Modelling and prediction of financial trading networks: An application to the NYMEX natural gas futures market

Brenda Betancourt

Abel Rodríguez

University of California, Santa Cruz, U.S.A.

Naomi Boyd

West Virginia University

Summary. Over the last few years there has been a growing interest in using financial trading networks to understand the microstructure of financial markets. Most of the methodologies developed so far for this purpose have been based on the study of descriptive summaries of the networks such as the average node degree and the clustering coefficient. In contrast, this paper develops novel statistical methods for modeling sequences of financial trading networks. Our approach uses a stochastic blockmodel to describe the structure of the network during each period, and then links multiple time periods using a hidden Markov model. This structure allows us to identify events that affect the structure of the market and make accurate short-term prediction of future transactions. The methodology is illustrated using data from the NYMEX natural gas futures market from January 2005 to December 2008.

1. Introduction

Financial trading networks are directed graphs in which nodes correspond to traders participating in a financial market, and edges represent pairwise buy-sell transactions among them that occurr within a period of time. Financial trading networks contain important information about patterns of order execution in order-driven markets; hence, they provide insights into aspects of market microstructure such as market frictions, trading strategies, and systemic risks.

Consider first the role of financial trading networks in understanding the effect of market frictions on market microstructure. In the absence of market frictions, we could expect orders from different traders to be matched randomly. However, real trading networks often exhibit features such as elevated transitivity or preferential attachment among certain groups of actors (Adamic et al., 2010; Boyd et al., 2013), which are inconsistent with random matching. In the

case of open-outcry markets, these features can be partially explained by sociological factors (for example, see Zaloom, 2004). Alternative explanations include the effect of different market roles (e.g., liquidity providers/takers) or trading strategies (e.g., long vs. short strategies), see Ozsoylev et al., 2010 or Hatfield et al., 2012.

Financial trading networks also provide information that is key in the assessment of systemic risks. Analysis of the evolution of financial trading networks can aid in tests of financial market stability (or fragility as it may be) by financial regulators to ensure that events such as a large trader failures do not serve to destabilize financial markets. In the event of a large trader failure, an understanding of their network will help guide regulators through the process of unwinding their positions and may dictate whether those positions are unwound in the open market or through a transfer to a suitable counterparty (Boyd et al., 2011). Financial trading networks can also be used to identify important traders that play a critical role in the market (for example, by acting as de facto market makers or liquidity providers). In addition, they can also help us identify frequent counterparties of specific traders which may aid in regulatory oversight by federal agencies and market exchanges alike; price distortion and manipulation may be more likely between frequent counterparties than by one agent acting in isolation (Harris et al., 1994).

The literature on the mathematical modeling of financial trading networks is limited. Theoretical approaches that explain the structure of a financial network as the outcome of a game have recently been developed (e.g., see Ozsoylev et al., 2010 and Hatfield et al., 2012), but they are of limited practical applicability. Most of the empirical work on trading networks has focused on the use of summary statistics such as degree distributions, average betweenness and clustering coefficients (Newman, 2003; Adamic et al., 2010; Boyd et al., 2013). These type of approaches provide some interesting insights into market microstructure, but suffer from two main drawbacks. First, the summary statistics to be monitored need to be carefully chosen to ensure that relevant features of the market are captured. Although some of the game-theoretic work mentioned before might provide some insights into which network summaries should be monitored, the choice is typically difficult and the selection is often incomplete. Second, and more importantly, approaches of this type are not helpful in predicting future interactions among traders.

In this paper we move beyond descriptive network summaries to focus on stochastic models for array-valued data that place a probability distribution on the full network. The statistical literature on stochastic models for individual networks is well developed. The simplest such model is the class Erdös-Rényi model (Erdös and Rényi, 1959), which assumes that interactions among any two traders occur independently and with constant probability that is independent of the identity of the traders. This class of models, although well studied from a theoretical perspective, is too simplistic to accommodate most realistic networks. As an alternative, Frank and Strauss (1986) proposed the class of exponentially

weighted random graphs, also called p^* models. These models formalize the use of summary measures by including them as sufficient statistics in exponential-family models. On the other hand, Holland and Leinhardt (1981) proposed the class of p_1 models, which extend generalized linear models to array-valued data. A related approach was introduced in Hoff et al. (2002) using the concept of latent social space models in which the probability of a link between nodes increases as they occupy closer positions in latent social space.

The model discussed in this paper extends the class of stochastic blockmodels first introduced in Wang and Wong (1987) to account for time dependence. Stochastic blockmodels rely on the concept of structural equivalence to identify groups of traders (which we shall refer to as *trading communities* in the context of this application) with similar interaction patterns. Indeed, the partitions induced by stochastic blockmodels are driven not only by the internal relations within a particular group of traders, but also by the interactions among these groups (Wasserman and Faust, 1994). Model-based stochastic blockmodels have been developed as array-valued extensions of traditional mixture models. For example, Nowicki and Snijders (2001) proposed a simple Bayesian model that uses a finite mixture model and a Dirichlet prior for the probabilities of the latent classes. An extension of this model that relies on infinite mixture models based on the Dirichlet process have been proposed by Kemp et al. (2006) and Xu et al. (2006). More recently Airoldi et al. (2008) introduced the idea of mixed membership stochastic blockmodels for binary networks wherein the actors can belong to more than one latent class to explore subjects with multiple roles in the network.

In this paper we propose modeling the dynamics of financial trading networks using an extension of the Bayesian infinite-dimensional model of Kemp et al. (2006). The model we propose accounts for dependence of the network structure over time and incorporates more general hierarchical priors on the interaction probabilities as well as the partition structure. To account for changes in market microstructure over time, the blockmodels associated with different time periods are linked through a hidden Markov model. In finance, regime switching models have been used in many contexts such as applications to model stock returns (Guidolin and Timmermann, 2005; Kim et al., 2001; Perez-Quiroz and Tim-mermann, 2000), in asset allocation (Ang and Bekaert, 2002a), business cycles (Filardo, 1994), and interest rates (Ang and Bekaert, 2002b). As we show in our illustration, by developing a dynamic, fully probabilistic model for array-valued data we are able to monitor structural changes in market microstructure while at the same time making more accurate short-term predictions of future trading patterns. The resulting model is similar in spirit to the one described in Ro-driguez (2011), but allows for additional flexibility and enhanced interpretability.

To motivate the structure of our model, consider a time series of weekly binary trading networks generated from proprietary transactions among 71 traders in the NYMEX natural gas futures market between January 2005 and December 2008 (for more details about this dataset, see Section 4). Figure 1 presents

time series plots of the corresponding clustering coefficients (which measure the tendency of traders to establish transitive relationships) and assortativity coefficients (which measures the tendency of traders to interact with other traders that are similar to themselves). From these plots it is clear that there is a change in market microstructure on September 2006 (which corresponds the date of introduction of an electronic trading system in this market). In addition, there is evidence of two additional structural changes around January 2007 and June 2008. This suggests that a Markov switching model in which the state of the system is assumed to be constant over short periods of time could be a reasonable model for this type of data.

Figure 2 presents a matrix representation for the trading network associated with the week of February 22, 2005 (traders have been reordered to make the graph easier to read). The graph suggests the existence of groups of traders that are structurally equivalent, including a large group of inactive traders that do not participate on transactions during this particular week, as well as small group of traders with a high number of intra-group and a relatively low number of inter-group transactions. This suggests that a stochastic block model might be a reasonable model for individual trading networks.

The rest of the paper is organized as follows: Section 2 provides a detailed description of our modeling approach and discusses its main properties. Section 3 describes the associated computational algorithm. Section 4 presents the analysis of the NYMEX natural gas futures market data. Finally, Section 5 discusses some limitations of the model and possibilities for further extensions and applications.

