Bräuning, Falk; Koopman, Siem Jan

Working Paper

The Dynamic Factor Network Model with an Application to Global Credit-Risk

Tinbergen Institute Discussion Paper, No. 16-105/III

Provided in Cooperation with:
Tinbergen Institute, Amsterdam and Rotterdam

TI 2016-105/III
Tinbergen Institute Discussion Paper

The Dynamic Factor Network Model with an Application to Global Credit-Risk

Falk Bräuning[a]
Siem Jan Koopman[b]

[a] *Federal Reserve Bank of Boston, United States;*
[b] *Faculty of Economics and Business Administration, VU University Amsterdam, and Tinbergen Institute, The Netherlands.*

The Dynamic Factor Network Model with an Application to Global Credit Risk*

Falk Bräuning[(a)] and *Siem Jan Koopman*[(b)]

[(a)] Federal Reserve Bank of Boston

[(b)] Vrije Universiteit Amsterdam, Tinbergen Institute, CREATES

November 23, 2016

Abstract

We introduce a dynamic network model with probabilistic link functions that depend on stochastically time-varying parameters. We adopt the widely used 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 is straightforward and transparent by nature. However, parameter estimation, signal extraction of the 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. A Monte Carlo study is carried out to provide evidence of how well the methods work. In an empirical study, we use the novel framework to analyze a database of significance-flags of Granger causality tests for pair-wise credit default swap spreads of 61 different banks from the United States and Europe. Based on our model, we recover two groups that we characterize as "local" and "international" banks. The credit-risk spillovers take place between banks, from the same and from different groups, but the intensities change over time as we have witnessed during the financial crisis and the sovereign debt crisis.

Keywords: Network Analysis, Dynamic Factor Models, Blockmodels, Credit-Risk Spillovers.

JEL Codes: C32, C58, G15.

*Emails: `falk.braeuning@bos.frb.org` and `s.j.koopman@vu.nl`. We thank Botao Wu for excellent 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.

1 Introduction

In recent years, economists have become increasingly interested in analyzing social and economic networks; an overview of the early literature is provided by Jackson (2008). These networks arise naturally in many fields of economics. Key examples are the input-output matrix of an economy, international trade relationships, and interbank lending networks. The relevance of studying such networks is evident, since the interactions of economic agents, such as firms, households, and governments, are generally at the core of economic activity. However, the econometric tools necessary for the empirical analysis of data that reflect pairwise interactions between economic agents, which we refer to as network data, are still in their infancy. In particular, most existing statistical methods focus on static networks and do not address models of the dynamic behavior of time-varying networks; see also the discussions in Kolaczyk (2009).

In this study, we develop a dynamic modeling framework for analyzing network data. In particular, we propose a dynamic factor network model to analyze time-varying heterogeneity in a binary link formation. Our 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 we cast into state space form. 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. Our proposed modeling framework is flexible, but also parsimonious, and it allows for the analysis of various 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).

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 our model formulation allows for efficient simulated maximum likelihood estimation, using importance sampling procedures to integrate 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 for which an approximating linear, Gaussian, state space model is used to draw samples of the latent factors. 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 to the correct conditional distribution.

The method of importance sampling is detailed in 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 treatment of both 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.

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) and Airoldi et al. (2008). The core-periphery models are special, non-stochastic versions of blockmodels, and they have gained interest in the financial literature. For example, the structure of the German interbank lending market has been described as blocks of core and peripheral banks by Craig and von Peter (2014). Similar studies have been carried out for Italian and Dutch data by Fricke and Lux (2015) and van Lelyveld and in 't Veld (2014), respectively. We contribute to this literature by introducing a time-varying link density in a stochastic blockmodel using dynamic latent factors. However, we do not base the challenging task of parameter estimation on an approximate model, which is the common approach to inference for stochastic blockmodels. Instead, we develop importance sampling methods to perform maximum likelihood parameter estimation based on the true model. Related inference procedures for (static) network models with random effects include Bayesian parameter estimation procedures as discussed in Hoff et al. (2002). 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. 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.

A Monte Carlo study is carried out 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. The results also suggest that the efficiency of the estimator increases in both the time and cross-section dimensions. Finally, 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).

In the empirical study, we analyze credit-risk spillovers between 61 main financial firms throughout the world from 2007 through 2015. To construct the credit-risk spillover networks, we follow Billio et al. (2012) to compute all significant positive Granger causalities from all pairwise bivariate VAR models of the daily changes of the CDS spreads of these 61 financial firms. To measure the time evolution, we estimate these Granger-causalities for 115 monthly subsamples and analyze the time evolution using the proposed dynamic factor network model. Our key findings are as follows. The estimated latent block structure suggests that we can partition the nodes into two groups. The first group is associated with large U.S. banks (such as Bank of America and Goldman Sachs) and large European banks (such as Deutsche Bank or Credit Agricole). The second group consists of the remaining (mostly) European banks, which supposedly have a smaller global scope. The estimated latent factors suggest a strong increase in spillovers within and between all groups during the 2008 financial crisis and the sovereign debt crisis. The heterogeneous effects show that spillovers from global to local banks are most sizeable, while spillovers from local to global banks are generally smaller. Moreover, spillovers among local European banks became elevated during the European debt crisis and again from 2014 until the end of our sample period.

The remainder of the paper is structured 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 measure time-varying credit-risk spillovers in the financial industry, and 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* $\mathcal{L}_t$.[1] The existence of a link between nodes i and j, $i, j \in \mathcal{N} = \{1, 2, ..., N\}$, where N is the number of nodes, at discrete time periods $t = 1, ..., T$ 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, the direction of the relationship does not matter and we have $y_{i,j,t} = y_{j,i,t}$.[2] 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 $N^2 \times 1$.

We are mainly concerned with the treatment of directed networks, although our framework easily accommodates undirected networks; see the discussion in Appendix A. We model each link between any two nodes i and j in the network at time t by the binary random variable

$$y_{i,j,t} \sim \mathcal{B}(\phi_{i,j,t}), \tag{1}$$

where $\mathcal{B}(\cdot)$ is the Bernoulli distribution and the probability of success $\phi_{i,j,t} \in (0, 1)$ is formulated in terms of the logistic transformation of the signal $\theta_{i,j,t} \in \mathbb{R}$, that is

$$\phi_{i,j,t} = \frac{\exp(\theta_{i,j,t})}{1 + \exp(\theta_{i,j,t})}. \tag{2}$$

[1] In graph theory, nodes are also referred to as vertexes, and links are referred to as edges. Our exposition is based on dynamic networks with a constant number of nodes, but in Appendix A we show that our framework also allows for time-varying network sizes.

[2] 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 $y_{i,i,t} = 0 \, \forall i \in \mathcal{N}$.

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

$$\theta_{i,j,t} = z'_{i,j} f_t, \tag{3}$$

where $f_t \in \mathbb{R}^M$ is a vector of common dynamic factors and $z_{i,j}$ is a pair-specific factor loading with $z_{i,j} \in \mathbb{Z}^M := \{\{0,1\}^M : \sum_{m=1}^M z_{i,j,m} = 1\}$ with dimension $M \geq 1$, where $z_{i,j,k}$ is the kth element of column vector $z_{i,j}$ for $k = 1, \ldots, M$. Hence, $z_{i,j}$ selects an element of f_t. The signal specification given in Equation (3) separates the cross-section and the time series features of our model by means of the $M \times 1$ vectors $z_{i,j}$ and f_t, respectively. Both vectors are assumed to be independent random vectors that come from certain parametric distributions.

2.1 Cross-Sectional Features of 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 multinomial random vector that can be written as a Kronecker tensor product,

$$z_{i,j} = z_{i \to j} \otimes z_{j \leftarrow i}, \tag{4}$$

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 $k = 1, \ldots, K$.[3] We model the link-specific loadings as a function of two node-specific vectors. In particular, for each node i in the network, the

[3]Since each node can be in one of the K groups, each pair of nodes can be in $M = K^2$ groups. Moreover, since we are modeling directional interactions, we have $z_{j \leftarrow i} \neq z_{i \leftarrow j}$ and $z_{i \leftarrow k} \neq z_{i \leftarrow l}$.

latent pair-specific group membership vectors are modeled by the hierarchical structure

$$z_{i\to j}, z_{i\leftarrow j} \overset{iid}{\sim} \mathcal{MN}(\pi_i) \text{ for all } j, \quad \pi_i = \frac{\exp(\gamma_i)}{\sum_{k=1}^{K}\exp(\gamma_{i,k})}, \quad \gamma_i \overset{iid}{\sim} \mathcal{N}(\mu, \Sigma_\gamma), \tag{5}$$

where $\mathcal{MN}(\cdot)$ indicates the multinomial distribution with multinomial probability vector π_i that is specified as an element-wise logistic transformation of the $K \times 1$ vector γ_i, with γ_i being drawn from a multivariate Gaussian distribution. The $K \times 1$ mean vector μ and the $K \times K$ variance matrix Σ_γ are unknown.[4] The model structure in terms of Equation (4) can be directly expressed as $z_{i,j} \sim \mathcal{MN}(\delta_{i,j})$ with probability vector $\delta_{i,j} = \pi_i \otimes \pi_j$.

