Noname manuscript No.
(will be inserted by the editor)

Spatial dependence in stock returns - Local normalization and VaR forecasts

Thilo A. Schmitt · Rudi Schäfer · Dominik Wied · Thomas Guhr

Received: date / Accepted: date

Abstract We analyze a recently proposed spatial autoregressive model for stock returns and compare it to a one-factor model and the sample covariance matrix. The influence of refinements to these covariance estimation methods is studied. We employ power mapping and the shrinkage estimator as noise reduction techniques for the correlations. Further, we address the empirically observed time-varying trends and volatilities of stock returns. Local normalization strips the time series of changing trends and fluctuating volatilities. As an alternative method, we consider a GARCH fit. In the context of portfolio optimization, we find that the spatial model and the shrinkage estimator have the best match between the estimated and realized risk measures.

Keywords GARCH · one-factor model · power mapping · spatial autoregressive model

PACS G17 · C33

Thilo A. Schmitt
Fakultät für Physik, Universität Duisburg-Essen, 47048 Duisburg, Germany
Tel.: + 49-203-3794729
E-mail: thilo.schmitt@uni-due.de

Rudi Schäfer
Fakultät für Physik, Universität Duisburg-Essen, 47048 Duisburg, Germany

Dominik Wied
Fakultät Statistik, TU Dortmund, 44221 Dortmund, Germany

Thomas Guhr
Fakultät für Physik, Universität Duisburg-Essen, 47048 Duisburg, Germany

1 Introduction

The covariance matrix of stock returns plays a crucial role for portfolio optimization, see Markowitz (1952). Finding the portfolio weights that yield minimum risk given a desired portfolio return requires the best possible estimation of the future covariance matrix. A natural approach is to estimate the sample covariance matrix using historical data, assuming that the recent past is a good predictor for the future. The length of the historical time series which can be used for covariance estimation is often rather short. In particular in emerging markets, the total length of the available time series may be the limiting factor. Another consideration is the non-stationarity of the financial markets, see, e.g., Longin and Solnik (1995), Bekaert and Harvey (1995) and Münnix et al (2012): The correlation structure changes with time. Hence, to achieve a decent estimate of the current or future covariance matrix, we should only take into account rather recent data. However, since the sample covariance matrix for n assets requires $n(n+1)/2$ parameters to be estimated, the finiteness of the time series leads to a considerable amount of measurement noise, see Laloux et al (1999), Bouchaud and Potters (2009), Plerou et al (1999) and Plerou et al (2002). As pointed out by, e.g., Pafka and Kondor (2002, 2003), this has dire consequences for portfolio optimization, but can be mitigated to a large extent by noise reduction techniques. Here we will concentrate on one such technique, the power mapping, which has been introduced in Guhr and Kälber (2003) and further studied in Schäfer et al (2010). We also take the shrinkage estimator of Ledoit and Wolf (2003) as a reference into account.

An alternative approach to reducing the noise in sample covariance matrices is to consider a model for the correlation or covariance matrix which entails fewer parameters. Many models have been proposed to reduce the number of parameters which have to be estimated, see Pantaleo et al (2011). Here we consider a simple one-factor model, see Sharpe (1963), where $2n+1$ parameters have to be estimated. In addition, we study the spatial autoregressive model for stock returns, which was recently introduced by Arnold et al (2013). One particular feature of the spatial model is its ability to produce reliable Value-at-Risk (VaR) forecasts. This is partly due to the fact that the model captures a lot of dependence with a small number of parameters. It involves only $n+3$ parameters, 3 for the dependence and n parameters describing the individual volatilities.

A specific issue we want to address in this paper concerns the empirically observed time-varying trends and volatilities of financial time series. We are going to study the influence of sudden changes in local trends and volatilities on the covariance estimation methods described above. In particular, we investigate how the estimation of the spatial parameters is affected. We suggest the following refinements to substantially improve covariance estimation methods: The well known GARCH(1,1) model (see Bollerslev (1986) and Bollerslev et al (1988)) can be utilized to remove the fluctuating volatilities in the return time series and to predict future volatilities. This approach is compared to a local normalization method, recently introduced by Schäfer and Guhr (2010), and a short-term historical prediction of the individual volatilities. Moreover, we apply the above mentioned power mapping to reduce the noise in the correlation matrices. We study the influence of these refinements on each of the covariance estimation methods and the implications for portfolio optimization. To this end we consider the stocks of the Euro Stoxx 50 and use the out-of-sample realized portfolio variances as a risk measure. We compare the results to the predicted portfolio variances. We analyze the VaR forecast quality in more detail by comparing it to predicted and realized variances. Comparing different models with respect to their VaR forecast ability is a quite common approach in the literature, see e.g., Santos et al (2013). While it were basically possible to include still other models like a multivariate DCC model, see Engle (2002), in the study or to account for possible structural breaks in the model parameters (in the spirit of Wied (2013)), we have decided to focus on the present models in order to keep the presentation clear. In fact, the current analysis is in our opinion sufficient for the main results: It is extremely important for covariance estimation to take into account time-varying trends and volatilities of financial time series and the measurement noise. And while using GARCH residuals and volatility forecasts yields comparable results, the combination of local normalization and short term historical volatilities requires much shorter time series.

The paper is structured as follows: Section 2 presents the above mentioned methods for estimating the covariance matrix of stock returns. In Section 3 we discuss the refinements to these methods, which aim at removing changes in local trends and volatilities, improving the quality of volatility predictions, and reducing estimation noise. Section 4 details the portfolio optimization technique and the data set under consideration. In Section 5 we discuss the results of the covariance estimation and VaR forecasts, and we summarize our findings in Section 6.

