Università degli Studi di Milano

Data Science and Economics

Bayesian approach to CDS analysis

Ivo Bonfanti
985892
Guido Giacomo Mussini
988273

Abstract

This project aims at modelling financial contagion between the main European banks via Bayesian analysis. The financial risk is proxied by the change in credit default swap (CDS) spreads over the time, while the underlying model is a Bayesian Vector Autoregressive (BVAR). For the coefficients, to handle the typical VAR over-parameterization, a Minnesota prior distribution is considered, while an Inverse-Wishart distribution incorporates prior beliefs about the second moment of the residual term. Once derived the conditional posterior distributions for both parameters, a Gibbs sampler algorithm is implemented to simulate realizations. The obtained results are eventually used to forecast and to extract a network describing financial interconnectedness.

ACADEMIC YEAR 2022-2023

Contents

1 Data Set

A credit default swap is a financial derivative providing a promise of payment, by the protection seller to the protection buyer, in case a contingent event takes place, e.g. the bankruptcy of a reference entity. Similarly to an insurance contract, the purchaser pays a fixed premium to the issuer to swap his credit risk against a third entity. The CDS spread, determining the premium to be paid, is somewhat an intuitive measure of how the reference entity is perceived by the market in terms of financial soundness. Indeed, if the entity's default is considered more likely, the related spread increases implying a higher premium.

The collected spreads series[1] make reference to 14 among the major, and geographically most representative, European banks, covering a time span ranging from November 2008 to April 2023. Data are reported on a daily basis. Figure 1 reports the banks involved in the study while figure 2 the CDS spreads series.

2 Bayesian Vector Autoregressive

The vector autoregressive (VAR) is a well established model, widely employed in the macroeconometrics literature for estimating and forecasting multivariate time series. The model is specified as follow,

$$\mathbf{y_t} = \boldsymbol{\beta_t} + B_1\mathbf{y_{t-1}} + B_2\mathbf{y_{t-2}} + \cdots + B_p\mathbf{y_{t-p}} + \boldsymbol{\epsilon_t}, \quad t = 1, 2, \ldots, T \tag{1}$$

where $\mathbf{y_t}$ is a $n \times 1$ vector of endogenous variables, in the present case, the CDS spreads of the 14 banks at time t, $\boldsymbol{\epsilon_t}$ is a $n \times 1$ vector of error terms $\boldsymbol{\epsilon_t} \overset{\text{iid}}{\sim} N(0, \Sigma)$, $\boldsymbol{\beta_t}$ is a $n \times 1$ vector of intercepts, $[B_1 \ldots B_p]$ are $n \times n$ matrices of coefficients and p is the number of lags. For a better understanding vectors are reported in bold lowercase, matrices in capital and vectors of matrices in bold capital. For the sake of clarity, the explicated form of the VAR model is reported here below,

$$\begin{aligned} y_{1t} &= \beta_{1t} + \beta_{11}^{(1)} y_{1,t-1} + \cdots + \beta_{1n}^{(1)} y_{n,t-1} + \beta_{11}^{(2)} y_{1,t-2} + \cdots + \beta_{1n}^{(2)} y_{n,t-2} + \cdots + \beta_{1n}^{(p)} y_{n,t-p} + \epsilon_{1t} \\ &\vdots \\ y_{nt} &= \beta_{nt} + \beta_{n1}^{(1)} y_{1,t-1} + \cdots + \beta_{nn}^{(1)} y_{n,t-1} + \beta_{n1}^{(2)} y_{1,t-2} + \cdots + \beta_{nn}^{(2)} y_{n,t-2} + \cdots + \beta_{nn}^{(p)} y_{n,t-p} + \epsilon_{nt} \end{aligned} \tag{2}$$

[1] CDS MM 5y senior. Source: Refinitiv Eikon

Moreover, we can represent a VAR(p) as a multivariate regression model,

$$\boldsymbol{y_t} = \boldsymbol{B X_t} + \boldsymbol{\epsilon_t} \tag{3}$$

where $\boldsymbol{X_t} = (I_n \otimes [\boldsymbol{y'_{t-1}}, \cdots, \boldsymbol{y'_{t-p}}])$ and $\boldsymbol{B} = [\boldsymbol{\beta_t}, B_1 \dots B_p]$. Finally, by stacking observation over time we get,

$$\boldsymbol{Y} = \boldsymbol{B X} + \boldsymbol{E} \tag{4}$$

the typical representation for Bayesian analysis, where $E \overset{\text{iid}}{\sim} N(0, I_T \otimes \Sigma)$.