2.2 Time Series Features of Model

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

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

where Φ is the $M \times M$ autoregressive coefficient matrix and ξ_t is treated as an $M \times 1$ independent Gaussian innovation vector with mean zero and variance matrix Σ_ξ. The initial distribution for f_1 is also assumed to be a Gaussian vector with mean zero and variance matrix V, where V is the stationary solution for the matrix equation $V = \Phi V \Phi' + \Sigma_\xi$ (for a discussion on initial conditions, see Durbin and Koopman 2012). The dynamic process for f_t is effectively a vector autoregressive process of order 1, that is, a VAR(1), but it can be extended toward any linear dynamic system that relies on Gaussian innovations including higher-order vector autoregressive moving average processes.

[4] More precisely, we only have $K - 1$ degrees of freedom because the logistic transformation is not identified otherwise. In the exposition we abstract from making this explicit for ease of notation.

2.3 Discussion of Model

The specification of the vectors $z_{i,j}$ and f_t completes the description of the 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 network data, such as core-periphery structures, potential clustering, homophily, and stochastic equivalence; see the discussion, for example, in Wasserman and Faust (1994), Kolaczyk (2009) and Jackson (2008). In particular, our model allows for clustering, which is not accommodated by other latent factor network models, as described in Hoff et al. (2002). In our modeling framework, two pairs 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.

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. However, their model is concerned with ordinary panel data and does not apply to the analysis of network data. In this paper, 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 stochastic Gaussian 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.

Overall, the modeling choice introduces a powerful, yet parsimonious and stylized modeling framework for the time-varying, cross-sectional dependency in network data. It allows us to construct a computationally efficient importance sampling procedure for maximum likelihood parameter estimation. In particular, from an econometric inference point of view, our goal is to estimate the parameters of the model and to obtain conditional distributions of the latent link- and node-specific group memberships as well as the latent dynamic factors. 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, i.e., $Z = [z'_{1,1}, ..., z'_{i,j-1}, z'_{i,j}, ..., z'_{i,N}, z'_{i+1,1}, ...]$. 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

In this section, we discuss the estimation of the parameter vector ψ that collects the elements of the vector and matrices μ, Σ_γ, Φ, and Σ_ξ from model Equations (5) and (6) by the method of maximum likelihood. The likelihood function is defined as $L(\psi) = p(y; \psi)$, where $p(y; \psi)$ 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

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

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

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

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 m-th 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

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

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

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

where the expectation is with respect to $p(f,z,\gamma) = p(f)p(z|\gamma)p(\gamma)$. Equation (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) for a general discussion. The importance sampling representation of the likelihood function is given by

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

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 $g(\cdot|y)$ and estimate the likelihood as

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

This estimator is consistent, and conditions under which the central limit theorem applies are discussed by Geweke (1989) and can be empirically tested as discussed by Monahan (1993) and Koopman et al. (2009); see also the discussion in Appendix C. We can then use $\hat{L}(\psi)$ to obtain the simulated maximum likelihood estimator $\hat{\psi}$ as

$$\hat{\psi} = \arg\max_{\psi} \hat{L}(\psi; y), \tag{8}$$

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 such matters.

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.[5]

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.[6] Formally, this factorization gives the two equations

$$g_\gamma(\gamma_i) = p(\gamma_i|\langle z_{i\to\cdot}\rangle_{g_z}, \langle z_{i\leftarrow\cdot}\rangle_{g_z}) \tag{9}$$

$$g_z(z_{i\to j}, z_{j\leftarrow i}) = p(z_{i\to j}, z_{j\leftarrow i}|y_{i,j,t}, f, \langle\gamma_i\rangle_{g_\gamma}, \langle\gamma_j\rangle_{g_\gamma}), \tag{10}$$

where $\langle\cdot\rangle_f$ denotes the expected Markov blanket, that is, all children, parents, and co-parents

[5] 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). In Appendix B, we present a brief introduction for the unfamiliar reader.

[6] 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.

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

$$g_{z_{i,j}}(z_{i\to j}, z_{j\leftarrow i}) \propto p(z_{i\to j}|\langle\gamma_i\rangle_{g_\gamma})p(z_{j\leftarrow i}|\langle\gamma_j\rangle_{g_\gamma}) \prod_{t=1}^{T} p(y_{i,j,t}|z_{i,j}, f_t), \tag{11}$$

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 $\delta_{i,j} = \pi_i \otimes \pi_j$.

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 u-th element of $z_{i\to j,u}$, and, similarly, $z_{j\leftarrow i,v}$ is the v-th 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 k-th 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 m-th element of the vector $\hat{z}_{i,j}$, with $m = 1, ..., M$. 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

$$y_{m,t}^* = c_{m,t}^* + f_{m,t} + u_{m,t}^*, \qquad u_{m,t}^* \overset{iid}{\sim} \mathcal{N}(0, \sigma^2_{u^*_{m,t}}), \tag{12}$$

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 Equation (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 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 $K^2 << N^2$. 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

Figure 1: Algorithm to Construct the Importance Sampler, with ψ given

```
Initialize f̂ and γ̂ at random values;
while not converged do
    Obtain variational approximation of p(z, γ|y, f̂);
    while not converged do
        Compute marginal distribution g_{z_{i,j}}(z_{i→j}, z_{j←i}) ∀i, j;
        Compute marginal distribution g_{γ_i}(γ_i) ∀i;
    end
    Obtain approximation of p(f|y, ẑ, γ̂);
    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 f̂;
        Apply Kalman filter and smoother to approximating model to update value
        for mode f̂;
    end
end
```

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. This procedure iterates until convergence, where in each step of the Newton-type algorithm, a new variational approximating model and a new Gaussian approximating model are computed. The algorithm presented in Figure 1 summarizes these steps for the construction of the importance sample.

4 Monte Carlo Study

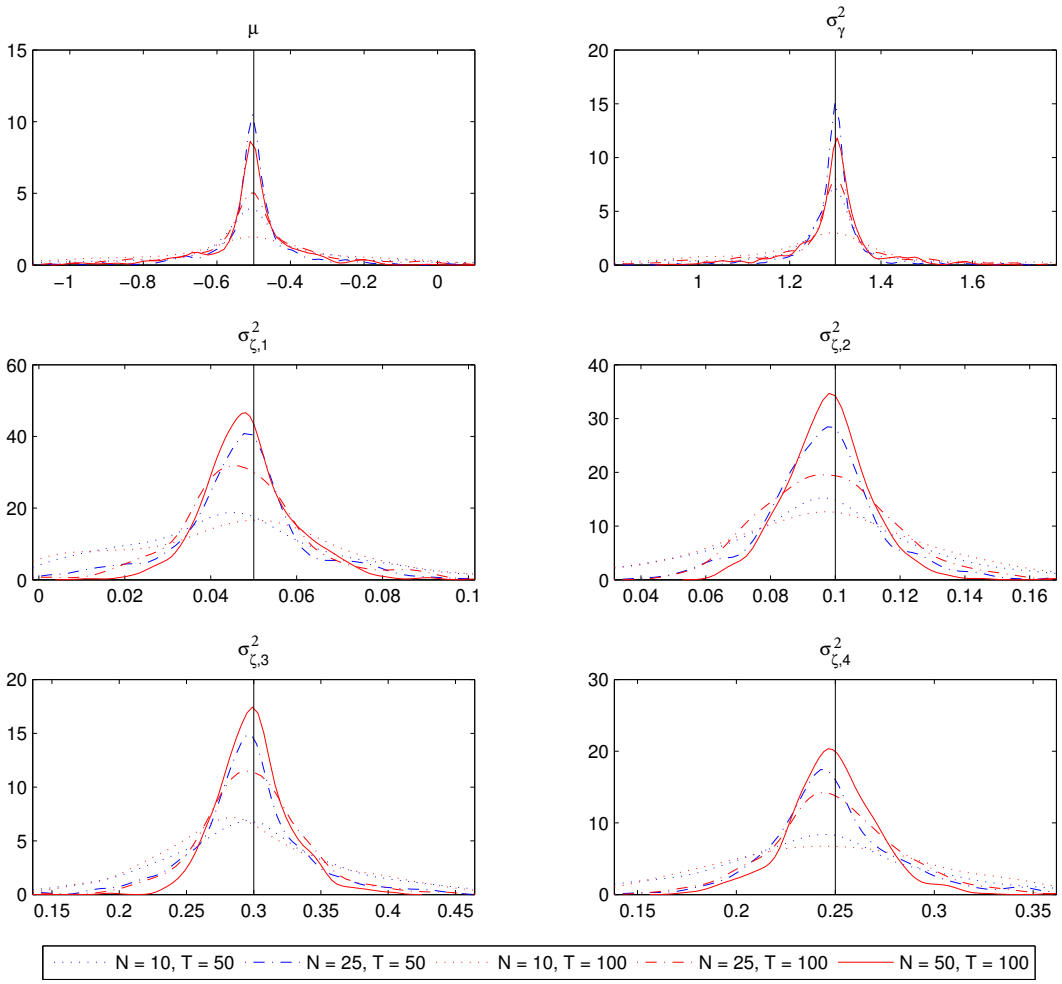
We conducted a simulation study to analyze the finite sample behavior of our proposed maximum likelihood estimation procedure for the dynamic factor network model. In this Monte Carlo study, we consider a dynamic factor network model as a data generation process that includes features resembling those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

In particular, 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 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 only need to 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 $N = 25, 50, 100$ and, for each network size, time series lengths of $T = 50, 100, 250$. For computational purposes, each simulation exercise consists of 500 Monte Carlo repetitions. The log likelihood function is computed using 500 simulated importance samples.