2 Covariance estimation

The covariance matrix is a crucial input parameter for many risk assessing methods in finance, such as portfolio optimization. The estimation of the covariance matrix is a non-trivial task due to the time-varying trends and volatilities of financial time series. The natural way is to calculate the sample covariance matrix from the time series. As usual, we use the definition $\widehat{\mathrm{Cov}}(y_t) = \frac{1}{T-1}\sum_{t=1}^{T}(y_t - \bar{y})(y_t - \bar{y})'$, where $\bar{y}$ is the mean of the time series. The sample covariance matrix requires the estimation of $n(n+1)/2$ parameters. If the covariance matrix is calculated on a short time horizon it contains a great degree of noise. For larger time horizons the predictive power of the covariance matrix decreases as the market constantly changes. In Section 2.1 and 2.2 we discuss two models that require less parameters to be estimated to determine the covariance matrix.

2.1 One-factor model

The simple one-factor model radically reduces the amount of parameters to estimate by assuming that the change of all assets is tied to one factor, e.g., a market index. It was first used to improve portfolio optimization by Sharpe (1963) and decreases the amount of parameters to estimate to $2n+1$. A comprehensive description of the model is given by Jorion (2007), p. 192 ff. The one-factor model assumes that the stock returns can be described by

$$y_t = \alpha + \beta y_{m,t} + \eta_t \quad , \tag{1}$$

where the n-dimensional vector y_t contains the returns for all stocks $1 \ldots n$ at time t. The scalar $y_{m,t}$ describes the market return, e.g., is calculated from a stock market index. The vector $\beta = (\beta_1, \ldots, \beta_n)$ contains a constant for each stock which must be estimated, for example with a linear fit (ordinary least squares-estimator) separately for all stocks. The fixed intercepts' vector α can be neglected in the context of risk estimation as it contains no randomness. Assuming that the error terms in η_t are uncorrelated to each other the covariance matrix of y_t is given by

$$\mathrm{Cov}(y_t) = \beta\beta'\sigma_m^2 + D_\eta \tag{2}$$

with a matrix $D_\eta = \mathrm{diag}(\sigma_{\eta 1}^2, \ldots, \sigma_{\eta n}^2)$ that contains the variances of η_t on its diagonal and which can be estimated by standard ordinary least squares methods. The one-factor model requires an estimation of n entries for the β vector and the n diagonal elements of D_η, plus one for the market volatility σ_m.

Given the parameter estimates, we can directly obtain a parametric estimate for $\mathrm{Cov}(y_t)$.

2.2 Spatial dependence model

The spatial dependence model introduced by Arnold et al (2013) is based on the assumption that a lot of the cross-sectional dependence between the stock returns can be captured by three different types of dependence: A general dependence, dependence within industrial branches and dependence based on geographic locations. For an overview of spatial dependence modeling see Anselin (1988), Cressie (1993) and LeSage and Pace (2009). Formally, we have the spatial autoregressive model

$$y_t = \rho_g W_g y_t + \rho_b W_b y_t + \rho_l W_l y_t + \varepsilon_t \quad , \quad t = 1, \ldots, T \tag{3}$$

where ρ_g is a scalar parameter measuring the general dependence, ρ_b is a scalar parameter measuring the dependence between industrial branches and ρ_l is a scalar parameter measuring the dependence based on geographic locations. W_g, W_b and W_l are spatial weighting matrices. The stochastic component in this model stems from the error vector ε_t. Given the basic model assumption it is plausible to assume that its covariance matrix has uncorrelated entries although heteroscedasticity is allowed.

The non-diagonal elements of the matrix W_g are set to the normalized market capitalization of the corresponding assets. For the matrices W_b and W_l the element in the i-th row and j-th column is non-zero if the i-th and j-th asset are in the same branch (W_b) or country (W_l). The non-zero elements are set to the normalized market capitalization of the asset in each row.

If the three parameters are known, the covariance matrix of the vector y_t is given by

$$\mathrm{Cov}(y_t) = (I_n - \rho_g W_g - \rho_b W_b - \rho_l W_l)^{-1} \Sigma (I_n - \rho_g W_g' - \rho_b W_b' - \rho_l W_l')^{-1} \quad , \tag{4}$$

where Σ is the covariance matrix of the error term ε_t and I_n is an $n \times n$ identity matrix. The error terms are assumed to be uncorrelated, so all off-diagonal elements of Σ are zero. This leads to n additional parameters. The model uses $n + 3$ parameters which are best estimated by a two-step procedure that is based on the generalized methods of moments (GMM) approach, see Arnold et al (2013) and in addition Lee and Liu (2009) and Lin and Lee (2010). Again, given the parameter estimates, we can directly obtain a parametric estimate for $\mathrm{Cov}(y_t)$.

3 Refined methods of covariance estimation

We discuss four approaches to enhance the predictive capabilities of the methods discussed in Section 2. The GARCH residuals (Section 3.1) and local normalization (Section 3.2) reduce the empirically observed time-varying trends and volatilities of return time series with regard to the volatility. The power mapping method discussed in Section 3.3 is aimed at decreasing the noise in a correlation matrix. In Section 3.4 we explore additional methods to estimate the volatilities of the individual stocks.

3.1 GARCH residuals

The return time series has a fluctuating volatility which can lead to estimation errors in parameters derived from the time series. To improve the estimation it is desirable to remove these fluctuations from the return time series. This is possible by modeling the returns with a GARCH process introduced by Bollerslev (1986) as a generalization of the ARCH process invented by Engle (1982). We fit the GARCH(1,1)-model

$$X_t = \sigma_t \varepsilon_t \tag{5}$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \tag{6}$$

to the historical data to estimate the parameters α_0, α_1 and β_1. Here, $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a strong white noise process with $\operatorname{var} \varepsilon_t = 1$ and $E[\varepsilon_t] = 0$. The conditional variances σ_t^2 can replicate the fluctuating volatilities in empirical time series. Hansen and Lunde (2005) have shown that in most cases a GARCH(1,1) is sufficient to capture the return time series. Then we use the GARCH residuals

$$\varepsilon_t = \frac{X_t}{\sigma_t} \tag{7}$$

to receive a return time series, where the volatility fluctuations are removed to the degree the return time series fits the GARCH process.

For the comparison in Section 5, we use a rolling window of $T = 100$ trading days to estimate model parameters. This window is too small for the GARCH fit to converge. Therefore we use a rolling window of $T_{\mathrm{GARCH}} = 1000$ trading days to estimate the GARCH parameters. We emphasize that this larger window is only used to estimate the GARCH parameters.