2. Modeling Approach

2.1. Stochastic blockmodels for financial trading networks

We encode a financial trading network among n traders using an $n \times n$ binary matrix $\mathbf{Y} = [y_{i,j}]$, where $y_{i,j} = 1$ if trader i sold at least one contract to trader j, and $y_{i,j} = 0$ otherwise. Since we focus on proprietary trading (i.e., transactions carried out by the traders with their own money, rather that their clients'), we adopt the convention $y_{i,i} \equiv 0$, as traders do not buy from themselves. Note that treating the network as binary ignores information about the transactions such as the number, maturities, and prices of the contracts. We proceed in this way for two reasons. First, in some markets (i.e., black pools) the prices and number of contracts might not be disclosed, making it impossible to apply more general models. Second, even if available, this extra information provides limit additional information about the identity of counterparties subject to contagion risks. Nonetheless, the framework we describe here can be easily extended to more general types of weighted networks.

A stochastic blockmodel for $\mathbf{Y}$ assumes that its entries are conditionally independent given two sets of parameters: a vector of discrete indicators $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)$, where $\xi_i = k$ if and only if trader i belongs to community $k =$
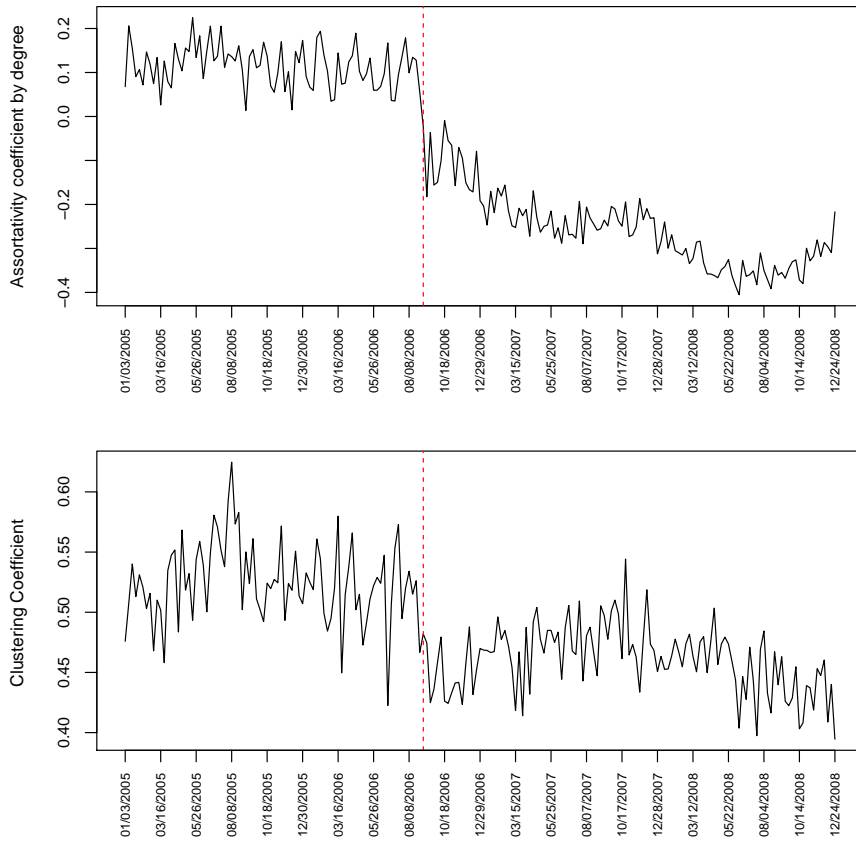
4

Fig. 1. Clustering coefficients and assortativity coefficient for weekly trading networks in the NYMEX natural gas futures market between January 2005 and December 2008. The vertical dashed line indicates the date in which electronic trading was introduced in this market.
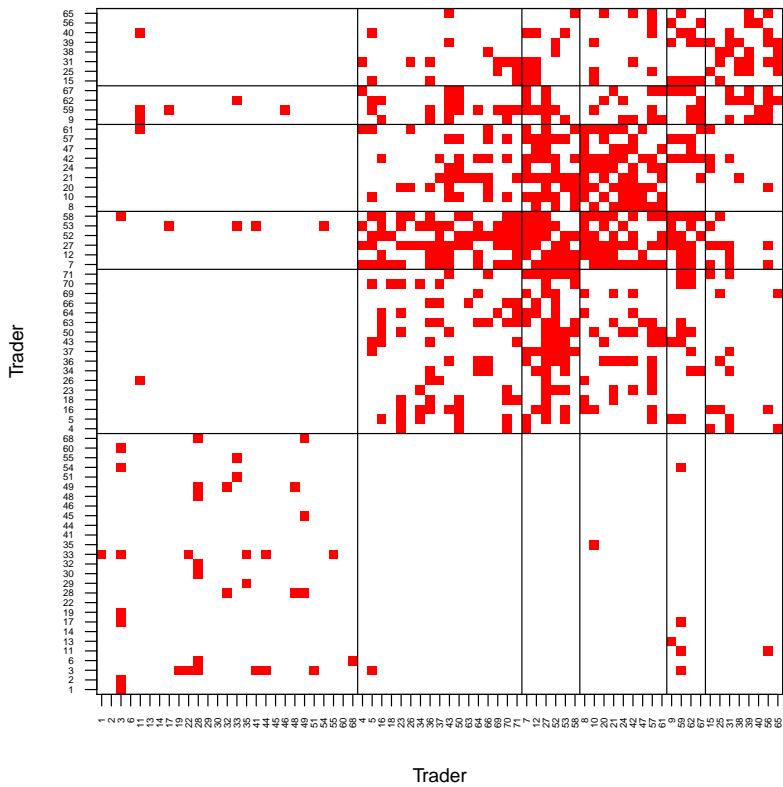
Fig. 2. Incidence matrix for the trading network associated with the week of February 22, 2005. The solid lines suggest one possible partition of the traders into groups of structurally equivalent nodes.

6

$1, \ldots K$, and a $K \times K$ matrix $\boldsymbol{\Theta} = [\theta_{k,l}]$ such that $\theta_{k,l}$ represents the probability that a member of community k sells a contract to a member of community l. Therefore,

$$y_{i,j} \mid \boldsymbol{\xi}, \boldsymbol{\Theta} \sim \mathsf{Ber}(\theta_{\xi_i, \xi_j}).$$

Note that K represents the maximum potential number of trading communities allowed a priori. A posteriori, the effective number of trading communities K^* present in the sample could potentially be smaller than K.

A Bayesian formulation for this model is completed by eliciting prior distributions for K, $\boldsymbol{\xi}$, and $\boldsymbol{\Theta}$. In the sequel we set $K = \infty$ and let the indicators be independent a priori where

$$\mathsf{Pr}(\xi_i = k \mid \mathbf{w}) = w_k, \qquad\qquad k = 1, 2, \ldots,$$

and the vector of weights $\mathbf{w} = (w_1, w_2, \ldots)$ is constructed so that

$$w_k = v_k \prod_{s<k} \{1 - v_s\}, \qquad\qquad v_k \sim \mathsf{Beta}(1 - \alpha, \beta + \alpha k), \qquad (1)$$

for $0 \leq \alpha < 1$ and $\beta > -\alpha$. Note that, by setting $K = \infty$, the model allows for the effective number of components K^* to be as large as the number of traders n, for any n.

