Modelling different kinds of spatial dependence in stock returns

Matthias Arnold, Sebastian Stahlberg and Dominik Wied*

Fakultät Statistik, TU Dortmund, Germany

January 27, 2014

Abstract

The paper modifies previously suggested GMM approaches to spatial autoregression in stock returns. Our model incorporates global dependencies, dependencies inside industrial branches and local dependencies. As can be seen from Euro Stoxx 50 returns, this combination of spatial modelling and finance allows for superior risk forecasts in portfolio management.

Keywords. GMM estimation, heteroscedasticity, spatial dependence, stock returns, Value at Risk

JEL subject classifications: C13, C51, G12.

*Corresponding author: Dominik Wied, Fakultät Statistik, TU Dortmund, 44221 Dortmund, Germany, email: wied@statistik.tu-dortmund.de, Phone: +49/231/755 3869, Fax: +49/231/755 5284.

1 Introduction and summary

Although spatial modelling of dependence structures has become very popular over the last years (see e.g. Anselin, 1988, Cressie, 1991, LeSage and Pace, 2009, Baltagi and Pirotte, 2011 and the references therein), it is not yet very popular in financial applications. There is some literature on information spillovers where proximity to innovation clusters or patent activity plays an important role (see e.g. Boasson and MacPherson, 2001 or Boasson et al., 2005) and in which stock performance is used as a measure for economic success. Another contribution is Eckel et al. (2011) who measure the effects of geographical distance on stock market correlation via a regression approach. Fernandez (2011) uses similarities of financial indicators to define spatial linkages of stocks and estimates a spatial version of the capital asset pricing model. Asgharian et al. (2011) also consider different linkages like economic and monetary integration between countries to explain the propagation of country specific shocks to other countries. The two latter papers are most closely connected to our approach.

However, while both of them are mainly interested in investigating which spatial linkages are the most relevant ones, our main contribution is in risk management. We modify previously suggested spatial autoregressive models for stock returns in order to compare three different kinds of spatial dependence and show by an out-of-sample study of Euro Stoxx 50 returns that this can lead to more accurate forecasts for risk measures than standard approaches like a factor model or the sample covariance matrix.

Our model includes three different types of spatial dependence. The first one is a general dependence which affects all stocks in the same way like previous performances of stock markets in the USA or Asia. The second one is global in nature and applies to firms that belong to the same industrial branch: Since global input factors like commodity prices should have a similar effect on firms belonging to the same branch, the corresponding stock returns should display a similar behavior. The third one is a local form of dependence: Firms that are located in the same country should display similar behavior because they are exposed to the same institutional conditions like regulatory frameworks or the country-specific business cycle. It is natural to assume such a dependence structure and we can expect it to catch a lot of cross sectional dependence despite its small number of parameters.

Many papers in the literature propose alternative kinds of spatial autoregressive models. Among them are LeSage and Pace (2008) and Lee and Liu (2010) who consider models with several weighting matrices, Lin and Lee (2010), Kelejian and Prucha (2010) and Anselin (1988) who consider heteroscedastic errors or Badinger and Egger (2011) who consider both of these aspects. The latter authors provide a general higher-order cross-sectional spatial model with additional exogenous variables, spatial error autocorrelation and heteroscedastic innovations. Our approach overlaps with their approach: It is partly a simplification of it fitting to our empirical question, i.e. we provide a simple 3-order spatial lag model with no exogenous variables or spatial error autocorrelation. However, as a new contribution we also consider a time component what makes it necessary to also allow for serial dependence in the data.

The parameters of the model are estimated by an easy to implement two stage procedure. We circumvent the large number of variance parameters by choosing our moment conditions in such a way that the variance parameters are not needed. The parameters of spatial dependence are estimated by GMM similar to Kelejian and Prucha (1999) and Kapoor et al. (2007). In a second step, given these GMM estimates, estimation of the variance parameters is straightforward. Using results of Hansen (1982), the GMM estimators of the correlation parameters can be shown to be consistent and asymptotically normal, which allows for asymptotic confidence intervals. In addition, we show how to estimate the variance parameters consistently. Our two stage procedure should not be confused with the two stage procedure of e.g. Badinger and Egger (2011) who estimate the common regression parameters in two stages but due to the missing time component cannot estimate the variance parameters consistently.

2 Two step estimation procedure

In this section, we present the spatial autoregressive model for stock returns and prove some results on the parameter estimates.

For $t = 1, \dots, T$, let y_t be an n-dimensional random vector. In the cross-sectional dimension, the components of y_t are assumed to be spatially correlated where we allow for three different kinds of spatial dependence:

$$y_t = \rho_g W_g y_t + \rho_b W_b y_t + \rho_l W_l y_t + \varepsilon_t,\ t = 1, \dots, T. \tag{1}$$

ρ_g denotes the general dependence parameter, ρ_b the parameter of dependence inside branches and ρ_l the local dependence parameter. W_g, W_b and W_l are the respective weighting matrices which are specified later. For the rest of this section and the proofs, we change the notation for ease of exposition, i.e. we write $W_1 := W_g, W_2 := W_b, W_3 := W_l$. Let A' be the transpose of a given matrix A.