Essentially, the model relates linearly the i^{th} bank's spread with those of the same i^{th} bank at past lags, as well as with those of the others $n-1$ banks. The parameters of interest are $\boldsymbol{B} = [\boldsymbol{\beta_t}, B_1 \dots B_p]$ and the variance-covariance matrix Σ, this last representing the second moment of the distribution of the error term. While in classical statistics these parameters $\boldsymbol{\theta}$ are considered fixed, in Bayesian statistics they are thought as random variables. The insidious but profound difference among the two schools' ways of capturing the parameters' nature, lead to different methodologies. Indeed, while frequentists maximize the sample likelihood to obtain a point estimate of the parameters of interest $\widehat{\boldsymbol{\theta}}$, Bayesians update their parameters' prior beliefs exploiting the new information retrieved from sample $\boldsymbol{Y}$, thus obtaining a posterior probability distribution $\mathcal{P}(\boldsymbol{\theta}|\boldsymbol{Y})$. This last approach grounds its foundation on the Bayes's theorem,

$$\mathcal{P}(A|B) = \frac{\mathcal{P}(B|A)\mathcal{P}(A)}{P(B)} \Rightarrow \Pi(\boldsymbol{\theta}|\boldsymbol{Y}) = \frac{\mathcal{P}(\boldsymbol{\theta})\mathcal{L}(\boldsymbol{Y}|\boldsymbol{\theta})}{P(\boldsymbol{Y})} \propto \mathcal{P}(\boldsymbol{\theta})\mathcal{L}(\boldsymbol{Y}|\boldsymbol{\theta}) \tag{5}$$

That said, a Bayesian vector autoregressive (BVAR) is structured exactly like a typical VAR. The only difference lies in the methodological approach adopted to estimate the parameters. In the following subsections, a step by step derivation of the posterior probability distributions is provided, starting from the definition of the priors and of the likelihood function.

2.1 Likelihood Function

Since the VAR(p) can be seen as a multivariate regression model, and given the normality assumption on the error term from equation (4), then the likelihood function can be expressed as $\mathcal{L}(\boldsymbol{Y}|\boldsymbol{B},\Sigma) \sim N(\boldsymbol{BX}, I_T \otimes \Sigma)$. Here below the explicated form,

$$\mathcal{L}(\boldsymbol{Y}|\boldsymbol{B},\Sigma) \propto |\Sigma|^{-\frac{T}{2}} exp\{-\frac{1}{2}(\boldsymbol{Y}-\boldsymbol{XB})'\Sigma^{-1}(\boldsymbol{Y}-\boldsymbol{XB})\} \tag{6}$$

2.2 Prior Distributions

Two priors have to be defined in order to find the conditional posterior distributions, one for each parameter. A Minnesota and an Inverse Wishart best represent our prior beliefs about $\boldsymbol{B}$ and Σ respectively.

2.2.1 Minnesota Prior

The Minnesota prior distribution, in our case adopted for the coefficients $\boldsymbol{B}$, was introduced by Litterman[1] as a shrinkage method to mitigate the overparametrization that can occur in models with a large number of dependent variables. It assumes that all the n time series which compose the model are a priori considered as univariate random walks, as they typically work well in forecasting macroeconomics series. This prior is a special case of a normal distribution where the first and the second moment depend on a restricted number of hyperparameters. In the following paragraphs its settings are illustrated.

First, recall from (2) the explicated form of the i^{th} equation,

$$y_{it} = \beta_{it} + \beta_{i1}^{(1)} y_{1,t-1} + \cdots + \beta_{in}^{(1)} y_{n,t-1} + \beta_{i1}^{(2)} y_{1,t-2} + \cdots + \beta_{in}^{(2)} y_{n,t-2} + \cdots + \beta_{in}^{(p)} y_{n,t-p} + \epsilon_{it} \tag{7}$$

