The dynamic factor network model with an application to international trade

Bräuning, Falk; Koopman, Siem Jan

published in
Journal of Econometrics
2020

DOI (link to publisher)
10.1016/j.jeconom.2019.10.007

document version
Publisher's PDF, also known as Version of record

citation for published version (APA)
Bräuning, F., & Koopman, S. J. (2020). The dynamic factor network model with an application to international trade. *Journal of Econometrics*, *216*(2), 494-515. https://doi.org/10.1016/j.jeconom.2019.10.007

Journal of Econometrics 216 (2020) 494–515

Contents lists available at ScienceDirect

Journal of Econometrics

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

The dynamic factor network model with an application to international trade☆

Falk Bräuning [a], Siem Jan Koopman [b,c,*]

[a] Federal Reserve Bank of Boston, United States of America
[b] Vrije Universiteit Amsterdam, Tinbergen Institute, The Netherlands
[c] CREATES, Aarhus University, Denmark

ARTICLE INFO

Article history:
Received 26 November 2016
Received in revised form 19 August 2019
Accepted 14 October 2019
Available online 30 November 2019

JEL classification:
C32
C58
F14

Keywords:
Network analysis
Dynamic factor models
Blockmodels
International trade

ABSTRACT

We introduce a dynamic network model with probabilistic link functions that depend on stochastically time-varying parameters. We adopt a blockmodel framework and allow the high-dimensional vector of link probabilities to be a function of a low-dimensional set of dynamic factors. The resulting dynamic factor network model has a basic and transparent structure. However, parameter estimation, signal extraction of stochastic loadings and dynamic factors, and the econometric analysis generally are intricate tasks for which simulation-based methods are needed. We provide feasible and practical solutions to these challenging tasks, based on a computationally efficient importance sampling procedure to evaluate the likelihood function. An extensive Monte Carlo study demonstrates the performance of our method in finite samples, both under correct and incorrect model specifications. In an empirical study, we use the novel framework to analyze global patterns of banana exports and imports. We identify groups of heavy and light traders in this highly active commodity market and their time-varying trade probabilities.

1. Introduction

Networks arise naturally in many fields of economics given that the interactions of economic agents, including households, firms, financial institutions, and governments, are generally at the core of economic activity. It is therefore not surprising that economists have in recent years become increasingly interested in analyzing economic networks, with applications in finance, macroeconomics, international trade, and other fields.[1] In addition to the academic interest, the

☆ We are grateful to Yacine Ait-Sahalia (the editor) and two anonymous referees for excellent comments and suggestions. We thank Kovid Puria and Botao Wu for valuable research assistance. S.J. Koopman acknowledges support from CREATES, the Center for Research in Econometric Analysis of Time Series (DNRF78) at Aarhus University, Denmark, funded by the Danish National Research Foundation. The views expressed in this paper do not necessarily reflect those of the Federal Reserve Bank of Boston or the Federal Reserve System.

* Corresponding author.
E-mail addresses: falk.braeuning@bos.frb.org (F. Bräuning), s.j.koopman@vu.nl (S.J. Koopman).

[1] For example, in macroeconomics, business cycle fluctuations can be viewed as the result of a propagation of macroeconomic shocks through input–output and trading networks; see the influential work of Acemoglu et al. (2012, 2015) with evidence that network-based propagation has quantifiable implications for the macroeconomy. Chaney (2014) argues that the dynamics in international trade relationships and their frictions rely mostly on the international networks of exporters from which new trading partners originate. Hence, the dynamic modeling of the networks of exporters is of key importance for analyzing international trade fluctuations. In finance, systemic risk monitoring crucially relies on the analysis of funding flows of large nationwide banks and smaller periphery banks, and the financial spillover effects of sudden changes in liquidity can be effectively measured by models with dynamic network structures; see Blasques et al. (2018) and the references therein.

https://doi.org/10.1016/j.jeconom.2019.10.007
0304-4076/

policy relevance of economic networks is widespread; illustrations include the detection of international cartel trading (which is clearly an obstacle to free trade), diagnosis of irregularities or changes in financial networks (which is of key interest to central banks), and monitoring of movements in trading blocs (which is relevant for international trade policy). An overview of the early literature on network modeling in economics and its applications is provided by Jackson (2008). However, despite the growing research on economic networks, the econometric tools necessary for the empirical analysis of networks, in particular, for the analysis of dynamic networks, are still in their infancy; see also the discussions on empirical methods to analyze networks in Kolaczyk (2009).

We propose a dynamic factor network model to analyze time-varying heterogeneity in a binary link formation. Our model provides an econometric tool to applied researchers to cluster nodes into structurally equivalent groups and estimate both within- and between-group dynamic link probabilities, a task inherent to many relevant empirical applications, including the detection of cartel trading or systemic risk monitoring. The model formulation for the binary observation density is based on a signal that separates the cross-section from the time dimension. We incorporate dynamics into the model using a dynamic factor approach with a parametric law of motion for the latent factors that are specified as stochastic functions of time. We take into account the network structure by modeling the link-specific factor loadings as random vectors that arise from a stochastic blockmodel structure based on latent node-level variables. The associated link probabilities of the network are modeled as functions of latent dynamic stochastic variables. We show how to include observed covariates into our latent variable structures which can be a key feature in many empirical applications, for example, for testing economic hypotheses. Our proposed framework is flexible, but also parsimonious, and it allows for the modeling of relevant features in network data, such as the potential for clustering, the tendency of link formation among similar nodes (homophily notion), and the phenomenon that similar nodes may relate to other nodes in the same way (equivalence notion). Network structures can be subject to endogenous and interaction effects which can lead to the usual challenges in the identification of latent variables and parameters, including those related to the so-called reflection problem as formulated by Manski (1993). Our network modeling framework allows for hierarchical structures and can implicitly incorporate information on the composition of network groups or the network structure. Hence, these identification issues can be addressed as part of our flexible modeling framework and estimation procedure.

While our proposed dynamic factor model is straightforward and transparent by nature, parameter estimation, signal extraction of the dynamic factors and the stochastic loadings, and the econometric analysis generally are intricate tasks for which simulation-based methods are needed. We show that the model parameters can be estimated by simulated maximum likelihood estimation, using our computationally efficient importance sampling procedure which integrates out all random effects – in the time, node, and link dimension – from the joint distribution of the observed variables. For this purpose, we propose an importance sampler that disentangles the integration in the cross-section (nodes and links) and time dimension. In particular, we adopt an importance sampler based on an approximating linear Gaussian state space model for the latent factors. We specifically use our model structure to collapse the observation density into a tractable binomial approximating model without losing relevant information on the dynamic factors, similar to the collapsing of panel data in Jungbacker and Koopman (2015). For integration of the latent node- and link-specific random variables that arise from the stochastic block structure of the factor loadings, we devise an importance sampler based on a variational approximation (mean field approximation) to the correct conditional distribution.

We discuss theoretical properties of the estimator and carry out a Monte Carlo study to investigate the finite-sample properties of our novel importance sampling procedure. The results suggest that the estimators are well-centered around their associated true parameter values, even for networks with a small number of nodes, say 25 nodes. The results also suggest that the efficiency of the estimator increases in both the time and the cross-section dimensions. Moreover, the results show large computational gains of our proposed binomial collapsing, with computational gains growing fast with the size of the network. For example, for a moderate network with 100 nodes and 250 time periods, the construction of the importance sample with our proposed methods is 10 times faster than using the collapsing method of Jungbacker and Koopman (2015). An additional simulation study shows that our proposed estimation method performs well, as measured by relatively low mean square errors and likelihood losses, under a range of misspecifications for both cross-sectional and time series features of the model. This simulation study includes misspecifications for factor dynamics (in particular, structural breaks), distributional assumptions, and number of latent groups.

In an empirical application, we show how our proposed model can be used to investigate the global trade linkages within the banana import/export market. Studying international trade between 90 countries (8,010 distinct country pairs) from 1962–2014, we uncover different groups of country pairs and their associated trade dynamics. Our latent variable estimation results suggest that the international banana trade network consists of a few countries at the core of the network (hubs) which are heavily importing and exporting countries. Specifically, our analysis suggests that the core typically consists of larger countries and of countries close to the sea, with the Netherlands and U.S. as examples. Also producer countries that supply bananas to these hubs, such as Ecuador, are part of the core. However, most country-pairs do not trade directly with each other but trade via import–export hubs, which have become more important since the 1980s. We empirically find that international trade can be viewed as an approximate core–periphery structure and confirm earlier studies such as Snyder and Kick (1979). Our estimated link probabilities also reveal a drop in global banana trade after the great financial crisis in 2008/09. These empirical findings are confirmed by our model with additional latent groups. However, this extended model also reveals other dynamic features in the data. For example, some country pairs, such as Kenya and Germany, show a decline in their trade probabilities from the late 1980s.

Literature review

Our proposed dynamic factor network model is embedded in a rich literature on blockmodels, random effects, and mixture models. The blockmodel can be regarded as the current workhorse in the statistical analysis of time-invariant networks; see, among others, Wasserman and Faust (1994) and Nowicki and Snijders (2001). For more recent developments on blockmodels, we refer to Xing et al. (2008), Airoldi et al. (2008), and, for an application to economics, Craig and von Peter (2014). We contribute to this literature by introducing a time-varying link density in a stochastic blockmodel using dynamic latent factors. Such dynamics have been introduced in network models with random effects by Hoff et al. (2002) and Hoff (2009). Our model belongs to this class of models since it is assumed that the link probabilities are generated as functions of latent node-specific random vectors. For example, the function can rely on an arbitrary distance in a similar way, as it is used for random cross-section effects in panel models; see also Mesters and Koopman (2014). Since our model relies on discrete factor loadings, it is also related to models of grouped patterns of heterogeneity in panel data; see, for example, Bonhomme and Manresa (2015). More broadly, we can relate our model with its challenges in parameter estimation to the literature on mixture state space models; see Frühwirth-Schnatter (2006) for a textbook treatment. Alternative econometric treatments for dynamic networks can also be considered. For example, Aït-Sahalia et al. (2015) develop a dynamic model based on Hawkes self-exciting jump processes which is parsimonious and allows for contagion dynamics.

At the core of our estimation procedure is the method of importance sampling which is discussed in detail by Ripley (1987) and is explored for non-Gaussian, nonlinear, state space models by Shephard and Pitt (1997) and Durbin and Koopman (1997). Modifications, further developments, and applications in the context of state space time series models are presented by Richard and Zhang (2007), Jung et al. (2011), and Koopman et al. (2014). In these contributions, the main focus is on univariate and multivariate time series applications. For cross-section models with non-Gaussian and nonlinear features, importance sampling methods have been developed by Hajivassiliou (1990), Geweke (1991), Keane (1994), and Hajivassiliou et al. (1996). The simultaneous analysis of cross-section and time series dimensions by importance sampling methods has been treated by Liesenfeld and Richard (2010) and Mesters and Koopman (2014). Our method is closely linked with this branch of literature, but we adopt and modify these simulation-based methods in the novel context of network blockmodels combined with dynamic factor structures.

The treatment of our dynamic network model requires the handling of high-dimensional integrals, and we therefore must rely on computationally efficient importance sampling methods. For this purpose, we show that our dynamic binary network data can be collapsed into a low-dimensional binomial vector series, which is used to sample the time-varying effects from the importance densities. In particular, we collapse the cross-sectional dimension of the network data without compromising any information that is needed to sample the time-varying effects. Such transformations have been introduced by Jungbacker and Koopman (2015) and are used to simulate random effects in ordinary panel data models by Mesters and Koopman (2014) in order to achieve large computational savings when evaluating the Monte Carlo likelihood. Due to our specific model structure, we are able to improve upon these methods considerably by collapsing the cross-sectional dimension using transformations that do not involve any costly matrix inversion or decomposition method. Indeed, even for networks with a small number of nodes, the number of bilateral relationships that define the cross-sectional dimension is large. In such cases, the collapsing method of Jungbacker and Koopman (2015) is prohibitively costly. We also show how variational approximations can be built used as approximating model within importance sampling procedures.

A Bayesian approach to estimation in the context of our dynamic factor network model can also be taken. In particular, the Markov chain Monte Carlo method can be explored in the context of Bayesian analyses for network models and dynamic factor models; see Hoff et al. (2002) and Hoff (2009) and Sewell and Chen (2015) for application to network models, Lopes et al. (2008) for an application to models with spatial dependence, and Aguilar and West (2000) for dynamic factor models. The sequential Monte Carlo methods as reviewed by Creal (2012) may also be used in a Bayesian time series analysis. In our current study, we only explore modifications of importance sampling and Monte Carlo maximum likelihood methods for the dynamic factor network model. These methods may also be useful as part of a Bayesian analysis.

The remainder of our study is organized as follows. Section 2 formally describes the dynamic factor network model in detail. In Section 3, we develop our simulated maximum likelihood method for parameter estimation. Section 4 evaluates the performance of our estimation method in a simulation study. In Section 5, we apply our model to estimate the international banana trade network. Section 6 concludes.

2. Model specification

A dynamic network is described by a set of *nodes* $\mathcal{N}$ that are connected by a potentially time-varying set of *links* $\mathcal{L}_t$.[2] The existence of a link between nodes i and j, $i, j \in \mathcal{N} = \{1, 2, \ldots, N\}$, where N is the number of nodes, at discrete time periods $t = 1, \ldots, T$ is denoted by the binary variable $y_{i,j,t} \in \{0, 1\}$ that equals one if the link exists, and zero otherwise. In general, networks may be *directed* or *undirected*. For *directed* networks, a link from (sender) node i to (receiver) node j at time t does not imply a reciprocal link from node j to node i at time t, that is $y_{i,j,t} \neq y_{j,i,t}$. For *undirected* networks,

[2] In graph theory, nodes are also referred to as vertexes, and links are referred to as edges.

the direction of the relationship does not matter and we have $y_{i,j,t} = y_{j,i,t}$.[3] The dynamic network can be represented by a sequence of adjacency matrices $\{Y_t\}_{t=1}^T$, where each $N \times N$ adjacency matrix Y_t collects the link variables $y_{i,j,t}$; it has its (i, j) element equal to $y_{i,j,t}$ and has zeros on the main diagonal. The adjacency matrix Y_t provides a full representation of the network at time t. In our analysis, we focus on a vectorization of the adjacency matrix Y_t and define $y_t = vec(Y_t)$, where $vec(\cdot)$ is the vectorization operator that stacks the $N \times 1$ columns of Y_t on top of one another. The dimension of y_t is then $N^2 \times 1$.