Define $\rho := (\rho_g, \rho_b, \rho_l)'$. We maintain the following assumptions.

Assumption 1. *1. The sequence $(y_t : t \in \mathbb{Z})$ has zero mean, is stationary and ergodic.*

2. For $i \in \{1, 2, 3\}$, $r = 1, \dots, n$, $s = 1, \dots, n$, $W_{i,rs} \geq 0$, $W_{i,rr} = 0$.

3. For $i \in \{1, 2, 3\}$ and $r = 1, \dots, n$, $\sum_{s=1}^{n} W_{i,rs} = 1$.

4. The parameter space S is defined as $S = \{\rho \in \mathbb{R}^3, |\rho_g| + |\rho_b| + |\rho_l| < 1\}$.

5. For $t \in \mathbb{Z}$, $\mathsf{Cov}\,(\varepsilon_t) = diag\{\sigma_1^2, \dots, \sigma_n^2\} =: \Sigma$.

The spatial weight matrices W_1, W_2 and W_3 are known; the elements on the main diagonals are zero and the matrices are row-standardized. We assume that the whole amount of spatial dependence is captured by the three types of spatial dependence so that the innovations, i.e. the elements of ε_t, can be assumed to be uncorrelated. However, they may be heteroscedastic. In our model, we do not include any explanatory variables, but generalizations to cases where the expectation of y_t is not 0 are straightforward under additional assumptions on the regressors since the spatial correlation structure (1) could then be applied to the disturbances of the corresponding regression model. However, in the context of daily stock returns, the zero mean assumption is plausible, see also Aue et al. (2009).

Partly, our model is a simplification of Badinger and Egger (2011) who allow for an arbitrary number of weighting matrices, explicitely consider exogenous variables and additionally allow for autoregression in the error terms. However, they do not include a time component and therefore do not include any serial dependence as we do with our ergodicity assumption. While they provide asymptotics in n, we consider the case $T \to \infty$.

As long as Assumption 1.3 holds, the inverse of the matrix $(I_n - \rho_g W_1 - \rho_b W_2 - \rho_l W_3)$ exists and our model leads to

$$\begin{aligned} \mathsf{Cov}(y_t) &= (I_n - \rho_g W_1 - \rho_b W_2 - \rho_l W_3)^{-1} \Sigma (I_n - \rho_g W_1' - \rho_b W_2' - \rho_l W_3')^{-1} \\ &=: V. \end{aligned} \tag{2}$$

Of course, the parameters could be estimated by way of maximum likelihood. Assuming normality and independence over time, the likelihood function would be

$$L(\rho_g, \rho_b, \rho_l, \Sigma) = (2\pi)^{-\frac{nT}{2}} (\det V)^{-\frac{T}{2}} \exp\left(-\frac{1}{2} \sum_{t=1}^{T} y_t' V^{-1} y_t \right).$$

Altogether, our model contains $n+3$ parameters, the three correlation parameters ρ_g, ρ_b and ρ_l and n parameters of variance, σ_i^2. Thus, the calculation of the maximum likelihood estimates can be computationally expensive, especially if n is large. For a model with only one kind of spatial dependence but additional regressors, Lin and Lee (2010) show that maximum likelihood estimation is inconsistent if the heteroscedasticity is not taken into account.

As an alternative, we suggest a two step estimation procedure which is easy to compute. First, we estimate the correlation parameters by generalized method of moments along the lines of Kelejian and Prucha (1999) or Kapoor et al. (2007). We will show that this step does not depend on the parameters of variance. Second, given the estimated correlation parameters it is straightforward to estimate the variance parameters.

The GMM estimator for the correlation parameters uses the following three moment conditions:

$$\begin{aligned}
\mathsf{E}\left(\varepsilon_t' W_1 \varepsilon_t\right) &= \mathsf{tr}(W_1\Sigma) = 0, \\
\mathsf{E}\left(\varepsilon_t' W_2 \varepsilon_t\right) &= \mathsf{tr}(W_2\Sigma) = 0, \\
\mathsf{E}\left(\varepsilon_t' W_3 \varepsilon_t\right) &= \mathsf{tr}(W_3\Sigma) = 0.
\end{aligned}$$

Note that the variance parameters σ_i^2 do not enter the moment conditions. Replacing ε_t by

$$\varepsilon_t = \left(I_n - \rho_g W_1 - \rho_b W_2 - \rho_l W_3\right) y_t$$

and making use of the weak-sense stationarity of the y_t gives the theoretical system of equations

$$\Gamma\lambda + \gamma = 0,$$

where

$$\lambda := \lambda(\rho) := \left(\rho_g, \rho_b, \rho_l, \rho_g^2, \rho_b^2, \rho_l^2, \rho_g\rho_b, \rho_g\rho_l, \rho_b\rho_l\right)'$$

and for $i, j \in \{1, 2, 3\}$, the elements of $\Gamma \sim (3 \times 9)$ and $\gamma \sim (3 \times 1)$ are defined by

$$\begin{aligned}
\Gamma_{i,j} &= \mathsf{E}\left(-y_t' \left(W_i + W_i'\right) W_j y_t\right), && (3) \\
\Gamma_{i,3+j} &= \mathsf{E}\left(y_t' W_j' W_i W_j y_t\right), && (4) \\
\Gamma_{i,7} &= \mathsf{E}\left(y_t' W_1' \left(W_i + W_i'\right) W_2 y_t\right), && (5) \\
\Gamma_{i,8} &= \mathsf{E}\left(y_t' W_1' \left(W_i + W_i'\right) W_3 y_t\right), && (6) \\
\Gamma_{i,9} &= \mathsf{E}\left(y_t' W_2' \left(W_i + W_i'\right) W_3 y_t\right), && (7) \\
\gamma_i &= \mathsf{E}\left(y_t' W_i y_t\right).
\end{aligned}$$