3.2 Addressing local trends and changes in volatility: local normalization

Estimating the GARCH parameters requires a rather large time window. Therefore we use a second method called local normalization introduced by Schäfer and Guhr (2010). It removes local trends and changes in volatility without altering the cross-correlations between time series. The local average of a function is defined as

$$\langle f_t \rangle_m = \frac{1}{m} \sum_{j=0}^{m-1} f_{t-j\Delta t} \quad , \tag{8}$$

where Δt is the return interval. Then the locally normalized returns are given by

$$\rho_{mt} = \frac{r_t - \langle r_t \rangle_m}{\sqrt{\langle r_t^2 \rangle_m - \langle r_t \rangle_m^2}} \quad , \tag{9}$$

where we first subtract the local mean value $\langle r_t \rangle_m$ from the return r_t and then divide by the local volatility. As shown by Schäfer and Guhr (2010) a value of $m = 13$ yields optimal results for daily stock returns.

3.3 Noise-reduction

The correlation matrix of financial assets contains a significant amount of noise, which can be seen by comparing the eigenvalue density of a correlation matrix to a random matrix, see Laloux et al (1999). The part of small eigenvalues, called the bulk part, exhibits the same shape for both matrices. They only differ for larger eigenvalues, which can be associated to industrial branches. The natural method to reduce the noise would be to increase the length of the time series to calculate the correlation matrix from. This is not a feasible way to predict the future correlation matrix, because the relationships between companies constantly change, as they start competing on new markets or discontinue their activities in one field. Several methods have been proposed in the past to reduce the noise, while keeping the times series short, e.g., see Gopikrishnan et al (2001) and Giada and Marsili (2001).

Here we discuss the power mapping method introduced by Guhr and Kälber (2003) to reduce the noise in a correlation matrix. Every entry of the correlation matrix C is substituted by

$$C_{ij}^{(q)} = \text{sign}(C_{ij})\,|C_{ij}|^q \tag{10}$$

yielding the noise reduced correlation matrix $C^{(q)}$. Notice that the diagonal elements are equal to one and thus not affected by power mapping. In general, the optimal value for the parameter q depends on the time horizon T on which the correlation matrix is calculated, i.e., the degree of noise in the correlation matrix. However, power mapping is a very robust method which yields good results for a wide range of q values around the optimal one, as discussed in Schäfer et al (2010). Here, we use $q = 1.5$.

As a reference we also study the shrinkage method of Ledoit and Wolf (2003, 2004b). This method uses a single-index covariance matrix, i.e., a market covariance matrix, as a shrinkage target and provides a method to estimate the optimal shrinkage parameter from the historical return time series, see page 613 in Ledoit and Wolf (2003). For covariance matrices where the length of the time series is smaller than the dimension of the matrix an improved technique exists by Ledoit and Wolf (2004a), which is not necessary here. By its design the shrinkage target is estimated by a Sharpe-like one-factor model. In order to combine shrinkage estimator with other methods, for example, the spatial dependence model, it would be necessary to change the shrinkage target, yielding a completely new method. Therefore, we restrict ourselves to study the shrinkage method on its own and provide the results as a reference. However, we do also estimate the shrinkage parameter from the GARCH residuals and locally normalized returns instead of the original returns.

3.4 Volatility forecast

The correlation matrix needed for the power mapping method can be calculated from the covariance matrix by dividing each element of the covariance matrix

$$C_{ij} = \frac{\text{Cov}(y_t)_{ij}}{\sigma_i \sigma_j} \tag{11}$$

by the respective volatilities σ_i and σ_j. In case of the sample covariance matrix we use the standard deviations of the returns calculated on the rolling window of $T = 100$ trading days. For the one-factor and the spatial dependence model we use the model specific volatilities from the diagonal of the covariance matrix in Equation (2) and (4), respectively. If local normalization or GARCH residuals are used the volatilities are calculated from these time series.

At this stage we may apply power mapping to the correlation matrix if we choose to. Applying power mapping will change the correlation matrix C, otherwise the correlation matrix C remains the same between Equation (11) and (12).

Regardless of the fact whether power mapping is applied or not we need the covariance matrix for the portfolio optimization in Section 4.2. We proceed by calculating the covariance matrix from the correlation matrix

$$\mathrm{Cov}(y_t)_{ij} = C_{ij}\hat{\sigma}_i\hat{\sigma}_j \quad . \tag{12}$$

At this point we use the volatilities $\hat{\sigma}_{i,j}$ which are the historical volatilities in case of the sample covariance matrix or the volatilities of the corresponding model estimated from the original returns.

However, it is also possible to use other methods to estimate the volatilities. There is a plethora of possible methods or models to forecast volatility, see Poon and Granger (2003) for a review. Here, we limit ourselves to two common methods.

First, we calculate the standard deviation from the original returns, i.e., with no further methods applied to them, in a rolling window of $T_{\mathrm{vol}} = 14$ trading days. Here we assume that the volatility in the past three weeks is a better indicator for the future standard deviation compared to the longer horizon of $T = 100$ trading days.

Second, we can use the parameters from the GARCH fit of the original returns described in Section 3.1 to predict the volatilities for the next trading day according to Equation (6).

The reference results for the shrinkage estimator only use historical volatilities calculated from the original returns.

4 Application to portfolio optimization

4.1 The data set

We use the adjusted daily closing prices for a collection of 49 stocks contained in the Euro Stoxx 50. It includes companies from various countries in the eurozone and spans across different branches. We had to remove GDF Suez because of incomplete data due to the merger. The data is taken from Thomson Reuters Datastream. A complete list of the stocks including their industrial branch and country as used in the spatial dependence model is given in Table 1. Nokia and CRH from Finland and Ireland are put together in the country group “others” because groups are not allowed to contain only one entry to avoid singularities. The observation period ranges from January 2001 to May 2012. We calculate the logarithmic returns from the adjusted prices. Table 1 gives an overview of the used stocks.