The formulation in (1) is equivalent to the constructive definition of the Poisson-Dirichlet process (Pitman, 1995; Pitman and Yor, 1997), with $\alpha = 0$ leading to the Dirichlet process. Hence, the implied prior on the effective number of trading communities K^* and the size of those communities, $m_1, \ldots, m_{K^*}$, is given by

$$\frac{\Gamma(\beta + 1)}{(\beta + \alpha K^*)\Gamma(\beta + n)} \prod_{k=1}^{K^*} (\beta + \alpha k) \frac{\Gamma(m_k - \alpha)}{\Gamma(1 - \alpha)}.$$

Note that larger values of α or β favor a priori a larger effective number of K^*. Setting $\alpha = 0$ leads to the prior expected number of communities to grow logarithmically with n, while for $\alpha > 0$ the expected number components grows as a power of the number of traders.

Consider now building a prior on the matrix of interaction probabilities $\boldsymbol{\Theta}$. In this case we let

$$\theta_{k,l} \mid a_O, b_O, a_D, b_D \sim \begin{cases} \mathsf{Beta}(a_O, b_O) & k \neq l \\ \mathsf{Beta}(a_D, b_D) & k = l. \end{cases}$$

This prior is more general than those typically used in stochastic blockmodels, as it allows the distribution of the diagonal and off-diagonal elements of $\boldsymbol{\Theta}$ to have different hyperparameters. This ensures additional flexibility in terms of the implied degree distribution of the network, while still ensures that both $p(\mathbf{Y})$

and $p(\boldsymbol{\Theta})$ are jointly exchangeable, i.e., that the distributions are invariant to the order in which traders or communities are labeled (Aldous, 1981). In addition, it allows us to define an assortative index for the network as

$$\Upsilon = \log\left\{\mathsf{E}(\theta_{k,k} \mid a_D, b_D)\right\} - \log\left\{\mathsf{E}(\theta_{k,l} \mid a_O, b_O)\right\}$$

$$= \log\left\{\frac{a_D}{a_D + b_D}\right\} - \log\left\{\frac{a_O}{a_O + b_O}\right\},$$

and a cycle-type transitivity index

$$\chi = \mathsf{Pr}(y_{i,j} = 1 \mid y_{j,k} = 1, y_{k,i} = 1, a_O, b_O, a_D, b_D, \alpha, \beta) = \frac{\chi_N}{\chi_D},$$

where

$$\chi_N = \frac{(1-\alpha)(2-\alpha)}{(\beta+1)(\beta+2)} \frac{(a_D+2)(a_D+1)a_D}{(a_D+b_D+2)(a_D+b_D+1)(a_D+b_D)}$$

$$+ 3\frac{(1-\alpha)(\beta+\alpha)}{(\beta+1)(\beta+2)} \frac{a_D}{(a_D+b_D)} \left(\frac{a_O}{a_O+b_O}\right)^2$$

$$+ \frac{(\beta+\alpha)(\beta+2\alpha)}{(\beta+1)(\beta+2)} \left(\frac{a_O}{a_O+b_O}\right)^3,$$

and

$$\chi_D = \frac{(1-\alpha)(2-\alpha)}{(\beta+1)(\beta+2)} \frac{(a_D+1)a_D}{(a_D+b_D+1)(a_D+b_D)}$$

$$+ 2\frac{(1-\alpha)(\beta+\alpha)}{(\beta+1)(\beta+2)} \frac{a_D}{(a_D+b_D)} \frac{a_O}{(a_O+b_O)}$$

$$+ \frac{(\beta+\alpha)(\beta+\alpha+1)}{(\beta+1)(\beta+2)} \left(\frac{a_O}{a_O+b_O}\right)^2.$$

These two indexes are model-based alternatives to assortativity by degree and the clustering coefficients discussed in Figure 1 (Rodriguez and Reyes, 2013).

2.2. Hidden Markov models for time series of financial trading networks

We are interested in extending the hierarchical blockmodel described in Section 2.1 to model a time series of financial trading networks $\mathbf{Y}_1, \ldots \mathbf{Y}_T$. The extension is built with two goals in mind. First, we are interested in identifying events associated with structural changes in the network and, therefore, in the microstructure of the market. Second, we aim at making short-term predictions about the structure of the network in future periods. For these reasons, we focus our attention on the use of hidden Markov models for network data. Hidden Markov models are widely used in financial (e.g., see Ryden et al., 1998 and

references therein) and biological (e.g., Yau et al., 2011 and references therein) applications where there is interest in identifying structural changes in the system under study. Hence, they represent a natural alternative in this context.

More specifically, consider now a sequence $\mathbf{Y}_1, \ldots, \mathbf{Y}_T$ of binary trading networks observed over T consecutive time intervals, where all networks are associated with a common set of n traders. In addition, let $\zeta_1, \ldots, \zeta_T$ be a sequence of unobserved state variables such that $\zeta_t = s$ indicates that the market is in state $s \in \{1, 2, \ldots, S\}$ during period $t = \{1, 2, \ldots, T\}$. Each state has associated with it a vector of community indicators $\boldsymbol{\xi}_s = (\xi_{1,s}, \ldots, \xi_{n,s})$ with $\xi_{i,s} \in \{1, 2, \ldots, K\}$ and a matrix of interaction probabilities $\boldsymbol{\Theta}_s = [\theta_{k,l,s}]$ representing, respectively, the grouping of traders into trading communities and the probabilities of trades occurring between communities when the system is in state s. Analogously to our previous discussion, S and K represent the maximum number of states and the maximum number of trading communities allowed by the model a priori. A posteriori, the effective number of states S^* and the effective number of communities on each state $K_1^*, \ldots, K_S^*$ is potentially smaller than S and K, respectively.

Conditionally on the state parameters, observations are assumed to be independent, i.e.,

$$y_{i,j,t} \mid \zeta_t, \{\boldsymbol{\xi}_s\}, \{\boldsymbol{\Theta}_s\} \sim \text{Ber}(y_{i,j,t} \mid \theta_{\xi_{i,\zeta_t}, \xi_{j,\zeta_t}, \zeta_t}).$$

Hence, the joint likelihood for the data can be written as

$$p\left(\{\mathbf{Y}_t\} \mid \{\boldsymbol{\zeta}_t\}, \{\boldsymbol{\xi}_s\}, \{\boldsymbol{\Theta}_s\}\right) = \prod_t^T \prod_{i=1}^n \prod_{\substack{j=1 \\ j \neq i}}^n \theta_{\xi_{i,\zeta_t}, \xi_{j,\zeta_t}, \zeta_t}^{y_{i,j,t}} \left(1 - \theta_{\xi_{i,\zeta_t}, \xi_{j,\zeta_t}, \zeta_t}\right)^{1-y_{i,j,t}}$$

$$= \prod_{s=1}^S \prod_{k=1}^K \prod_{l=1}^K \prod_{(i,j,t) \in A_{k,l,s}} \theta_{k,l,s}^{y_{i,j,t}} \left(1 - \theta_{k,l,s}\right)^{1-y_{i,j,t}},$$

where $A_{k,l,s} = \{(i,j,t) : i \neq j, \zeta_t = s, \xi_{i,\zeta_t} = k, \xi_{j,\zeta_t} = l\}$ is the set of observations associated with the interactions between communities k and l in state s.