In many economic settings, the interest lies in modeling directional links, and, thus, we focus on the treatment of directed networks in the main part of this paper. For example, in our empirical application in international trade a link between country i and j represents a trade flow from exporting country i to importing country j. However, we highlight that our framework is flexible enough to easily accommodate, with only minor adjustments, undirected networks as well as bipartite networks, i.e., networks where the set of nodes can be portioned into two disjoint sets of sender and receiver nodes (for example, banks and firms, firms and locations). Similarly, our framework allows for the treatment of missing data and dynamic networks with a time-varying number of nodes. We discuss these model extensions in the Supplementary Appendix. In general, we model each link between any two nodes i and j in the network at time t by the binary random variable

$$y_{i,j,t} \sim \mathcal{B}(\phi_{i,j,t}), \qquad \phi_{i,j,t} = \frac{\exp(\theta_{i,j,t})}{1 + \exp(\theta_{i,j,t})}, \tag{1}$$

where $\mathcal{B}(\cdot)$ is the Bernoulli distribution and the probability of success $\phi_{i,j,t} \in (0, 1)$ which is formulated in terms of the logistic transformation of the signal $\theta_{i,j,t} \in \mathbb{R}$. A key part of our model is the specification for the time-varying link-specific signal $\theta_{i,j,t}$ that takes the multiplicative form given by

$$\theta_{i,j,t} = z_{i,j}' f_t = \sum_{m=1}^{M} z_{i,j,m} f_{t,m}, \tag{2}$$

where $f_t \in \mathbb{R}^M$ is a vector of common dynamic factors, that is $f_t = (f_{t,1}, \ldots, f_{t,M})'$, and $z_{i,j}$ is a vector of pair-specific factor loadings, that is $z_{i,j} \in \mathbb{Z}^M := \{\{0, 1\}^M : \sum_{m=1}^M z_{i,j,m} = 1\}$, for dimension $M \geq 1$, and with $z_{i,j,m}$ as the mth element of column vector $z_{i,j}$, for $m = 1, \ldots, M$. Hence $z_{i,j}$ selects a specific element of f_t. The signal specification for $\theta_{i,j,t}$ in Eq. (2) separates the cross-section ($z_{i,j}$) and the time series (f_t) features of our model. The $M \times 1$ vectors $z_{i,j}$ and f_t are assumed to be independent random vectors that come from certain parametric distributions.[4]

2.1. Cross-sectional features of the model

We incorporate the cross-sectional network dependencies by means of the pair-specific loading vector $z_{i,j}$ and by adapting a stochastic blockmodel structure that has become popular in sociological and statistical network analysis; see, for example, Wasserman and Faust (1994) and Nowicki and Snijders (2001). In particular, we assume that $z_{i,j}$ is a M-dimensional multinomial random vector that we can express by the Kronecker tensor product as given by

$$z_{i,j} = z_{i\to j} \otimes z_{j\leftarrow i}, \tag{3}$$

where the $K \times 1$ vector $z_{i\to j} \in \mathbb{Z}^K := \{\{0, 1\}^K : \sum_{k=1}^K z_{i\to j,k} = 1\}$ is an indicator vector that determines the latent group membership of sender node i in the directional interaction with receiver node j, where $z_{i\to j,k}$ is the kth element of $z_{i\to j}$, for $k = 1, \ldots, K$. Similarly, the $K \times 1$ vector $z_{j\leftarrow i} \in \mathbb{Z}^K := \{\{0, 1\}^K : \sum_{k=1}^K z_{j\leftarrow i,k} = 1\}$ is an indicator vector that determines the latent group membership of receiver node j in the directional interaction with sender node i, where $z_{j\leftarrow i,k}$ is the kth element of $z_{j\leftarrow i}$, for $k = 1, \ldots, K$. By construction, we have $M = K^2$. Since each node can be in one of the K groups in each interaction, each pair of nodes can be in $M = K^2$ groups. Moreover, as we are modeling directional interactions, we have $z_{j\leftarrow i} \neq z_{i\leftarrow j}$ and $z_{i\leftarrow k} \neq z_{i\leftarrow l}$, for all $i, j, k, l = 1, \ldots, K$.

We model the link-specific loadings $z_{i,j}$ as a function of two node-specific vectors. In particular, for each node i in the network, the latent pair-specific group membership vectors are modeled by the hierarchical structure

$$z_{i\to j}, z_{i\leftarrow j} \overset{iid}{\sim} \mathcal{MN}(\pi_i), \quad \text{for } j = 1, \ldots, K, \tag{4}$$

where $\mathcal{MN}(\cdot)$ indicates the multinomial distribution with $K \times 1$ multinomial probability vector $\pi_i \in (0, 1)^K$. We treat the probability vector π_i as a random variable generated by the vector model specification with the element-wise logistic function as given by

$$\pi_i = \frac{\exp(\gamma_i)}{\sum_{k=1}^K \exp(\gamma_{i,k})}, \qquad \gamma_i = B \cdot X_i + \eta_i, \qquad \eta_i \overset{iid}{\sim} \mathcal{N}(0, \Sigma_\gamma), \tag{5}$$

[3] In our treatment, the first index relates to the sender node, the second to the receiver node. This distinction is only relevant for directed links. For notational convenience, but without loss of generality, we abstract from self-loops in this paper and we set $y_{i,i,t} = 0 \,\forall i \in \mathcal{N}$.

[4] Such a separation in cross-sectional and time effects is discussed earlier in the context of Bayesian spatial dynamic factor models; see, for example, Lopes et al. (2008) and references therein.

where $\gamma_i \in \mathbb{R}^K$ is the vector of auxiliary coefficient variables $(\gamma_{i,1}, \ldots, \gamma_{i,K})'$ and it is modeled as the node-specific regression model with $K \times K_B$ coefficient matrix B containing the regression parameters for the $K_B \times 1$ vector of explanatory variables X_i consisting of observed covariates, and η_i is a normally distributed disturbance vector with zero mean and $K \times K$ variance matrix Σ_γ. It follows that each element in μ_i is modeled as a linear transformation of the same set of observables. The inclusions of node-specific covariates or exclusive restrictions can be accomplished easily by imposing zeros in matrix B. Hence the model can be used to formally test hypotheses of economic relevance. Furthermore we emphasize that each regressor can only have $K-1$ degrees of freedom because Eq. (5) implies that $\sum_{k=1}^{K} \pi_i = 1$. We do not make this explicit further for ease of notation. The model structure implied by Eq. (3) can be directly expressed as $z_{i,j} \sim \mathcal{MN}(\delta_{i,j})$ with probability vector $\delta_{i,j} = \pi_i \otimes \pi_j$.

The inclusion of regression effects for the random vector γ_i in our hierarchical model for the cross-sectional correlations can be an important ingredient for the economic understanding of the network structures and its dynamic evolutions. In an empirical setting, the econometrician may, for example, want to link the latent group membership of each node in the network to observed node-specific information and enables the measurement of the impact of economic and financial policies on the network structure. For example, in our empirical study, where we uncover the underlying dynamics of international trade patterns, it may be desirable to incorporate country-level information, such as geographic location, climate conditions or country size, in the econometric methodology to identify the membership of a country to some latent group in the network.[5]

2.2. Time series features of the model

To incorporate time-varying link probabilities in the network model, we use a dynamic latent factor approach. We assume that the $M \times 1$ vector of common factors $f_t \in \mathbb{R}^M$ that governs the dynamics of the link probabilities evolves according to a standard linear Gaussian state transition

$$f_t = C \cdot W_t + u_t, \qquad u_{t+1} = \Phi u_t + \xi_t, \qquad \xi_t \overset{iid}{\sim} \mathcal{N}(0, \Sigma_\xi), \tag{6}$$

for $t = 1, \ldots, T$, where the $M \times M_C$ matrix C consists of regression coefficients and the $M_C \times 1$ vector W_t contains time series explanatory variables, Φ is the $M \times M$ autoregressive coefficient matrix and ξ_t is the $M \times 1$ disturbance vector with mean zero and variance matrix Σ_ξ. The dynamic process for u_t is effectively a vector autoregressive process of order 1 as denoted by VAR(1). The initial distribution for u_1 is assumed to be Gaussian with mean zero and variance matrix V, where V is the stationary solution for the matrix equation $V = \Phi V \Phi' + \Sigma_\xi$, we refer to Durbin and Koopman (2012) for further discussions on initial conditions. The dynamic regression equation (6) for f_t (regression model with correlated errors) can be regarded as a special case of a state space model specification for f_t. Hence other linear dynamic specifications can also be considered for f_t in our framework. For example, the dynamic processes for u_t can be extended towards higher-order vector autoregressive moving average models or towards decomposition models with stochastic trend and cycle components. In all these cases, the model can be specified in a general state space formulation.

Although the model can be made more complex, we will show in our Monte Carlo study that a simple random walk process for f_t, that is with $C = 0$ and $\Phi = I$, where I is the identity matrix, provides a setting where complex dynamics in observed link probabilities, including structural breaks and sine-cosine curves, can be approximated accurately. However, when interest focuses on econometric modeling within an empirical setting, the model needs to have the ability to include more (and more complex) features. For example, in the context of international trade, the network dynamics are heavily inter-dependent due to changes in trade agreements between countries, the global business cycle, and technological changes in transportation modes that affect the composition of international trade opportunities over time. To channel these dependencies in a structural way, we require a modeling framework that is sufficiently flexible and general. We believe that with the model equations (4) and (6) and their implied generalizations, a flexible and general model specification is provided that is useful in many economic applications.

2.3. Discussion of model and assumptions

The specification of the vectors $z_{i,j}$ and f_t completes the description of our dynamic factor network model. The resulting model is effectively a mixture dynamic factor model, where the random vector $z_{i,j}$ selects the factors in f_t that are assigned to observation pair (i, j). The time-invariant component weights $\delta_{i,j}$ capture the relational structure inherent in network data, where links involving the same node i are likely to be correlated. The blockmodel formulation for the factor loadings is flexible enough to represent stylized facts in empirical network data, such as skewed degree distributions, core–periphery structures, potential clustering, homophily, and stochastic equivalence; see the discussion of these concepts, for example, in Wasserman and Faust (1994), Kolaczyk (2009) and Jackson (2008).[6] In our modeling framework, two pairs

[5] Countries which are similar in terms of, say, geography and climate may exhibit similar trade patterns in agricultural products. For example, bananas are typically produced and exported by countries close to the equator.

[6] In particular, our model allows for clustering, which is not accommodated by other latent factor network models, as described in Hoff et al. (2002).

of nodes, say (i, j) and (k, l), are stochastically equivalent when $z_{i,j} = z_{k,l}$ because they share the same link probability by construction.

These network dependencies are also relevant in many economic applications, such as when modeling global trade linkages between countries in our empirical study. For example, previous studies have shown that international trade networks reveal an approximate core–periphery structure, where core countries bilaterally trade more with other countries in the core than with peripheral countries that have less trade with other peripheral countries; see, for example, Snyder and Kick (1979). In our model, the key parameters that govern the implied network properties are related to the cross-sectional distribution of latent groups and dynamic factors. The distribution of nodes across different groups is governed by the group probability vectors $\{\pi_i\}$, which depend on the parameters μ and Σ_γ. Furthermore, the covariates X_i can inform the econometric method about latent group memberships depending on observed information (in a similar way as a random versus fixed coefficient approach). The distribution of nodes across groups and the within- and between-link probabilities govern the strength of the (stochastic) core–periphery structure. Similarly, homophily can be easily encompassed in our model and can be made dependent on time-varying link probabilities. These modeling features are relevant since, for example, homophily is typically due to stronger trade linkages between the European Union countries while its dependence can be due to trade agreements or other forms of increased economic integration. Our modeling options introduce a powerful, yet parsimonious and stylized modeling framework for the time-varying, cross-sectional dependency in network data.[7]

The core model proposed in this paper assumes independence between the cross-sectional factor loadings and the factors to highlight the key features and estimation procedure in an as simple as possible setup. However, the model can be extended further along several dimensions that may be relevant in empirical applications; for example, by including a conditional dependence between pair characteristics (cross-sectional features) and the node dynamics (time series feature). The vectors $z_{i,j}$ can represent a mechanism for selecting different dynamic factors on the basis of pair characteristics to introduce a cross-dependence between cross-section and time series features. The selection mechanism can also be based on other (possibly time-varying) covariates related to the pairs. Furthermore, the dynamic specification of the common factor f_t can be made conditional on a function of pair probabilities and characteristics. Since the estimation procedure below iterates between the conditioning on cross-sectional and on time series features, these extensions can be implemented without further complications. However, the performances of the methods under these generalizations need to be verified in more detail.

Our modeling framework allows the construction of a computationally efficient importance sampling procedure for maximum likelihood parameter estimation. In the analysis, we will make use of many of the well-established tools in modern state space analysis. It is therefore insightful to write the model using matrix notation. In particular, we can write the signal for all link variables at time t in compact form as $\theta_t = Zf_t$, where the $N^2 \times K^2$ selection matrix Z stacks the row vectors $z'_{i,j}$ below one another, that is $Z = [z'_{1,1}, \ldots, z'_{i,j-1}, z'_{i,j}, \ldots, z'_{i,N}, z'_{i+1,1}, \ldots]$. We then obtain $y_t \sim p(y_t|\theta_t)$, which resembles the close relationship of the observation density of a nonlinear non-Gaussian state space model as discussed in Durbin and Koopman (2012).