Table 1 The data set

Automobile	BMW (Germany), Daimler (Germany), VW (Germany)
Basic industry	Arcelor Mittal (Benelux), CRH (Ireland), Saint-Gobain (France), Vinci (France)
Consumer electronics	Nokia (Finland), Philips (Benelux), SAP (Germany), Schneider (France), Siemens (Germany)
Consumer Retail	Anheuser Busch (Benelux), Carrefour (France), Danone (France), Inditex (Spain), L'Oreal (France), LVMH (France), Unilever (Benelux)
Energy	E.ON (Germany), ENEL (Italy), ENI (Italy), Iberdrola (Spain), RWE (Germany), Repsol (Spain), Total (France)
Finance	AXA (France), Allianz (Germany), BNP (France), Banco Bilbao (Spain), Banco Santander (Spain), Deutsche Bank (Germany), Deutsche Brse (Germany), Generali (Italy), ING (Benelux), Intesa (Italy), Mnchener Rck (Germany), Socit Gnrale (France), Unicredit (Italy), Unibail-rodamco (France)
Pharma and chemicals	Air Liquide (France), BASF (Germany), Bayer (Germany), Sanofi (France)
Telecom and media	Deutsche Telekom (Germany), France Telecom (France), Telecom Italia (Italy), Telefonica (Spain), Vivendi (France)

4.2 Portfolio optimization

We compare the effects of the methods discussed in Section 3 on the covariance estimation techniques of Section 2. For each covariance matrix, we perform a portfolio optimization to determine the minimum variance portfolio, see Markowitz (1952) and also Markowitz (1959) and Elton et al (2006).

We estimate each covariance matrix $\mathrm{Cov}(y_t) =: V$ on a rolling window of 100 trading days. The covariance matrix yields the portfolio weights

$$\omega = \frac{\hat{V}^{-1}\tau}{\tau'\hat{V}^{-1}\tau} \tag{13}$$

for the minimum variance portfolio, where τ is a vector composed of ones. The predicted portfolio variance is then

$$\hat{\sigma}^2_{\text{port}} := \left(\tau'\hat{V}^{-1}\tau\right)^{-1} . \tag{14}$$

The estimated portfolio variance is then used to calculate a Gaussian Value-at-Risk (VaR) for a given α-quantile u_α

$$\widehat{\mathrm{VaR}}_\alpha = u_\alpha \sqrt{\hat{\sigma}^2_{\text{port}}} . \tag{15}$$

In this setup it is possible to calculate the VaR on a daily basis and compare it to the out-of-sample realized portfolio returns.

In addition, we discuss two portfolio metrics. First, the Sharpe ratio, see Sharpe (1994),

$$S_p = \frac{\frac{1}{T}\sum_{t=1}^{T} \hat{r}_{a,t} - \hat{r}_{b,t}}{\sqrt{\mathrm{var}(\hat{r}_{a,t} - \hat{r}_{b,t})}} \quad , \tag{16}$$

where $\hat{r}_{a,t}$ is the portfolio return and $\hat{r}_{b,t}$ is the portfolio return of a portfolio with homogeneous weights at time t. Second, we calculate the portfolio turnover

$$T_p = \frac{1}{K} \sum_{k=1}^{K} \frac{1}{T-1} \sum_{t=1}^{T-1} |\omega_{k,t+1} - \omega_{k,t}| \quad , \tag{17}$$

which is the average of changes for the portfolio weights after each restructuring of the portfolio.

5 Results

5.1 Spatial parameters

Figure 1 shows the influence of local normalization on the parameter estimation for the spatial dependence model. The three parameters of the spatial dependence model are calculated for a rolling time window of $T = 250$ days. The dashed lines show the parameters calculated from the original returns. The solid lines present the three parameters with local normalization applied to the returns. Especially during the financial crisis of 2008 a strong jump is noticeable in the general and branch spatial parameters. This coincides with the peaking volatility during this turbulent time. The artifact which has the same width as the rolling window vanishes when applying the local normalization. We note that the use of GARCH residuals yields similar results.

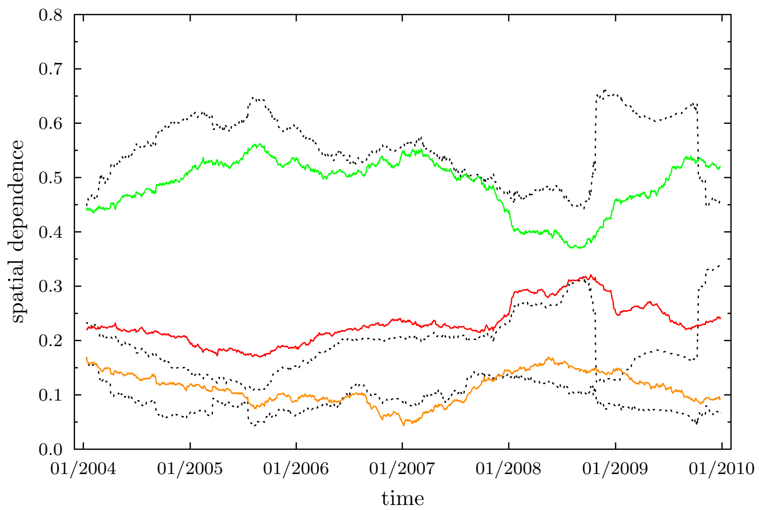

Fig. 1 The parameters ρ_g, ρ_b and ρ_l are shown from top to bottom estimated at each trading day for an interval of 250 days. The solid lines are with local normalization applied to the returns, while the dotted lines are estimated from the unaltered returns.

5.2 Portfolio variances

First, we discuss the impact of each method presented in Section 3 on the realized portfolio variances. Then, we examine how combinations of the methods affect the risk assessment. The realized portfolio variances are given in Table 2. The table is structured as follows: The second column states which returns were used. We can use the original returns, the GARCH residuals or the locally normalized returns. The volatility forecast method is specified in the third column. The fourth column indicates whether or not power mapping was used to suppress the estimation noise. The last three columns show the results of the realized portfolio variances for the spatial dependence model, the one-factor model and the sample covariance matrix. Variances smaller than 0.0009 are underlined.