To account for the persistence in network structure illustrated in Figure 1, we assume that the evolution of the system indicators follows a first-order Markov process with transition probabilities

$$p(\zeta_t = s \mid \zeta_{t-1} = r, \{\boldsymbol{\pi}_r\}) = \pi_{r,s},$$

where $\boldsymbol{\pi}_r = (\pi_{r,1}, \ldots, \pi_{r,S})$, the r-th row of the transition matrix $\boldsymbol{\Pi} = [\pi_{r,s}]$, must satisfy $\sum_{s=1}^S \pi_{r,s} = 1$. A natural prior for $\boldsymbol{\pi}_r$ is a symmetric Dirichlet distribution,

$$\boldsymbol{\pi}_r \mid \gamma \sim \text{Dir}\left(\frac{\gamma}{S}, \frac{\gamma}{S}, \ldots, \frac{\gamma}{S}\right).$$

Note that, as $S \to \infty$, the induced distribution of transitions over states is equivalent to that generated by a Dirichlet process prior with concentration

parameter γ (for example, see Green and Richardson, 2001). Therefore the model is similar in spirit to the infinite hidden Markov model discussed in Rodriguez (2011) (see also Teh et al., 2006). However our construction does not couple the values of $\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots$ through a common centering probability. This is in contrast to the infinite hidden Markov model, where all transition probabilities into the same state are assumed to be similar. Indeed, the structure of the infinite hidden Markov model implies that if state s is highly persistent (i.e., $\pi_{s,s}$ is close to one), then the probability of transitioning from any other states into state s will also tend to be large, a property that is unappealing when modeling financial trading networks. Since γ plays an important role in controlling the number of effective states S^*, its value is estimated from the data by assigning an exponential prior to it and carrying out a sensitivity analysis to evaluate the impact of our prior choice on model performance.

The specification of the model is completed by eliciting hierarchical priors on the state-specific parameters $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_S$ and $\boldsymbol{\Theta}_1, \ldots, \boldsymbol{\Theta}_S$. Following the discussion in Section 2.1, we let

$$\mathsf{Pr}(\xi_{i,s} = k \mid \mathbf{w}_s) = w_{k,s}, \qquad k = 1, 2, \cdot,$$

where $w_{k,s} = v_{k,s} \prod_{h<k}\{1 - v_{h,s}\}$ are weights constructed from a sequence $v_{1,s}, v_{2,s}, \ldots$ where $v_{k,s} \sim \mathsf{Beta}(1 - \alpha_s, \beta_s + k\alpha_s)$. Again, since the hyperparameters α_s and β_s play a critical role in controlling the number of expected trading communities, they are assigned independent hyperpriors $\alpha_s \sim p(\alpha_s)$ and $\beta_s \sim p(\beta_s)$. A natural choice is to assign α_s a uniform prior on the unit interval and β_s an exponential prior, while carrying out a sensitivity analysis that involves priors that favor small values of α_s as well as priors that favor both lower and higher values for β_s.

Similarly, the interaction probabilities are assigned priors

$$\theta_{k,l,s} \mid a_{s,O}, b_{s,O}, a_{s,D}, b_{s,D} \sim \begin{cases} \mathsf{Beta}(a_{s,O}, b_{s,O}) & k \neq l \\ \mathsf{Beta}(a_{s,D}, b_{s,D}) & k = l. \end{cases}$$

where $\{a_{s,O}\}$, $\{b_{s,O}\}$, $\{a_{s,D}\}$, and $\{b_{s,D}\}$ are independent and gamma distributed with shape parameter c and unknown rates d_O, e_O, d_D and e_D, which are in turn assigned exponential priors with means λ_d and λ_e.

3. Computation

The joint posterior distribution for the model is proportional to

$$\begin{aligned} p\left(\{\mathbf{Y}_t\} \mid \{\zeta_t\}, \{\boldsymbol{\xi}_s\}, \{\boldsymbol{\Theta}_s\}\right) & p\left(\{\boldsymbol{\Theta}_s\} \mid \{a_{s,O}\}, \{b_{s,O}\}, \{a_{s,D}\}, \{b_{s.D}\}\right) \\ & p\left(\{a_{s,O}\} \mid d_O\right) p\left(\{b_{s,O}\} \mid e_O\right) p\left(\{a_{s,D}\} \mid d_D\right) p\left(\{b_{s,D}\} \mid e_D\right) \\ & p(d_O)\, p(e_O)\, p(d_D)\, p(e_D)\, p\left(\{\zeta_t\} \mid \boldsymbol{\Pi}\right) p\left(\boldsymbol{\Pi} \mid \gamma\right) p\left(\gamma\right) \\ & \qquad p\left(\{\boldsymbol{\xi}_s\} \mid \{\alpha_s\}, \{\beta_s\}\right) p\left(\{\alpha_s\}\right) p\left(\{\beta_s\}\right). \quad (2) \end{aligned}$$

This posterior distribution is not analytically tractable. Therefore, we implemented a Markov chain Monte Carlo algorithm (Robert and Casella, 2005) that simulates a dependent sequence of random draws from the target distribution. Given initial values for the parameters, these are successively updated from their full conditional distributions. Standard Markov chain theory ensures that, after an appropriate burn-in, the values of the parameters generated by the algorithm are approximately distributed according to (2).

To derive the algorithm, we rely on the fact that the joint posterior distribution in (2) can be factorized as

$$
p\left(\{\boldsymbol{\Theta}_s\} \mid \{\boldsymbol{\xi}_s\}, \{\zeta_t\}, \{a_{s,O}\}, \{b_{s,O}\}, \{a_{s,D}\}, \{b_{s,D}\}, \{\mathbf{Y}_t\}\right) \times
$$
$$
p\left(\{\boldsymbol{\xi}_s\}, \{\zeta_t\}, \{a_{s,O}\}, \{b_{s,O}\}, \{a_{s,D}\}, \{b_{s,D}\},\right.
$$
$$
\left. d_O, e_O, d_D, e_D, \{\alpha_s\}, \{\beta_s\}, \gamma \mid \{\mathbf{Y}_t\}\right) \quad (3)
$$

Since the values of $\theta_{k,l,s}$ are conditionally independent a posteriori given the observations, the indicators $\{\zeta_t\}$ and $\{\xi_{i,s}\}$, and the prior parameters $\{a_{s,O}\}$, $\{b_{s,O}\}$, $\{a_{s,D}\}$ and $\{b_{s,D}\}$, the first term in (3) is easy to sample from. Furthermore, conditionally on the other parameters in the model, the state indicators $\zeta_1, \ldots, \zeta_T$ are sampled jointly using a forward-backward algorithm (Rabiner, 1986), while the full conditional distribution for each collection of indicators $\xi_{1,s}, \ldots, \xi_{n,s}$ is sampled using a collapsed (marginal) Gibbs sampler (Neal, 2000). Details of the algorithm are discussed in Appendix A.

3.1. Estimation and prediction

Given a sample from the previous Markov chain Monte Carlo algorithm,

$$
\left(\{\boldsymbol{\Theta}_s^{(b)}\}, \{\boldsymbol{\xi}_s^{(b)}\}, \{\zeta_t^{(b)}\}, \{a_{s,O}^{(b)}\}, \{b_{s,O}^{(b)}\}, \{a_{s,D}^{(b)}\},\right.
$$
$$
\left. \{b_{s,D}^{(b)}\}, d_O^{(b)}, e_O^{(b)}, d_D^{(b)}, e_D^{(b)}, \{\alpha_s^{(b)}\}, \{\beta_s^{(b)}\}, \gamma^{(b)}\right), \qquad b = 1, \ldots, B,
$$

obtained after an appropriate burn-in period, point and interval estimates for model parameters can be easily obtained by computing the empirical mean and/or the empirical quantiles of the posterior distribution. For example, posterior co-clustering probabilities, $\omega_{t,t'} = \mathsf{Pr}(\zeta_t = \zeta_{t'} \mid \{\mathbf{Y}_t\})$ can be estimated as

$$
\hat{\omega}_{t,t'} = \mathsf{Pr}(\zeta_t = \zeta_{t'} \mid \{\mathbf{Y}_t\}) \approx \frac{1}{B} \sum_{b=1}^{B} \mathbb{I}(\zeta_t^{(b)} = \zeta_{t'}^{(b)}),
$$

where $\mathbb{I}(\cdot)$ denotes the indicator function. The estimates can be arranged into a co-clustering matrix $[\hat{\omega}_{t,t'}]$, which can in turn be used to identify the state of the system at each time period through a decision-theoretic approach (e.g., see Lau and Green, 2007). A similar procedure can be used to identify communities on each period.