Let G and g be the empirical counterparts of Γ and γ, i.e., for $i \in \{1, 2, 3\}$, $j \in \{1, \ldots, 9\}$, $G_{i,j}$ and g_i are given by $\Gamma_{i,j}$ and γ_i with the expectation operator replaced by a sample average, respectively. The GMM estimator for ρ_g, ρ_b and ρ_l is defined as

$$\hat{\rho}_{GMM} := (\hat{\rho}_g, \hat{\rho}_b, \hat{\rho}_l)'_{GMM} := \arg\min_{\rho \in S} ||G\lambda + g||.$$

The theoretical term $\Gamma\lambda + \gamma$ is equal to zero for the true parameter values. Our GMM estimator is calculated by finding the values for ρ_g, ρ_b and ρ_l for which the corresponding empirical system $G\lambda + g$ is closest to zero. Compared to the ML estimator of the model parameters, this GMM estimator is easy to calculate: We just have to minimize $||G\lambda + g||$ with respect to ρ_g, ρ_b and ρ_l. Even for large n, this is easy to handle. In particular, the parameters of variance σ_i^2 are not needed to calculate the GMM estimator for the correlation parameters. The following theorem states consistency and asymptotic normality of the GMM estimator for $T \to \infty$. For the proof, we need an additional assumption.

Assumption 2. *1. The true parameter $\rho_0 \in S$ is the unique solution of the theoretical system of equations, i.e.*

$$\Gamma\lambda + \gamma = 0 \Leftrightarrow \rho = \rho_0.$$

2. The matrix $\mathsf{E}\left(\frac{\partial(G\lambda+g)}{\partial\rho}(y_1, \rho_0)\right) = d_0 = \Gamma\lambda^{(1)}$ *with*

$$\lambda^{(1)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 2\rho_g & 0 & 0 \\ 0 & 2\rho_b & 0 \\ 0 & 0 & 2\rho_l \\ \rho_b & \rho_g & 0 \\ \rho_l & 0 & \rho_g \\ 0 & \rho_l & \rho_b \end{pmatrix}.$$

exists, is finite and has full rank.

3. For

$$f(y_t, \rho_0) = \begin{pmatrix} \varepsilon_t' W_1 \varepsilon_t \\ \varepsilon_t' W_2 \varepsilon_t \\ \varepsilon_t' W_3 \varepsilon_t \end{pmatrix},$$

it holds that, for $j \to \infty$, $\mathsf{E}(f(y_t, \rho_0)|f(y_{t-j}, \rho_0), f(y_{t-j-1}, \rho_0), \ldots)$ converges in mean square to zero and that, for

$$\begin{aligned} v_j &:= \mathsf{E}(f(y_t, \rho_0)|f(y_{t-j}, \rho_0), f(y_{t-j-1}, \rho_0), \ldots) \\ &\quad -\mathsf{E}(f(y_t, \rho_0)|f(y_{t-j-1}, \rho_0), f(y_{t-j-2}, \rho_0), \ldots), \end{aligned}$$

the infinite sum $\sum_{j=0}^{\infty} \mathsf{E}(v_j' v_j)^{1/2}$ *is finite.*

Theorem 2.1. *Under Assumptions 1 and 2, for* $T \to \infty$*,*

1. $\hat{\rho}_{GMM} \to_p \rho_0$

2. $\sqrt{T}(\hat{\rho}_{GMM} - \rho_0) \to_d N(0, d_0^{-1} S_W (d_0^{-1})')$ *with*

$$S_W = \sum_{t=-\infty}^{\infty} \mathsf{E}(f(y_1, \rho_0) f(y_t, \rho_0)').$$

Remark. For $i, j \in \{1, 2, 3\}$, the entries of $\mathsf{E}(G) = \Gamma$ given in (3)-(7) can be calculated as

$$\begin{aligned}
&\Gamma_{i,j} = \mathsf{tr}\left(\left(W_i + W_i'\right) W_j V\right), \Gamma_{i,3+j} = \mathsf{tr}\left(W_j' W_i W_j V\right), \\
&\Gamma_{i,7} = \mathsf{tr}\left(W_1'\left(W_i + W_i'\right) W_2 V\right), \Gamma_{i,8} = \mathsf{tr}\left(W_1'\left(W_i + W_i'\right) W_3 V\right) \text{ and} \\
&\Gamma_{i,9} = \mathsf{tr}\left(W_2'\left(W_i + W_i'\right) W_3 V\right).
\end{aligned}$$

To calculate confidence intervals in finite samples, d_0 and S_W can be replaced by consistent estimates. We suggest using G and the estimated correlation parameters for the matrix d_0 as well as a kernel-based variance estimator for S_W, see e.g. de Jong and Davidson (2000). The latter requires choosing a kernel and a bandwidth.