The Minnesota distribution assumes that $\beta_{ij}^{(s)} \sim N(\mu_{ij}^{(s)}, \Omega_{ij}^{(s)})$, with $s = 1 \ldots p$, where the first moment is,

$$\mu_{ij}^{(s)} = \begin{cases} 1 \text{ if } i = j \text{ and } s = 1 \\ 0 \text{ otherwise} \end{cases}$$

while the second moment is structured as follow,

$$\Omega_{i,j}^{(s)} = \begin{cases} \omega^2 \text{ if } s = 1 \text{ and } i = j \\ (\omega \cdot f(s^{-1}))^2 \text{ if } s \geq 2 \text{ and } i = j \\ (\kappa \cdot \omega \cdot \tau_{i,j} \cdot f(s^{-1}))^2 \text{ if } i \neq j \end{cases}$$

where $f(s^{-1})$[3] is a deterministic function describing the decay rate of the variance term. The most common choice, the one we adopted, is the identity function which leads to a linear decay: $f(s^{-1}) = s^{-1}$. $\tau_{i,j} = \frac{\tau_i}{\tau_j}$ is a correction term for the scale of the series i with respect to the series j, while κ is a parameter handling the importance of the different series in forecasting. Indeed, by setting $\kappa < 1$, implies $\Omega_{ii}^{(s)} > \Omega_{ij}^{(s)}$, thus suggesting a greater confidence that $\beta_{ii}^{(s)} = 0$ with respect $\beta_{ij}^{(s)} = 0$, definitely reasonable for these data. A value of $\kappa = 0.5$ when $\omega = 0.2$ is suggested in the literature. [4]

Table 1 represents the probability distribution of the i^{th} equation's coefficients of a VAR(3) model. It can be seen that all distributions are centered in zero, except for the one referring to the own first lag. Also, note how the distributions flatten toward the mean for larger lags, i.e. they have a lower variance and therefore they are less informative, given the null mean. Finally, observe how the prior tends to give more importance to own lags than to other variables. The equations paired to the figure describe the distributions of the coefficients, assuming a linear decay rate.

The Minnesota prior is therefore chosen to estimate $\boldsymbol{B}$ as: $\mathcal{P}(\boldsymbol{B}) \sim N(\mu, \Omega)$, with explicit form:

$$\mathcal{P}(\boldsymbol{B}) = \frac{1}{\sqrt{2\pi\Omega}} exp\{-\frac{1}{2}\ (\boldsymbol{B} - \mu)'\Omega^{-1}(\boldsymbol{B} - \mu)\} \tag{8}$$

where μ and Ω are, respectively, the prior mean and the prior variance-covariance matrix of $\boldsymbol{B}$.

2.2.2 Inverse Wishart

According with the literature, the Inverse Wishart is well suited for incorporating prior beliefs about parameter Σ, $\mathcal{P}(\Sigma) \sim IW(\zeta, \nu)$, in explicit form:

$$\mathcal{P}(\Sigma) = \frac{|\zeta|^{\frac{\nu}{2}}}{2^{\frac{\nu n}{2}}\Gamma_n(\frac{\nu}{2})}|\Sigma|^{-\frac{\nu+n+1}{2}} exp\{-\frac{1}{2}tr(\zeta\Sigma^{-1})\}, \tag{9}$$

where ζ is the scale matrix and ν the degrees of freedom.

2.3 Posterior Distributions

Assuming a linear decay, the parameters that have to be defined are essentially four, κ and ω for the Minnesota, ν and ζ for the Inverse-Wishart.

The literature suggests $\nu = n + 1$, $\zeta = I_{[n\times n]}$, where $I_{[n\times n]}$ is an identity matrix of dimensionality n. Note that other choices are possible, ζ could be taken equals to $10 \cdot I_{[n\times n]}$, or estimated by using the OLS estimator. Again, the values chosen for κ and ω were, respectively, $\kappa = 0.5$ and $\omega = 0.2$

In the next sections the conditional posterior distributions for $\boldsymbol{B}$ and Σ are derived from the prior distributions and from the likelihood function.