In Figure 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 are centred around the true parameter values (as indicated by a vertical 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

Figure 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.

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

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.

parameters, relative to their true values in the simulation study. 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 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

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

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.

was used for simulating data. The results indicate that for the latent factor f, we obtain smaller RMSE and MAE values when we increase the network size N and keep the time series dimension T 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 with a very small $T = 50$ and when increasing N does not monotonically decrease the RMSE, but does decrease the MAE). We can conclude that for all three latent variables, when increasing both N and T, we obtain more precise estimates in terms of RMSE and MAE.

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 Appendix D. 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.

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

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.

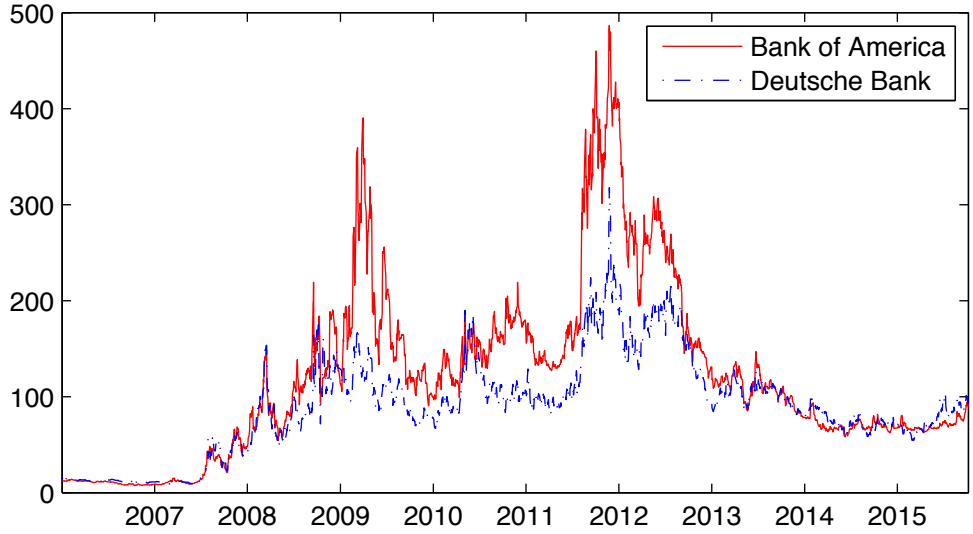
5 A Study into Global Credit-Risk Spillovers Networks

For our empirical study on global credit-risk spillovers, we consider the dynamic factor network model to analyze credit-risk spillovers in the global financial industry. We are interested in the characteristics of networks of financial institutions, in particular banks, and how these characteristics have evolved over time, in particular during the recent financial crisis and sovereign debt crisis. We study the credit-risk spillovers using time series data on credit default swaps (CDS) and financial instruments that allow the buyer of the contract to insure against default risk of an underlying reference entity by paying an insurance premium. It is referred to as the CDS spread to the seller of the contract.

5.1 Data Description

We used the database Markit to collect daily CDS spreads for contracts written on five-year senior unsecured debt for the period from January 1, 2006, to September 30, 2015. The cross-section of firms we consider is restricted by consistent data availability throughout our sample period. In total, our final data cover $N = 61$ global financial institutions, of which 12 have their headquarters in the United States and 49 in Europe, including commercial banks, investment banks, insurance companies, and investment funds. For simplicity, we refer to these institutions as banks or firms. As an illustration of two CDS spreads, Figure 3 presents the CDS spreads for Deutsche Bank and Bank of America within our sample period. The graphs show strong comovements between the two series. In particular, we clearly observe an increased cost of in-

Figure 3: Examples of Credit Default Swap (CDS) Spreads



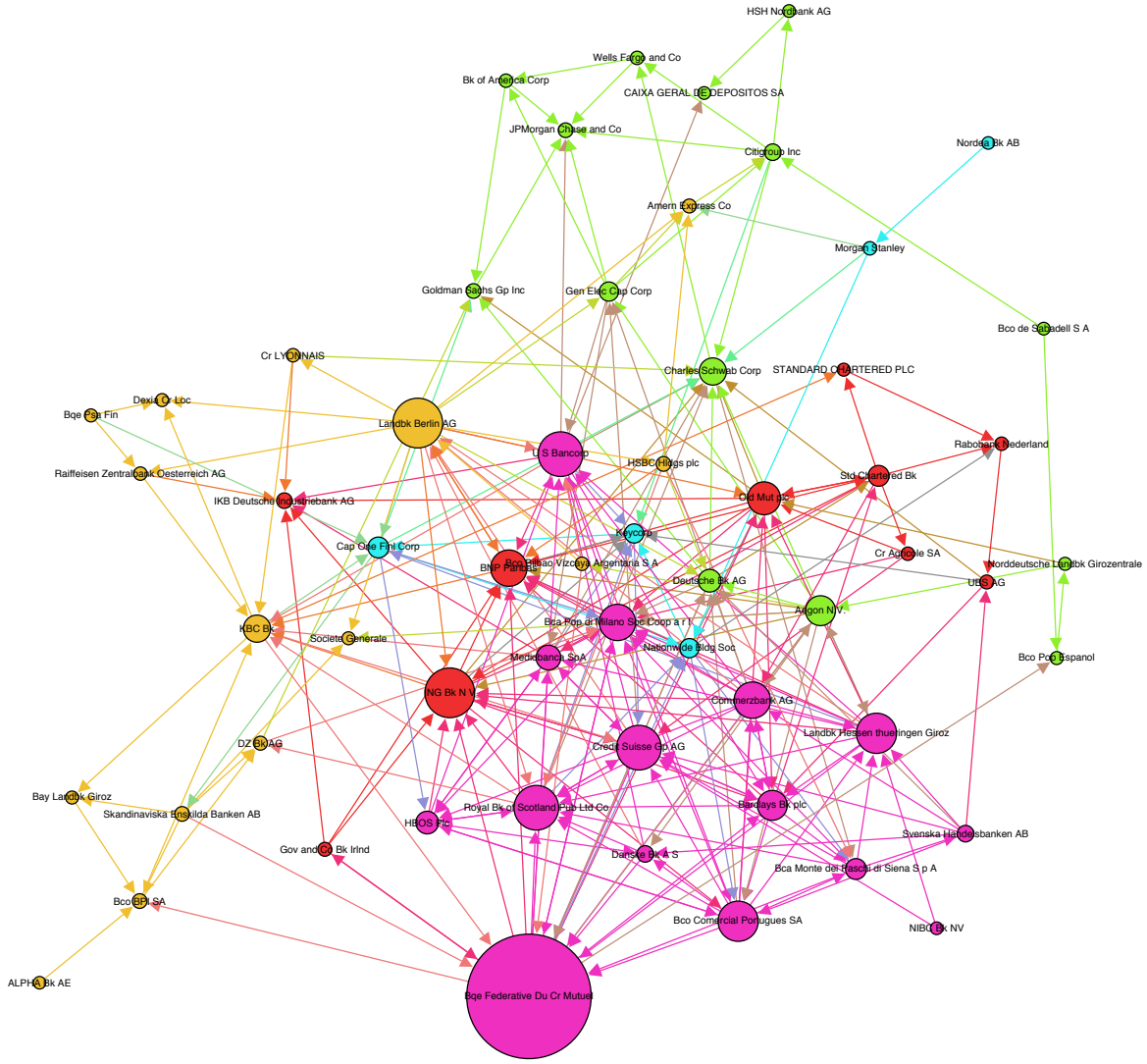
The graphs depict the 5-year CDS spread (in basis points) of Bank of America and Deutsche Bank, respectively, during the sample period from January 1, 2006, to September 30, 2015.
Source: Markit.

suring against default during the 2008/09 financial crisis and during the height of the European sovereign debt crisis in mid 2011.

From these daily CDS data we constructed a monthly sequence of binary, directed networks that capture all significant credit-risk spillovers among the considered firms. More specifically, we measure credit-risk spillovers by means of directional Granger causality tests between all two pairs of changes in CDS spreads that are statistically significant in our sample. The dependent variable takes the value $y_{i,j,t} = 1$ if there is a Granger-causal positive spillover from increases in the CDS spread of firm i to the CDS spread of firm j as measured in month t, and $y_{i,j,t} = 0$ otherwise. In this way of constructing credit-risk spillover networks, we follow Billio et al. (2012), who also study network relations among financial firms using covariance decompositions. In Appendix E, we provide a more detailed discussion on how we construct the sequence of binary networks.