11

The samples from the posterior distribution can also be used as the basis for prediction. For this purpose, note that the probability that trader i sells at least one security to trader j in the unobserved period $T + 1$ can be estimated by

$$\mathsf{E}\left(y_{i,j,T+1} \mid \{\mathbf{Y}_t\}\right) \approx \frac{1}{B} \sum_{b=1}^{B} \pi^{(b)}_{\zeta_T^{(b)}, s}\, \theta_{\xi_{i,s}^{(b)}, \xi_{j,s}^{(b)}}.$$

Using a simple 0/1 utility function, a future sell trade from trader i to trader j is predicted as $\hat{y}_{i,j,T+1} = \mathbb{I}(\mathsf{E}\{y_{i,j,T+1} \mid \{\mathbf{Y}_t\}\} > f)$, for some threshold f that reflects the relative cost associated false positive and false negative links.

4. An application to the NYMEX natural gas futures market

In this section we analyze the sequence of $T = 201$ weekly financial trading networks constructed from proprietary trades in the natural gas futures market on the New York Mercantile Exchange (NYMEX) between January 2005 and December 2008 that were introduced in Section 1. As discussed in Section 2, directed binary networks were constructed by setting $y_{i,j,t} = 1$ if there was at least one transaction in which trader i sold an option to trader j during week t. One particularity of this market is that futures were traded on the New York Mercantile Exchange (NYMEX) only through traditional open-outcry trades until September 5, 2006, and as a hybrid market that included electronic trading conducted via the CME Globex platform after that date.

A total of 970 unique traders participate in proprietary transactions at least once over the four years to December 2008. However, this list includes traders that either abandoned proprietary trading or went bankrupt during the period under study, as well as traders that entered the market after January 2005. Indeed, only between 240 and 340 traders participated in trades each week (see Figure 3). Since we have limited information about the times at which different traders entered or left the market, our analysis focuses on 71 traders we identified as being present in the market (although not necessarily active) during the whole period. Traders were anonymized and are identified in the plots using numbers.

The results presented in this section are based on 100,000 iterations collected after a burn-in period of 10,000 iterations. Convergence of the algorithm was diagnosed using the single-chain approach discussed in Geweke (1992) and by a visual evaluation of trace plots. We monitored the log-likelihood function, as well as the number of active states S^* and the mean and variance over time of the assortativity and transitivity indexes $\{\Upsilon_t\}$ and $\{\chi_t\}$. In terms of hyperparameters, the maximum number of states is set to $S = 30$, the prior means for γ and $\{\beta_s\}$ are assigned exponential priors with unit mean, and the priors for d_O, e_O, d_D and e_D are exponential distributions with mean 2. This specification implies that, a priori, $\mathsf{E}(\Upsilon_t) = 0$ for all $t = 1, \ldots, T$, so that we favor neither assortative nor dissasortive trading communities a priori.
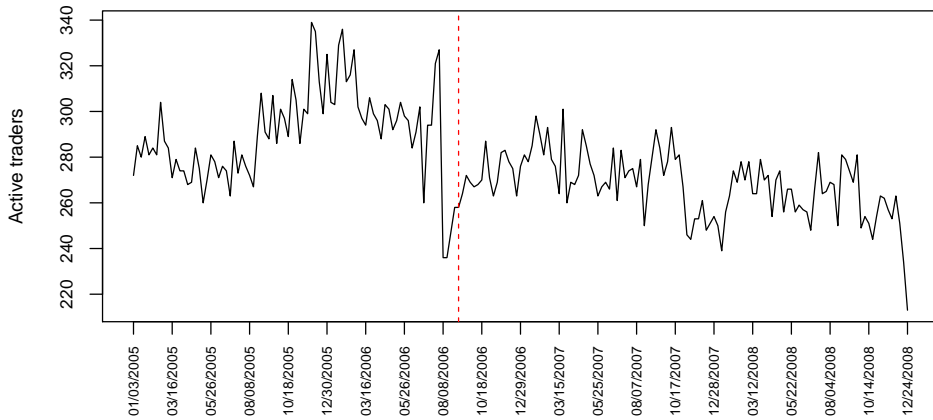
Fig. 3. Number of active traders each week in the NYMEX natural gas future market between January 2005 and December 2008.
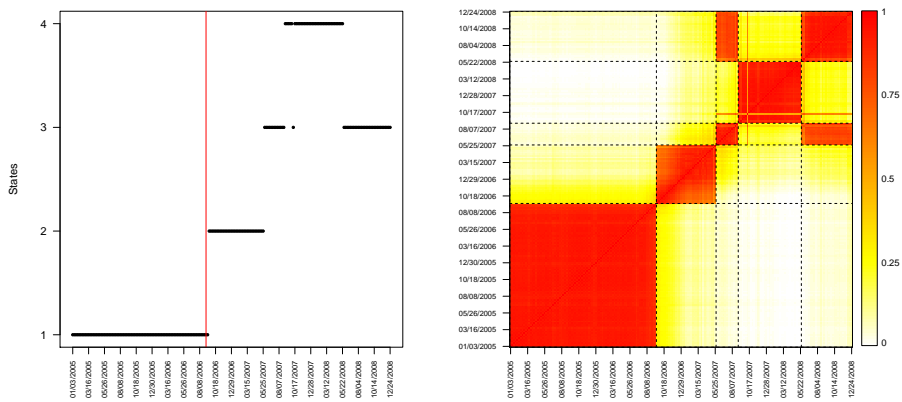

Fig. 4. In the left panel, point estimate of the states for the 201 weeks observed for the trading network with the vertical line indicating the introduction of the electronic platform on week 85. On the right, mean posterior pairwise incidence matrix for the weekly networks, illustrating the uncertainty associated with this point estimate.

13

4.1. Identifying changes in market microstructure

Figure 4 presents the posterior estimate of the co-clustering matrix for the latent states $\zeta_1, \ldots, \zeta_T$, along with a point estimator for the grouping of networks into states (recall Section 3.1). This point estimator suggests that the structure of the trading networks alternates between four highly persistent states. The first state runs between early January 2005 and early September 2006, when the electronic market is introduced. The presence of a change point at this date is not surprising in light of the descriptive analysis presented in Section 1. The second state runs between early September 2006 and early May 2007, when the system transitions to a new state for a short period of 3 months. After that, the system seems to transition to a fourth state in early August 2007 (interestingly, the beginning of the recent financial crises), where it stays for 37 weeks before returning to the third state in early June 2008 (which coincides with some of the largest drops in the S&P500 energy sector index over the last 13 years). Also, it is clear from the heatmap that, although there is some uncertainty associated with this point estimate of the system states (mostly in time of the transitions between states three and four), this uncertainty is relatively low.

Figure 5 shows estimates of the community structure associated with two different weeks, that of October 11, 2005 ($t = 40$) and that of November 14, 2007 ($t = 145$). We selected these dates because they are representative of states 1 and 4. Note that, although there are some similarities, the overall structure of the communities is quite different. State 1 is characterized by a large group of 25 mostly inactive traders, while all other traders tend to fall, for the most part, into singleton clusters. On the other hand, while state 4 also exhibits a number of singleton clusters, it also shows a number of small communities comprising between 5 and 10 traders each.