2.3.1 Full conditional posterior distribution for $\boldsymbol{B}$

The conditional posterior distribution for $\boldsymbol{B}$ [6] can be derived by substituting (6) and (8) in (5). After the required computation, the result is,

$$\Pi(\boldsymbol{B}|\Sigma, \boldsymbol{Y}) \propto \ exp\{-\frac{1}{2}[(\boldsymbol{B'} - \tilde{\Omega}\tilde{\boldsymbol{B}})'\tilde{\Omega}^{-1}(\boldsymbol{B'} - \tilde{\Omega}\tilde{\boldsymbol{B}})]\} \tag{10}$$

where $\tilde{\boldsymbol{B}} = \tilde{\Omega}(\Omega^{-1}\boldsymbol{B} + \Sigma^{-2}\boldsymbol{X'Y})$ and $\tilde{\Omega} = (\Omega^{-1} + \Sigma^{-2}\boldsymbol{X'X})^{-1}$

Therefore,

$$\Pi(\boldsymbol{B}|\Sigma, \boldsymbol{Y}) \propto N(\tilde{\boldsymbol{B}}, \tilde{\Omega}) \tag{11}$$

2.3.2 Full conditional posterior distribution for Σ

Substituting (6) and (9) in (5) the conditional posterior distribution for Σ can be derived as,

$$\begin{aligned}\Pi(\Sigma|\boldsymbol{B}, \boldsymbol{Y}) \propto \frac{|\zeta + \sum_{t=1}^{T}(\boldsymbol{Y_t} - \boldsymbol{X_t B})(\boldsymbol{Y_t} - \boldsymbol{X_t B})'|^{\frac{\nu+T}{2}}}{2^{\frac{(\nu+T)n}{2}}\Gamma_n(\frac{\nu+T}{2})}|\Sigma|^{-\frac{\nu+T+n+1}{2}} \cdot \\ \cdot\, exp\{-\frac{1}{2}tr(\zeta + \Sigma(\boldsymbol{Y_t} - \boldsymbol{X_t B})(\boldsymbol{Y_t} - \boldsymbol{X_t B})'\Sigma^{-1})\}\end{aligned} \tag{12}$$

Therefore,

$$\Pi(\Sigma|\boldsymbol{B}, \boldsymbol{Y}) \sim iW(\nu + T, \zeta + \Sigma(\boldsymbol{Y_t} - \boldsymbol{X_t B})(\boldsymbol{Y_t} - \boldsymbol{X_t B})') \tag{13}$$

3 Gibbs Sampler

Once obtained the conditional posteriors from the non conjugate priors and from the likelihood function, realizations from the full posterior via the Gibbs sampling algorithm can be simulated. The Gibbs sampler is a Markov chain Monte Carlo (MCMC) method implemented in R as follow:

Algorithm 1: Gibbs Sampler

1 Initialize $Draws\boldsymbol{B}$ and $Draws\Sigma$ as empty matrices of dimensionality $[n \cdot (n \cdot p + 1) \times 9000]$ and $[(n \cdot n) \times 9000]$ respectively.
2 Set the counter $i = 1$.
3 Initialize $\Sigma^{(0)} = I_{[n \times n]}$
4 For i in 1:10000 draw,
   (I) $\boldsymbol{B}^{(i)}$ from the posterior $\Pi(\boldsymbol{B}^{(i)}|\Sigma^{(0)}, \boldsymbol{Y})$
   (II) $\Sigma^{(i+1)}$ from the posterior $\Pi(\Sigma^{(i+1)}|\boldsymbol{B}^{(i)}, \boldsymbol{Y})$
5 If i>1000,
   (I) Store $\boldsymbol{B}^{(i)}$ in $Draws\boldsymbol{B}$.
   (II) Store $\Sigma^{(i+1)}$ in $Draws\Sigma$.
6 Change the counter i to $i + 1$ and return to step 4.
7 Exit from the loop.

The number of lag p considered is 4 and, as can be noted from step 5, the burn-in period involves the first thousand iterations.

4 Results

Once the Gibbs sampler reached convergence, estimates of the coefficient matrix $\mathbf{B}$ and of the variance-covariance matrix $\boldsymbol{\Sigma}$ were obtained by averaging the posterior draws.[9]. $\mathbf{B}$ is a $[n \times (n \cdot p+1)]$ matrix reporting all the BVAR(4) coefficients. For the sake of clarity, if we consider a system with 2 banks V and W and 2 lags, the matrix will have the form:

	V_1	V_2	W_1	W_2	Constant
V	...	...	...	...	...
W	...	...	...	...	...

Finally, $\boldsymbol{\Sigma}$ is a $[n \times n]$ variance-covariance matrix of the error terms.