As an example of our network structure based on directional Granger causality tests, Figure 4 depicts the inferred credit-risk spillover network as defined by the observations $y_{i,j,t}$ in March 2006. The individual nodes are scaled according to their total degree. These nodes correspond

Figure 4: Credit-Risk Spillover Network as Measured in March 2006



We present an example of a credit-risk spillover network for 61 financial institutions. The directional Granger causality tests are computed based on changes in daily CDS spreads during the period from January 1, 2006, through March 31, 2006. Each directed link represents a significant credit-risk spillover from the sender node to the receiver node. Nodes are scaled by the number of significant spillovers they receive from or contribute to other nodes in the network. This graph is created with the Gephi software; the positioning and coloring of the nodes are based on Gephi preset values.
Source: Authors' calculations.

to banks, which either induce or receive spillover effects from relatively many other banks (arrows indicate the direction of the spillover). While such a static network visualization helps to understand the structure of the relationships in the credit-risk spillover network at a given point in time, there are clear limitations to this approach, due to the number of potential links. For example, when we present a similar network graph during the 2007/08 financial crisis when spillovers increased, all nodes would have been cluttered, making visual inspection very hard. Moreover, a static network visualization is by nature only able to describe the network structure at a given point in time. It does not address the dynamic changes in the observed relationships. We therefore analyze the credit-risk spillovers by using our proposed dynamic factor network model that provides a dimension reduction by clustering banks and by modeling between- and within-cluster credit-risk spillover dynamics.

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 of credit-risk spillovers. We used the maximum likelihood importance sampling procedure as outlined in Section 3 for parameter estimation. In this empirical study, we set the number of clusters to $K = 2$ so that we allow for four pair-specific groups. 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. For the maximum likelihood estimation of the 6×1 parameter vector, we adopt $M = 5000$ importance samples to estimate the likelihood function via simulation. Table 4 presents the parameter estimates, together with the standard errors, which are based on the White sandwich estimate of the asymptotic variance matrix of ψ.

The estimation results show a negative estimate for μ and a large estimate for the variance parameter σ_γ^2, which we take as an indication that the distribution of the individual γ_i's is shifted to the left and is highly dispersed. However, neither estimate is 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 that they

Table 4: Model Parameter Estimates

Parameter	Estimate	Std. Error
μ	–1.2702	1.2527
σ_γ	6.1398	4.7817
$\sigma_{\xi,1}$	0.5454	0.0486
$\sigma_{\xi,2}$	0.6407	0.3005
$\sigma_{\xi,3}$	0.8407	0.3684
$\sigma_{\xi,4}$	0.7969	0.0730
Avg. Log-Likelihood	–0.2961	
K (# Groups)	2	
N (# Nodes)	61	
T (# Periods)	115	
Total Number of Observations	427,915	
Hill Test Statistic (with $\kappa = 2S^{1/3}$) = 3.7724 ; $\text{Crit}_{\alpha=0.05} = 1.645$		

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 sample includes 61 financial institutions and spans the period from March 2006 to September 2015, at a monthly frequency.

cannot be treated as being constant over time. The average log-likelihood value is maximized at a value of -0.2961. 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 Appendix C. The value of the Hill statistic is reported in Table 4, together with its 5 percent critical value. 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.

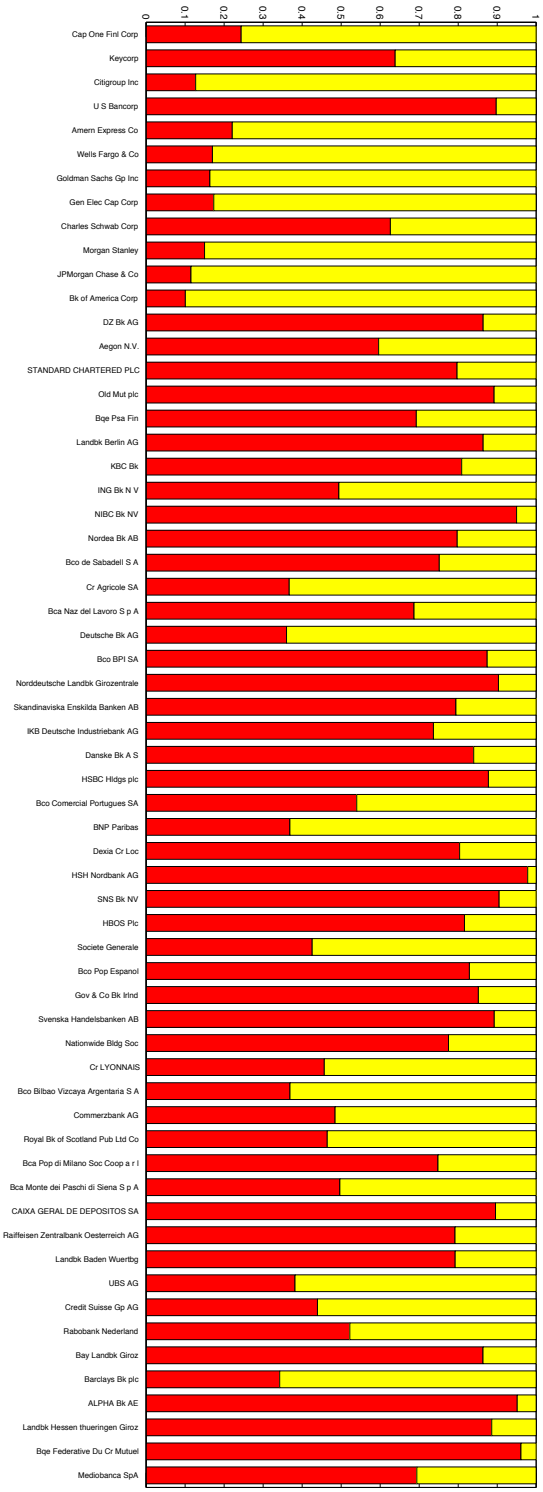
5.3 Network Estimation Results

The conditional distributions of all latent factors in our model provide us with insights into the bank- and pair-specific roles and with information about the model-implied latent group memberships of individual banks. Also, from the conditional distribution we can compute the smoothed between- and within-group link probabilities. In Figure 5, we present for each of the 61 banks 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 10 percent that Bank of America (row 12) belongs to group 1 (dark/red bar), and a probability of about 90 percent that Bank of America belongs to group 2 (light/yellow bar). The estimated structure implied by the model appears to allocate a high probability of group 2 membership to other large U.S. banks, including Goldman Sachs and J.P. Morgan. However, not all U.S. banks are associated with group 2. For instance, U.S. Bancorp has a high estimated probability of belonging to group 1. Furthermore, some larger European banks, including Deutsche Bank and Credit Agricole, have a high probability of belonging to group 2 as well. Many other European banks—in particular, the domestically focused and public banks such as the German Landesbanks or the Spanish saving banks (cajas)—have a very high probability of belonging to group 1. In many cases, banks are equally linked with both groups of banks. Overall, and without much subtlety, we may label group 1 as "local banks" and group 2 as "global or international banks."