Table 2 Realized portfolio variances

	returns	volatility forecast	noise reduction	realized portfolio variances sdep	1-factor	sample
1	original	hist	no	0.000241	0.000093	0.000121
2	GARCH	hist	no	0.000171	0.000171	0.000115
3	normalized	hist	no	0.000172	0.000148	0.000127
4	original	hist	power mapping	0.000256	0.000087	0.000086
5	GARCH	hist	power mapping	0.000133	0.000207	0.000086
6	normalized	hist	power mapping	0.000135	0.000190	0.000093
7	original	GARCH	no	0.000100	0.000087	0.000126
8	GARCH	GARCH	no	0.000087	0.000142	0.000119
9	normalized	GARCH	no	0.000087	0.000121	0.000132
10	original	GARCH	power mapping	0.000097	0.000086	0.000098
11	GARCH	GARCH	power mapping	0.000084	0.000160	0.000095
12	normalized	GARCH	power mapping	0.000084	0.000143	0.000101
13	original	hist†	no	0.000101	0.000086	0.000116
14	GARCH	hist†	no	0.000091	0.000138	0.000119
15	normalized	hist†	no	0.000088	0.000119	0.000120
16	original	hist†	power mapping	0.000099	0.000084	0.000092
17	GARCH	hist†	power mapping	0.000085	0.000156	0.000093
18	normalized	hist†	power mapping	0.000086	0.000139	0.000097
19	original	hist	shrinkage	-	-	0.000076
20	GARCH	hist	shrinkage	-	-	0.000077
21	normalized	hist	shrinkage	-	-	0.000077

Using the GARCH residuals (row 2) or local normalization (row 3) yields similar effects for all models. The spatial dependence model benefits from both methods in a similar way, while the portfolio variances for the one-factor model get worse. There is no significant effect on the sample covariance matrix. Power mapping (row 4) slightly increases the portfolio variance in case of the spatial dependence model.

The one-factor model improves a little bit, but is very good from start. The strength of power mapping unveils when applied to the sample covariance matrix. Here power mapping greatly reduces the realized portfolio variance. However, it does not match the lowest observed portfolio variances obtained by the shrinkage estimator (row 19-21). The GARCH predicted volatilities (row 7) enormously decrease the portfolio variance for the spatial dependence model. They have a minor positive effect on the one-factor model, while a minuscule negative effect on the sample covariance matrix.

By combining different methods it is possible to further reduce the portfolio variance. The GARCH predicted volatilities with either GARCH residuals or local normalization (row 8,12 and 9,11) improve the spatial dependence model to be on par with the sample covariance matrix. The one-factor model does not benefit and the sample covariance matrix only gets better if power mapping (row 11,12) or the shrinkage estimator (row 19-21) is used. The effect of power mapping is marginal in case of the spatial dependence model.

Estimating the shrinkage parameter from locally normalized returns (row 20), or alternatively GARCH residuals (row 21), does not significantly affect the realized portfolio variances. The same behavior is observable for power mapping, where local normalization and GARCH residuals only have a minor impact.

Again, we notice that the one-factor model works best if used together with the original returns (row 10), while the spatial dependence model benefits from the GARCH volatility forecast. Local normalization or the GARCH residuals improve the spatial dependence model (row 5,6), but not to the same extent as a better volatility forecast (row 7-12 and 13-18).

If we shorten the time interval on which the volatility is estimated (denoted by a dagger in the table) we can observe comparable results to the GARCH predicted volatilities (row 13-18 and 7-12). In particular, local normalization in combination with the shorter historical volatilities achieve matching results compared to GARCH residuals with GARCH predicted volatilities and requires far shorter time series (row 12 and 18).

With regard to the realized portfolio variances we can conclude that the spatial dependence model works best in combination with methods that improve the volatility estimation, like GARCH predicted volatilities or local normalization with a shorter horizon for the volatility calculation. The one-factor works best without any refinements with the exception of minor improvements in combination with power mapping. For the sample covariance matrix only power mapping or the shrinkage estimator is required to achieve the best performance.

For comparing the predicted variances and the Sharpe ratios of the different return series resulting from the different optimizations and a benchmark series, respectively, we performed a simple bootstrap-based test in the spirit of Ledoit and Wolf (2008). In the first step, for each series, the difference of the quantity of interest (predicted variance or Sharpe ratio) between the series and the benchmark is calculated. This difference yields the nominator of the test statistic. For the denominator, we first obtain $B = 1000$ new bivariate time series of length T, respectively, which are obtained by a non-overlapping block bootstrap similarly to Wied (2015+). This means that we draw with replacement blocks of bivariate vectors of the original time series. The block length is $T^{1/3}$. For each of the B series, we calculate the quantity of interest. The denominator of the test statistic is then given by the empirical standard deviation of the B quantities. Similarly to Wied (2015+), we make use of the asymptotic normality of this test statistic. Thus, the approximate p-value of the one-sided tests is given by $F(T)$ for the variance and $1 - F(-T)$ for the Sharpe ratio test, where T is the test statistic and F is the distribution function of the standard normal distribution.

We utilize two different scenarios as a benchmark in case of the variances. First, we compare the variances to the variances of a portfolio with homogeneous weights. We find that all models and methods beat the homogeneous portfolio with one exception. The spatial dependence model with power mapping applied does not yield significantly lower portfolio variances compared to the homogeneous portfolio. Second, with the Shrinkage estimator as a benchmark none of the improved models provide lower portfolio variances.