With these results we are able, as illustrated in next sections, to forecast the series and to extract a network representing the interconnectedness of the interbank market.

4.1 Forecast

Since the output provided by the Gibbs sampler algorithm is a realization of a Markov chain it suffers of autocorrelation. However, given the limited number of lags considered, the sample autocorrelations should be small. Nevertheless, as a precautionary measure, the thinning method has been implemented with a value of 20.

The prediction has been made by using the function *predict* of the *bvartool* package, which can produce posterior draws for standard BVAR models with independent normal-Inverse Wishart priors.[8]. The forecasts are reported in the appendix in figures 3, 4 and 5.

4.2 Financial Interconnectedness

4.2.1 Adjacency matrix

The financial interconnectedness between banks can be extracted from an adjacency matrix derived by the coefficient matrix **B**. At first, the column from **B** reporting the intercepts is removed, thus obtaining a $[n \times (n \cdot p)]$ matrix. Secondarily, the columns reporting values of different lags of the same bank have to be combined. A weighted mean has been used to accomplish this task, by assigning higher weights to the most recent lags. Defining lag = T, the mean associated to the coefficient G is the weighted mean $\bar{M}$ of all the G_t, $t = 1 \dots T$ lagged coefficients. Mathematically:

$$\bar{M} = \sum_{t=1}^{T} G_t \frac{1}{2^t}, \text{ since } \sum_{t=1}^{T} \frac{1}{2^t} \to 1 \text{ if } T \to +\infty \text{ for } t \in \mathbb{N}. \tag{14}$$

This method ensures that the first lags are more informative than the others and that for a finite number of lags, the weights sum can be approximated[2] to 1. This operation reduces the dimensionality of the matrix **B** to $[n \times n]$. As a result, the matrix represents the influence of a change in the CDS spread of a bank to all the other banks in the system.

Ultimately, a threshold has been applied to **B**, in order to discard the weaker relations and define the network links. The chosen threshold is the total mean of the matrix **B**, called $\bar{C}$. Mathematically:

$$\forall c_{i,j} \in C, c_{i,j} = \begin{cases} c_{i,j} \text{ if } c_{i,j} \geq \bar{C} \\ 0 \text{ otherwise} \end{cases}, \forall i = 1 \ldots n, j = 1 \ldots n \tag{15}$$

4.2.2 Network

The resulting network is a directed weighted graph with banks as nodes. Since the interconnection level is derived from the coefficients' mean of the BVAR model, the value of the edge $E_{A,B}$ represents the effect on the bank B of a change in the CDS spread of bank A.

The nodes opacity is proportional to the weighted-in degree, while nodes size is proportional to the weight-out degree. This means that, according to the data, the bank depending most on the others is by far Sweden bank, while those affecting most the others are Société Générale followed by Credit Suisse. The graph is illustrated in figure 6.

[2] $\sum_{t=1}^{T} \frac{1}{2^t} \sim 1$ for finite T > 1 with T $\in \mathbb{N}$. With lag $= 4$, the series sum up to 0.9375

References

[1] Minnesota prior original paper
Litterman, Robert B. "Forecasting with Bayesian Vector Autoregressions: Five Years of Experience." Journal of Business & Economic Statistics, vol. 4, no. 1, 1986, pp. 25–38.
`https://doi.org/10.2307/1391384`

[2] Minnesota prior introduction
Fabio Canova: "Methods for applied macroeconomic research"

[3] Decacy term in minnesota prior
Fabio Canova: "Methods for applied macroeconomic research"

[4] Minnesota prior model
James D.Hamilton. "TimeSeries Analysis" Ch. XII, pp. 351-362.

[5] Minnesota Prior Figure
Fabio Canova: "Methods for applied macroeconomic research"

[6] full conditional posterior distribution for beta
Ciccarelli and Rebucci. "Bayesian VARs: A survay of the recent literature with an application to the european monetary system" Ch. III, p. A.
`https://www.imf.org/external/pubs/ft/wp/2003/wp03102.pdf`

[7] BVAR
James D.Hamilton. "TimeSeries Analysis" Ch. XII, pp. 351-362.