Figure 6 shows time series plots for the estimates of the assortativity and transitivity indexes $\Upsilon_1, \ldots, \Upsilon_T$ and $\chi_1, \ldots, \chi_T$. Recall that these quantities are model-based alternatives to the assortativity by degree and the clustering coefficient presented in Figure 1. Both sets of plots share some common features, revealing mild assortativity and higher transitivity before September 2006 and highly disassortative networks with lower transitivity afterwards. This makes sense because we would expect that the introduction of an electronic market would limit the effect of social connections among traders (which tend to be assortative and transitive) and favor connections based on differential trending strategies (which tend to be disassortative).

Finally, Table 1 shows point estimates and credible intervals associated with some hyperparameters in the model, both a priori and a posteriori. In all cases, the posterior estimates appear to be more concentrated and be centered around different values than the prior.
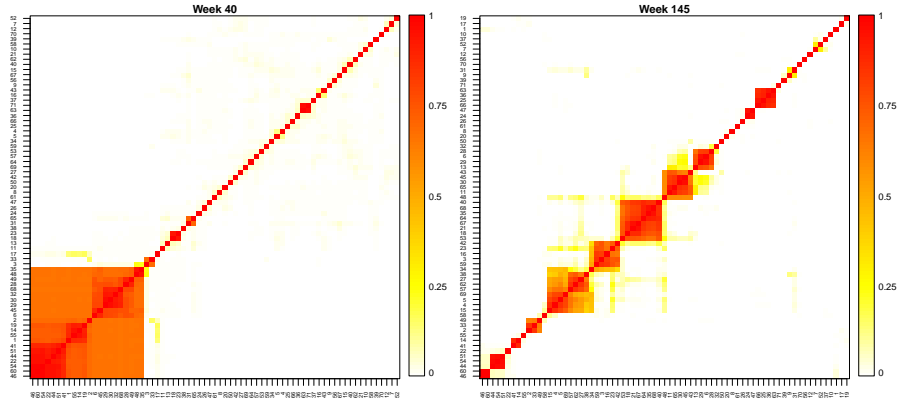
14

Fig. 5. Mean posterior pairwise incidence matrices of traders for $t = 40$ from state 1 and $t = 145$ from state 4.

Table 1. Prior and posterior point estimates and credibility intervals for some model hyperparameters.

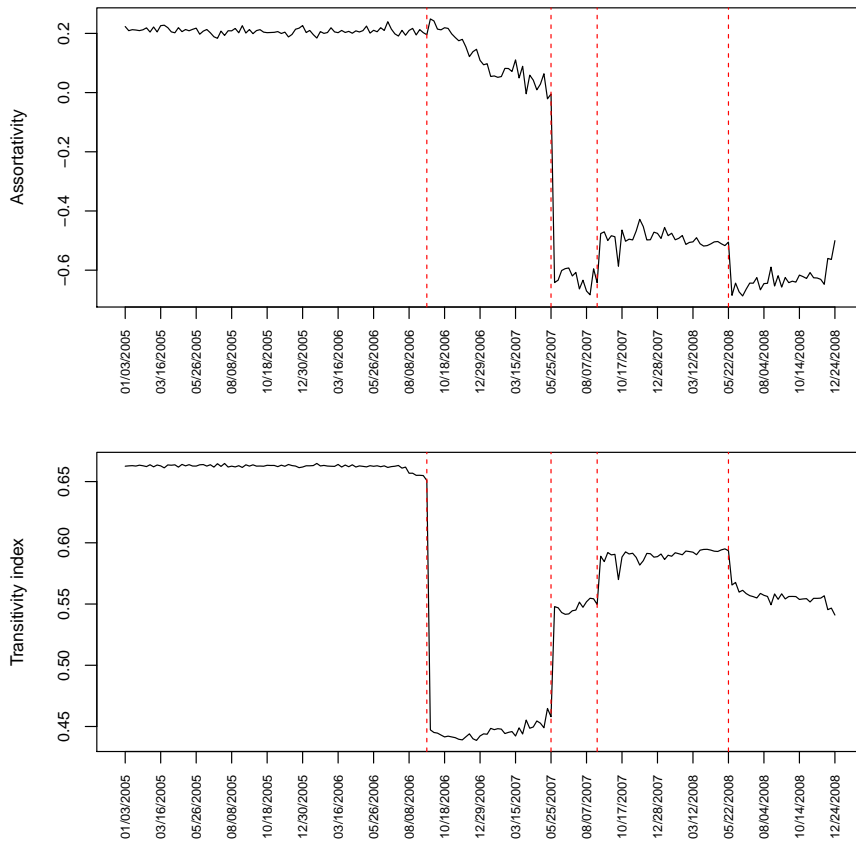| Parameter	Posterior mean	Posterior 95% credible interval	Prior mean	Prior 95% credible interval
α_{40}	0.748	$(0.198, 0.942)$	0.500	$(0.025, 0.975)$
α_{100}	0.524	$(0.240, 0.851)$	0.500	$(0.025, 0.975)$
α_{145}	0.587	$(0.264, 0.785)$	0.500	$(0.025, 0.975)$
β_{40}	1.225	$(0.034, 4.526)$	1.000	$(0.025, 3.689)$
β_{100}	2.518	$(0.107, 7.563)$	1.000	$(0.025, 3.689)$
β_{145}	1.223	$(0.034, 4.510)$	1.000	$(0.025, 3.689)$
γ	0.344	$(0.090, 0.814)$	1.000	$(0.025, 3.689)$

Fig. 6. Time series plot for assortativity and transitivity indexes. The vertical line represents the transitions across states identified from Figure 4.
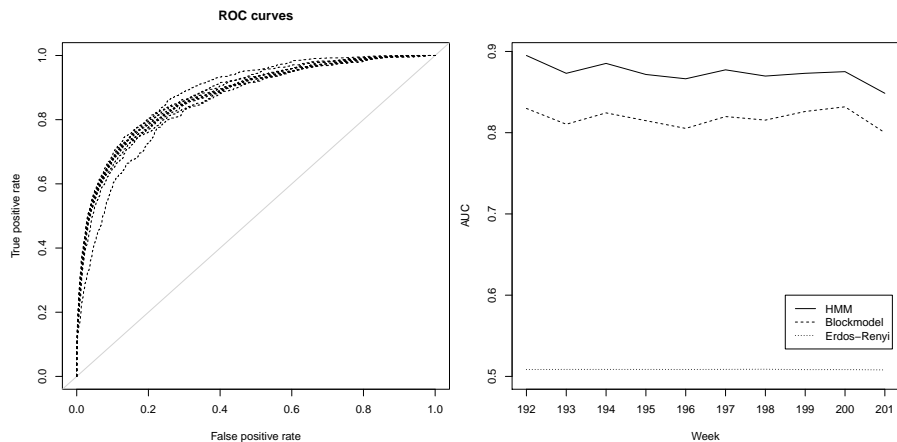
16

Fig. 7. The left panel shows ten operating characteristic curves associated with one-step-ahead out of sample predictions from our hidden Markov model. The right panel shows a time series plot of the area under the receiver operating characteristic curves for three different models.

