambda'_{N,i}H'^{-1}\left(\tilde{f}_t - H'f_t\right).$$

From the properties of the least squares method, we know $\hat{\lambda}'_{N,i} - \lambda'_{N,i}H'^{-1} = O_p(T^{-1/2})$. Furthermore, Lemma A.1 and Theorem 1 imply

$$\begin{aligned} \lambda'_{N,i}H'^{-1}\left(\tilde{f}_t - H'f_t\right) &= O_P(N^{-\alpha})\left[O_p(\frac{N^{4\alpha}}{C_{NT}^2}) + O_p(N^{\alpha-1/2})\right] \\ &= O_p(\frac{N^{3\alpha}}{C_{NT}^2}) + O_p(N^{-1/2}). \end{aligned}$$

Thus, we obtain

$$\hat{e}_{it} - e_{it} = O_p(T^{-1/2}) + O_p(\frac{N^{3\alpha}}{C_{NT}^2}) + O_p(N^{-1/2}).$$

PS show if $\hat{e}_{it} - e_{it} = O_p(C_{NT}^{-1})$, the probability of detecting zero factors in $\hat{e}_t$ approaches one using the criteria of Bai and Ng (2002) when $N, T \to \infty$. To ensure this condition, we need $\alpha \in [0, 1/6]$, as well as, some constraint on limiting behavior of N relative to T. In particular, it is required that $N/T \to K < \infty$. Note this condition ensures $N^{1-\varepsilon}/T \to K < \infty$ for $\varepsilon > 0$ required for consistency of the principal component method in Lemma A.2.

These results ensure the argument of PS holds even if factors are not strong as long as α, N, and T satisfy the conditions mentioned. Then, the theorem simply follows Theorem 1 of PS, that completes the proof. *Q.E.D.*