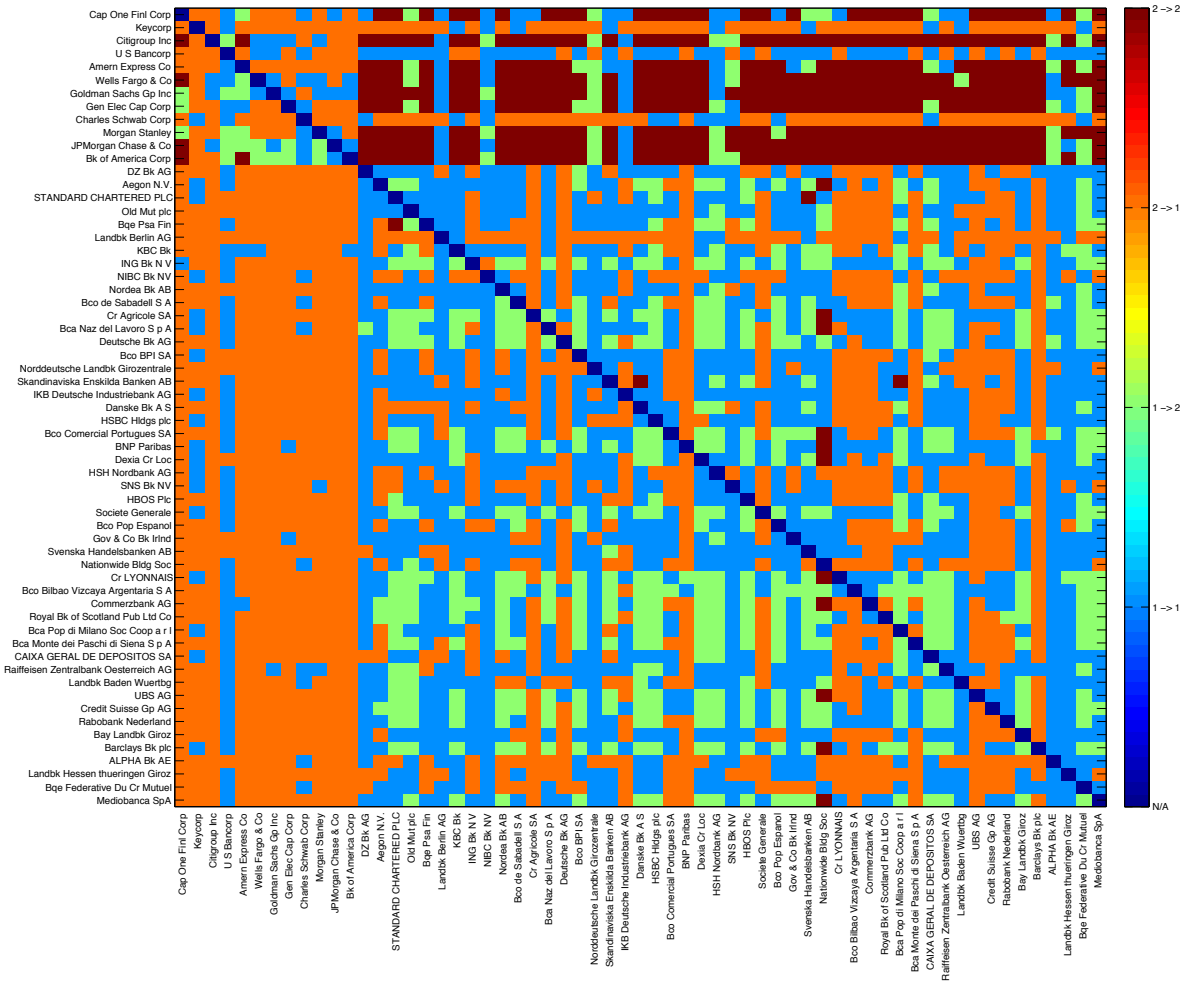
In terms of credit-risk spillovers, an increase in the CDS spread of Bank of America may affect the CDS spread of J.P. Morgan differently than the way this increase may affect the CDS spread of HSH Nordbank, a German public bank. Hence, credit-risk spillovers and their dynamics are pair specific. This key feature of network spillovers is incorporated in our proposed dynamic factor network model through the multiplicative interaction ($\delta_{i,j} = \pi_i \otimes \pi_j$) of the bank-specific stochastic group memberships π_i. In Figure 6, we present for each bank-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 sender banks (who initiate the spillover) are on the horizontal axis, while the receiver banks (who are the targets of the spillover) are on the vertical axis. For example, when Bank of America's CDS spread increases and we are interested in its impact on the J.P. Morgan's CDS spread, the model and the parameter estimates assign the highest probability that Bank of America belongs to group 1 and J.P. Morgan belongs to group 1 in this interaction. On the other, hand when we trace the impact on DZ Bank, our empirical model assigns the highest probability that both Bank of America and DZ Bank belong to group 2 in this interaction. However, when we consider the impact of a shock to the CDS spread of HSH Nordbank on DZ Bank, DZ Bank can be characterized as being a member of group 1.

Figure 5: Estimated Mean of Group-Membership Vectors at the Bank Level



We present for each bank 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).
Source: Authors' calculations.

Figure 6: Estimated Mode of Group Membership at the 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 bank pairs (i, j) where both bank i and bank j are allocated to group 1; green entries correspond to bank pairs where bank i is allocated to group 1 and bank j is allocated to group 2; orange entries correspond to bank pairs where bank i is allocated to group 2 and bank j is allocated to group 1; dark red entries correspond to bank pairs where both bank i and bank j are allocated to group 2.
Source: Authors' calculations.

We may conclude from Figure 6 that the pair-specific estimated modal group memberships imply a block structure for the credit-risk spillover network, where large U.S. banks as a shock sender and most European banks as a shock receiver tend to behave in a structurally similar way (the mostly dark red block). But also, the large U.S. and European banks (global banks) behave structurally similarly to shocks in most European bank CDS spreads (the predominately orange blocks). Within each of the diagonal blocks there is, however, substantial heterogeneity in the modal group membership. This can be concluded from the less-uniformly coloured cells among both U.S. bank pairs and European bank pairs. For instance, the results from our empirical model appear to suggest that the CDS spread of Credit Agricole, Deutsche Bank, or Barclay's reacts more similarly to increases of most European CDS spreads, than the way that CDS spreads of most U.S. banks react to other European banks (when in color, these are the orange elements in the matrix). Nevertheless, the larger European banks, including Credit Agricole, Deutsche Bank, and Barclay's, react differently to increases in CDS spreads of U.S. banks than other main U.S. banks do.

5.4 Dynamic Factor Estimation Results

The dynamic factors are associated with the probabilities of spillovers in credit-risk between and within groups, at each time point t. It is of interest to investigate how these factors have changed over time during our sample period. In Figure 7 we present the estimated mean link probabilities and their 95 percent confidence intervals, over time and for each of the four pair-specific group combinations. The time series graphs show how likely a significant credit-risk spillover is for each group and how these probabilities evolve over time within our sample period. The first panel presents the within-group link probabilities of banks from group 1 to other banks from group 1, corresponding to the bank pairs colored blue in Figure 6. This link probability is generally low (about 0.05 to 0.3), but it shows a clear increase during the beginning of the 2007/08 financial crisis and—after a decrease in 2009—during a second period of elevated spillovers in 2010. The second period corresponds to the beginning of the European sovereign debt crisis. The second panel depicts the across-group link probability from banks of group 1 to banks of group 2, corresponding to bank pairs colored green in Figure 6. The link

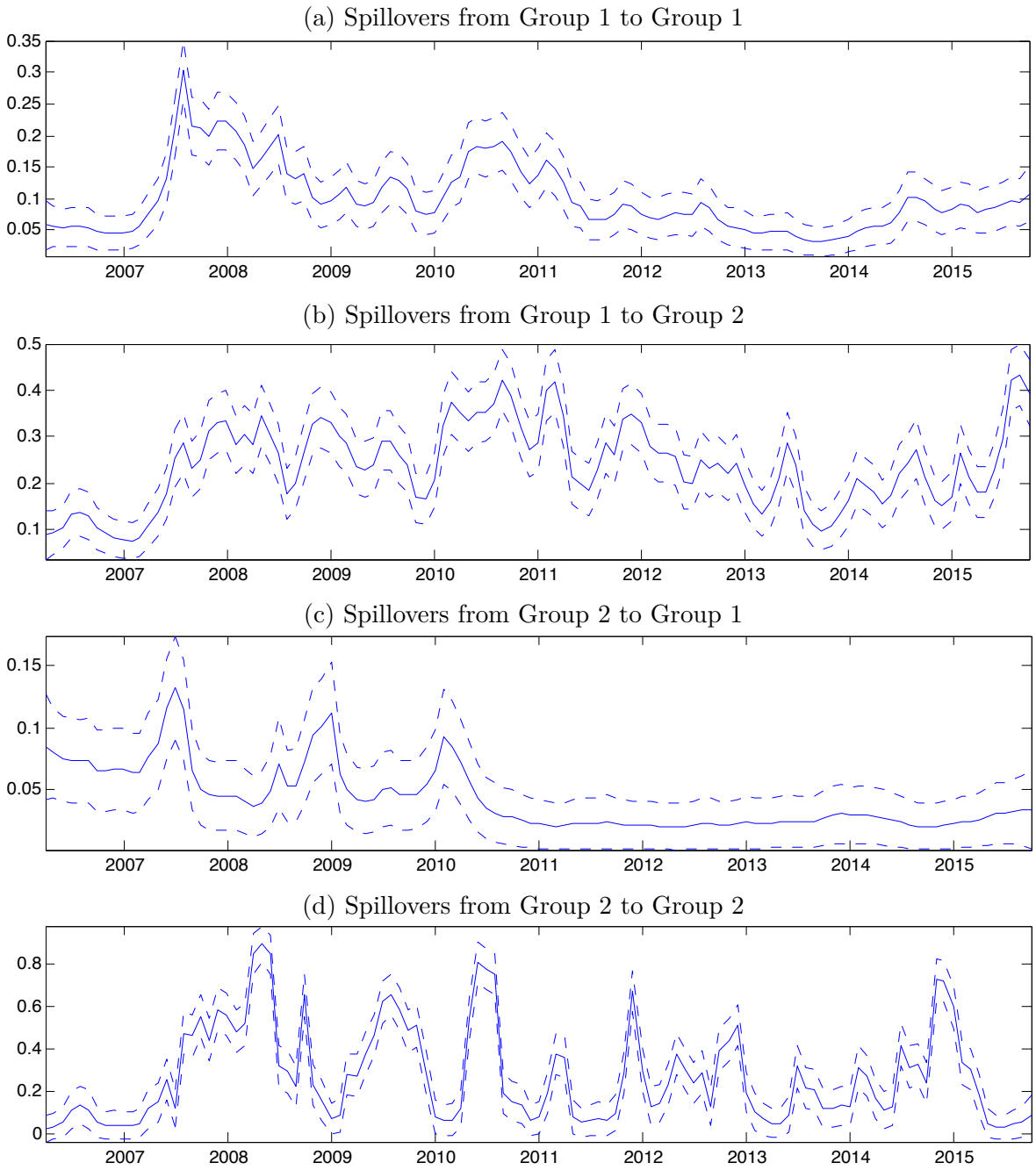
probability for spillovers from group 1 to group 2 increases with the 2007/09 financial crisis but shows a stronger and persistent increase during the European sovereign debt crisis. This is an intuitive finding, since the bank pairs, which are grouped in this cluster, are typically European institutions. It is also interesting that the second panel reveals a steady increase in spillovers from 2014 toward the end of the sample.

