CUSUM-type testing for changing parameters in a spatial autoregressive model for stock returns

Dominik Wied*

TU Dortmund

January 27, 2014

Abstract

The paper suggests a CUSUM-type test for time-varying parameters in a recently proposed spatial autoregressive model for stock returns and derives its asymptotic null distribution as well as local power properties. As can be seen from Euro Stoxx 50 returns, a combination of spatial modelling and change point tests might allow for superior risk forecasts in portfolio management.

Keywords: Brownian Bridge; Fluctuation test; GMM estimation; Spatial dependence; Stock returns

*Fakultät Statistik, CDI-Gebäude, TU Dortmund, 44221 Dortmund. Email: wied@statistik.tu-dortmund.de, Phone: +49 231 755 5419.

1. Introduction

Spatial modelling of stock returns has recently become an important field of research in the econometrics and statistics literature. Some recent approaches are Fernandez (2011), who uses similarities of financial indicators to define spatial linkages of stocks and estimates a spatial version of the capital asset pricing model, and Asgharian et al. (2011), who consider different linkages like economic and monetary integration between countries to explain the propagation of country specific shocks to other countries.

Arnold et al. (2011) propose a spatial autoregressive model which is partly an extension of previous models proposed by Badinger and Egger (2011). It contains both a time dimension and a cross-section dimension; in the latter it allows for distinguishing between general dependencies, dependencies inside branches and local dependencies. As can be seen in an out of sample study of Euro Stoxx 50 returns, this model can lead to forecasts for risk measures which are superior to standard approaches like a factor model or the sample covariance matrix.

In the out of sample study, the correlation parameters used for estimating the covariance matrix of the returns are simply estimated by a rolling window of 100 days. However, the question arises if the spatial correlation parameters can be assumed to be constant over time and which data of the past can be used to estimate the parameters. Arnold et al. (2011) find empirical evidence against this hypothesis, i.e. they identify increasing general dependence during the financial crisis in 2008.

The present paper suggests a method to formally answer the question. Considering the model from Arnold et al. (2011), we propose a new formal CUSUM-type statistical test for constancy of spatial dependence over time. It compares successive parameter estimates which are obtained by GMM estimation. There are comparable CUSUM tests in the literature, e.g. Brown et al. (1975) test for parameter constancy in linear regression models, Sowell (1996) proposes optimal tests for parameter instability in the GMM framework using test statistics that base on the GMM objective functions, Manly and Mackenzie (2000) and Manly and Mackenzie (2003) propose CUSUM methods for environmental monitoring, Lee et al. (2003) propose a general framework for parameter constancy tests in time series models and Wied et al. (2012) propose a non-parametric test for constant Pearson correlation. However, as far as the author knows, a special test for changing parameters in a spatial autoregressive model does not exist in the literature up to now. The new test does not assume a particular distribution of the random variables and does not assume potential break points to be known a priori, two properties which it shares with other CUSUM-tests. It has considerable power in small samples. Combining the backtesting study in Arnold et al. (2011) with the test procedure slightly improves the former results.

2. Model and testing procedure

For $t \in \mathbb{Z}$, let y_t be a n-dimensional random vector which will describe stock returns in the following. The observation period is $t = 1, \ldots, T$; for the asymptotic results, we keep n fixed and let T converge to ∞.
The spatial autoregressive model, which we consider in this paper, basically builds on the idea that the returns are spatially correlated in the cross-sectional dimension and that the spatial dependence can be separated reasonably into three different parts. The first type of dependence is a general one which affects all stocks in the same way. The second type captures dependencies between different industrial branches and the third type captures dependencies between different countries. We can expect that the three types capture much of the spatial dependence with a small number of parameters.
These ideas lead to the model

$$y_t = \rho_{1,t} W_1 y_t + \rho_{2,t} W_2 y_t + \rho_{3,t} W_3 y_t + \varepsilon_t, \ t = 1, \ldots, T, \tag{2.1}$$

where $\rho_{1,t}$ denotes the general dependence parameter, $\rho_{2,t}$ the parameter of dependence inside industrial branches and $\rho_{3,t}$ the local dependence parameter. W_1, W_2 and W_3 are known and row-standardized weighting matrices which reflect the spatial dependencies.

The diagonal entries of all these matrices are zero. Furthermore, all off-diagonal elements of W_1 are non-zero, the off-diagonal elements of W_2 are non-zero if the corresponding firms belong to the same industrial branch and the off-diagonal elements of W_3 are non-zero if the corresponding firms belong to the same country. For defining the non-zero entries, it is useful to consider the "a priori influence" of one stock return to another for which the market capitalization is a good approximation (see the discussion after Theorem 2).

Due to the assumption regarding the spatial dependencies we can assume that the different elements of the error vector ε_t are uncorrelated; however, they may be heteroscedastic. We assume that ε_t has zero mean which is plausible for stock returns. The model, which can be denoted as SAR(3,0), does not include any explanatory variables.

Now, denote $\rho_t = (\rho_{1,t}, \rho_{2,t}, \rho_{3,t})'$, where A' denotes the transpose of a matrix or a vector A. We are interested in structural changes in the spatial parameters, i.e. we consider the null hypothesis

$$H_0 : \rho_1 = \ldots = \rho_T \text{ vs. } H_1 : \exists t \in \{1, \ldots, T-1\} : \rho_t \neq \rho_{t+1}.$$

We opt to use the convenient CUSUM procedure for this testing problem, i.e. we first estimate ρ_t successively from the data and then compare the estimates with the estimate from the whole data set. The estimation relies on the GMM-approach which is motivated by the fact that the ML-approach might be problematic due to the possible large number of variance parameters to be estimated. Instead, with GMM, the n variance parameters (which may depend on other parameters as well) and their specific structure are not necessary for estimating ρ_t in a first step, but can be estimated separately in a second step, if necessary.