3. Maximum likelihood estimation

A major challenge is the estimation of the parameter vector ψ that collects the elements of the vector and matrices μ, Σ_γ, Φ, and Σ_ξ from the model equations (4) and (6) by the method of maximum likelihood.[8] The likelihood function is defined as $L(\psi) = p(y; \psi)$, where $p(y; \psi)$[9] is the joint density of all observations $y = \{y_t\}_{t=1}^T$ obtained from integrating out the latent variables from the complete data likelihood

$$L(\psi) = p(y) = \iiint p(y, f, z, \gamma)\, \mathrm{d}f\, \mathrm{d}z\, \mathrm{d}\gamma,$$

with the latent variable sets given by $f = \{f_t\}_{t=1}^T$, $z = \{z_{i,j}\}_{i,j=1}^N$, and $\gamma = \{\gamma_i\}_{i=1}^N$. Consider the contribution of observation $y_{i,j,t}$ to the complete data likelihood conditional on the vector of common factors f_t, which is given by

$$p(y_{i,j,t}, z_{i,j}, \gamma_i, \gamma_j|f_t) = \sum_{m=1}^{M} p(y_{i,j,t}|f_t, z_{i,j,m} = 1)\delta_{i,j,m} p(\gamma_i)p(\gamma_j),$$

[7] Our model is related to Koopman and Mesters (2017), who propose a stochastic dynamic factor model that treats the factor loadings as a composite of Gaussian random vectors. Their model is concerned with ordinary panel data and does not apply to the analysis of network data. We also model the time-varying heterogeneity by a dynamic factor model with both stochastic loadings and factors. However, in our model we propose discrete random factor loadings generated by a stochastic block structure that are different from Gaussian dynamic factor models, to account for the network-specific modeling challenges. Moreover, our model is specific for the analysis of binary data, which prevents an analytical solution to the filtering and smoothing problem.

[8] For ease of exposition, we focus on the case of no observables. Hence we assume $B = 0$ and $C = 0$. When estimation of elements of matrices B and C is required, we include these elements in parameter vector ψ which is estimated using the simulation-based method that we develop here.

[9] We will suppress the dependency of $p(y; \psi)$ on ψ in the notation, that is $p(y; \psi) \equiv p(y)$.

where $\delta_{i,j,m} = (\pi_i \otimes \pi_j)_m = \Pr(z_{i,j,m} = 1|\gamma_i, \gamma_j)$ equals the probability that the observation (i, j) is in the mth group. We have used Bayes' rule and the independence assumptions of the model, while we have integrated out z analytically to obtain the expression above.

Given the model structure, we can further factorize the complete data likelihood in a product of conditional densities across pairs and time periods. We then have the likelihood function as

$$L(\psi) = \int\int \prod_{t=1}^{T} p(f_t|f_{t-1}) \prod_{i=1}^{N}\prod_{j=1}^{N} \left[\sum_{m=1}^{M} \delta_{i,j,m} p(y_{i,j,t}|f_t, z_{i,j,t,m} = 1)\right] p(\gamma_i)p(\gamma_j)\,\mathrm{d}f\,\mathrm{d}\gamma,$$

which we can write more compactly by writing the sums as integrals as

$$L(\psi) = \int\int\int p(y|f, z)p(f)p(z|\gamma)p(\gamma)\,\mathrm{d}f\,\mathrm{d}z\,\mathrm{d}\gamma = \mathbb{E}p(y|f, z), \tag{7}$$

where the expectation is with respect to $p(f, z, \gamma) = p(f)p(z|\gamma)p(\gamma)$. Eq. (7) shows that the likelihood function can be interpreted as an expectation of $p(y|f, z)$. However, the evaluation of the likelihood function involves a high-dimensional integral, which does not have a closed-form solution. In particular, integration with respect to the cross-sectional latent variable γ gives rise to various problems and is analytically intractable.

We propose to evaluate the likelihood function by Monte Carlo integration, which replaces the integration with an average of conditional densities that relies on draws from sampling distributions. A simple frequency-based estimator of the likelihood function is given by

$$\bar{L}(\psi) = \frac{1}{S}\sum_{s=1}^{S} p(y|f^{(s)}, z^{(s)}),$$

where $z^{(s)}$ and $f^{(s)}$, for $i = 1, \ldots, S$, denote samples that are jointly drawn from the density $p(f)p(z|\gamma)p(\gamma)$. To draw a sample for $f^{(s)}$, we first require a sample for $\gamma^{(s)}$. This estimator is consistent but can be highly inefficient and might require a very large number of draws before convergence is achieved. It is numerically more efficient to adopt the method of importance sampling, which relies on generating draws that are more often in high probability regions of the latent variables; see Ripley (1987, Chapter 5) for a general discussion. The importance sampling representation of the likelihood function is given by

$$L(\psi) = \int\int\int p(y|f, z, \gamma)\frac{p(f, z, \gamma)}{g(f, z, \gamma|y)} g(f, z, \gamma|y)\,\mathrm{d}f\,\mathrm{d}z\,\mathrm{d}\gamma,$$

where $g(f, z, \gamma|y)$ denotes the importance density function. The likelihood function then reduces to an expectation with respect to the importance density, $\mathbb{E}_g[p(y|f, z)\,p(f, z, \gamma)/g(f, z, \gamma|y)]$, which can be numerically evaluated via Monte Carlo integration. Hence, we sample $f^{(s)}$, $z^{(s)}$ and $\gamma^{(s)}$ from the importance density $g(\cdot|y)$, independently over $s = 1, \ldots, S$, and estimate the likelihood as

$$\hat{L}(\psi) = \frac{1}{S}\sum_{s=1}^{M} p(y|f^{(s)}, z^{(s)})\frac{p(f^{(s)}, z^{(s)}, \gamma^{(s)})}{g(f^{(s)}, z^{(s)}, \gamma^{(s)}|y)}, \tag{8}$$

which is a consistent estimator due to the law of large numbers; see Ripley (1987, Section 5.2). The conditions under which a central limit theorem applies are discussed by Geweke (1989) and can be empirically tested as discussed by Monahan (1993) and Koopman et al. (2009); we provide more details in the Supplementary Appendix. The asymptotic properties for any Monte Carlo estimator rely generally on core results from the theory of empirical processes; see the treatments in Pollard (1984, Chapter 2) and van der Vaart (1998, Chapter 19).

We use $\hat{L}(\psi)$ to obtain the simulated maximum likelihood estimator $\hat{\psi}$ as the result of

$$\hat{\psi} = \arg\max_{\psi} \hat{L}(\psi; y), \tag{9}$$

where the maximization with respect to the parameter vector ψ is done numerically. To let the simulated estimator $\hat{L}(\psi; y)$ be a smooth function of ψ, we use the same initial random seed value when generating the necessary samples, for each likelihood evaluation. In practice, the logarithm (log) of the likelihood function is maximized for numerical purposes. The log of $\hat{L}(\psi; y)$ is a biased estimator of $\log L(\psi; y)$ but a standard bias correction based on a first-order Taylor expansion can be adopted for this purpose; see Durbin and Koopman (2012) for a discussion on these matters.

The maximum likelihood estimator of ψ has the usual asymptotic properties because we are maximizing a consistent estimator of the log likelihood function; see the discussion in Newey and McFadden (1994). The likelihood function $L(\psi)$ is for the $n \times 1$ data vector y with $n = N^2 \cdot T$. Hence the estimator $\hat{\psi}$ is a $\sqrt{n}$-consistent estimator of ψ. The divergences of N^2 and T are equally fast. The model treats both the cross-section and the time series dimensions simultaneously. For the purpose of constructing the importance density, the cross-section and time series parts are treated separately (but conditional upon each other). However, the resulting importance density is used to evaluate the likelihood function of the unified data set y with n observations. The property of asymptotic normality for the estimator $\hat{\psi}$ also follows from the

standard central limit theorem. Given our likelihood-based approach using the full likelihood function, hypothesis testing can be based on likelihood-ratio, Wald, and Lagrange Multiplier test statistics which have the usual limiting properties.

Next, we discuss the construction of a computationally fast importance sampler from which it is easy to sample. The computational speed and convenience are important given the high dimensions that we are faced with, even for moderate networks, as the number of links grows quadratically in the number of nodes. We construct two importance samplers simultaneously: one for the cross-sectional latent variables, and one for the dynamic factors in the time dimension. Using the independence assumption from the model, it is justified to separate the treatments for (z, γ) and f. Therefore, we propose to sample (z, γ) from $g(z, \gamma|y)$ and f from $g(f|y)$. However, $g(z, \gamma|y)$ and $g(f|y)$ still depend on f and (z, γ), respectively, by means of the data y. For this purpose, we fix the random variables to their conditional modal values $\hat{f}$ and $(\hat{z}, \hat{\gamma})$ for obtaining the respective samplers $g(z, \gamma|y, \hat{f})$ and $g(f|y, \hat{z}, \hat{\gamma})$.

3.1. The cross-sectional importance sampler

For constructing the cross-sectional importance sampler $g(z, \gamma|y, \hat{f})$, with $\hat{f}$ being the modal value, we propose to use the generalized mean field method as a variational approximation to the true conditional distribution $p(z, \gamma|y, \hat{f})$. Variational approximation methods are deterministic techniques for making approximate inference in statistical models by maximizing a lower bound of the log likelihood. In particular, these methods rely on choosing a family of variational distributions and then selecting a particular distribution from this family by minimizing the Kullback–Leibler divergence between the true conditional p and the approximate conditional g with respect to the variational parameters. In the Supplementary Appendix, we present a brief introduction for the unfamiliar reader.[10]

In our context, we use a generalized mean field approximation, where we approximate the intractable conditional density $p(z, \gamma|y, \hat{f})$ by a product of simpler marginals from which it is easy to sample. We therefore factorize the approximate conditional density into clusters for (γ_i and $z_{i\to j}, z_{j\leftarrow i}$) as

$$g(\gamma, z|y, \hat{f}) = \prod_{i=1}^{N} g_\gamma(\gamma_i) \prod_{i=1}^{N}\prod_{j=1}^{N} g_z(z_{i\to j}, z_{j\leftarrow i}),$$

where $g_\gamma(\cdot)$ and $g_z(\cdot)$ are the variational marginal distribution for each cluster. We thereby follow Xing et al. (2003), who show that under the generalized mean field approximation the optimal solution of each cluster's marginal distribution is isomorphic to the true conditional distribution of the cluster given its expected Markov blanket.[11] Formally, this factorization gives the two equations

$$g_\gamma(\gamma_i) = p(\gamma_i|\langle z_{i\to\cdot}\rangle_{g_z}, \langle z_{i\leftarrow\cdot}\rangle_{g_z}) \tag{10}$$

$$g_z(z_{i\to j}, z_{j\leftarrow i}) = p(z_{i\to j}, z_{j\leftarrow i}|y_{i,j,t}, f, \langle\gamma_i\rangle_{g_\gamma}, \langle\gamma_j\rangle_{g_\gamma}), \tag{11}$$

where $\langle\cdot\rangle_f$ denotes the expected Markov blanket, that is, all children, parents, and co-parents of the cluster, under a variational distribution f. The two equations define a fixed point for $g_\gamma(\cdot)$ and $g_z(\cdot)$. The solution to the fixed-point problem is obtained by sequentially updating the parameters of the marginal distribution of one cluster, while fixing the parameters of the marginals of all other clusters.

The updating formula for the cluster marginal for $z_{i\to j}, z_{j\leftarrow i}$ given $\langle\gamma_i\rangle_{g_\gamma}, \langle\gamma_j\rangle_{g_\gamma}$ is obtained as a multinomial distribution with K^2 possible outcomes

$$g_{z_{i,j}}(z_{i\to j}, z_{j\leftarrow i}) \propto p(z_{i\to j}|\langle\gamma_i\rangle_{g_\gamma}) p(z_{j\leftarrow i}|\langle\gamma_j\rangle_{g_\gamma}) \prod_{t=1}^{T} p(y_{i,j,t}|z_{i,j}, f_t), \tag{12}$$

with the (uv)-th element of the $K^2 \times 1$ multinomial probability vector $\tilde{\delta}_{i,j}$ given by

$$\tilde{\delta}_{i,j,uv} = \frac{1}{c} \exp(\langle\gamma_{i,u}\rangle_{g_\gamma} + \langle\gamma_{j,v}\rangle_{g_\gamma}) \prod_{t=1}^{T} \left(\frac{\exp(f_{t,uv})}{1+\exp(f_{t,uv})}\right)^{y_{i,j,t}} \left(\frac{1}{1+\exp(f_{t,uv})}\right)^{1-y_{i,j,t}},$$

where c is a normalizing constant that ensures that the probabilities sum up to one, that is, $\sum_{u=1}^{K}\sum_{v=1}^{K} \tilde{\delta}_{i,j,uv} = 1$. Here, we have introduced an injective mapping between the double-digit index (uv), with $u, v = 1, \ldots, K$ and the single-digit index $m = 1, \ldots, M = K^2$, that is determined by the Kronecker product $\delta_{i,j} = \pi_i \otimes \pi_j$.

[10] Variational approximation methods are by now widely employed in computer science; compare, for example, Jordan et al. (1999) and Xing et al. (2003) in the context of blockmodels. Variational approximation methods recently also started to attract attention in statistics and econometrics, see Ormerod and Wand (2010).

[11] In directed graphical models, the Markov blanket is the set of parents, children and co-parents of a given node (co-parents are nodes that have a child in common with the given node). In the undirected case, the Markov blanket is simply the set of neighbors of a given node.