The third panel presents the dynamics for spillovers from group 2 to group 1, corresponding to bank pairs coloured in orange in Figure 6. The smoothed link probability exhibits only three distinct spikes during the summer of 2007, the autumn of 2008, and the beginning 2010; each of these short periods is associated with strong market turbulence. Overall, this panel reveals a very low link probability, in particular after 2010. Given the decomposition in Figure 6, this finding suggests that, in the second half of our sample, most American banks and some European banks, including Deutsche Bank and Credit Agricole, were relatively isolated from increases in CDS spreads of many other European banks. The fourth panel corresponds to the link probabilities between banks of group 2, corresponding to bank pairs colored in dark red in Figure 6. The link probabilities for these bank pairs, which are mostly comprised of a global sender and a local receiver bank, are generally high; in some periods it reaches values of more than 0.8. However, this time series graph also shows large fluctuations.

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, since many features of the model are treated by stochastic variables. For the cross-section dimension of the model, we adopted the stochastic blockmodel structure that is widely used for analyzing network data. For the time series dimension of the model, we 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 simulation-based maximum likelihood method. The simulations are directed toward 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

Figure 7: Smoothed Credit-Risk Spillover Probabilities



We present the smoothed mean and 95 percent confidence intervals of the within- and between-group link probabilities, where a link indicates a significant credit-risk spillover as measured by Granger-causality in CDS spread changes.
Source: Authors' calculations.

collapsing the high-dimensional observational vectors to achieve a feasible estimation procedure without losing any information. In a Monte Carlo study we show that the small-sample properties of our proposed estimation procedure are favorable, from both an estimation accuracy and a computational efficiency perspective. In the empirical study we analyze credit-risk spillovers between 61 main financial firms from the United States and Europe, from 2007 through 2015. In partitioning the nodes into two groups, we found that the first group is associated with large U.S. and European banks, while the second group consists of more local banks with a less global scope. The latent factors show a strong increase in spillovers during the 2008 financial crisis and the sovereign debt crisis. The spillovers from global to local banks were most sizeable, while the spillovers from local to global banks were larger during the European debt crisis and from 2014 onward. Further extensions and more precision in our network structure need to be considered to observe links at the country level and within common bank characteristics. Such developments would be of special interest to financial regulators and supervisors.

References

Airoldi, E. M., Blei, D. M., Fienberg, S. E., and Xing, E. P. (2008). Mixed Membership Stochastic Blockmodels. *Journal of Machine Learning Research*, 9:1981–2014.

Billio, M., Getmansky, M., Lo, A. W., and Pelizzon, L. (2012). Econometric Measures of Connectedness and Systemic Risk in the Finance and Insurance Sectors. *Journal of Financial Economics*, 104(3):535–559.

Bonhomme, S. and Manresa, E. (2015). Grouped Patterns of Heterogeneity in Panel Data. *Econometrica*, 83(3):1147–1184.

Craig, B. R. and von Peter, G. (2014). Interbank Tiering and Money Center Banks. *Journal of Financial Intermediation*, 23(3):322–347.

Durbin, J. and Koopman, S. J. (1997). Monte Carlo maximum likelihood estimation for non-Gaussian state space models. *Biometrika*, 84(3):669–684.

Durbin, J. and Koopman, S. J. (2002). A Simple and Efficient Simulation Smoother for State Space Time Series Analysis. *Biometrika*, 89(3):603–616.

Durbin, J. and Koopman, S. J. (2012). *Time Series Analysis by State Space Methods: Second Edition*. Oxford Statistical Science Series. Oxford University Press, Oxford, UK.

Fricke, D. and Lux, T. (2015). Core-Periphery Structure in the Overnight Money Market: Evidence from the e-MID Trading Platform. *Computational Economics*, 45(3):359–395.

Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. Springer Verlag, New York.

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. Proceedings of the 23rd Symposium on the Interface. Seattle Washington, April 21-24, 1991*, pages 571–578.

Hajivassiliou, V. (1990). Smooth Simulation Estimation of Panel Data LDV Models. Mimeo, Yale University.

Hajivassiliou, V., McFadden, D., and Ruud, P. (1996). Simulation of multivariate normal rectangle probabilities and their derivatives. Theoretical and computational results. *Journal of Econometrics*, 72:85–134.

Hoff, P. D., Raftery, A. E., and Handcock, M. S. (2002). Latent Space Approaches to Social Network Analysis. *Journal of the American Statistical Association*, 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., and Saul, L. K. (1999). An Introduction to Variational Methods for Graphical Models. *Machine Learning*, 37(2):183–233.

Jung, R. C., Liesenfeld, R., and Richard, J. F. (2011). Dynamic Factor Models for Multivariate Count Data: An Application to Stock-Market Trading Activity. *Journal of Business and Economic Statistics*, 29(1):73–85.

Jungbacker, B. and Koopman, S. J. (2015). Likelihood-based Dynamic Factor Analysis for Measurement and Forecasting. *The Econometrics Journal*, 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*. Springer Publishing Company, Incorporated, 1st edition.

Koopman, S. J., Lucas, A., and Scharth, M. (2014). Numerically accelerated importance sampling for nonlinear non-Gaussian state space models. *Journal of Business and Economic Statistics*, 33(1):114–127.

Koopman, S. J. and Mesters, G. (2017). Empirical Bayes Methods for Dynamic Factor Models. *Review of Economics and Statistics*, forthcoming.

Koopman, S. J., Shephard, N., and Creal, D. D. (2009). Testing the assumptions behind importance sampling. *Journal of Econometrics*, 149:2–11.

Liesenfeld, R. and Richard, J. F. (2010). Efficient Estimation of Probit Modeld with Correlated Errors. *Journal of Econometrics*, 156:367–376.

Mesters, G. and Koopman, S. J. (2014). Generalized Dynamic Panel Data Models with Random Effects for Cross-Section and Time. *Journal of Econometrics*, 180(2):127–140.

Monahan, J. F. (1993). Testing the Behaviour of Importance Sampling Weights. *Computer Science and Statistics: Proceedings of the 25th Annual Symposium on the Interface*, pages 112–117.

Monahan, J. F. (2001). *Numerical Methods of Statistics*. Cambridge University Press, Cambridge, UK.

Nowicki, K. and Snijders, T. A. B. (2001). Estimation and Prediction for Stochastic Block-structures. *Journal of the American Statistical Association*, 96(455):1077–1087.

Ormerod, J. T. and Wand, M. P. (2010). Explaining Variational Approximations. *The American Statistician*, 64(2):140–153.

Richard, J. F. and Zhang, W. (2007). Efficient high-dimensional importance sampling. *Journal of Econometrics*, 141:1385–1411.

Ripley, B. D. (1987). *Stochastic Simulation*. Wiley, New York.

Shephard, N. and Pitt, M. K. (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika*, 84(3):653–667.

van Lelyveld, I. and in 't Veld, D. (2014). Finding the Core: Network Structure in Interbank Markets. *Journal of Banking and Finance*, 49(C):27–40.

Wasserman, S. and 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., and Song, L. (2008). A State-Space Mixed Membership Blockmodel for Dynamic Network Tomography. *The Annals of Applied Statistics*, 4(2):535–566.

Xing, E. P., Jordan, M. I., and Russell, S. J. (2003). A generalized mean field algorithm for variational inference in exponential families. In *Uncertainty in Artificial Intelligence*, pages 583–591.

APPENDICES

A Model Extensions

A.1 Missing Data and Time-Varying Network Size

The proposed model easily accommodates missing data values and networks of time-varying size. Networks of time-varying size are present if the set of nodes in the network changes over time such that we have a time-varying set of nodes $\mathcal{N}_t$, with $|\mathcal{N}_t| = N_t$ number of nodes at time t. Missing data of pairwise measurements are present when for one or more periods we do not observe the value of $y_{i,j,t}$ for at least one pair (i, j). These missing observations do not enter the likelihood function and are simply omitted in the computation of the simulated likelihood given in Equation 13. For the construction of the importance sampler, similar changes are necessary.