The proof for Theorem 2.1 is basically an application of Hansen (1982). Details are given in the appendix. Simulation results which are not reported here but are available from the corresponding author upon request show that the estimation method works well even in small samples.

Given the estimates for the correlation parameters, estimation of the parameters of variance in the second step is straightforward: We just take the averages over the estimated $\hat{\varepsilon}_{i,t}^2$:

$$\begin{aligned}
\hat{\sigma}_i^2 &:= \frac{1}{T} \sum_{t=1}^{T} \hat{\varepsilon}_{i,t}^2 := \frac{1}{T} \sum_{t=1}^{T} \left[\left(I_n - \hat{\rho}_g W_1 - \hat{\rho}_b W_2 - \hat{\rho}_l W_3\right) y_t\right]_i^2 \\
&= \frac{1}{T} \sum_{t=1}^{T} e_i' \left(I_n - \hat{\rho}_g W_1 - \hat{\rho}_b W_2 - \hat{\rho}_l W_3\right) y_t y_t' \left(I_n - \hat{\rho}_g W_1' - \hat{\rho}_b W_2' - \hat{\rho}_l W_3'\right) e_i,
\end{aligned}$$

where e_i is the i-th unit vector. Consistency then mainly follows with the ergodic theorem:

Theorem 2.2. *Under assumptions 1 and 2, for* $T \to \infty$ *and* $i = 1, \ldots, n$, $\hat{\sigma}_i^2 \to_p \sigma_i^2$.

3 Application to stock returns

We analyze spatial dependencies in the daily stock returns of the Euro Stoxx 50 members in the composition of January 2010 for the period from 2003 until 2009, using adjusted stock prices from Datastream which we transfer to log returns.
Table 1 shows the partitioning of the Euro Stoxx 50 members into branches and countries.

- Table 1 here -

Nokia and CRH are the only representatives of their home countries, respectively, but in order to avoid singularities, groups must not consist of only one member. We consider two different groupings. In model 1, we impose a group called „others“ for Finland and Ireland, where only one company is part of the Euro Stoxx 50, respectively. In model 2, we put Nokia and CRH to the Benelux group which would then be labeled „small countries“. According to these groupings, the adjacency matrices are constructed in the following way.
The off-diagonal elements of the general adjacency matrix W_g are chosen as the weights of the firms in the Euro Stoxx 50. In W_b and W_l, the element in the i^{th} row and j^{th} column is nonzero if the corresponding stocks belong to the same branch (W_b) or country (W_l). In each row, the nonzero entries again consist of the stock weights in the Euro Stoxx 50. Finally, the three adjacency matrices are row-standardized. There are two reasons for choosing the non-zero elements proportional to market capitalizations: First, we expect big stocks to have stronger influence on "neighboring" stock returns than small stocks. Second, this provides the following economic interpretation. For each day t, $t = 1, \dots, T$, the spatial autoregressive model

$$y_t = \rho_g W_g y_t + \rho_b W_b y_t + \rho_l W_l y_t + \varepsilon_t \tag{8}$$

regresses the stock returns on three components: The weighted market return of the same day (as measured by $W_g y_t$), the weighted market return of the respective industrial branches ($W_b y_t$) and the weighted local market return of the respective countries ($W_l y_t$).

The elements of the innovation vector ε_t are heteroscedastic, but there is no spatial correlation among the innovations. We checked this assumption by performing Moran's I tests on the model residuals for the matrices W_g, W_b and W_l, respectively. These tests do not reveal evidence for spatial correlation among the innovations.
For $\rho_b = \rho_l = 0$, the spatial model (8) would correspond to a one factor model

$$y_{it} = \alpha_i + \beta_i y_{mt} + \eta_{it}$$

with constant β_i for all stocks, where y_{mt} is the market return on day t. The spatial model replaces the 50 different β_i of the factor model by only two additional spatial lags, one for dependencies inside industrial branches and one for local dependencies.

3.1 Evolution of spatial dependencies

We estimate the dependence parameters ρ_g, ρ_b and ρ_l on rolling windows of 250 trading days and use part 2 of Theorem 2.1 to calculate pointwise asymptotic 95% confidence intervals, thereby using the Bartlett kernel and bandwidth $[\log(T)]$ for the long-run variance estimator.
Figure 1 shows the results.

- Figure 1 here -

For the whole data set, general dependence is the largest in both models. In model 2, where all companies of small countries are put together in one group, local dependence is smaller than in model 1 (perhaps because the group of small countries contains countries which are not really locally connected), whereas general dependence increases correspondingly. Dependence inside industrial branches is basically the same for both models. This leads to the fact that in model 2, dependence inside branches is almost always higher than local dependence. Also in model 1, dependence inside branches is mostly either similar or higher than local dependence.
The pointwise asymptotic confidence intervals very rarely include zero for the three kinds of dependence over the whole time span. For both models, general dependence strongly increases after insolvency of Lehman Brothers in September 2008, whereas dependence between branches decreases correspondingly. This effect goes along with

much wider confidence intervals for these two dependencies. It seems to be more difficult to distinguish between general dependence and dependence inside industrial branches in times of crises. Comparing models 1 and 2, we conclude that the results are very similar so that sensitivity with respect to the choice of weight matrices seems to be limited.