We can also compute the expectation of the group indicators $z_{i\to j}$ and $z_{j\to i}$ under the variational approximation as

$$\langle z_{i\to j,u}\rangle_{g_z} = \sum_{v=1}^{K} \tilde{\delta}_{i,j,uv}, \quad \text{and} \quad \langle z_{j\leftarrow i,v}\rangle_{g_z} = \sum_{u=1}^{K} \tilde{\delta}_{i,j,uv},$$

where $z_{i\to j,u}$ is the uth element of $z_{i\to j,u}$, and, similarly, $z_{j\leftarrow i,v}$ is the vth element of $z_{j\leftarrow i,v}$. Moreover, for the updating formula of γ_i we have

$$\begin{aligned} g_{\gamma_i}(\gamma_i) &\propto p(\gamma_i|\mu, \Sigma_\gamma) p(\langle z_{i\to\cdot}\rangle_{g_z}, \langle z_{i\leftarrow\cdot}\rangle_{g_z}|\gamma_i) \\ &= \mathcal{N}(\gamma_i; \mu, \Sigma_\gamma) \exp(\langle m_i\rangle'_{g_z}\gamma_i - (2N-2)C(\gamma_i)), \end{aligned}$$

where the kth element of $\langle m_i\rangle_{g_z}$ is given as $\langle m_{i,k}\rangle_{g_z} = \sum_{j\neq i}(\langle z_{i\to j,k}\rangle_{g_z} + \langle z_{i\leftarrow j,k}\rangle_{g_z})$, and $C(\gamma_i) = \log(\sum_{k=1}^{K}\exp(\gamma_{i,k}))$ is the normalization constant that prevents a close-form solution for the integration of $g_{\gamma_i}(\gamma_i)$. We approximate $C(\gamma_i)$ by a second-order Taylor expansion around $\hat{\gamma}_i$, given by $C(\gamma_i) \approx C(\hat{\gamma}_i) + g'(\gamma_i - \hat{\gamma}_i) + 1/2(\gamma_i - \hat{\gamma}_i)'H(\gamma_i - \hat{\gamma}_i)$, where g denotes the gradient vector and H is the Hessian matrix. Using this approximation and some basic algebra, we derive at the approximating variational density $g_{\gamma_i}(\gamma_i) = \mathcal{N}(\tilde{\gamma}_i, \tilde{\Sigma}_{\gamma_i})$ with mean

$$\tilde{\gamma}_i = \mu + \tilde{\Sigma}_{\gamma_i}\left(\langle m_i\rangle_{g_z} - (2N-2)[g - H(\hat{\gamma}_i - \mu)]\right),$$

and variance matrix

$$\tilde{\Sigma}_{\gamma_i} = (\Sigma_\gamma^{-1} + (2N-2)H)^{-1}.$$

The parameter updating continues until convergence. This algorithm guarantees convergence to a local optimum; see the discussion in Xing et al. (2008).

3.2. The time series importance sampler

To sample the latent dynamic factors, we specify $g(f|y, \hat{z}, \hat{\gamma})$ as the conditional density obtained from a linear Gaussian approximating model that we derive conditional on the modal values $\hat{z}, \hat{\gamma}$. Conditional on a value of z, such as the mode, we can write the model as a partial nonlinear non-Gaussian state space model given by

$$y_t \sim p(y_t|\theta_t), \qquad \theta_t = Zf_t, \qquad f_{t+1} = \Phi f_t + \xi_t, \qquad \xi_t \sim \mathcal{N}(0, \Sigma_\xi),$$

where the matrix Z is fixed for any given value of z, such as the mode. The construction of an importance density for θ_t conditional on all observations y, and on $(\hat{z}, \hat{\gamma})$, relies on an approximating linear Gaussian state space model for which the standard Kalman filtering and smoothing recursions can be applied. Specifically, we consider here an approximation method based on the equalization of the conditional mode and the variance matrix of the smoothing densities $p(\theta|y)$ and $g(\theta|y)$, where $g(\theta|y)$ refers to the smoothed approximating Gaussian density; see Shephard and Pitt (1997) and Durbin and Koopman (1997). On the basis of this method, we maintain the VAR(1) specification for f_t, but use a linear Gaussian observation equation

$$y_{i,j,t} = c_{i,j,t} + \hat{z}'_{i,j}f_t + u_{i,j,t},$$

where $u_{i,j,t} \overset{iid}{\sim} \mathcal{N}(0, \sigma^2_{u_{i,j,t}})$ and $c_{i,j,t}$ is the intercept. The unknown values are the intercept $c_{i,j,t}$ and the variance $\sigma^2_{u_{i,j,t}}$, and both are treated as parameters in solving the set of equations

$$\frac{\partial \log p(y_{i,j,t}|\theta_{i,j,t})}{\partial \theta_{i,j,t}} = \frac{\partial \log g(y_{i,j,t}|\theta_{i,j,t})}{\partial \theta_{i,j,t}}, \quad \frac{\partial^2 \log p(y_{i,j,t}|\theta_{i,j,t})}{\partial \theta_{i,j,t}\partial \theta_{i,j,t}} = \frac{\partial^2 \log g(y_{i,j,t}|\theta_{i,j,t})}{\partial \theta_{i,j,t}\partial \theta_{i,j,t}}.$$

The solutions can be obtained using an iterative process, and during this process the variables $c_{i,j,t}$ and $\sigma^2_{u_{i,j,t}}$ become functions of the conditional mean (and mode) $\hat{f}$. The iterations continue until a level of convergence with respect to $\hat{f}$; for more details we refer to the discussion in Durbin and Koopman (2012). After convergence of this iterative procedure, we adopt the resulting linear Gaussian state space model and use it for simulating samples for f from the importance sampler. These draws are obtained from the computationally efficient simulation smoother as developed by Durbin and Koopman (2002).

A particular challenge in network analysis is that the dimension of the observation vector y_t grows quadratically in the number of nodes N, since all N^2 bilateral relationships in directed networks are modeled. This makes a direct computation of the approximating linear model computationally infeasible even for a small number of nodes N, as the Kalman filter and smoother requires inversion of a full $N^2 \times N^2$ matrix for each time index t. We therefore collapse the large observation vector into a low-dimensional vector such that the observation equation of the associated state space model changes. However, the collapse is done in a way that maintains the information needed to compute the conditional density of the latent factors, and it does not alter the dynamic equation for f_t. The Kalman filter and smoother computations only consider the low-dimensional collapsed system of equations.

For dynamic factor models, Jungbacker and Koopman (2015) have proposed to collapse the observation vector by pre-multiplying it with an appropriate non-singular matrix. Since, for a given fixed value of z, our linearized dynamic factor network model corresponds to a linear Gaussian state space model, we can apply this collapse method. For a panel data model with random effects in cross-section and time, the Jungbacker and Koopman (2015) method has been successfully used by Mesters and Koopman (2014) to achieve significant computational savings. However, in our case, this collapse method is still computationally demanding.

3.3. A more efficient time series importance sampler

For our dynamic factor network model, for which the rows of the matrix Z are multinomially distributed, it is possible to develop a computationally more-efficient collapse method that is solely based on addition operators and does not involve any matrix multiplication, including inversions and decompositions. We develop the method next. To obtain the smoothed estimate of the $M \times 1$ latent vector f_t and its corresponding variance matrix, we can represent the high-dimensional observation vector y_t by a low-dimensional vector of dimension $M \times 1$, without loss of any information. Given the modal value $\hat{z}$ of z, we construct the vector $y_t^* = (y_{1,t}^*, \ldots, y_{M,t}^*)'$ with elements $y_{m,t}^* = \sum_{i,j} y_{i,j,t} \hat{z}_{i,j,m}$, where $\hat{z}_{i,j,m}$ denotes the mth element of the vector $\hat{z}_{i,j}$, with $m = 1, \ldots, M$. It follows that $y_{m,t}^*$ is binomially distributed with $\sum_{i,j} \hat{z}_{i,j,m}$ number of trials and probability of success $\exp(f_{m,t})/(1 + \exp(f_{m,t}))$ in each trial. This follows immediately, as conditional on the mode of z (the most likely group memberships), we can identify to which component a particular observation $y_{i,j,t}$ belongs (with highest probability) and then count the number of successes in each of the M components. The importance sampler is then obtained from a linearization of the collapsed binomial model with observation equation

$$y_{m,t}^* = c_{m,t}^* + f_{m,t} + u_{m,t}^*, \qquad u_{m,t}^* \overset{iid}{\sim} \mathcal{N}(0, \sigma^2_{u_{m,t}^*}), \tag{13}$$

where intercept $c_{m,t}^*$ and variance $\sigma^2_{u_{m,t}}$ are treated as unknown parameters. The dynamic equation for f_t remains as in Eq. (6). The procedure for computing the mode $\hat{f}$ is similar to the description in the previous section, but is now computed directly in terms of f_t rather than $\theta_t = Zf_t$. Further details of computing the mode are provided by Durbin and Koopman (2012). The linearization based on the collapsed model given here reduces the dimension of the observation vector in the state space model from N^2 to $M = K^2$ without losing any relevant information needed to extract the conditional density of the latent factors f_t that govern the group-specific link probability dynamics. Hence, our modification significantly reduces the computations for the Kalman filter and smoother recursions, as in typical applications $K^2 \ll N^2$. In practice, and even for moderate values of N, the standard Kalman filter implementation will be computationally too demanding. Moreover, our methods allow us to avoid the large computational costs of matrix multiplications, inversions, and decomposition that are needed in the method proposed by Jungbacker and Koopman (2015).

3.4. The algorithm

The procedure for constructing the importance sampler for the dynamic factor network model can be viewed as an algorithm that incorporates both the cross-section and time series dimensions. The cross-sectional part is represented by the variational density $g_z(z_{i\to j}, z_{j\leftarrow i})$ that still depends on the dynamic factors f_t; we fix this part at its current estimate of the mode. The time series part is represented by the approximating linear Gaussian density $g(f|y, z)$, where we set z at its current estimate of the mode. The procedure contains recursive steps which continue until convergence. In each step of the Newton-type algorithm, a new variational approximating model and a new Gaussian approximating model are obtained. The algorithm displayed in Fig. 1 summarizes these steps for the construction of the importance sampling distribution. We notice that, in contrast to other simulation-based methods, no data are simulated for the construction of the importance sampling distribution. Instead, the algorithm iterates over parameters of variational and Gaussian distributions to find tractable approximations of the intractable conditional distributions of the latent variables (given the data) with the aim to maximize similarity between these distributions.

3.5. Discussion of other estimation methods

Other estimation methods can also be considered. For example, we can replace the stochastic variables by fixed effects, or flexible functions of these, and adopt the fixed effects estimation method. The latent group probabilities and the dynamic factors can be proxied by appropriate transformations and functions of fixed effects. As a result, the fixed effects typically enter the model in a nonlinear fashion, leading to a non-standard fixed effects estimation method. However, despite the general availability of such alternative approaches to estimation, it is far from clear how a regression-based method can be modified to address the generality and flexibility of latent groups and dynamic factors in our current model specification.

Alternatively, we can consider an estimation method based on the Markov chain Monte Carlo (MCMC) framework; see Gilks et al. (1995) and Gamerman and Lopes (2006) for a general treatment and Lopes et al. (2008) for an

Initialize $\hat{f}$ and $\hat{\gamma}$ at random values;
while *not converged* **do**
  Obtain variational approximation of $p(z, \gamma|y, \hat{f})$;
  while *not converged* **do**
    Compute marginal distribution $g_{z_{i,j}}(z_{i\to j}, z_{j\leftarrow i})\ \forall i, j$;
    Compute marginal distribution $g_{\gamma_i}(\gamma_i)\ \forall i$;
  end
  Obtain approximation of $p(f|y, \hat{z}, \hat{\gamma})$;
  Collapse observation vector y_t to low-dimensional multinomial vector y_t^*;
  while *not converged* **do**
    Compute approximating linear-Gaussian state space model for y_t^*, conditional on current mode $\hat{f}$;
    Apply Kalman filter and smoother to approximating model to update value for mode $\hat{f}$;
  end
end

Fig. 1. Algorithm to construct the importance sampler, with ψ given.

implementation of a spatial dynamic factor model. By adopting the MCMC method, the analysis is carried out from a Bayesian perspective. Given that both maximum likelihood and MCMC methods rely on simulations for the latent variables and the dynamic factors, we believe that the two methods have many similarities. The challenge within an MCMC analysis is to manage the number of iterations necessary for integrating out the parameter vector and the latent variables jointly, before some level of convergence of the chain is obtained. The challenge within our approach is the maximization of the likelihood that is evaluated via the importance sampling method. From our simulation and empirical studies below, we have found that (*i*) the number of simulations S for the likelihood evaluation in (8) can be set to relatively low values, and (*ii*) the maximization of the Monte Carlo likelihood estimate with respect to parameter vector ψ has been a smooth process, in almost all cases. It is our view that both estimation methods are complementary. An example of an MCMC analysis for a dynamic network model with latent variables is recently provided by Sewell and Chen (2015).

4. Monte Carlo study

We present and discuss the results of a simulation study with the aim to analyze the finite sample behavior of our proposed maximum likelihood estimation procedure for the dynamic factor network model. In this Monte Carlo study, we consider a dynamic factor network model as a data generation process that includes features resembling those from our empirical study in Section 5.