4.2. Network prediction

As we discussed in the introduction, besides identifying points of change in the microstructure of the market, one of our goals is to predict future trading partnerships. To assess the predictive capabilities of the model we ran an out-of-sample crossvalidation exercise where we held out the last ten weeks in the dataset and made one-step-ahead predictions for the structure of the held-out networks. More specifically, for each $t = 191, 192, \ldots, 200$ we use the information contained in $\mathbf{Y}_1, \ldots, \mathbf{Y}_t$ to estimate the model parameters and obtain predictions for $\hat{\mathbf{Y}}_{t+1}$ for different values of the threshold f. Each of these predictions is compared against the observed network $\mathbf{Y}_{t+1}$, the number of false and true positives computed, and a receiver operating characteristic (ROC) curve is constructed. For comparison purposes, the same exercise was repeated first by fitting a single blockmodel (which corresponds to taking $S = 1$) and then by fitting a single Erdös-Renyi model (which corresponds to $S = 1$ and $K_1 = 1$) to the whole collection of networks. Figure 7 shows the ten operating characteristic curves associated with one-step-ahead out of sample predictions from our hidden Markov model, along with estimates of the area under the receiver operating characteristic curves (AUC) for all three models. The proposed hidden Markov model shows superior predictive capabilities, with AUCs between 0.85 and 0.9.

4.3. Sensitivity analysis

To assess the effect of our prior choice on posterior inference we conducted a sensitivity analysis where the model was fitted with somewhat different priors. In particular, we used independent Beta priors with mean $1/10$ and variance $9/1100$ for each α_s, as well as exponential priors with means $1/3$ and 3 for each β_s. On the other hand, a exponential priors with mean 2 were also used for d_O, e_O, d_D and e_D. Although inferences on the community structure were somewhat affected by prior choices, inferences on the state parameters as well as the assortativity and transitivity indexes and the predictive performance were essentially unchanged.

5. Discussion

We have presented a class of hidden Markov models for financial trading networks that have clear potential for market regulatory oversight. Key application of these models include identifying specific events (such as large trader failures or specific changes in market rules) that affect market stability, as well as identifying frequent trading counterparties that might be likely collusion partners or particularly at risk in case of bankruptcies.

In this paper we have focused on models for binary networks where only the presence/absence of transactions over a week is recorded. However, when information about volumes is available, the model can be easily extended to incorporate this information.

Although the use of a hidden Markov model allows us to account for time dependence and is useful for identifying structural changes in the system, a structure that assumes abrupt changes in the network might be too restrictive for predictive purpose. In the future we plan to evaluate models based on fragmentations and coagulations (e.g., see Bertoin, 2006) that allow for smooth evolution in the community structure, as well as extensions of auto logistic models that might allow for improved predictions.

A. Appendix A

Here, we provide the details of the MCMC algorithm discussed in 3. The algorithm proceeds by updating the model parameters from the following full conditional distributions:

(a) For each $i = 1, \ldots, n$ and occupied states s, $\xi_{i,s} = k$ with probability

$$\Pr(\xi_{i,s} = k \mid \cdots, \mathbf{Y})$$

$$= \begin{cases} (m_k^{-i} - \alpha_s) \prod_{l=1}^{K^*_{s,-i}} \dfrac{p(\{y_{i,j,t}:(i,j,t) \in A^i_{k,l,s}\})}{p(\{y_{i,j,t}:(i,j,t) \in A^{-i}_{k,l,s}\})} \dfrac{p(\{y_{j,i,t}:(i,j,t) \in A^i_{k,l,s}\})}{p(\{y_{j,i,t}:(i,j,t) \in A^{-i}_{k,l,s}\})}, & k \leq K^*_{s,-i} \\[2em] (\beta_s + \alpha_s K^*_{s,-i}) \prod_{l=1}^{K^*_{s,-i}} p(\{y_{i,j,t}:(i,j,t) \in A^{-i}_{l,s}\}) \\[1em] \qquad\qquad\qquad\qquad p(\{y_{j,i,t}:(i,j,t) A^{-i}_{l,s}\}), & k = K^*_{s,-i} + 1, \end{cases}$$

where $K^*_{s,-i} = \max_{j \neq i}\{\xi_{j,s}\}$, $m_k^{-i} = \sum_{j \neq i} \mathbb{I}_{(\xi_{j,s} = k)}$,

$$A^{-i}_{k,l,s} = \{(i',j',t) : i' \neq j' \neq i, \zeta_t = s, \xi_{i',\zeta_t} = k, \xi_{j',\zeta_t} = l\},$$
$$A^i_{k,l,s} = \{(i',j',t) : i' = i, \zeta_t = s, \xi_{j',\zeta_t} = l\} \bigcup A^{-i}_{k,l,s},$$
$$A^{-i}_{l,s} = \{(j,t) : j \neq i, \zeta_t = s, \xi_{j,\zeta_t} = l\},$$

and the marginal predictive distribution, $p(\{y_{i,j,t} : (i,j,t) \in A\})$ is given by

$$\frac{\Gamma(\sum_A y_{i,j,t} + a_s)\Gamma(|A| + b_s - \sum_A y_{i,j,t})}{\Gamma(a_s + b_s + |A|)} \frac{\Gamma(a_s + b_s)}{\Gamma(a_s)\Gamma(b_s)}.$$

and $|A|$ is the number of elements in A.

(b) Since the prior for $\theta_{k,l,s}$ is conditionally conjugate, we update these parameters for $k, l \in \{1, \ldots, K^*_s\}$ by sampling from

$$\theta_{k,l,s} \mid \cdots, \mathbf{Y} \sim \text{Beta}\left(\sum_{A_{k,l,s}} y_{i,j,t} + a_s, m_{k,l,s} + b_s - \sum_{A_{k,l,s}} y_{i,j,t}\right)$$

for $A_{k,l,s} = \{(i,j,t) : i \neq j, \zeta_t = s, \xi_{i,\zeta_t} = k, \xi_{j,\zeta_t} = l\}$ and $m_{k,l,s} = |A_{k,l,s}|$.

(c) Since the prior for the transition probabilities is conditionally conjugate, the posterior full conditional for $\boldsymbol{\pi}_r$, $r = 1, \ldots, S$ is the Dirichlet distribution

$$p(\boldsymbol{\pi}_r \mid \cdots, \mathbf{Y}) = \prod_{s=1}^{S} \pi_{r,s}^{\gamma/S + n_{rs} - 1}$$

for $n_{rs} = |\{t : \zeta_{t-1} = r, \zeta_t = s\}|$.

(d) The posterior full conditional of γ is

$$p(\gamma \mid \cdots, \mathbf{Y}) \propto p(\gamma) \prod_{s=1}^{S} \frac{\Gamma(\gamma)}{\Gamma(\gamma + n_s)} \gamma^{L_s}$$

where $n_s = |\{t : \zeta_t = s\}|$ and $L_s = \sum_r \mathbb{I}_{n_{s,r} > 0}$ for $n_{s,r} = |\{t : \zeta_{t-1} = s, \zeta_t = r\}|$. Since this distribution has no standard form, we update γ using a random walk Metropolis-Hastings algorithm with symmetric log-normal proposal,

$$\log\{\gamma^{(p)}\} \mid \gamma^{(c)} \sim \mathsf{Normal}\left(\log\{\gamma^{(c)}\}, \kappa_\gamma^2\right)$$

where κ_γ^2 is a tuning parameter chosen to get an average acceptance rate between 30% and 40% .