3.2 Risk estimation

We investigate the utility of our spatial approach to model stock returns by comparing the accuracy of predicted Values at Risk (VaR). Replacing the unknown parameters in (2) by their estimates yields an estimate $\hat{V}_{spat}$ for the stock returns' covariance matrix V. This estimate can be compared to alternative estimates of V. This study considers two alternative estimates of V: The sample covariance matrix $\hat{V}_{samp} = \frac{1}{T-1}\sum_{t=1}^{T}(y_t - \bar{y})(y_t - \bar{y})'$ and the estimated covariance matrix resulting from a one factor model described in e.g. Jorion (2001), p. 169 ff. The latter one models the return of stock i by

$$y_{it} = \alpha_i + \beta_i y_{mt} + \eta_{it}, \tag{9}$$

where y_{mt} is the return of a market portfolio and the idiosyncratic terms η_{it} may be heteroscedastic with variances $\sigma^2_{\eta i}$, but are assumed to be uncorrelated across assets. The one factor model implies that the stock returns' covariance matrix can be written as

$$V_{fact} = \beta\beta'\sigma^2_m + D_\eta,$$

where $\beta = (\beta_1, \ldots, \beta_n)'$ and $D_\eta = diag(\sigma^2_{\eta 1}, \ldots, \sigma^2_{\eta n})$. To estimate V_{fact}, we take the returns of the Euro Stoxx 50 index as market returns and estimate (9) separately for all stocks.

Each of the three models suggests a different vector of portfolio weights to minimize portfolio variance. The minimizing weights are given by

$$\frac{\hat{V}^{-1}\tau}{\tau'\hat{V}^{-1}\tau},$$

where τ denotes a vector of ones. The three different approaches to model dependencies in stock returns can thus be compared in the following way: For each of the models, the resulting covariance matrix is calculated on a rolling window of 100 days. This provides minimal variance portfolio weights as well as an estimate for the corresponding portfolio variance, which is given by

$$\hat{\sigma}^2_{port} := \left(\tau' \hat{V}^{-1} \tau\right)^{-1}.$$

The resulting Gaussian VaR at level α is

$$\widehat{VaR}_\alpha := u_\alpha \sqrt{\hat{\sigma}^2_{port}},$$

where u_α is the α-quantile of the standard normal distribution. Alternatively, one could use quantiles from some heavy tailed distribution. We stay with the normal quantiles for two reasons. On the one hand, the portfolio returns are weighted averages of 50 single returns so that deviations from the normal distribution should be smaller than for single stock returns. On the other hand, the choice of some other distribution would affect all models in the same way so that the comparison of the models would remain the same.

For each α and each of the three models, we thus get daily updated estimated VaR. We compare these with the realized portfolio returns of the following day. For a good model, the percentage of days where the realized portfolio return is smaller than $\widehat{VaR}_\alpha$ should be close to α. Consequently, we assess model performance by comparing α to the share of days where the portfolio return falls below $\widehat{VaR}_\alpha$. Figure 2 shows the results for $\alpha \in (0, 0.05)$.

- Figure 2 here -

Indeed, the spatial model seems to be more adequate to estimate risk than the other two approaches. Consider e.g. the estimated VaR for $\alpha = 0.01$. For the spatial model, portfolio returns fall below $\widehat{VaR}_\alpha$ in 2.3% of all days, whereas this happens much more frequently for the one factor model (6.9%) or the sample covariance matrix (12.1%). This pattern can be found for all values of α considered here. For $\alpha = 0.05$, the actual quantiles are 12.7% for the one factor model and 19.3% for the sample covariance matrix, whereas the portfolio returns fall below the $\widehat{VaR}_\alpha$ produced by the spatial model in only 6.3% of all cases. We conclude that the spatial approach can be useful to measure risk adequately.

4 Discussion

We have proposed a spatial autoregressive model for financial data, which is a rather novel approach in the literature, and have seen that a lot of cross sectional dependence is captured by it.

The model contains three spatial lags which represent the market returns of (i) the whole market, (ii) the industrial branches and (iii) the countries. In an out-of-sample study to forecast VaR for minimum variance portfolios, the model performs surprisingly well compared to a one factor model or the sample covariance matrix.

The reason why the spatial approach seems to be more adequate to estimate risk might be that it is indeed able to capture a lot of cross sectional dependence while at the same time, the model is comparatively sparse so that sampling errors for the spatial model are smaller than for the other two models. To this end, note that the number of unknown parameters to be estimated is different for the three models: The sample covariance matrix does not imply any structure at all so that $50 * 51/2 = 1275$ parameters have to be estimated, whereas the one factor model includes only $50 + 50 + 1 = 101$ unknown parameters. This number is reduced even further by the spatial model which imposes more structure and consists of only $50 + 3 = 53$ parameters. This is due to the fact that only three correlation parameters have to be estimated instead of 50 β_i's.