Table 3 Relative predicted portfolio variances in percent

		volatility	noise	$\frac{\text{realized}-\text{predicted}}{\text{predicted}}$ in %		
	returns	forecast	reduction	sdep	1-factor	sample
1	original	hist	no	57	203	426
2	GARCH	hist	no	64	45	362
3	normalized	hist	no	73	53	397
4	original	hist	power mapping	80	195	180
5	GARCH	hist	power mapping	32	206	174
6	normalized	hist	power mapping	39	234	184
7	original	GARCH	no	137	256	581
8	GARCH	GARCH	no	141	847	515
9	normalized	GARCH	no	144	452	564
10	original	GARCH	power mapping	108	260	283
11	GARCH	GARCH	power mapping	119	1330	267
12	normalized	GARCH	power mapping	121	755	276
13	original	hist†	no	323	401	903
14	GARCH	hist†	no	297	953	877
15	normalized	hist†	no	233	480	727
16	original	hist†	power mapping	186	319	353
17	normalized	hist†	power mapping	187	765	359
18	GARCH	hist†	power mapping	183	1345	350
19	original	hist	shrinkage	-	-	111
20	GARCH	hist	shrinkage	-	-	114
21	normalized	hist	shrinkage	-	-	114

We compare the predicted variances with the realized portfolio variances shown in Table 3. Values smaller than 150 are underlined. The prediction error of the sample covariance matrix is greatly reduced by the use of power mapping or the shrinkage estimator. Local normalization yields a small enhancement for the spatial dependence model, while there is no improvement for the one-factor model and the sample covariance matrix if used in combination with power mapping. The one-factor model and the sample covariance matrix are not very well suited to predict the realized variance, while maintaining competitive realized portfolio variances. The spatial dependence model and the shrinkage estimator are the best predictors for the realized portfolio variances.

5.3 VaR forecast

We calculate the VaR forecast according to Equation (15) on a daily basis for each α. The probability that the realized portfolio return is smaller than the VaR forecast is shown in Figure 2 for $\alpha \in (0, 0.5]$. The probability is calculated from all trading days in the observation period. For a perfect model the probability $P(y_{\text{port}} < \widehat{\text{VaR}}_\alpha)$ should be equal to α, which is indicated by a straight line.

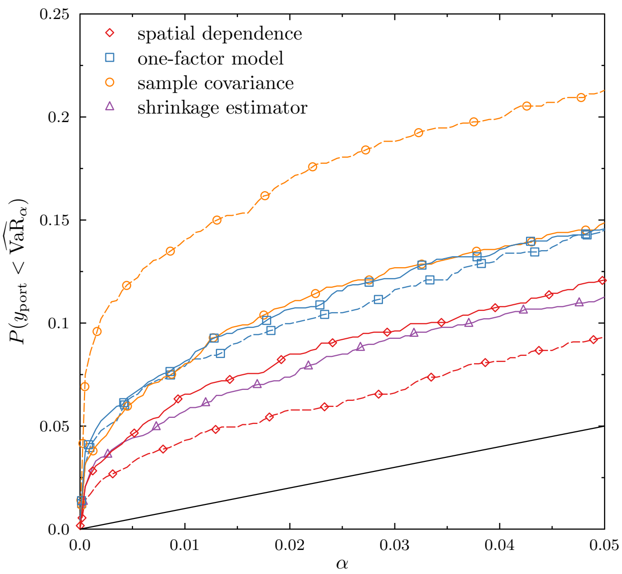

Fig. 2 The probability of a portfolio return being smaller than the Value-at-Risk given a fixed quantile of α. The dashed lines show the results without any refinements for the spatial dependence model (diamond), the one-factor model (square) and the sample covariance matrix (circle). The solid lines show the effect of improved covariance estimation methods (see text for details). The shrinkage estimator (triangle) is a noise-reduced refinement of the sample covariance matrix (circle).

We do not show the results for each case presented in Table 2; instead we limit ourselves to one case per model, where the realized portfolio variances are lowest. For the spatial dependence model (diamond) we present the VaR forecast with GARCH residuals, GARCH-predicted volatilities and applied power mapping. The one-factor model (squares) uses a combination of original returns, historical volatilities and power mapping, while the sample covariance matrix (circle) uses GARCH residuals instead, historical volatilities and power mapping. In addition, we show the sample covariance matrix with the shrinkage estimator applied, where the shrinkage parameter is estimated from the original returns. The dashed lines show the original models without any refinements while the solid lines show the refined cases.

Without any refinements the spatial dependence model produces the best VaR forecasts in the observed period, while the results for the one-factor model and the sample covariance matrix are rather poor in comparison. The large improvement with regard to the realized portfolio variances leads to a poorer VaR forecast. In contrast the sample covariance matrix not only gains better realized variances by applying power mapping and using GARCH residuals but also leads to a better VaR forecast. For the one-factor model there is no significant change to the predictive power of the VaR forecast. The shrinkage estimator provides a serious improvement for the sample covariance matrix with regard to the VaR prediction, which is slightly better compared to the spatial model with applied refinements.

Nonetheless, the sample covariance matrix with power mapping and the one-factor model do not surpass the spatial dependence model and the shrinkage estimator in their risk estimation.

5.4 Portfolio turnover and Sharpe ratios

Table 4 shows the portfolio turnover, see Equation (17), for all cases. The unmodified spatial dependence and one-factor model have a very small turnover compared to the sample covariance matrix (row 1). However, only the one-factor model shows a competitive variance for the realized portfolio returns. Improving the portfolio variance in case of the spatial dependence model results at best in a three times increase of the portfolio turnover (row 11,12). The biggest impact stems from the alternative volatility forecast methods. While they improve the portfolio variance for the spatial dependence model they have a severe effect on the portfolio turnover. The one-factor model with power mapping applied results in the smallest observed portfolio turnover, while also providing a competitive portfolio variance (row 4). The shrinkage method doubles the portfolio turnover and provides a slightly better portfolio variance (row 19). In general, power mapping has a positive effect on the portfolio turnover, while the improved volatility forecasts result in an elevated portfolio return. The shrinkage yields a one third smaller portfolio turnover compared to power mapping.