(e) The posterior full conditional of the pairs $(a_{s,O}, b_{s,O})$ and $(a_{s,D}, b_{s,D})$ has the following general form:

$$p(a_s, b_s \mid \cdots, \mathbf{Y}) \propto p(a_s \mid d)p(b_s \mid e) \prod_{k=1}^{S} \prod_{l=1}^{S} p(y_{i,j,t} \mid A_{k,l,s}, m_{k,l,s})$$

for the marginal predictive $p(y_{i,j,t} \mid A_{k,l,s}, m_{k,l,s})$ as defined in step (b), $A_{k,l,s} = \{(i,j,t) : i \neq j, \zeta_t = s, \xi_{i,\zeta_t} = k, \xi_{j,\zeta_t} = l\}$ and $m_{k,l,s} = |A_{k,l,s}|$. Since no direct sampler is available for this distribution, we update each pair using a random walk Metropolis-Hastings algorithm with bivariate log-normal proposals,

$$\left(\log\{a_s^{(p)}\}, \log\{b_s^{(p)}\}\right)^t \mid \left(a_s^{(c)}, b_s^{(c)}\right)^t \sim$$

$$\mathsf{Normal}\left[\left(\log\{a_s^{(c)}\}, \log\{b_s^{(c)}\}\right)^t, \mathbf{\Sigma}_{ab}\right]$$

where $\mathbf{\Sigma}_{ab}$ is a tuning parameter matrix chosen independently for diagonal and off-diagonal pairs of parameters.

(f) The parameters of the Poisson-Dirichlet process (α_s, β_s) can be jointly updated using the algorithm described in Escobar and West (1995).

(g) The posterior full conditional distributions for the hyperparameters d_O, e_O, d_D, and e_D correspond to gamma distributions with shape parameter $(cS^* + 1)$ and rate parameters $(\sum_{S^*} a_{s,O} + \lambda_d)$, $(\sum_{S^*} b_{s,O} + \lambda_e)$, $(\sum_{S^*} a_{s,D} + \lambda_d)$, $(\sum_{S^*} b_{s,D} + \lambda_e)$, respectively.

References

Adamic, L., C. Brunetti, J. H. Harris, and A. A. Kirilenko (2010). Trading networks. Technical report, MIT Sloan School of Management.

Airoldi, E., D. M. Blei, S. E. Fienberg, and E. P. Xing (2008). Mixed membership stochastic blockmodels. *Journal of Machine Learning Research* **9**, 1981–2014.

Aldous, D. J. (1981). Representations for partially exchangeable arrays of random variables. *Journal of Multivariate Analysis* **11**, 581–598.

Ang, A. and G. Bekaert (2002a). International asset allocation with regime shifts. *Review of Financial Studies* **15**, 1137–1187.

Ang, A. and G. Bekaert (2002b). Regime switches in interest rates. *Journal of Business Economics and Statistics* **20**, 163–182.

Bertoin, J. (2006). *Random fragmentation and coagulation processes*. Cambridge studies in advanced mathematics. Cambridge University Press.

Boyd, N., B. Betancourt, and A. Rodríguez (2013). On the effect of electronic trading on financial market microstructure. Technical report, University of California, Santa Cruz.

Boyd, N., J. Harris, and A. Nowak (2011). The role of speculators during periods of financial distress. *Journal of Alternative Investments* **14**, 10–25.

Erdös, P. and A. Rényi (1959). On random graphs. *Publicationes Mathematicae* **6**, 290–297.

Escobar, M. and M. West (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association 90*, 577–588.

Filardo, A. (1994). Business cycle phases and their transitional dynamics. *Journal of Business, Economics, and Statistics* **12**, 299–308.

Frank, O. and D. Strauss (1986). Markov graphs. *Journal of the American Statistical Association 81*, 832–842.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In J. M. Bernardo, J. Berger, A. P. Dawid, and A. F. M. Smith (Eds.), *Bayesian Statistics 4*, Oxford, pp. 169–193. Oxford University Press.

Green, P. and S. Richardson (2001). Modelling heterogeneity with and without the Dirichlet process. *Scandinavian Journal of Statistics* **28**, 355–375.

Guidolin, M. and A. Timmermann (2005). Economic implications of bull and bear regimes in uk stock and bond returns. *The Economic Journal* **115**, 111–143.

Harris, J., W. Christie, and P. Schulta (1994). Why did NASDAQ market makers stop avoiding odd eighth quotes? *Journal of Finance 49*, 1841–1860.

Hatfield, J. W., S. D. Kominers, A. Nichifor, M. Ostrovsky, and A. Westkamp (2012). Stability and competitive equilibrium in trading networks. Technical report, Standford University.

Hoff, P., A. Raftery, and M. Handcock (2002). Latent space approaches to social network analisys. *Journal of the American Statistical Association 97*, 1090–1098.

Holland, P. and K. Leinhardt (1981). An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association 76*, 33–65.

Kemp, C., J. B. Tenenbaum, T. Griffiths, T. Yamada, and N. Ueda (2006). Learning systems of concepts with an infinite relational data. In *Proceedings of the 22nd Annual Conference on Artificial Intelligence(UAI 2006)*, Cambridge, MA, USA.

Kim, C., J. Morley, and C. Nelson (2001). Does an international tradeoff between risk and return explain mean reversion in stock prices? *Journal of Empirical Finance* **8**, 403–426.

Lau, J. W. and P. Green (2007). Bayesian model based clustering procedures. *Journal of Computational and Graphical Statistics* **16**, 526–558.

Neal, R. M. (2000). Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics* **9**, 249–265.

Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review 45*, 167–256.

Nowicki, K. and T. A. B. Snijders (2001). Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association* **96**, 1077–1087.

Ozsoylev, H. N., J. Walden, and R. Yavuz, M. D. Bildik (2010). Investor networks in the stock market. Technical report, University of California, Berkeley.

Perez-Quiroz, G. and A. Timmermann (2000). Firm size and cyclical variations in stock returns. *Journal of Finance* **55**, 1229–1262.

Pitman, J. (1995). Exchangeable and partially exchangeable random partitions. *Probability Theory and Related Fields 102*, 145–158.

Pitman, J. and M. Yor (1997). The two-parameter poisson-dirichlet distribution derived from a stable subordinator. *Annals of Probability 25(2)*, 855–900.

Rabiner, L. (1986). An introduction to hidden markov models. *IEEE ASSP Magazine*, 4–15.

Robert, C. and G. Casella (2005). *Monte Carlo Statistical Methods 2nd ed.* New York: Springer.

Rodriguez, A. (2011). Modeling the dynamics of social networks using Bayesian hierarchical blockmodels. *Statistical Analysis and Data Mining*, In press.

Rodriguez, A. and P. E. Reyes (2013). On the statistical properties of random blockmodels for network data. Technical report, University of California - Santa Cruz.

Ryden, T., T. Terasvirta, and S. Asbrink (1998). Stylized facts of daily return series and the hidden Markov model. *Journal of Applied Econometrics 13*, 217–244.

Teh, Y. W., M. I. Jordan, M. J. Beal, and D. M. Blei (2006). Sharing clusters among related groups: Hierarchical Dirichlet processes. *Journal of the American Statistical Association* **101**, 1566–1581.

Wang, Y. and G. Wong (1987). Stochastic blockmodels for directed graphs. *Journal of the American Statistical Association 82*, 8–19.

Wasserman, S. and K. Faust (1994). *Social Network Analysis: Methods and Applications.* New York: Cambridge University Press.

Xu, Z., V. Tresp, K. Yu, and H. Kriegel (2006). Infinite hidden relational models. In *Proceedings of the 22nd Annual Conference on Artificial Intelligence(UAI 2006)*, Cambridge, MA, USA.

Yau, C., O. Papaspiliopoulos, G. O. Roberts, and C. C. Holmes (2011). Bayesian non-parametric hidden markov models with applications in genomics. *Journal of the Royal Statistical Society, Series B 73*, 37–57.

Zaloom, C. (2004). Time, space and technology in financial networks. In M. Castells (Ed.), *The Network Society: A Cross-cultural Perspective*, pp. 198–216. Edward Elgar Publishing Limited.