4.1. Baseline results: Correct specification

We consider a model for a directed network with $K = 2$ latent groups for the nodes so that each pair of nodes may be assigned to four different clusters ($M = K^2 = 4$), and the vector of dynamic factors f_t has dimension 4×1. The dynamic factors are generated as four independent random walk processes. Hence, we have $\Phi = I_{4\times 4}$ and a diagonal variance matrix Σ_ξ, which we set as $\Sigma_\xi = \text{diag}(\sigma^2_{\xi,1}, \sigma^2_{\xi,2}, \sigma^2_{\xi,3}, \sigma^2_{\xi,4}) = \text{diag}(0.5, 0.1, 0.03, 0.25)$ where diag() represents a diagonal matrix with its values on the leading diagonal given by the arguments. The parameter values for the normal distribution of γ_i are given by $\mu = -0.5$ and $\Sigma_\gamma = \sigma^2_\gamma = 1.3$. We further specify a $(K-1)$-dimensional normal distribution, as it is justified to fix one element of γ_i to an arbitrary value (we set it to zero). We perform our simulation study for different network sizes with $N = 25, 50, 100$ nodes and, for each network size, time series lengths of $T = 50, 100, 250$.[12] Each simulation exercise consists of 500 Monte Carlo repetitions. The log likelihood function is computed using $S = 500$ simulated importance samples.

In Fig. 2, we present the kernel density estimate for each parameter of the model. The density estimate is based on the 500 maximum likelihood estimates from a simulation exercise, with specific values for N and T. For all six parameters, the finite sample distribution of the estimates is centered around the true parameter values (as indicated by a vertical

[12] These dimensions are relevant in many empirical applications in economics, such as international trade analysis or sectoral input–output analysis. Our estimation procedure can handle large networks with $N > 100$. However, the estimation process of larger networks is computationally intensive, given that our maximum likelihood procedure requires the computation of the likelihood contribution of $N(N-1)$ links for each $t = 1, \ldots, T$. While we found the estimation of networks with up to $N \approx 250$ to be feasible, large-scale simulation studies based on $N > 100$ are clearly computationally demanding.

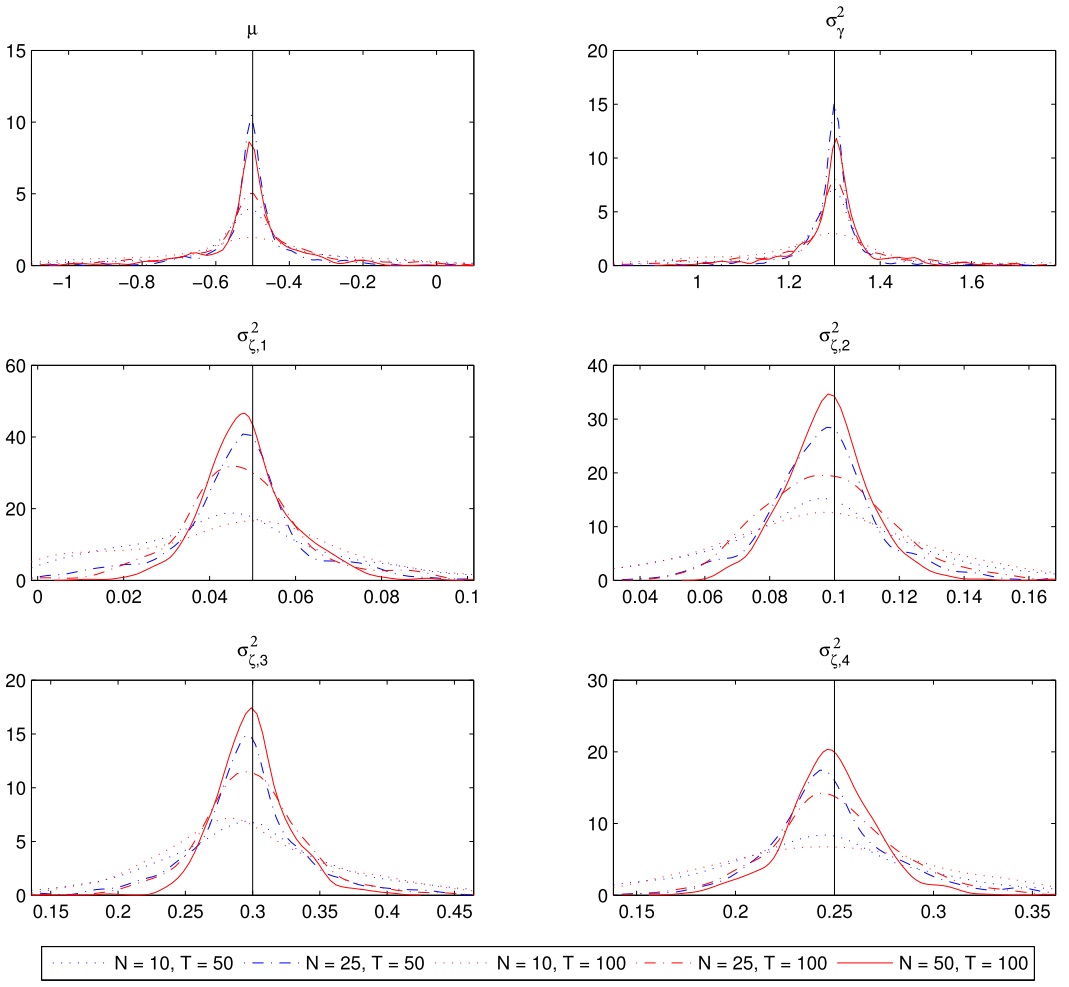

Fig. 2. Finite sample distribution of ML estimator. We present the kernel density estimates for the simulated maximum likelihood estimators in finite samples, using our importance sampling procedure with $S = 500$ importance samples; see discussion in Section 3. Simulation results are based on 500 Monte Carlo replications.
Source: Authors' calculations.

line in the graph) that we have adopted in our Monte Carlo study. Moreover, for an increasing network size N and time series dimension T, the density becomes more concentrated around the true parameter value. A particular finding is that for a small network of size $N = 10$ (which is indicated by a dotted line in each graph), the sample distribution of the parameter estimates has a large variance for all three time series dimensions.

Table 1 presents more-detailed results of the Monte Carlo study. In particular, it depicts the root mean square error (RMSE) and the mean absolute error (MAE) for each of the six parameters, relative to their true values in the simulation study (we focus on RMSE and MAE, given that bias and standard deviation become apparent in Fig. 2). The results for the four variance parameters of the random walk processes, $\sigma^2_{\xi,i}$, for $i = 1, 2, 3, 4$, show that, for a given network size N, a larger time series dimension T reduces both the RMSE and the MAE, in most cases. In particular, comparing the results for $T = 250$ with $T = 50$, we encounter a clear reduction in the RMSEs. Similarly, for a given time series dimension T, the RMSE and MAE of the variance parameter strictly decrease as the network size becomes larger. The results for the parameters μ and σ^2_γ, which, govern the cross-sectional dependence, reveal that, for a given time series dimension T, an increasing network size N leads to RMSE and MAE reductions in all cases. Interestingly, however, when we hold the network size N constant and increase the time series dimension T, we do not find these reductions, but instead we find increases in the RMSE and MAE. This phenomenon may indicate that the importance sampler may need to account for possible heterogeneity when the time series dimension becomes larger.

Table 2 analyzes the fit of the approximating conditional densities for the latent variables f_t, γ_i, and $z_{i,j}$. This table presents the RMSE and MAE values as measures of divergence between the estimated mode of each of the approximating distributions relative to its true value that was used for simulating data. The results indicate that for the latent factor f_t, we obtain smaller RMSE and MAE values when we increase the network size N and keep the time series dimension T

Table 1
Monte Carlo results for maximum likelihood estimation.

		$T = 50$			$T = 100$			$T = 250$		
		$N = 25$	$N = 50$	$N = 100$	$N = 25$	$N = 50$	$N = 100$	$N = 25$	$N = 50$	$N = 100$
$\sigma^2_{\xi,1}$	RMSE	0.0171	0.0246	0.0096	0.0164	0.0105	0.0078	0.0097	0.0070	0.0043
	MAE	0.0111	0.0071	0.0059	0.0118	0.0080	0.0058	0.0078	0.0056	0.0035
$\sigma^2_{\xi,2}$	RMSE	0.0187	0.0132	0.0102	0.0199	0.0144	0.0104	0.0122	0.0088	0.0060
	MAE	0.0135	0.0090	0.0070	0.0159	0.0102	0.0079	0.0098	0.0070	0.0047
$\sigma^2_{\xi,3}$	RMSE	0.0429	0.0292	0.0216	0.0426	0.0308	0.0265	0.0296	0.0199	0.0172
	MAE	0.0300	0.0182	0.0126	0.0304	0.0212	0.0183	0.0221	0.0152	0.0133
$\sigma^2_{\xi,4}$	RMSE	0.0333	0.0260	0.0200	0.0319	0.0231	0.0206	0.0229	0.0171	0.0138
	MAE	0.0241	0.0162	0.0113	0.0245	0.0173	0.0148	0.0181	0.0132	0.0108
μ	RMSE	0.1131	0.0629	0.0377	0.1800	0.1265	0.0755	0.2529	0.1744	0.1191
	MAE	0.0653	0.0299	0.0185	0.1179	0.0769	0.0467	0.1994	0.1353	0.0922
σ^2_{γ}	RMSE	0.0836	0.0401	0.0328	0.1412	0.0882	0.0542	0.1944	0.1267	0.0892
	MAE	0.0448	0.0198	0.0135	0.0881	0.0542	0.0328	0.1479	0.0983	0.0686

Note: The table shows the root mean square errors (RMSE) and the mean absolute errors (MAE) between the true parameter values and the estimated parameter values using the ML importance sampling estimator with $S = 500$ importance weights as outlined in Section 3. The simulation results are based on 500 Monte Carlo replications.

Table 2
Monte Carlo results for latent variable estimation.

		$T = 50$			$T = 100$			$T = 250$		
		$N = 25$	$N = 50$	$N = 100$	$N = 25$	$N = 50$	$N = 100$	$N = 25$	$N = 50$	$N = 100$
f_t	RMSE	0.0743	0.0514	0.0403	0.0811	0.0498	0.0338	0.1165	0.0723	0.0525
	MAE	0.1164	0.0809	0.0582	0.1094	0.0703	0.0439	0.1283	0.0822	0.0530
γ_i	RMSE	0.0723	0.0675	0.0911	0.0556	0.0289	0.0147	0.0536	0.0272	0.0136
	MAE	0.1635	0.1219	0.0908	0.1441	0.1050	0.0740	0.1400	0.0999	0.0716
$z_{i,j}$	RMSE	0.0223	0.0208	0.0230	0.0144	0.0082	0.0066	0.0113	0.0055	0.0028
	MAE	0.0863	0.0837	0.0792	0.0357	0.0304	0.0276	0.0120	0.0070	0.0048

Note: We present the root mean square error (RMSE) and the mean absolute error (MAE) between the estimated mode of the latent variables f, z, and γ as implied by the approximating importance sampling distribution with $S = 500$ importance weights and the true realized values in each simulation. The simulation results are based on 500 Monte Carlo replications.

Table 3
Relative computational time of different importance samplers.

	$T = 50$	$T = 100$	$T = 250$
$N = 25$	0.3884	0.3599	0.3458
$N = 50$	0.2480	0.2223	0.1871
$N = 100$	0.1531	0.1435	0.1073

Note: We present the relative computing times for constructing the importance sampling distribution of the dynamic factors with $S = 500$ importance weights. The ratio is expressed as the computing time for the collapsing method based on the binomial distribution in the numerator and the Jungbacker and Koopman (2015) collapsing method in the denominator.

constant. However, when we keep N constant and increase T, we do not have monotonically decreasing RMSE and MAE values. For the cross-sectional latent variables γ_i and $z_{i,j}$ we have that increasing N and/or T reduces the RMSE and MAE clearly (except for small $T = 50$ and when increasing N does not monotonically decrease the RMSE, but does decrease the MAE). Overall, when increasing both N and T, we obtain more precise estimates.

Table 3 compares the computational cost of constructing the importance sampler for the latent dynamic factor by showing the ratio of the computing times between the importance samplers, based on the collapsed observational model of Jungbacker and Koopman (2015) and based on our newly approximating binomial model. Further results on computing times are presented in the Supplementary Appendix. As discussed, our proposed collapsing method transforms the binary observation vector into a binomial observation vector. Therefore, it does not rely on any matrix inversions and related decompositions. As a result, our method is computationally faster and the evidence is given in Table 3. The computational gains become specifically large when the number of nodes N gets larger because the number of potential links in the network grows quadratically in N. The computations for the existing collapsing method rely on $N^2 \times N^2$ matrices; this is the main reason for its slow speed.

In the Supplementary Appendix, we present and discuss in detail results under correct model specification for $K > 2$ number of latent groups. Overall, the evidence shows that the estimation procedure behaves well in finite samples under correct model specification, also for larger number of latent groups.

4.2. Further results: Model misspecification

In addition to simulation studies under correct model specification, we have performed various other simulation studies to test the ML estimator's behavior under a range of different model misspecifications, including misspecification of the dynamic link probabilities and cross-sectional dependence. We summarize here the main findings and refer to the Supplementary Appendix for a detailed discussion of the results.

With regards to the dynamic misspecification in the link probabilities, we show that our proposed dynamic network estimation procedure with a random walk factor specification to be flexible enough to account for DGPs with richer dynamics. In particular, we simulate data under a DGP where not all factors follow a Gaussian random walk, as in our estimated model, but allow the true factors to follow more complex dynamics. For example, estimated link probabilities do recover closely structural breaks in the true data generating process, with the errors in the underlying estimated dynamic factors being small, even with relatively small networks of $N = 50$ nodes. Similarly, a random walk specification is able to capture well the situation where the true factors follow a sine curve. Moreover, we also show that mean squared and absolute errors of the estimation factors are small when the factor innovations in the true DGP do not follow a Gaussian distribution, but are t-distributed innovations. In all those cases, also bias and standard deviation of the model parameters are small, similar to the mean squared and absolute errors of the estimated cross-sectional latent variables.