[8] Predict function
`https://cran.r-project.org/web/packages/bvartools/bvartools.pdf`

[9] Coefficient and covariance matrices
`https://www.r-econometrics.com/timeseries/bvar/`

Appendix

$$\beta_{ii}^{(1)} \sim N(1, \omega^2) \qquad \beta_{ij}^{(1)} \sim N(0, (\kappa \cdot \omega \cdot \tau_{ij})^2)$$

$$\beta_{ii}^{(2)} \sim N(0, (\tfrac{\omega}{2})^2) \qquad \beta_{ij}^{(2)} \sim N(0, (\tfrac{\kappa \cdot \omega \cdot \tau_{ij}}{2})^2)$$

$$\beta_{ii}^{(3)} \sim N(0, (\tfrac{\omega}{3})^2) \qquad \beta_{ij}^{(3)} \sim N(0, (\tfrac{\kappa \cdot \omega \cdot \tau_{i,j}}{3})^2)$$

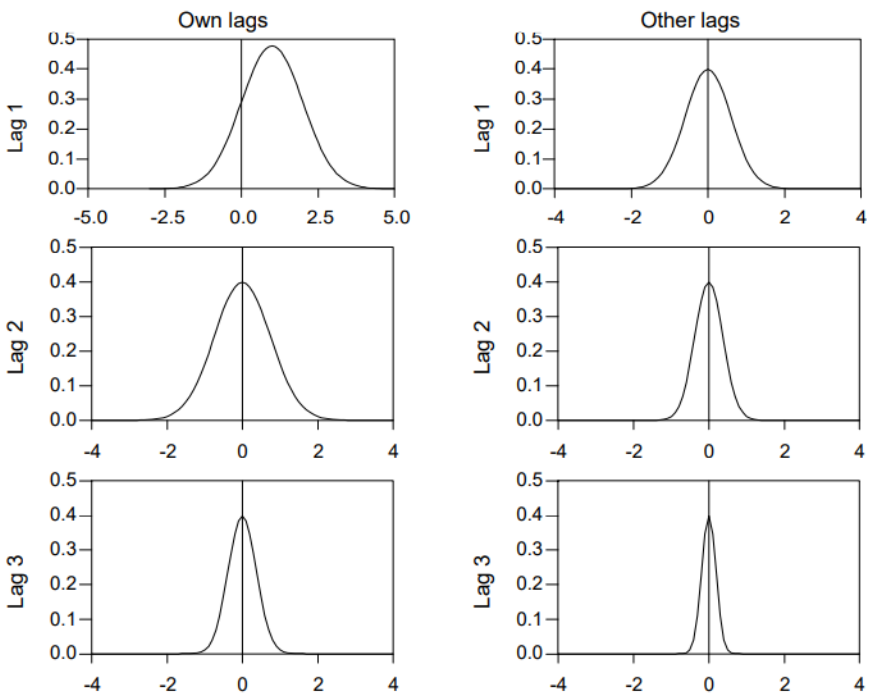

Table 1: Distributions' shapes associated to the different coefficients[5].