First, for construction of the time series sampler we have to adjust the number of observations in each group. Let $m_{i,j,t} = 1$ if $y_{i,j,t}$ is missing, and zero otherwise, that is, when it is observed. We then have $y^*_{k,t} = \sum_{i,j}(1 - m_{i,j,t})y_{i,j,t}\hat{z}^k_{i,j}$, which is again binomially distributed, but due to the missing data with $\sum_{i,j}(1 - m_{i,j,t})\hat{z}^k_{i,j}$ number of trials for each group k at time t. The probability of success in each trial remains $\exp(f_{k,t})/(1 + \exp(f_{k,t}))$, similar to the case when all $y_{i,j,t}$ are observed. Second, for construction of the cross-sectional sampler using the variational approximation, we simply omit the missing data point from the product in Equation 11, by writing $\prod_{t=1}^{T}(1 - m_{i,j,t})p(y_{i,j,t}|z_{i,j}, f_t)$. No further adjustment is needed.

The treatment of networks with a time-varying number of nodes is a problem of particular relevance in empirical analysis. In fact, however, time-varying networks are simply a special case of missing data problems, as we observe no relation to any other node for a node that is not part of the network. Formally, we can define the set $\mathcal{N} = \cup_{t=1}^{T}\mathcal{N}_t$ to contain all nodes that have ever been part of any network $\mathcal{N}_t$. Then, we simply work with a network between the nodes in $\mathcal{N}$, and if node i is not part of the network at time t, we have a missing observations for any bilateral relation involving node i, that is, $m_{i,j,t} = m_{j,i,t} = 1 \; \forall j \in \mathcal{N}_t$. Importantly, we must distinguish between a node that is not part of the network and a node that is part of the network but not linked to other nodes.

A.2 Undirected Networks

It is straightforward to use our model to analyze undirected networks where $l_{i,j,t} = l_{j,i,t}$ by definition. Modeling an undirected network effectively reduces the dimension of the observation vector y_t to $N(N-1)/2$. A minor adjustment must be made for the factor loading Z and the dynamic factors f_t. For the construction of each row $z_{i,j}$ recall the hierarchical structure given in Equation 4. Due to the reciprocal nature of links, our model requires that, for node i in group k and node j in group k', the pair must have the same link probability as for node i in group k' and node j in group k. Therefore, at the pair level, each observation loads effectively on one out of $K(K+1)/2$ dynamic factors as compared with K^2 factors in the directed network model. We can easily incorporate this feature by maintaining the hierarchical structure for z, letting f_t denote a $K(K+1)/2$ vector of dynamic factors, and introducing the selection matrix T that maps the K^2 vector $z_{i,j}$ to a $K(K+1)/2$ vector that selects one of the $K(K+1)/2$ dynamic factors.

Consider an example of a model with $K = 2$ groups. Then, the second element being one, $z_{i,j,2} = 1$, refers to the situation that node i is in group 1 and node j is in group 2. The third element being one, $z_{i,j,3} = 1$, refers to the situation that node i is in group 2 and node j is in group 1. In the undirected network model both link probabilities must be the same, and hence we can write the signal for observation $y_{i,j,t}$ as

$$\theta_{i,j,t} = z_{i,j} T f_t \quad \text{with} \quad T = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

B The Variational Approximation

The variational approximation method is based on an approximation of the likelihood $p(y)$. Let α denote the vector that collects the elements of the cross-sectional state variables z and π. We then can rewrite the log likelihood for any arbitrary density $g(\cdot)$ indexed by some parameter $\theta \in \Theta$ as

$$\begin{aligned}
\log p(y) &= \log p(y) \int g(\alpha) \mathrm{d}\alpha = \int g(\alpha) \log p(y) \mathrm{d}\alpha \\
&= \int g(\alpha) \log \frac{p(y,\alpha)/g(\alpha)}{p(\alpha|y)/g(\alpha)} \mathrm{d}\alpha \\
&= \int g(\alpha) \log \frac{p(y,\alpha)}{g(\alpha)} \mathrm{d}\alpha + \int g(\alpha) \log \frac{g(\alpha)}{p(\alpha|y)} \mathrm{d}\alpha.
\end{aligned}$$

The term $\int g(\alpha) \log \frac{g(\alpha)}{p(\alpha|y)} \mathrm{d}\alpha \geq 0$ is the Kullback-Leibler divergence between $g(\cdot)$ and $p(\cdot|y)$ denoted by, $D(g(\cdot)||p(\cdot|y))$. The inequality is binding if and only if $g(\cdot) = p(\cdot|y)$ almost everywhere. The variational density $g(\cdot)$ is then obtained by minimizing $D(g(\cdot)||p(\cdot|y))$ over the class of densities with respect to the variational parameters θ.

From the above derivation we can also rewrite the minimization problem as a maximization problem of the lower bound on the likelihood as

$$p(y) \geq \exp \int g(\alpha) \log \frac{p(y,\alpha)}{g(\alpha)} \mathrm{d}\alpha,$$

which shows that the accuracy of the approximation depends on the tightness of the bound. Note that we could have obtained the lower bound on the log-likelihood directly by applying Jensen's inequality.

The variational density $g(\cdot)$ is typically chosen such that it belongs to a tractable class of densities. The most common restriction that we also apply in this paper is that $g(\cdot)$ factors according to a graph

$$g(\alpha) = \prod_{c=1}^{C} g_c(\alpha_c)$$

for components α_c. Variational approximations using the factorization restriction are also known as mean field methods, or variational Bayes, compare Ormerod and Wand (2010).

The factorization restriction allows one to obtain solutions for each product density in terms of the others, which leads to an iterative solution scheme for the variational parameters. Specifically it can easily be shown from the upper bound of the likelihood that the updating equations of the mean field method are

$$
\begin{aligned}
g_1(\alpha_1) &\leftarrow \frac{\exp(E_{-\alpha_1} \log p(y,\alpha))}{\int \exp(E_{-\alpha_1} \log p(y,\alpha)) \mathrm{d}\alpha_1} \\
&\vdots \\
g_C(\alpha_C) &\leftarrow \frac{\exp(E_{-\alpha_C} \log p(y,\alpha))}{\int \exp(E_{-\alpha_C} \log p(y,\alpha)) \mathrm{d}\alpha_1}
\end{aligned}
$$

where the expectations are with respect to all variational densities except $g_c(\cdot)$. The iteration is initialized by some starting values and repeated until convergence to a local optimum. The procedure is repeated several times with random starting values.

C Importance Sampling Diagnostics

Using the specific importance sampling distributions derived in the last subsections, we can rewrite the simulated likelihood function as

$$
\hat{L}(\psi) = g(y|\hat{z}) \frac{1}{M} \sum_{s=1}^{S} p(y|f^{(s)}, z^{(s)}) \frac{p(z^{(s)}|\gamma^{(s)}) p(\gamma^{(s)})}{g(y|f^{(s)}, \hat{z}) g(z^{(s)}, \gamma^{(s)}|\hat{f}, y)}, \tag{13}
$$

where we made use of Bayes' rule and the fact that the marginals $p(f)$ and $g(f)$ are the same. We can rewrite the estimated likelihood equivalently as

$$
\hat{L}(\psi) = g(y|\hat{z}) \frac{1}{S} \sum_{s=1}^{S} w_s, \tag{14}
$$

where the importance sampling weights w_s are given as

$$
w_s = \frac{p(y|f^{(s)}, z^{(s)}) p(z^{(s)}|\gamma^{(s)}) p(\gamma^{(s)})}{g(y|f^{(s)}, \hat{z}) g(z^{(s)}, \gamma^{(s)}|\hat{f}, y)}. \tag{15}
$$

An important assumption that allows consistent and asymptotically normal importance sampling maximum likelihood estimation requires that the first two moments of the importance

weights exist, see Geweke (1989). In particular, failure of the existence of the variance of the importance weights can lead to unstable and slow convergence as the central limit theorem fails to hold.

To assess the existence of the variance of the importance sampling weights in our empirical application, we use several formal and informal diagnostics checks. Specifically, we use the test proposed by Monahan (1993) (see also Monahan 2001) based on the tail behavior of the distribution of the weights, which is modeled as $1 - F(w) = Cw^{-1/\xi}(1 + Dw^{-1} + o(w^{-1}))$, for $w > 0$ and constant parameters C,D. The variance of the importance weights exists if $\xi \leq 1/2$. Monahan (1993) then uses the Hill estimator of the parameter ξ to test the null hypothesis $H_0 : \xi \geq 1/2$ against the alternative hypothesis $H_1 : \xi < 1/2$. We outline here the key steps of the testing procedure:

1. For a given sample of importance weights $\{w_s\}_{s=1}^{S}$, compute the order statistics $w_{(1)} \leq w_{(2)} \leq ... \leq w_{(S)}$.

2. Choose the tuning parameter κ that determines how many order statistics are used to compute the hill estimator; typical choices are $\kappa = 2S^{1/3}$ and $\kappa = 4S^{1/3}$.

3. Compute the Hill estimator $\xi^{Hill} = \frac{1}{\kappa}\sum_{j=1}^{\kappa} \log w_{(S-j+1)} - \log w_{(S-\kappa)}$.

4. Compute the test statistic $\sqrt{\kappa/\xi^2}(\hat{\xi}^{Hill} - \xi) \xrightarrow{d} \mathcal{N}(0, 1)$.

5. For given significance level α, reject H_0 if the test statistic exceeds the critical value.

In our empirical application, we investigate the importance sampling weights at our final parameter estimate to ensure that the maximum likelihood estimator is well behaved.