The Sharpe ratios with a homogeneously weighted portfolio as reference are presented in Table 5. The original returns yield poor results for the Sharpe ratio in case of the spatial dependence model, where even negative ratios occur (row 1, 4, 10). The one-factor model has a high Sharpe ratio out of the box (row 1) and yields the highest Sharpe ratio if combined with the GARCH volatility forecast (row 7). The shrinkage method provides a higher Sharpe ratio com-

Table 4 Portfolio turnover

	returns	volatility forecast	noise reduction	portfolio turnover sdep	1-factor	sample
1	original	hist	no	0.14	0.11	0.69
2	GARCH	hist	no	0.10	0.50	0.67
3	normalized	hist	no	0.11	0.62	0.71
4	original	hist	power mapping	0.05	0.09	0.29
5	GARCH	hist	power mapping	0.07	0.35	0.27
6	normalized	hist	power mapping	0.08	0.42	0.29
7	original	GARCH	no	0.83	0.43	1.27
8	GARCH	GARCH	no	0.59	0.28	1.22
9	normalized	GARCH	no	0.58	0.50	1.34
10	original	GARCH	power mapping	0.68	0.32	0.66
11	GARCH	GARCH	power mapping	0.46	0.19	0.62
12	normalized	GARCH	power mapping	0.45	0.35	0.67
13	original	hist†	no	0.93	0.59	1.35
14	GARCH	hist†	no	0.77	0.34	1.31
15	normalized	hist†	no	0.75	0.53	1.39
16	original	hist†	power mapping	0.83	0.48	0.83
17	GARCH	hist†	power mapping	0.66	0.25	0.79
18	normalized	hist†	power mapping	0.64	0.39	0.83
19	original	hist	shrinkage	-	-	0.20
20	GARCH	hist	shrinkage	-	-	0.19
21	normalized	hist	shrinkage	-	-	0.19

Table 5 Sharpe ratio

	returns	volatility forecast	noise reduction	sharpe ratio sdep	1-factor	sample
1	original	hist	no	-0.0081	0.0118	-0.0001
2	GARCH	hist	no	0.0058	-0.0037	-0.0085
3	normalized	hist	no	0.0090	0.0080	-0.0034
4	original	hist	power mapping	-0.0219	0.0114	0.0049
5	GARCH	hist	power mapping	0.0040	-0.0079	-0.0012
6	normalized	hist	power mapping	0.0077	0.0111	0.0016
7	original	GARCH	no	0.0029	0.0140	-0.0021
8	GARCH	GARCH	no	0.0097	0.0009	-0.0063
9	normalized	GARCH	no	0.0115	0.0012	0.0004
10	original	GARCH	power mapping	-0.0015	0.0129	0.0039
11	GARCH	GARCH	power mapping	0.0070	0.0022	0.0018
12	normalized	GARCH	power mapping	0.0092	0.0022	0.0044
13	original	hist†	no	0.0014	0.0071	0.0050
14	GARCH	hist†	no	0.0081	0.0057	-0.0052
15	normalized	hist†	no	0.0095	0.0066	0.0016
16	original	hist†	power mapping	0.0005	0.0091	0.0079
17	GARCH	hist†	power mapping	0.0075	0.0074	0.0023
18	normalized	hist†	power mapping	0.0089	0.0085	0.0052
19	original	hist	shrinkage	-	-	0.0075
20	GARCH	hist	shrinkage	-	-	0.0093
21	normalized	hist	shrinkage	-	-	0.0089

pared to power mapping (row 4, 19). In most cases power mapping lowers the Sharpe ratio if used with the spatial dependence or one-factor model. In case of the sample covariance matrix power mapping has in each case a positive effect on the Sharpe ratio. The shrinkage method yields a better Sharpe ratio if the shrinkage parameter is estimated from the locally normalized returns or GARCH residuals (row 20, 21). On the other hand local normalization and GARCH residuals in combination with the sample covariance matrix or the one-factor model yield lower results for the Sharpe ratio. Without noise reduction applied the sample covariance matrix has rather poor results for the Sharpe ratio.

We conduct the same bootstrap approach as discussed in section 5.2. However, the homogeneous portfolio is not an option here, because it is already used as the reference in the calculation of the Sharpe ratio. Therefore, we only use the Shrinkage estimator as a benchmark and find that the three models do not yield significantly higher Sharpe ratios. This is even true for the Shrinkage estimator calculated from GARCH residuals or locally normalized returns (row 20, 21).

6 Conclusion

We compare three approaches for covariance estimation: The spatial dependence model, a one-factor model and the sample covariance matrix. As a benchmark for the quality of the covariance estimation, we use portfolio optimization. The realized portfolio variances and the relative prediction error are scrutinized. In addition to the original approaches, we investigate several refinement methods. An estimation error can arise from fluctuating volatilities; they can be removed from the return time series either by employing a GARCH fit and using the residuals, or by using a local normalization method. Volatilities of the individual return time series can be better predicted using a short-term historical estimate. This is due to the slowly decaying autocorrelation of empirical volatilities. Alternatively, we can use the volatility predictions of the GARCH fits. There is a large statistical estimation error, if the length of the time series is not much larger than the parameters to be estimated. This measurement noise can be reduced by noise reduction techniques such as power mapping.

Given that a noise reduction of the covariance matrix has no big impact, we draw the conclusion that the spatial dependence model captures the correlation between assets well. However, the realized portfolio variances can be immensely reduced by combining the spatial dependence model with better methods for volatility forecasting, and using locally normalized returns for the regression. The cost for this is that the prediction errors increase (with the exception of locally normalized returns in combination with historical volatility forecast and power mapping). We conjecture that the large impact of different methods for volatility forecasting are due to the fact that the dependence structure is basically captured by weighting matrices whereas the only stochastic component is the error vector. The one-factor model produces quite good realized portfolio variances on its own: This may be attributed to the fact that the market factor is an additional stochastic component (one can interpret the term $\rho_g W_g y_t$ in the spatial model as a market factor as well, but in this case, there is no additional randomness). It works best with the original returns and should not be used with GARCH residuals or locally normalized returns. Slight improvements are possible with better volatility forecast methods or noise reduction. The sample covariance matrix suffers from noise due to the finite length of the time series. Noise reduction methods such as power mapping are sufficient to achieve results that are equally good or better in case of the shrinkage estimator compared to the other approaches.

Local normalization and the GARCH residuals effectively remove fluctuations in the volatility and reduce estimation artifacts for the spatial parameters. With the right choice of refinements all three approaches are capable of producing good realized portfolio variances, though the spatial dependence model and the shrinkage estimator provide the smallest prediction error for portfolio variances and VaR forecasts.