The results suggest that in order to estimate risk accurately, it pays off to impose a sparse structure for the stock returns' covariance matrix.

Acknowledgement:

Financial support by Deutsche Forschungsgemeinschaft (SFB 823, Statistik nichtlinearer dynamischer Prozesse) is gratefully acknowledged. We are grateful to the editor Badi H. Baltagi, an anonymous associate editor and two anonymous referees for helpful comments and suggestions.

References

T. Amemiya. Regression analysis when the dependent variable is truncated normal. *Econometrica*, 41(6):997–1016, 1973.

L. Anselin. *Spatial Econometrics: Methods and Models*. Kluwer, Dordrecht, Boston and London, 1988.

H. Asgharian, W. Hess, and L. Liu. A spatial analysis of international stock market linkages. *working paper*, 2011.

A. Aue, S. Hörmann, L. Horvath, and M. Reimherr. Break detection in the covariance structure of multivariate time series models. *Annals of Statistics*, 37(6B): 4046–4087, 2009.

H. Badinger and P. Egger. Estimation of higher-order spatial autoregressive cross-section models with heteroskedastic disturbances. *Papers in Regional Science*, 90(1):213–235, 2011.

B. Baltagi and A. Pirotte. Seemingly unrelated regressions with spatial error components. *Empirical Economics*, 40(1):5–49, 2011.

V. Boasson and A. MacPherson. The role of geographic location in the financial and innovation performance of publicly traded pharmaceutical companies: empirical evidence from the United States. *Environment and Planning A*, 33(8):1431–1444, 2001.

V. Boasson, E. Boasson, and A. MacPherson. Firm value and geographic competitive advantage: Evidence from the US pharmaceutical industry. *Journal of Business*, 78(6):2465–2495, 2005.

N. A. C. Cressie. *Statistics for Spatial Data*. John Wiley & Sons, New York, 1991.

R.M. de Jong and J. Davidson. Consistency of kernel estimators of heteroscedastic and autocorrelated covariance matrices. *Econometrica*, 68(2):407–424, 2000.

S. Eckel, G. Löffler, A. Maurer, and V. Schmidt. Measuring the effects of geographical distance on stock market correlation. *Journal of Empirical Finance*, 18(2): 237–247, 2011.

V. Fernandez. Spatial linkages in international financial markets. *Quantitative Finance*, 11(2):237–245, 2011.

L.P. Hansen. Large sample properties of generalized method of moments estimators. *Econometrica*, 50:1029–1054, 1982.

R.I. Jennrich. Asymptotic properties of non-linear least squares estimators. *Annals of Mathematical Statistics*, 40:633–643, 1969.

P. Jorion. *Value at Risk: The new benchmark for managing financial risk*. McGraw-Hill, 2 edition, 2001.

M. Kapoor, H. H. Kelejian, and I. R. Prucha. Panel data models with spatially correlated error components. *Journal of Econometrics*, 140(1):97–130, 2007.

H. H. Kelejian and I. R. Prucha. A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review*, 40: 509–533, 1999.

H. H. Kelejian and I. R. Prucha. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics*, 157(1):53–67, 2010.

L.F. Lee and X. Liu. Efficient GMM estimation of high order spatial autoregressive models with autoregressive disturbances. *Econometric Theory*, 26(1):187–230, 2010.

J. LeSage and R. K. Pace. *Introduction to Spatial Econometrics*. CRC Press, Boca Raton, London and New York, 2009.

J. P. LeSage and R. K. Pace. Spatial econometric modeling of origin-destination flows. *Journal of Regional Science*, 48(5):941–967, 2008.

X. Lin and L.F. Lee. GMM estimation of spatial autoregressive models with unknown heteroskedasticity. *Journal of Econometrics*, 157(1):34–52, 2010.

B.M. Poetscher and K. Prucha. Basic structure of the asymptotic theory in dynamic nonlinear econometric models, Part I: Consistency and approximation concepts. *Econometric Reviews*, 10:125–216, 1991.

A Proofs

Proof of Theorem 2.1.1

This follows by standard arguments as e.g. presented in Poetscher and Prucha (1991), Amemiya (1973) or Jennrich (1969), using the uniform convergence of $G\lambda+g$ to $\Gamma\lambda+\gamma$ and the identifiability condition.

Proof of Theorem 2.1.2

For the asymptotic normality, we apply Theorem 3.1 of Hansen (1982), which provides a general result about the asymptotic normality of GMM estimators. Assumption 3.1 of Hansen (1982) is fulfilled by our dependence assumption, Assumption 3.2 by our Assumption 1.4, Assumptions 3.3 and 3.4 by the fact that $\Gamma\lambda+\gamma$ is a polynomial in ρ and our Assumption 2.2, Assumption 3.6 by choosing the unity matrix. Assumption 3.5 is fulfilled because of our Assumption 2.3 and the fact that $S_W^0 := \mathsf{E}\left(f(y_1,\rho_0)f(y_1,\rho_0)'\right)$ exists and is finite, which can be seen by the following calculations:

For $i, j \in \{1, 2, 3\}$,

$$\begin{aligned}
S^0_{W,ij} &= \mathsf{E}\left(\varepsilon_1' W_i \varepsilon_1 \varepsilon_1' W_j \varepsilon_1\right) \\
&= \mathsf{E}\left(\sum_{r=1}^n \sum_{s=1}^n \left(\varepsilon_{1,r}\varepsilon_{1,s} W_{i,rs}\right) \cdot \sum_{u=1}^n \sum_{v=1}^n \left(\varepsilon_{1,u}\varepsilon_{1,v} W_{j,uv}\right)\right) \\
&= \sum_{r=1}^n \sum_{s=1}^n \sum_{u=1}^n \sum_{v=1}^n W_{i,rs} W_{j,uv} \mathsf{E}\left(\varepsilon_{1,r}\varepsilon_{1,s}\varepsilon_{1,u}\varepsilon_{1,v}\right) \\
&= \sum_{r=1}^n W_{i,rr} W_{j,rr} \mathsf{E}\left(\varepsilon_{1,r}^4\right) + \sum_{r=1}^n \sum_{s=1}^n W_{i,rr} W_{j,ss} \sigma_r^2 \sigma_s^2 \\
&+ \sum_{r=1}^n \sum_{s=1}^n W_{i,rs} W_{j,rs} \sigma_r^2 \sigma_s^2 + \sum_{r=1}^n \sum_{s=1}^n W_{i,rs} W_{j,sr} \sigma_r^2 \sigma_s^2 \\
&- 3 \sum_{r=1}^n W_{i,rr} W_{j,rr} (\sigma_r^2)^2 \\
&= \sum_{r=1}^n \sum_{s=1}^n W_{i,rs} W_{j,rs} \sigma_r^2 \sigma_s^2 + \sum_{r=1}^n \sum_{s=1}^n W_{i,rs} W_{j,sr} \sigma_r^2 \sigma_s^2,
\end{aligned}$$

where the last line follows from the fact that $W_{i,rr} = 0$ for all i and r. Since this is a finite sum of entries of the weighting matrices and variances of ε (which are both bounded), Assumption 3.5 follows.

Now, we get the convergence

$$\sqrt{T}(\hat{\rho}_{GMM} - \rho_0) \to_d N(0, d_0^{-1} S_W (d_0^{-1})').$$

The matrix d_0 is given by

$$\mathsf{E}\left(\frac{\partial f}{\partial \rho}(y_1, \rho_0)\right) = \mathsf{E}\left(\frac{\partial (G\lambda + g)}{\partial \rho}(y_1, \rho_0)\right),$$

which anon is equal to $\mathsf{E}(G)\lambda^{(1)} = \Gamma\lambda^{(1)}$. The expression for the entries in Γ in the Remark then follows from the independence of the y_t and a formula for the expectation of quadratic forms, namely

$$\mathsf{E}(\varepsilon' A \varepsilon) = \mathsf{tr}(A\mathsf{Cov}(\varepsilon)) + \mathsf{E}(\varepsilon)' A \mathsf{E}(\varepsilon).$$

Weak consistency directly follows from the asymptotic normality. □

Proof of Theorem 2.2

First, we define the quantity

$$\tilde{\sigma}_i^2 = \frac{1}{T} \sum_{t=1}^T \varepsilon_{i,t}^2 = \frac{1}{T} \sum_{t=1}^T [(I_n - \rho_g W_1 - \rho_b W_2 - \rho_l W_3) y_t]_i^2.$$

We have

$$\hat{\sigma}_i^2 - \sigma_i^2 = (\hat{\sigma}_i^2 - \tilde{\sigma}_i^2) + (\tilde{\sigma}_i^2 - \sigma_i^2) =: A + B.$$

To show convergence of $\hat{\sigma}_i^2 - \sigma_i^2$ against 0 in probability, we thus have to show that A and B converge to 0 in probability. The latter follows from the ergodic theorem and the identity $\mathsf{E}(\varepsilon_{i,t}^2) = \sigma_i^2$. For the convergence of A, write

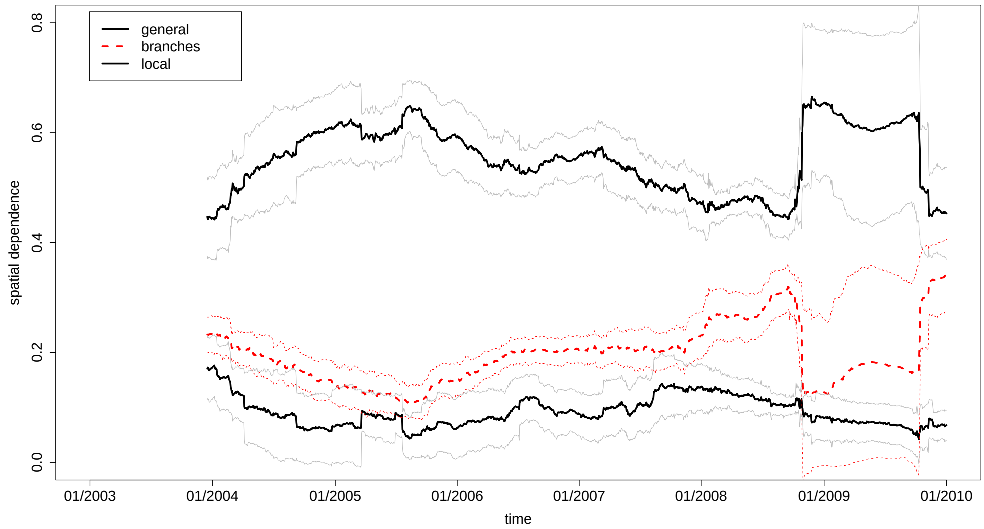
$$\begin{aligned} A = {} & e_i' \left(I_n - \hat{\rho}_g W_1 - \hat{\rho}_b W_2 - \hat{\rho}_l W_3\right) \frac{1}{T} \sum_{t=1}^{T} y_t y_t' \left(I_n - \hat{\rho}_g W_1' - \hat{\rho}_b W_2' - \hat{\rho}_l W_3'\right) e_i \\ & - e_i' \left(I_n - \rho_g W_1 - \rho_b W_2 - \rho_l W_3\right) \frac{1}{T} \sum_{t=1}^{T} y_t y_t' \left(I_n - \rho_g W_1' - \rho_b W_2' - \rho_l W_3'\right) e_i. \end{aligned}$$