To verify the performance of our estimation procedure under cross-sectional misspecification, we focus in our simulation studies on (i) misspecification of the number of latent groups K, and (ii) misspecification of the group heterogeneity (cross-sectional effects). That is we simulate data under a DGP with more/less groups than estimated with our model, or allow the cross-sectional effects in our DGP to have different distributions than in our estimated model. Our results show that in the case of overspecification in the number of latent groups, the likelihood differences are relatively small (less than 1 percent). On the other hand, when the true data is simulated with 3 latent groups, but the model is restricted with two groups, likelihood differences are larger (about 3 percent in relative terms) and the differences increase slightly when the network becomes larger. Similarly, the standard deviations of the estimates for the misspecified model increases. Finally, a misspecified distribution leads to larger likelihood differences of up to 6 percent in relative terms, again it coincides with larger standard deviations of the parameter estimates. Overall, based on these findings of the simulation study, we can conclude that the fit does not decrease dramatically under a wide range of model misspecification.

5. Empirical application: International banana trade

To show how our proposed model works in an empirical application using real data, we consider the dynamic factor network model to study international trade patterns and its dynamics. In particular, we use our model to analyze the global patterns of exports and imports of bananas, a healthy and delicious fruit that is widely enjoyed in many countries in the world. Banana trade has an economic significance with a total value of $8 billion annually. For some countries, the international trade with bananas represents a large share of income.[13] However, banana trade in the past 50 years has undergone substantial changes due to technological changes in transportation modes (refrigerated containers), innovations in pest control and the breeding of more resilient and higher-yielding varieties, as well as lower production costs and liberalization of trade agreements, amongst other changes in international trade; see Hallam (1995) for a discussion. Thus, discovering the dynamics of global banana trade is an economically important question that we address in this section.[14]

5.1. Data description

We use bilateral (country-pair level) trade data that contain annual export and import values (reported in USD billions) on 986 goods and 47,190 country-pairs from 1962–2014. The bilateral trade data organized by SITC4 Revision 2 codes are from The Center for International Data (1962–2000) and UN COMTRADE (2001–2014). The data from these two sources are compiled neatly by MIT's Observatory of Economic Complexity project; for more details, see Feenstra et al. (2005) and Gaulier and Zignago (2010). Since we focus on banana trade, we only keep the data for the country pairs that are involved in the banana trade (SITC code 0573, which includes, both fresh and dried, bananas and plantains). From the import and export value, we can directly measure the banana trade network by computing an indicator variable $l_{i,j,t}$ for bilateral trading activity; it equals unity if a country i reports positive exports with country j in year t, and zero otherwise. The banana trade network is directional: generally, we have $l_{i,j,t} \neq l_{j,i,t}$. We follow earlier studies such as Helpman et al. (2008) and fix the number of countries in the sample by keeping the countries that were actively trading in 1962 and still exist in 2014. As a result, we follow 90 countries and have 8,010 ordered country-pairs throughout our analysis.

13 https://web.stanford.edu/class/e297c/trade_environment/wheeling/hbanana.html.

14 We focus on bananas, because the banana trade network exhibits interesting patterns that are appropriate to be characterized by our model. The full trade network based on all goods, on the other hand, is too dense and shows less interesting dynamics. Of course, our model, could be applied to study other traded goods as well.

Fig. 3. The global banana trade network. *Note:* The graphs depict the global banana trade network from 1962–2014. During these years, we count the number of years each country pair trades with one another. The resulting edge weights of the graph are filtered to show only country pairs that trade with each other for more than 10 years. Additionally, the nodes are sized and colored according to a country (total) degree measure. The more trading partners a country has, the larger and darker the node color.
Source: MIT Observatory of Economic Complexity project and calculations from authors.

Fig. 3 visualizes the global banana trade network from 1962–2014 which is the object of our empirical study. To create the network graph for the entire sample period, the number of years each country pair trades with one another and visualize those aggregate trading flows on a Winkel tripel projection map. The resulting edge weights of the graph are filtered to show only country pairs that trade with each other for more than 10 years. Additionally, the nodes are sized and colored according to a country degree measure. The more trading partners a country has, the larger and darker the node color. The visualization reveals several interesting patterns. First, trade flows between South/Central America and Europe (and to some extent between South-East Asia and Europe) appear to be the most prominent from the graph. Second, several developed countries, such as the United States and the Netherlands, have a high degree, meaning that they are banana trading hubs. Third, within emerging markets trade (i.e., trade within large parts of Africa, Asia, and South/Central America) is relatively modest. These patterns provide evidence of clustering of country pairs in an approximate core–periphery structure.

The summary statistics reveal substantial heterogeneity in banana trade patterns, both at the country-pair-level and country-level. We next use our dynamic factor network model to analyze cross-sectional trading patterns and its dynamics evolution in more detail. In particular, our focus will be on uncovering the latent group memberships and the within and between trade link probabilities.

5.2. Parameter estimation results

We first report and discuss the parameter estimation results for the dynamic factor network model that we adopt for analyzing our data on international banana trade. We used the maximum likelihood importance sampling procedure as outlined in Section 3 for parameter estimation. In the main discussion of our empirical application, we set the number of clusters to $K = 2$ so that we allow for four pair-specific groups, but we also consider and discuss the case of $K = 3$ latent groups in a robustness analysis. For $K = 2$ groups, each node is characterized by a two-dimensional group membership vector γ_i, while each *pair of nodes* is characterized by a $K^2 = 4$-dimensional group membership vector $\delta_{i,j}$. Further, we model the dynamic factors as independent random walks. Hence, we set Φ equal to the identity matrix, and the variance matrix of the innovations Σ_ξ is set to a diagonal matrix. As a benchmark model, we consider the case with no covariates in the model; we have matrix $B = 0$. In this case, the maximum likelihood procedure estimates the 6×1 parameter vector ψ, and we adopt $M = 500$ importance samples to estimate the likelihood function via simulation methods.

As an extension to our benchmark model, we also consider the case when the latent group memberships depend on observed covariates; we have matrix $B \neq 0$. We incorporate observed country information variables which are important determinants of a country's latent group membership in the banana trade network. Specifically, we allow a country's latent group to be a function of (1) the country's size (in thousands of square kilometers), (2) the country's mean elevation (in

Table 4
Model parameter estimates.

$K = 2$ node groups					$K = 3$ node groups				
Parameter	Model (1)		Model (2)		Parameter	Model (3)		Model (4)	
	Estimate	Std. error	Estimate	Std. error		Estimate	Std. error	Estimate	Std. error
μ	−2.5521	0.1201	−2.4524	0.0253	μ_1	−0.4988	0.0086	−0.508	0.0000
σ^2_γ	1.2533	0.0972	1.0521	0.0514	μ_2	0.5432	0.0359	0.5218	0.0993
$\sigma^2_{\xi,1}$	0.2601	0.0253	0.2957	0.0145	$\sigma^2_{\gamma,1}$	1.3111	0.0536	1.3084	0.0000
$\sigma^2_{\xi,2}$	0.2225	0.0260	0.2387	0.0166	$\sigma^2_{\gamma,2}$	1.3114	0.0121	1.3404	0.0000
$\sigma^2_{\xi,3}$	0.1603	0.0183	0.1683	0.0042	$\sigma^2_{\xi,1}$	0.4686	0.0101	0.4835	0.0307
$\sigma^2_{\xi,4}$	0.3176	0.0742	0.2748	0.0075	$\sigma^2_{\xi,2}$	0.1136	0.0094	0.105	0.0000
$\beta_{1,1}$ (Size)			0.3056	0.0483	$\sigma^2_{\xi,3}$	0.0385	0.0002	0.0323	0.0000
$\beta_{1,2}$ (Elevation)			0.4102	0.1801	$\sigma^2_{\xi,4}$	0.2605	0.0016	0.2243	0.0213
$\beta_{1,3}$ (Distance to coast)			−0.2876	0.0210	$\sigma^2_{\xi,5}$	0.0541	0.0004	0.0617	0.0000
$\beta_{1,4}$ (Tropical area)			0.1367	0.1752	$\sigma^2_{\xi,6}$	0.099	0.0014	0.1075	0.0075
$\beta_{1,5}$ (Distance to equator)			0.0207	0.0220	$\sigma^2_{\xi,7}$	0.2817	0.0028	0.3094	0.0343
					$\sigma^2_{\xi,8}$	0.2621	0.0013	0.2671	0.0256
					$\sigma^2_{\xi,9}$	0.3105	0.0057	0.2885	0.0421
					$\beta_{1,1}$			0.7994	0.1046
					$\beta_{1,2}$			0.4932	0.0771
					$\beta_{1,3}$			0.6697	0.2078
					$\beta_{1,4}$			0.8185	0.2078
					$\beta_{1,5}$			0.3896	0.0968
					$\beta_{2,1}$			0.4997	0.0951
					$\beta_{2,2}$			0.6534	0.1536
					$\beta_{3,3}$			0.1572	0.1521
					$\beta_{3,4}$			0.2796	0.0465
					$\beta_{4,5}$			0.5761	0.0475
Avg Loglikelihood	−0.1015		−0.1014			−0.0945		−0.0941	
K (# groups)	2		2			2		2	
N (# nodes)	90		90			90		90	
T (# years)	53		53			53		53	
Total # observations	424,530		424,530			424,530		424,530	
Hill test statistic	3.982		4.20			3.126		4.258	

Note: We present the model parameter estimates with standard errors. The estimates are obtained from the maximum-likelihood importance sampling procedure outlined in Section 3, with $S = 5000$ importance samples. The Hill Estimator tests the null hypothesis that the importance weights have infinite variance, and is based on the $\kappa = 2S^{1/3}$ largest weights. The critical value for $\alpha = 0.05$ is 1.645. The sample includes 90 countries and spans the period from 1962 through 2014, at an annual frequency.

thousands of meters above sea level), (3), the country's mean distance to the nearest coastline or sea navigable river (in kilometers), (4) the percentage of the country's area that is defined as geographic tropics, and (5) the distance to the equator (in thousands of kilometers).[15] In this model with five covariates, $K = 2$ and $K_B = 5$, the matrix B reduces to a five dimensional vector of coefficients since for $K = 2$ we have one degree of freedom in the number of groups. We estimate the 11 parameters by our maximum likelihood procedure.

Table 4 presents the parameter estimate of ψ, together with its standard errors. The estimation results of our benchmark model (1) show a negative estimate for μ and a large estimate for the variance parameter σ^2_γ, which we take as an indication that the distribution of the individual γ_i's is shifted to the left and is highly dispersed. The standard errors also indicate that the estimates are significant at the 5 percent level. The variance parameters for the innovations of the four random walk processes are estimated to be relatively small, but they appear to be significantly different from zero. These estimates imply that the dynamic factors are slowly varying over time but cannot be treated as being constant over time. The average log-likelihood value is maximized at a value of −0.1015. The Hill statistic is a test for the null hypothesis of an infinite variance for the importance sampling weights; the null is indicative that the importance sampling method is not valid. The details of the Hill statistic are presented in the Supplementary Appendix. The value of the Hill statistic is reported in Table 4 and is larger than the 5 percent critical value of 1.645. We can conclude that the null hypothesis can be rejected. This finding supports the validity of our simulation-based estimation procedure as it requires the importance weights to have finite variance.

Model (2) includes our covariates. The estimation results show that despite the addition of the covariates we obtain quite similar parameter estimates for μ, σ^2_γ, and the variances of the four random walks. Moreover, the coefficient

[15] These data come from various sources and are made available online at https://www.pdx.edu/econ/country-geography-data.

estimates on the covariates reveal further information about the data. For example, the positive coefficient estimates for $\beta_{1,1}$ and $\beta_{1,2}$ suggest that larger countries and countries located in high altitudes have a larger μ_i value and thus a higher probability of being part of the second latent group. On the other hand, the negative coefficient estimate for a county's distance to coast suggests that countries without access to water transportation have a higher probability of being part of group 1. Interestingly, conditional on the other factors, the parameters on the share of tropical area in a given country and the distance to equator are not significantly different from zero. The Hill statistics also provide in this case evidence for the validity of our simulation-based estimation procedure.

5.3. Network estimation results

The conditional distributions of all latent factors in our model provide us with insights into the country- and country-pair-specific roles and with information about the model-implied latent group memberships of individual countries. From the conditional distribution we compute the smoothed between- and within-group link probabilities of banana trade throughout our sample period. In our discussion, we focus on the conditional distributions of latent factors obtained from Model (1). Indeed, we have found that the conditional distributions obtained from Model (2) that includes additional covariates are very similar. For example, the average absolute difference between the dynamic factor estimates from the model with covariates and those from the model without covariates is only 1.1 percent (the relative average difference is 3.45 percent). These small differences suggest that the inclusion of covariates, while statistically significant, for the latent group vectors do not lead to substantially different conditional distributions.

In Fig. 4, we present for each of the 90 countries in our sample the estimated probability of being in either the first or the second latent group. For instance, the smoothed estimates indicate a probability of about 15 percent that Brazil belongs to group 1 (dark/red bar), and a probability of about 85 percent that Brazil belongs to group 2 (light/yellow bar). On the other hand, Germany, for example, has a substantially larger probability of belonging to group 1 with a value of about 0.45. The estimated structure implied by the model appears to allocate a higher probability of group 1 membership to other European countries that are import and export hubs in the global banana trade network; for example, the Netherlands, France and Great Britain. However, also the United States have a higher probability of being in group 1. Smaller countries, which are not important in the banana trade network on neither the import or export side, such as Afghanistan or Qatar, are mostly allocated to group 2 (light/yellow bars). Hence we label group 1 as "strong banana traders" and group 2 as "low banana traders".

In terms of bilateral trade patterns, increases in the efficiencies of banana production and distribution (through the technological changes related to fertilization or transportation) may affect trade link probabilities differently between, say, Ecuador and the Netherlands than the way these technological changes may affect trade between Ecuador and Afghanistan. Hence, trade link probabilities and their dynamics are pair specific. This key feature of network data is incorporated in our proposed dynamic factor network model through the multiplicative interaction ($\delta_{i,j} = \pi_i \otimes \pi_j$) of the country-specific stochastic group memberships π_i. Indeed, as we will show, looking at individual country-pairs' trade probabilities is important to understand the full pattern of the banana trade network and its dynamics.