Appendix

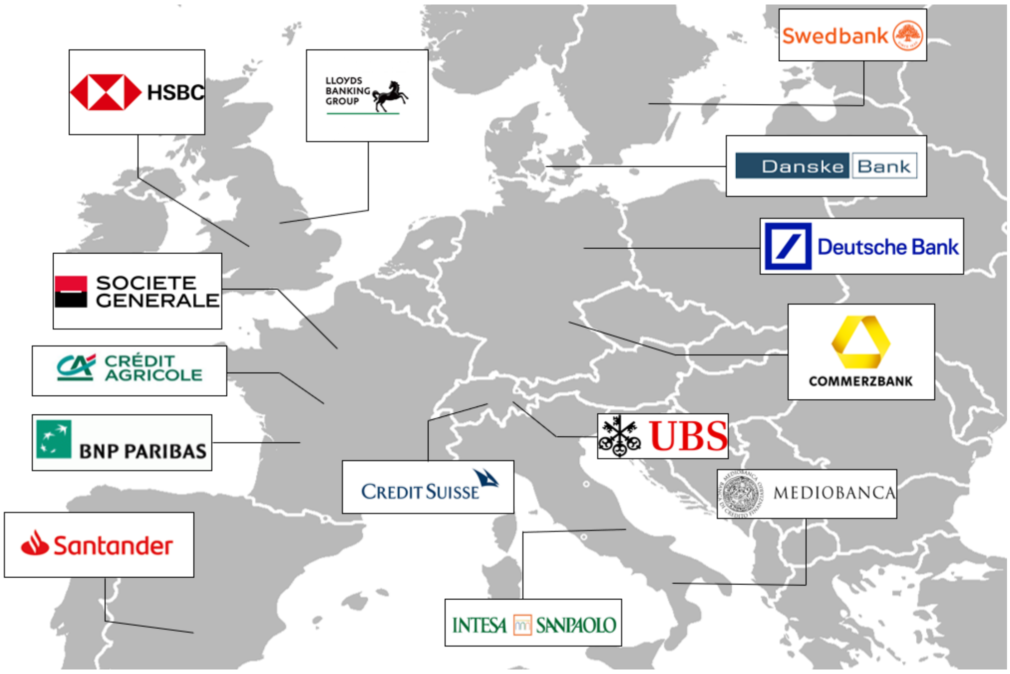

Figure 1: European banks included in the study and their nationality.