The theorem then follows from the consistency of $\hat{\rho}_{GMM}$ and the ergodic theorem yielding $\frac{1}{T} \sum_{t=1}^{T} y_t y_t' \to_p \mathsf{E}(y_1 y_1')$. □

Table 1: Partitioning of Euro Stoxx 50 members into branches and countries in model 1; groups „Benelux“ and „others“ are merged to the new group „small countries“ in model 2

Finance	Aegon, Allianz, AXA, Banco Bilbao, Banco Santander, BNP, Crédit Agricole, Deutsche Bank, Deutsche Börse, Generali, ING, Intesa, Münchener Rück, Société Générale, Unicredit
Automobil	Daimler, VW
Energy	Alstom, E.ON, ENEL, ENI, Iberdrola, Repsol, RWE, SUEZ, Total
Telecom and Media	Deutsche Telekom, France Telecom, Telecom Italia, Telefonica, Vivendi
Pharma and Chemicals	Air Liquide, BASF, Bayer, Sanofi
Consumer Electronics	Nokia, Philips, SAP, Siemens, Schneider
Consumer retail	Anheuser Busch, Carrefour, Danone, L'Oreal, LVMH, Unilever
Basic Industry	Arcelor Mittal, CRH, Saint Gobain, Vinci
Benelux	Aegon, Anheuser Busch, Arcelor, ING, Philips, Unilever
France	Air Liquide, Alstom, AXA, BNP, Carrefour, Crédit Agricole, France Telecom, Danone, L'Oreal, LVMH, Saint Gobain, Sanofi, Schneider, Société Générale, SUEZ, Total, Vinci, Vivendi
Germany	Allianz, BASF, Bayer, Daimler, Deutsche Bank, Deutsche Börse, Deutsche Telekom, E.ON, Münchner Rück, RWE, SAP, Siemens, VW
Italy	Generali, ENEL, ENI, Intesa, Telecom Italia, Unicredit
Spain	Banco Bilbao, Banco Santander, Iberdrola, Repsol, Telefonica
Others	CRH, Nokia

Figure 1: Three kinds of spatial dependence with pointwise confidence bounds, estimated in rolling windows of 250 trading days, model 1 (above) and model 2 (bottom)



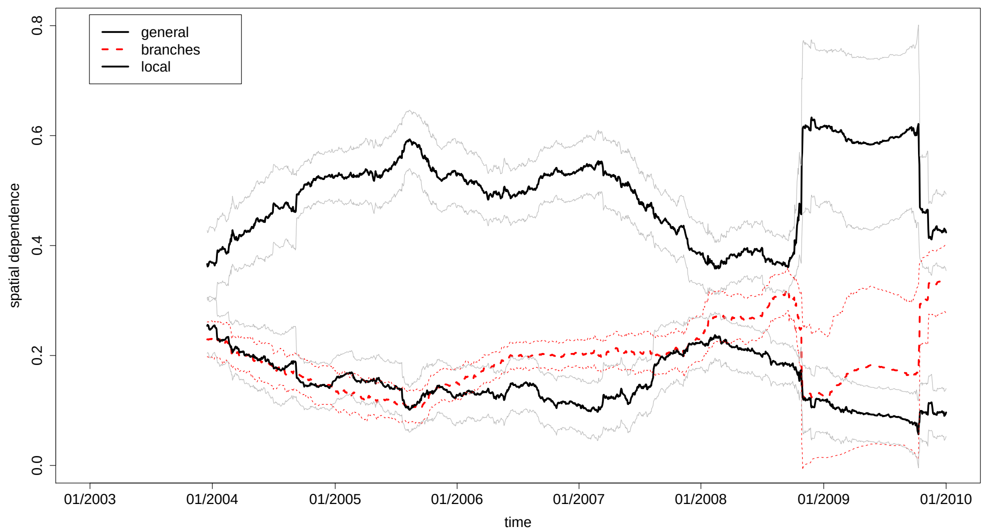


Figure 2: Estimated VaR for different models