D More Results for Collapsed Importance Sampler

Table 5: Monte Carlo Results for Jungbacker and Koopman (2015) collapsing

		$T = 50$			$T = 100$			$T = 250$		
		$N = 25$	$N = 50$	$N = 100$	$N = 25$	$N = 50$	$N = 100$	$N = 25$	$N = 50$	$N = 100$
μ	RMSE	0.1494	0.0931	0.0780	0.1985	0.1517	0.1311	0.2582	0.1800	0.1693
	MAE	0.1318	0.0772	0.0670	0.1417	0.1108	0.0876	0.2042	0.1492	0.1244
σ^2_γ	RMSE	0.1878	0.0150	0.0410	0.0221	0.0180	0.0141	0.0216	0.0177	0.0162
	MAE	0.4696	0.0225	1.3510	0.0432	0.0342	0.0296	0.1851	0.0373	0.0341
$\sigma^2_{\xi,1}$	RMSE	2.4618	0.6361	0.2530	2.1173	0.8701	0.4001	1.4542	0.9215	0.5709
	MAE	0.5970	0.7714	0.0547	1.3727	0.5910	0.4851	0.8926	0.5688	0.3422
$\sigma^2_{\xi,2}$	RMSE	0.6662	0.2692	0.3075	0.6492	0.2960	0.1913	0.5027	0.3124	0.2095
	MAE	0.0862	0.0494	0.0415	0.1345	0.1068	0.0898	0.1687	0.1342	0.1183
$\sigma^2_{\xi,3}$	RMSE	0.0683	0.0364	0.0301	0.0991	0.0745	0.0615	0.1286	0.1038	0.0835
	MAE	0.0229	0.0101	0.0116	0.0175	0.0141	0.0113	0.0174	0.0155	0.0142
$\sigma^2_{\xi,4}$	RMSE	0.0445	0.0154	0.0809	0.0341	0.0282	0.0230	0.0569	0.0336	0.0302
	MAE	0.5684	0.1171	0.0520	0.7840	0.2868	0.1225	0.8263	0.4560	0.2639
f_t	RMSE	0.1198	0.0845	0.0658	0.1477	0.1099	0.0782	0.2853	0.2279	0.1831
	MAE	0.1508	0.1054	0.0727	0.1592	0.1088	0.0700	0.2406	0.1615	0.1187
γ_i	RMSE	0.0964	0.0956	0.1152	0.0575	0.0315	0.0168	0.0508	0.0270	0.0139
	MAE	0.1696	0.1262	0.0945	0.1437	0.1053	0.0741	0.1398	0.1002	0.0715
$z_{i,j}$	RMSE	0.0270	0.0273	0.0274	0.0158	0.0097	0.0075	0.0114	0.0057	0.0028
	MAE	0.0930	0.0845	0.0831	0.0402	0.0321	0.0281	0.0126	0.0075	0.0047

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

E Construction of Dynamic Correlation Network

We compute the Granger causality-based correlation networks from the daily changes in CDS spreads of $N = 61$ firms for the period from January 1, 2006, to September 30, 2015 ($\tilde{T} = 2548$ observations). We thereby follow Billio et al. (2012) and first estimate for each bank pair (i, j), with $i, j = 1, ..., N$, the following regression

$$\Delta \log CDS_{i,\tau} = \alpha_{i,j} + \sum_{l=1}^{p} \beta^l_{i,j} \Delta \log CDS_{i,\tau-l} + \sum_{l=1}^{p} \rho^l_{i,j} \Delta \log CDS_{j,\tau-l} + \epsilon_{i,j,\tau}, \tag{16}$$

where $\epsilon_{i,j,\tau} \sim N(0, \sigma^2_{i,j,\tau})$ and the parameters are indexed by (i, j) to highlight that they are pair specific. Clearly, this is one equation of a bivariate vector autoregressive (VAR) model

with lag order p for the pair of banks (i, j). Billio et al. (2012) are interested in the parameters $\rho^l_{i,j}$ that determine the influence of a change in the CDS of firm j at lag l on the change in firm i after controlling for changes in its own CDS spreads. Suppose we consider a VAR(p) with lag length one, that is, $p = 1$. Then a statistically significant positive estimate $\hat{\rho}^1_{i,j}$ is evidence of Granger causality from firm j to firm i, and we would conclude that there are significant spillovers from increases in the riskiness of firm j to the riskiness of firm i. To determine the statistical significance of the estimate, Billio et al. (2012) compute the t-statistic $t_{\hat{\rho}^1_{i,j}}$ and test the hypothesis $H_0 : \ \rho^1_{i,j} \leq 0$ against the alternative $H_1 : \ \rho^1_{i,j} > 0$ by comparing $t_{\hat{\rho}^1_{i,j}}$ with the critical value c_α at given significance level α.

In our analysis, we do not impose a lag length of one, but select the lag length based on the Bayesian information criterion (BIC) for each bivariate VAR model that we run for each of the bank pairs in the sample. For a VAR(p) with more than one lag, we could use an F-test for Granger causality by testing the null hypothesis $H_0 : \ \rho^1_{i,j} = ... = \rho^l_{i,j} = ... = \rho^p_{i,j} = 0$ against the alternative that at least one coefficient is not equal to zero (Billio et al. (2012)). However, in our analysis of credit-risk spillovers we are specifically interested in the *direction* of Granger causality, information that is not revealed by the F-test. Therefore, to determine whether there are significant *positive* spillovers from shocks to one CDS spread to the other, we analyze for each bivariate VAR(p) the sign and significance of the implied impulse response functions, which allows us to determine the sign of the spillover at various lags. To compute the impulse responses, we follow the standard procedure and orthonormalize the shocks by a Choleski decomposition of the estimated covariance matrix in a way that the unit shock we are interested in has no contemporaneous effect on the the other variable in the bivariate system. We then compute the lower and upper bounds of the confidence bound of the impulse response function for a given significance level α, denoted by $IR^{(i,j)}_{lb,\alpha}(k)$ and $IR^{(i,j)}_{ub,\alpha}(k)$, where $k = 1, ..., 10$ denotes the periods after the shock; that is, we analyze spillovers up to two weeks following the shock. Based on this impulse response function, we define a significant positive credit-risk spillover from firm i to firm j if the lower bound of the impulse responses is in at least one period above zero, that is, $\max_{0<k<11}\{IR^{(i,j)}_{lb,\alpha}(k)\} > 0$, and the upper bound is always above zero, that is $\min_{0<k<11}\{IR^{(i,j)}_{lb,\alpha}(k)\} > 0$.

Unlike the static analysis of credit-risk spillovers by Billio et al. (2012), a major focus of this paper is to estimate the proposed dynamic factor model to analyze the time-varying heterogeneity of credit-risk spillovers in the global financial industry. To allow for time variation in the correlation structure of the CDS spreads, we therefore estimate Equation (16) using the daily CDS data on different subsamples indexed by $t = 1, ..., T$. This results in a sequence of estimated impulse responses that capture the dynamic changes in spillover effects among financial firms. In our choice of the length of the subsamples, we face a tradeoff between the number of observations in each subsample and the number of estimated parameters $\{\hat{\rho}^{l}_{i,j,t}\}$ that we can estimate given our overall sample length. In particular, we want sufficient observations in each subsample to conduct a meaningful inference, but, on the other hand, we do not want to make each subsample too long, as we want to capture changes in the dynamics and to have a sufficient number of estimates of the cross-correlation parameters $\rho_{i,j}$. We therefore choose to estimate Equation (16) for a window of three months and move the three-month window forward in time by one month. To construct the credit-risk spillover networks, we then compute the binary link variable that indicates a significant credit-risk spillover from firm i to firm j at monthly frequency as

$$y_{i,j,t} = \begin{cases} 1 & \text{if } \max_{0<k<11}\{IR^{(i,j,t)}_{lb,\alpha}(k)\} > 0 \text{ and } \min_{0<k<11}\{IR^{(i,j,t)}_{ub,\alpha}(k)\} > 0 \\ 0 & \text{otherwise,} \end{cases}$$

where we added the index t to $IR^{(i,j,t)}_{lb,\alpha}$ and $IR^{(i,j,t)}_{ub,\alpha}$ to highlight that the impulse response is estimated from subsample t. Moreover, we define $y_{i,j,t} = 0$ such that self loops are excluded.[7] The link variable $y_{i,j,t}$ hence indicates if during the three months that define period t the VAR analysis estimates a significant positive spillover in CDS spread changes from bank i to bank j. The time index for $y_{i,j,t}$ then runs from $t = 1, ..., T = 115$; where $T = 115$ corresponds to the number of periods in our sample (March 2006 to September 2015, at a monthly frequency). We use these binary data to estimate our dynamic factor network model.

[7] Observations $y_{i,i,t}$ do not enter the likelihood function for all $i = 1, ..., N$.