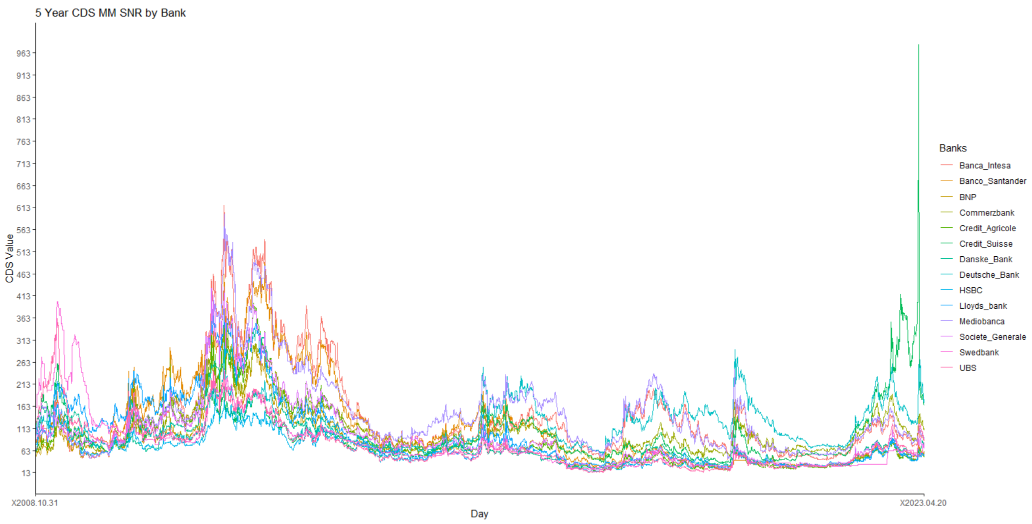

Figure 2: CDS time series of the considered banks

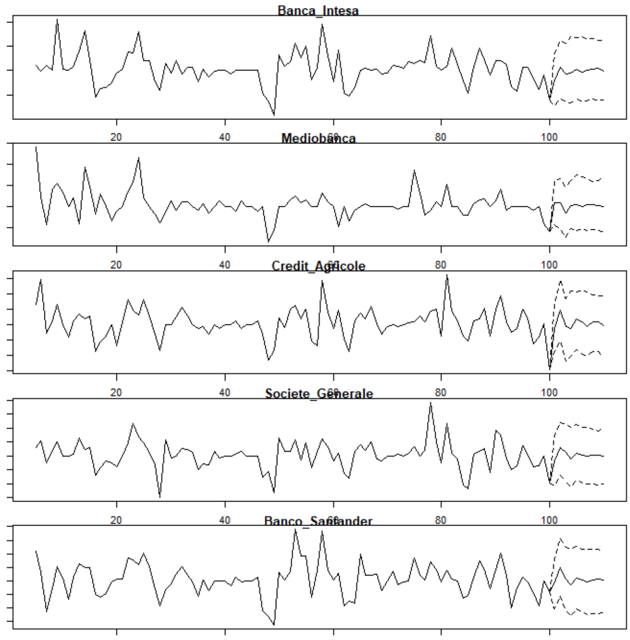

Figure 3: CDS spreads forecasts

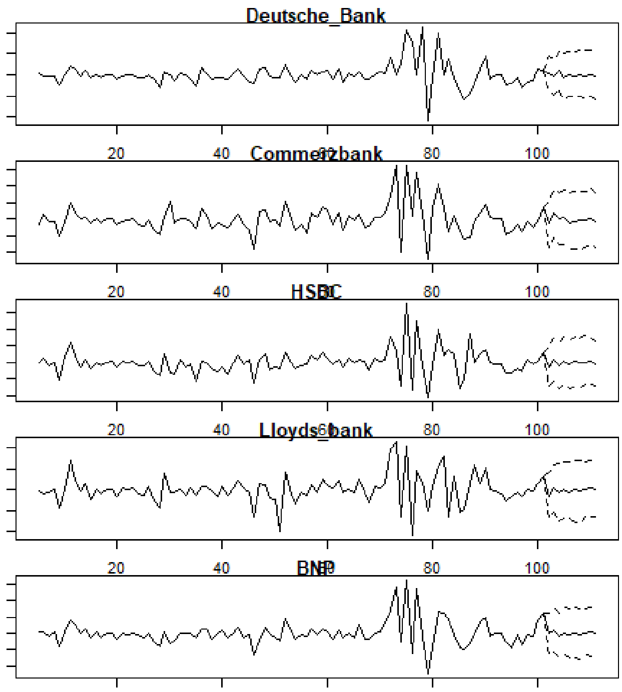

Figure 4: CDS spreads forecasts

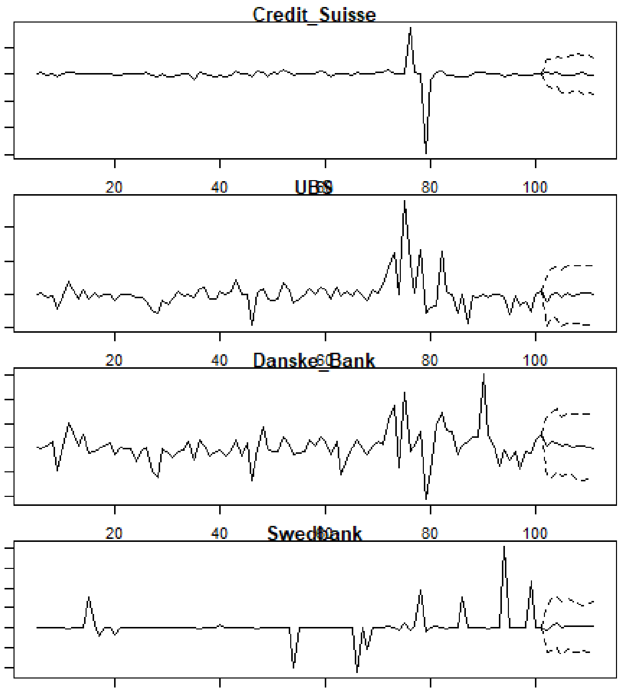

Figure 5: CDS spreads forecasts

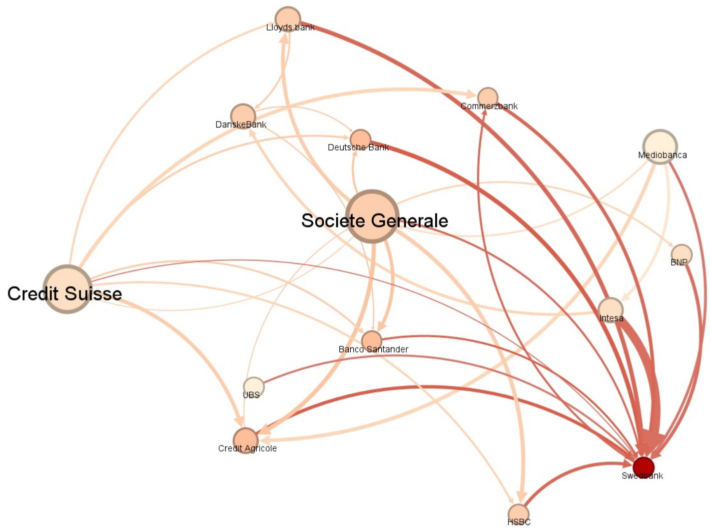

Figure 6: Interconnectedness of the European interbank market