In Fig. 5, we present for each country-pair in the sample the modal value (the highest estimated probability) of the group-membership vector at the pair level, $z_{i,j}$. In this figure, the exporting countries are on the vertical axis and the importing countries are on the horizontal axis. For example, when considering a banana export from Brazil to Germany, the model and the parameter estimates assign the highest probability that Brazil belongs to group 1 and Germany belongs to group 1 in this interaction (blue coloring). On the other hand, for an export from Brazil to Afghanistan, the model assigns the highest probability that both countries belong to group 2 in this interaction (yellow coloring). However, when considering an export to Afghanistan coming from Pakistan, we would characterize Afghanistan as being a member of group 1 (instead of group 2) and Pakistan a member of group 2 (orange coloring).

We may conclude from Fig. 5 that the pair-specific estimated modal group memberships imply an approximate block structure for the banana trade network, where South and Latin American exporting countries (such as Colombia, Ecuador, Honduras) and developed countries that are banana trade hubs (such as Germany, the Netherlands, the United States) tend to behave in a similar way overall (the green blocks, group 1 to group 2 trades). These trade patterns may be thought of as the periphery to core trades. The trade among country pairs from the developed world behave similarly overall (the predominately blue blocks, group 1 to group 1 trades). These pairs include core-to-core trades between trade hub countries. The graph also reveals that most country pairs consisting of exporters and importers that are not part of the trade hub nor the typical producer countries are also estimated as behaving in a similar way (yellow blocks, group 2 to group 2 trades). These pairs can be viewed as the peripheral pairs. Finally, there is also some heterogeneity in the periphery-to-periphery patterns as indicated by the orange blocks (group 2 to group 1 trades).

We further assess our proposed method's ability to recover the latent cross-sectional network structure estimated from empirical trade data using a Monte Carlo (MC) study. Specifically, we simulate $S = 500$ sets of network data, $y^{(s)}$, given the smoothed signal estimated from the empirical data, $E(\phi|y)$, and then, for each MC draw, we re-estimate the parameters and smoothed latent variable structure; in particular, the node- and link-specific group memberships π_i and $z_{i,j}$. Hence, we generate data that is similar to the observed empirical network and verify how close the estimated mean latent structures from the different simulated data sets are to those obtained from the empirical data. Our key statistic for assessing the "fit" is the mean absolute error (MAE) of the elements of the mean group membership vectors π_i and $z_{i,j}$ obtained in the MC and from the empirical data: $|E(\pi_{i,k}^{(s)}|y^{(s)}) - E(\pi_{i,k}|y)|$ and $|E(z_{i,j,m}^{(s)}|y^{(s)}) - E(z_{i,j,m}|y)|$.

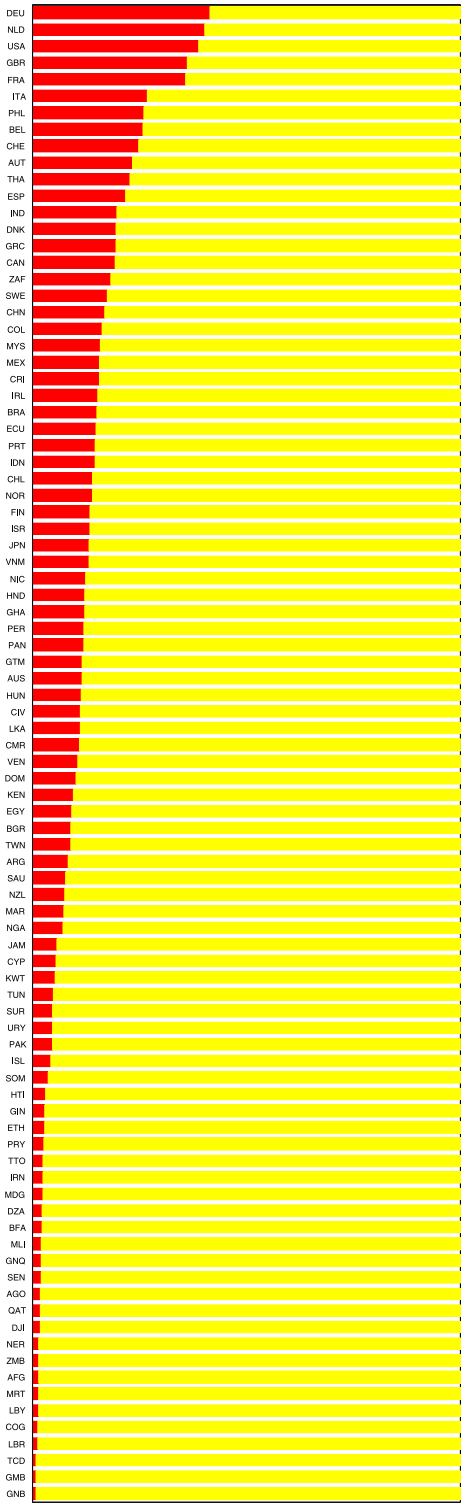

Fig. 4. Estimated mean of group-membership vectors at the country level. We present for each country i in the sample the estimated mean of the two-dimensional mixed group membership vector π_i. The dark red bar corresponds to the probability that bank i is allocated to the first group (the first element of π_i). The light yellow bar corresponds to the probability that bank i is allocated to the second group (the second element of π_i). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
Source: Authors' calculations.

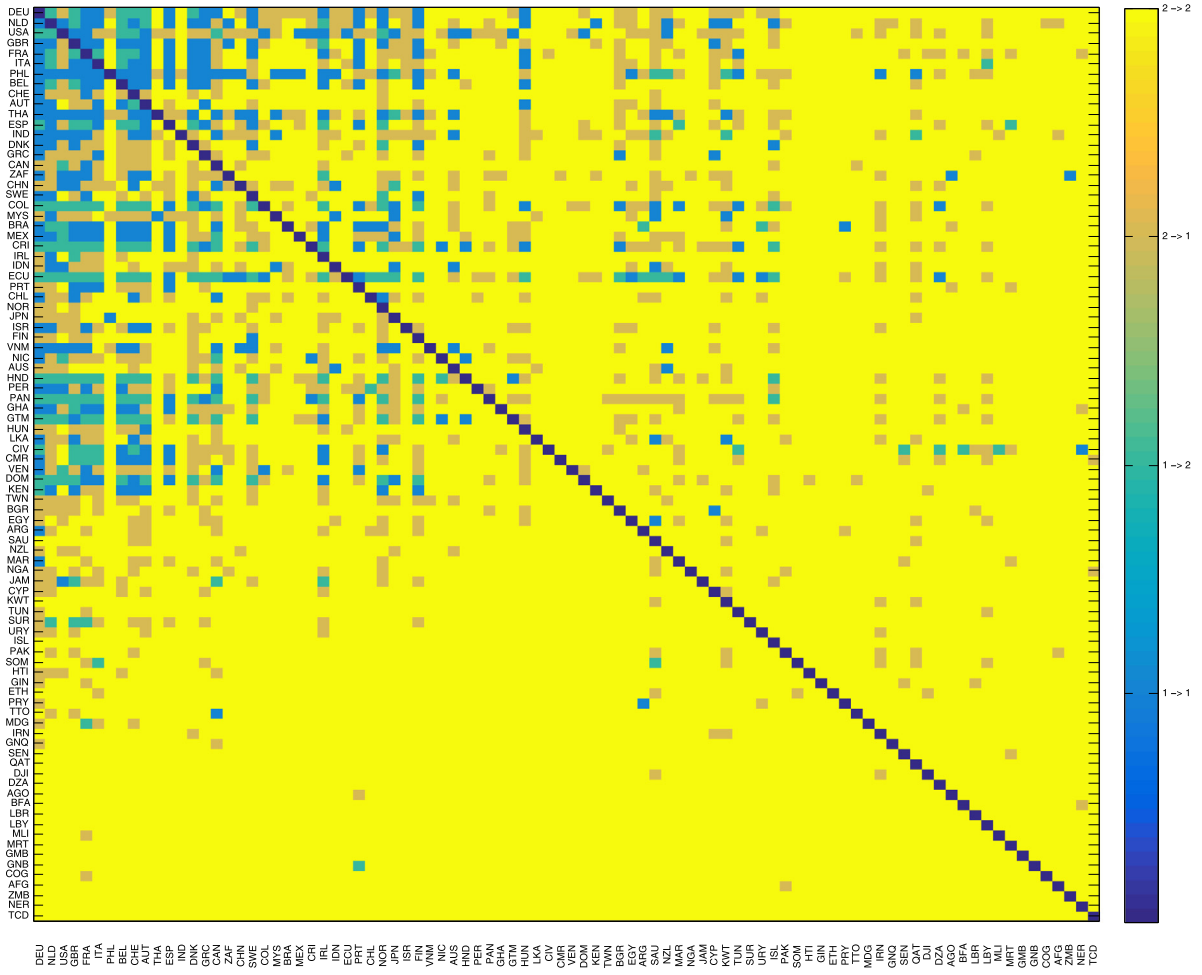

Fig. 5. Estimated mode of group membership at the country pair level. We present the estimated modal value of the pair-level group-membership vectors, $z_{i,j} \sim \mathcal{MN}(\delta_{i,j})$, with $\delta_{i,j} = \pi_i \otimes \pi_j$. The blue entries correspond to country pairs (i, j) where both country i and country j are allocated to group 1; green entries correspond to country pairs where country i is allocated to group 1 and country j is allocated to group 2; orange entries correspond to country pairs where country i is allocated to group 2 and country j is allocated to group 1; dark red entries correspond to country pairs where both country i and country j are allocated to group 2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
Source: Authors' calculations.

Table 5
MC evidence of errors in latent network structure.

	Percentiles of MAE				
	p50	p75	p90	p95	p99
Group membership at node level (π_i)	0.0020	0.0048	0.0078	0.0102	0.0206
Group membership at link level ($z_{i,j}$)	0.0000	0.0002	0.0035	0.0401	0.1782

Note: We present the distribution of mean absolute errors of the elements of the mean group membership vectors π_i and $z_{i,j}$ obtained in the MC and from the empirical data: $|E(\pi_{i,k}^{(s)}|y^{(s)}) - E(\pi_{i,k}|y)|$ and $|E(z_{i,j,m}^{(s)}|y^{(s)}) - E(z_{i,j,m}|y)|$. The MC study is based on $S = 500$ replications.

Table 5 presents the distribution of the MAEs (across links/nodes and MC draws). The results show that the errors in the estimated mean group membership probabilities are smaller than 1 percentage point in 90 percent of the cases, both at the link and node level. However, there are also some larger errors: for example, for 5 percent of the observations, the error in the node-level mean group membership is larger than 1 percentage point, and the error in the link-level mean group membership is larger than 4 percentage points. Overall, given the simulation study and the evidence from the previous section, we conclude that our model estimation recovers similar group membership vectors from the different MC draws.

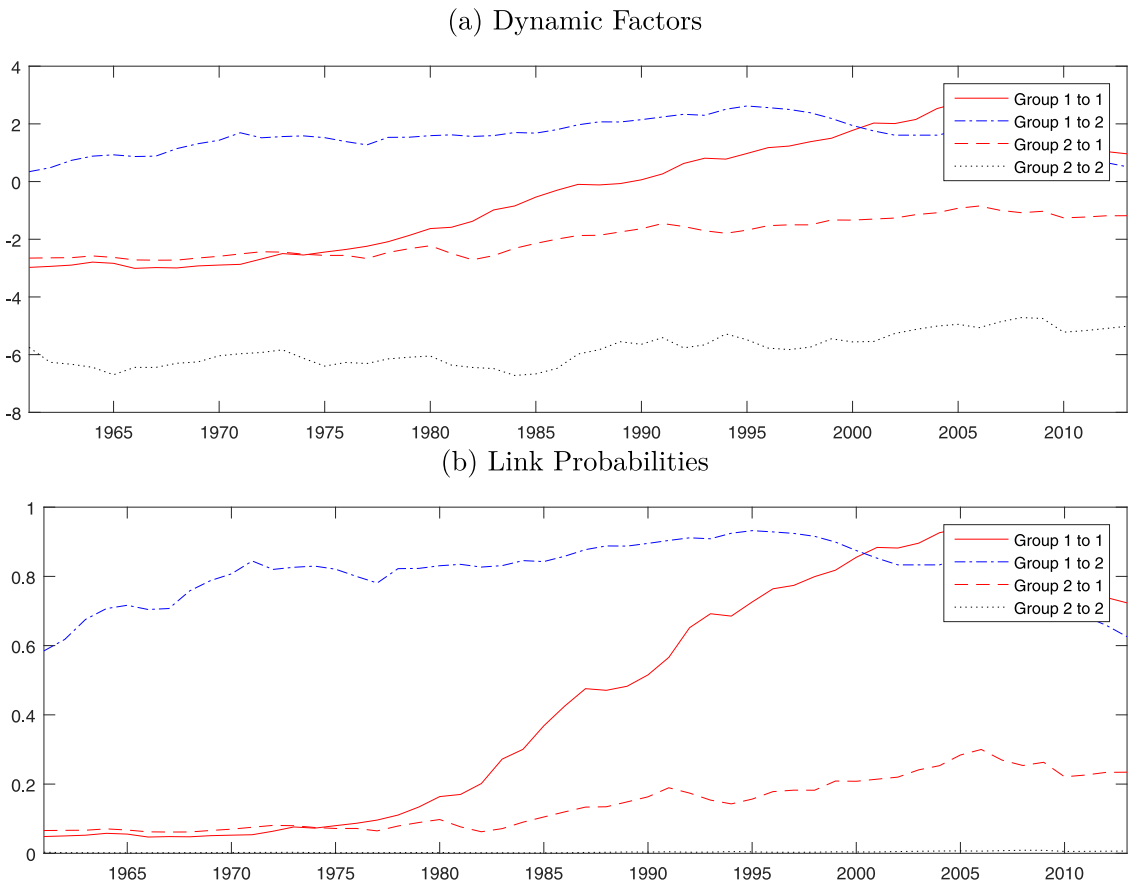

Fig. 6. Smoothed dynamic factors and trade link probabilities. We present the smoothed mean of the within- and between-group factors and link probabilities, where a link indicates a trade between two countries.
Source: Authors' calculations.