Acknowledgements This work was supported by Deutsche Forschungsgemeinschaft [SFB 823, project A1].

References

Anselin L (1988) Spatial Econometrics: Methods and Models. Studies in Operational Regional Science, Springer

Arnold M, Stahlberg S, Wied D (2013) Modeling different kinds of spatial dependence in stock returns. Empirical Economics 44(2):761–774

Bekaert G, Harvey C (1995) Time-varying world market integration. The Journal of Finance L(2):403–44

Bollerslev T (1986) Generalized Autoregressive Conditional Heteroskedasticity. Journal of Econometrics 31:307–327

Bollerslev T, Engle R, Wooldridge J (1988) A Capital Asset Pricing Model with Time-Varying Covariances. The Journal of Political Economy 96(1):116–131

Bouchaud JP, Potters M (2009) Theory of Financial Risk and Derivative Pricing: From Statistical Physics to Risk Management, 2nd edn. Cambridge University Press

Cressie N (1993) Statistics for spatial data. Wiley series in probability and mathematical statistics: Applied probability and statistics, J. Wiley

Elton EJ, Gruber MJ, Brown SJ, Goetzmann WN (2006) Modern Portfolio Theory and Investment Analysis, 7th edn. Wiley

Engle R (1982) Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. Econometrica: Journal of the Econometric Society 50(4):987–1007

Engle R (2002) Dynamic Conditional Correlation. J Bus Econ Stat 20(3):339–350

Giada L, Marsili M (2001) Data clustering and noise undressing of correlation matrices. Physical Review E 63(6):061,101

Gopikrishnan P, Rosenow B, Plerou V, Stanley H (2001) Quantifying and interpreting collective behavior in financial markets. Physical Review E 64(3):035,106

Guhr T, Kälber B (2003) A new method to estimate the noise in financial correlation matrices. Journal of Physics A: Mathematical and General 36(12):3009–3032

Hansen PR, Lunde A (2005) A forecast comparison of volatility models: does anything beat a GARCH(1,1)? Journal of Applied Econometrics 20(7):873–889

Jorion P (2007) Value at risk: the new benchmark for managing financial risk, 3rd edn. McGraw-Hill, New York

Laloux L, Cizeau P, Bouchaud JP, Potters M (1999) Noise Dressing of Financial Correlation Matrices. Physical Review Letters 83(7):1467–1470

Ledoit O, Wolf M (2003) Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. Journal of Empirical Finance 10(5):603–621, DOI 10.1016/S0927-5398(03)00007-0

Ledoit O, Wolf M (2004a) A well-conditioned estimator for large-dimensional covariance matrices. Journal of Multivariate Analysis 88(2):365–411, DOI 10.1016/S0047-259X(03)00096-4

Ledoit O, Wolf M (2004b) Honey, I shrunk the sample covariance matrix. The Journal of Portfolio Management pp 1–22

Ledoit O, Wolf M (2008) Robust performance hypothesis testing with the Sharpe ratio. Journal of Empirical Finance 15:850–859

Lee LF, Liu X (2009) Efficient GMM Estimation of High Order Spatial Autoregressive Models With Autoregressive Disturbances. Econometric Theory 26(01):187

LeSage J, Pace R (2009) Introduction to Spatial Econometrics. Statistics: a Series of Textbooks and Monographs, CRC PressINC

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

Longin FM, Solnik B (1995) Is the correlation in international equity returns constant: 19601990? Journal of International Money and Finance 14:3–26

Markowitz H (1952) Portfolio Selection. The Journal of Finance 7(1):77–91

Markowitz H (1959) Portfolio Selection: Efficient Diversification of Investment. Yale University Press

Münnix MC, Shimada T, Schäfer R, Leyvraz F, Seligman TH, Guhr T, Stanley HE (2012) Identifying states of a financial market. Sci Rep 2

Pafka S, Kondor I (2002) Noisy covariance matrices and portfolio optimization. Eur Phys J B 27:277–280

Pafka S, Kondor I (2003) Noisy covariance matrices and portfolio optimization II. Physica A 319:487–494

Pantaleo E, Tumminello M, Lillo F, Mantegna RN (2011) When do improved covariance matrix estimators enhance portfolio optimization? An empirical comparative study of nine estimators. Quantitative Finance 11(7):1067–1080

Plerou V, Gopikrishnan P, Rosenow B, Amaral L, Stanley H (1999) Universal and nonuniversal properties of cross correlations in financial time series. Phys Rev Lett 83:1471–1474

Plerou V, Gopikrishnan P, Rosenow B, Amaral L, Guhr T, Stanley H (2002) Random matrix approach to cross correlations in financial data. Phys Rev E 65:066,126

Poon S, Granger C (2003) Forecasting volatility in financial markets: A review. Journal of Economic Literature XLI(June):478–539

Santos A, Nogales F, Ruiz E (2013) Comparing Univariate and Multivariate Models to Forecast Portfolio Value-at-Risk. Journal of Financial Econometrics 11(2):400–441

Schäfer R, Guhr T (2010) Local normalization: Uncovering correlations in non-stationary financial time series. Physica A 389(18):3856–3865

Schäfer R, Nilsson NF, Guhr T (2010) Power mapping with dynamical adjustment for improved portfolio optimization. Quantitative Finance 10(1):107–119

Schäfer R, Nilsson NF, Guhr T (2010) Power mapping with dynamical adjustment for improved portfolio optimization. Quant Finance 10(1):107–119

Sharpe W (1963) A simplified model for portfolio analysis. Management science 9(2):277–293

Sharpe WF (1994) The Sharpe Ratio. The Journal of Portfolio Management

Wied D (2013) Cusum-type testing for changing parameters in a spatial autoregressive model for stock returns. Journal of Time Series Analysis 34(1):221–229

Wied D (2015+) A Nonparametric Test for a Constant Correlation Matrix. Econometric Reviews URL http://dx.doi.org/10.1080/07474938.2014.998152