5.4. Dynamic factor estimation results

The dynamic factors are associated with the bilateral banana trade probabilities between and within the estimated country-pair groups, at each time point t. It is of interest to investigate how these factors have changed over time during our sample period, given various changes that affect banana trade, such as changes in transportation (refrigerated containers since the 70s), higher-yielding varieties, or adjustment in international trade agreements. In Fig. 6, we present the estimated dynamic factors and the associated mean probabilities, over time and for each of the four pair-specific group combinations. The time series graphs show how likely a banana export is for each group and how these probabilities evolve over time within our sample period. The first panel presents the estimated factors and the second panel presents the associated link probabilities. We focus on the discussion of the link probabilities.

The solid red line represents the within-group link probabilities of countries from group 1 to other countries from group 1, corresponding to the country pairs colored blue in Fig. 5. This link probability is generally low at the beginning of our sample (below 10 percent), but it shows a clear increase since the 1980s to link probabilities of about 90 percent in the early 2000s, as well as a slight decrease in trade probabilities after the global financial crisis. The overall increase is a reflection of the rising number of (mostly) developing countries trading with new export countries. During this period we have also witnessed improvements in trade and storage technologies, the development of new more resistant banana varieties, and the establishment of trade agreements and the World Trade Organization (WTO) in 1994, all of which are likely to have fostered the banana trade, in particular across long distances and potentially through "hub" countries. Similar empirical findings have been found in earlier studies, also for other commodities; see the discussion in Helpman et al. (2008) and the references therein. The dashed–dotted blue line in Fig. 6 depicts the across-group link probability from countries of group 1 to countries of group 2, corresponding to country pairs colored green in Fig. 5. The link probability for spillovers from group 1 to group 2 is generally high (above 60 percent) throughout our sample period with some slight decrease since the mid-1990s. Moreover, link probabilities between group 2 countries and group 1 countries, corresponding to country pairs colored in orange in Fig. 5, are relatively low (10 to 20 percent) but have

been increasing since the 1980s (dashed red line). Finally, the within-group-2 link probabilities are very close to zero throughout the entire sample.

5.5. Discussion of results with $K = 3$ groups

To verify whether we have missed some empirical features in our analysis based on the model with $K = 2$, we next consider the same model with $K = 3$ latent node groups, that is $K^2 = 9$ link groups. The estimation results for the model with $K = 3$ are presented in Table 4 and it reveals that the log likelihood value does only increase modestly after adding one more latent group, for models with and without covariates. However, the results for a model with $K = 3$ does provide a more nuanced view of the latent country-pair group memberships and of the dynamic evolutions in the international banana trade network. We next discuss the key findings for this extended model.

In the Supplementary Appendix, we provide a Table for $K = 3$ groups and report the modal values of the pair-level group membership for each country pair in the sample, and a Table for the smoothed within and between trade link probabilities. For example, using the model with $K = 3$, we find that some country pairs, such as Philippines and Canada, or Ecuador and Spain (Group 1 to 1), are subject to an increased trade link probability from the late 1970s, while for other country pairs the increased trade probability are starting later, from the mid-1980s (Group 3 to 1). Those country-pairs include China and France, or Venezuela and the Netherlands. We also find here that in general trade probabilities decreased after the great recession in 2008/09. Interestingly, the results for a model with $K = 3$ groups show that some country pairs decline in trade in the late 1980s (Group 3 to 3); for example, Kenya and Germany, and Nicaragua and the Netherlands. This last finding has been missed by the estimation results for a model with $K = 2$ groups. Moreover, and consistent with our results for the model with $K = 2$ groups, we find that most country pairs have a trade probability close to zero.

6. Conclusion

We have developed a dynamic factor network model that is sufficiently general and flexible for all practical purposes. On the other hand, the parametric structure is parsimonious, with many features of the model being treated by stochastic variables. For the cross-sectional part of the model, we have adopted the stochastic blockmodel structure that is widely used for the analysis of network data. For the time series part of the model, we have adopted the dynamic factor model. The nodes of the network are linked as functions of the dynamic latent factors. The typically small number of parameters can be estimated by our proposed computationally efficient Monte Carlo maximum likelihood method. The simulations are carried out with the purpose of integrating out the latent node- and link-specific vectors and the dynamic factors, simultaneously. The method of importance sampling is key in our procedure, and we have proposed a key modification in collapsing the high-dimensional observational vectors to achieve a feasible estimation procedure without losing any information. In a simulation study we find that the small-sample properties of our parameter estimates are favorable under correct and incorrect model specifications, from estimation accuracy and computational efficiency perspectives. In the empirical study we analyze bilateral banana trade between 90 countries from 1962 through 2014. In partitioning the nodes into two groups, we find that the first group is associated with banana trade hub countries, which include key producing and distribution countries, while the second group is representing countries that have fewer trading partners, which are typically trade hub countries. The estimated link probabilities show a strong increase in trade starting from the 1980s onwards but trade slow down after the great recession from 2008/09.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2019.10.007.

References

Acemoglu, D., Akcigit, U., Kerr, W., 2015. Networks and the macroeconomy: An empirical exploration. Working Paper.
Acemoglu, D., Carvalho, V., Ozdaglar, A., Tahbaz-Salehi, A., 2012. The network origins of aggregate fluctuations. Econometrica 80 (5), 1977–2016.
Aguilar, O., West, M., 2000. Bayesian dynamic factor models and portfolio allocation. J. Bus. Econom. Statist. 18 (3), 338–357.
Airoldi, E.M., Blei, D.M., Fienberg, S.E., Xing, E.P., 2008. Mixed membership stochastic blockmodels. J. Mach. Learn. Res. 9, 1981–2014.
Aït-Sahalia, Y., Cacho-Diaz, J., Laeven, R.J., 2015. Modeling financial contagion using mutually exciting jump processes. J. Financ. Econ. 117 (3), 585–606.
Blasques, F., Brauning, F., van Lelyveld, I., 2018. A dynamic stochastic network model of the unsecured interbank lending market. J. Econom. Dynam. Control 90 (C), 310–342.
Bonhomme, S., Manresa, E., 2015. Grouped patterns of heterogeneity in panel data. Econometrica 83 (3), 1147–1184.
Chaney, T., 2014. The network structure of international trade. Amer. Econ. Rev. 104 (11), 3600–3634.
Craig, B.R., von Peter, G., 2014. Interbank tiering and money center banks. J. Financ. Intermed. 23 (3), 322–347.
Creal, D.D., 2012. A survey of sequential Monte Carlo methods for economics and finance. Econometric Rev. 31, 245–296.
Durbin, J., Koopman, S.J., 1997. Monte Carlo maximum likelihood estimation for non-Gaussian state space models. Biometrika 84 (3), 669–684.
Durbin, J., Koopman, S.J., 2002. A simple and efficient simulation smoother for state space time series analysis. Biometrika 89 (3), 603–616.
Durbin, J., Koopman, S.J., 2012. Time Series Analysis By State Space Methods, second ed. In: Oxford Statistical Science Series, Oxford University Press, Oxford, UK.

Feenstra, R., Lipsey, R., Deng, H., Ma, A.C., Mo, H., 2005. World trade flows: 1962-2000. NBER Working Papers 11040, National Bureau of Economic Research, Inc.
Frühwirth-Schnatter, S., 2006. Finite Mixture and Markov Switching Models. Springer Verlag, New York.
Gamerman, D., Lopes, H.F., 2006. Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference, second ed. Chapman and Hall, CRC Press, Boca Raton, FL.
Gaulier, G., Zignago, S., 2010. BACI: International trade database at the product-level. The 1994-2007 version. Working Papers 2010-23, CEPII research center.
Geweke, J., 1989. Bayesian inference in econometric models using Monte Carlo integration. Econometrica 57 (6), 1317–1339.
Geweke, J.F., 1991. Efficient Simulation from the Multivariate Normal and Student-t Distributions Subject to Linear Constraints. Computer Science and Statistics. in: Proceedings of the 23rd Symposium on the Interface. Seattle Washington, April (1991) 21-24, pp. 571–578.
Gilks, W., Richardson, S., Spiegelhalter, D. (Eds.), 1995. Markov Chain Monte Carlo in Practice. Chapman and Hall, CRC Press, Boca Raton, FL.
Hajivassiliou, V., 1990. Smooth Simulation Estimation of Panel Data LDV Models. Mimeo, Yale University, page.
Hajivassiliou, V., McFadden, D., Ruud, P., 1996. Simulation of multivariate normal rectangle probabilities and their derivatives. Theoretical and computational results. J. Econometrics 72, 85–134.
Hallam, D., 1995. the World Banana Economy. Springer Verlag, Dordrecht.
Helpman, E., Melitz, M., Rubinstein, Y., 2008. Estimating trade flows: Trading partners and trading volumes. Q. J. Econ. 123, 441–487.
Hoff, P.D., 2009. Multiplicative latent factor models for description and prediction of social networks. Comput. Math. Organ. Theory 15 (4), 261–272.
Hoff, P.D., Raftery, A.E., Handcock, M.S., 2002. Latent space approaches to social network analysis. J. Amer. Statist. Assoc. 97 (460), 1090–1098.
Jackson, M.O., 2008. Social and Economic Networks. Princeton University Press, Princeton, NJ, USA.
Jordan, M.I., Ghahramani, Z., Jaakkola, T.S., Saul, L.K., 1999. An introduction to variational methods for graphical models. Mach. Learn. 37 (2), 183–233.
Jung, R.C., Liesenfeld, R., Richard, J.F., 2011. Dynamic factor models for multivariate count data: An application to stock-market trading activity. J. Bus. Econom. Statist. 29 (1), 73–85.
Jungbacker, B., Koopman, S.J., 2015. Likelihood-based dynamic factor analysis for measurement and forecasting. Econom. J. 18 (2), C1–C21.
Keane, M.P., 1994. A computationally practical simulation estimator for panel data. Econometrica 62, 95–116.
Kolaczyk, E.D., 2009. Statistical Analysis of Network Data: Methods and Models, first ed. Springer Publishing Company, Incorporated.
Koopman, S.J., Lucas, A., Scharth, M., 2014. Numerically accelerated importance sampling for nonlinear non-Gaussian state space models. J. Bus. Econom. Statist. 33 (1), 114–127.
Koopman, S.J., Mesters, G., 2017. Empirical Bayes methods for dynamic factor models. Rev. Econ. Stat. 99 (3), 486–498.
Koopman, S.J., Shephard, N., Creal, D.D., 2009. Testing the assumptions behind importance sampling. J. Econometrics 149, 2–11.
Liesenfeld, R., Richard, J.F., 2010. Efficient estimation of probit modeld with correlated errors. J. Econometrics 156, 367–376.
Lopes, H.F., Salazar, E., Gamerman, D., 2008. Spatial dynamic factor analysis. Bayesian Anal. 3 (4), 759–792.
Manski, C.F., 1993. Identification of endogenous social effects: The reflection problem. Rev. Econom. Stud. 60 (3), 531–542.
Mesters, G., Koopman, S.J., 2014. Generalized dynamic panel data models with random effects for cross-section and time. J. Econometrics 180 (2), 127–140.
Monahan, J.F., 1993. Testing the Behaviour of Importance Sampling Weights. in: Computer Science and Statistics: Proceedings of the 25th Annual Symposium on the Interface, pp. 112–117.
Newey, W.K., McFadden, D.L., 1994. Large sample estimation and hypothesis testing. In: Engle, R.F., McFadden, D.L. (Eds.), HandBook of Econometrics. Elsevier, Amsterdam, pp. 2111–2245.
Nowicki, K., Snijders, T.A.B., 2001. Estimation and prediction for stochastic blockstructures. J. Amer. Statist. Assoc. 96 (455), 1077–1087.
Ormerod, J.T., Wand, M.P., 2010. Explaining variational approximations. Amer. Statist. 64 (2), 140–153.
Pollard, D., 1984. Convergence of Stochastic Processes. Springer, New York.
Richard, J.F., Zhang, W., 2007. Efficient high-dimensional importance sampling. J. Econometrics 141, 1385–1411.
Ripley, B.D., 1987. Stochastic Simulation. Wiley, New York.
Sewell, D.K., Chen, Y., 2015. Latent space models for dynamic networks. J. Amer. Statist. Assoc. 110 (512), 1646–1657.
Shephard, N., Pitt, M.K., 1997. Likelihood analysis of non-Gaussian measurement time series. Biometrika 84 (3), 653–667.
Snyder, D., Kick, E.L., 1979. Structural position in the world system and economic growth, 1955-1970: A multiple-network analysis of transnational interactions. Am. J. Sociol. 84 (5), 1096–1126.
van der Vaart, A., 1998. Asymptotic Statistics. van der Vaart A Cambridge University Press, Cambridge.
Wasserman, S., Faust, K., 1994. Social Network Analysis: Methods and Applications. Structural Analysis in the Social Sciences. Cambridge University Press, Cambridge, UK.
Xing, E.P., Fu, W., Song, L., 2008. A state-space mixed membership blockmodel for dynamic network tomography. Ann. Appl. Stat. 4 (2), 535–566.
Xing, E.P., Jordan, M.I., Russell, S.J., 2003. A generalized mean field algorithm for variational inference in exponential families. In: Uncertainty in Artificial Intelligence. pp. 583–591.