To be more precisely, the test statistic is some suitably transformed version of the process

$$(\hat{\rho}_j - \hat{\rho}_T, j = 1, \ldots, T)$$

where $\hat{\rho}_t := h(y_1, \ldots, y_t)$ is the estimator for ρ_t.

The GMM-estimator for ρ_t relies on the three moment conditions

$$\mathsf{E}(\epsilon_t' W_1 \epsilon_t) = \mathsf{E}(\epsilon_t' W_2 \epsilon_t) = \mathsf{E}(\epsilon_t' W_3 \epsilon_t) = 0$$

which are implied by the structure of the model. Replacing ε_t by $\varepsilon_t = (I_n - \rho_g W_1 - \rho_b W_2 - \rho_l W_3) y_t$ and replacing the theoretical quantities by their empirical counterparts leads to the estimator - based on the first j observations -

$$\hat{\rho}_j := (\hat{\rho}_{1,j}, \hat{\rho}_{2,j}, \hat{\rho}_{3,j})' := \arg\min_{\rho \in S} ||G_j \lambda + g_j||^2 = \arg\min_{\rho \in S} (G_j \lambda + g_j)'(G_j \lambda + g_j)$$

with the Euclidean norm $|| \cdot ||$.

Here, the empirical quantity G_j is a successive mean of $(3,9)$-matrices, defined as

$$G_j = \frac{1}{T} \sum_{t=1}^{j} f_G(\rho_t, y_t, W_1, W_2, W_3),$$

where, for $i, j \in \{1,2,3\}$, the elements of $f_G(y_t, W_1, W_2, W_3) = f_G(\rho_t, y_t, W_1, W_2, W_3)$ are defined as

$$\begin{aligned}
(f_G(y_t, W_1, W_2, W_3))_{i,j} &= -y_t'(W_i + W_i')W_j y_t, \\
(f_G(y_t, W_1, W_2, W_3))_{i,3+j} &= -y_t' W_j' W_i W_j y_t, \\
(f_G(y_t, W_1, W_2, W_3))_{i,7} &= -y_t' W_1'(W_i + W_i')W_2 y_t, \\
(f_G(y_t, W_1, W_2, W_3))_{i,8} &= -y_t' W_1'(W_i + W_i')W_3 y_t, \\
(f_G(y_t, W_1, W_2, W_3))_{i,9} &= -y_t' W_2'(W_i + W_i')W_3 y_t
\end{aligned}$$

and the empirical quantity g_j is a successive mean of $(3,1)$-vectors, defined as

$$g_j = \frac{1}{T} \sum_{t=1}^{j} f_g(y_t, W_1, W_2, W_3),$$

where, for $i \in \{1,2,3\}$, the elements of $f_g(y_t, W_1, W_2, W_3) = f_g(\rho_t, y_t, W_1, W_2, W_3)$ are defined as

$$(f_g(y_t, W_1, W_2, W_3))_i = y_t' W_i y_t.$$

Furthermore,

$$\lambda := \lambda(\rho) := \left(\rho_1, \rho_2, \rho_3, \rho_1^2, \rho_2^2, \rho_3^2, \rho_1\rho_2, \rho_1\rho_3, \rho_2\rho_3\right)'.$$

Under the null hypothesis, for the true parameter values,

$$\mathsf{E}[f_G(\rho_0, y_t, W_1, W_2, W_3)\lambda(\rho) + f_g(\rho_0, y_t, W_1, W_2, W_3)] =: \Gamma\lambda + \gamma = 0,$$

such that the $\hat{\rho}_j$ are consistent for $\rho_t (=: \rho_0 = (\rho_1, \rho_2, \rho_3))$ under the null hypothesis and the assumptions imposed below. Even for small j, ρ_0 can be well estimated if n is reasonably large. The assumptions for consistency are also needed for the derivation of the asymptotic null distribution of our test statistic and partly summarize the discussion in this section.

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

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

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

4. *The parameter space S is defined as $S = \{\rho \in \mathbb{R}^3, |\rho|_1 < 1\}$, where $|\cdot|_1$ denotes the 1-norm.*

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

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

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

7. *The matrix* $d_0 := \Gamma\lambda^{(1)}$ *with*

$$\lambda^{(1)'}(\rho_0) = \begin{pmatrix} 1 & 0 & 0 & 2\rho_1 & 0 & 0 & \rho_2 & \rho_3 & 0 \\ 0 & 1 & 0 & 0 & 2\rho_2 & 0 & \rho_1 & 0 & \rho_3 \\ 0 & 0 & 1 & 0 & 0 & 2\rho_3 & 0 & \rho_1 & \rho_2 \end{pmatrix}.$$

exists, is finite and has full rank.

8. *The process* $(f_G(y_t, W_1, W_2, W_3)\lambda(\rho) + f_g(y_t, W_1, W_2, W_3), t = 1, \ldots, T)$ *fulfills a functional central limit theorem, i.e. it holds for the process*

$$W_{T,S_W}(s) := \frac{1}{\sqrt{T}} \sum_{t=1}^{[sT]} [f_G(y_t, W_1, W_2, W_3)\lambda(\rho_0) + f_g(y_t, W_1, W_2, W_3)], s \in [0,1],$$

that, in $D([0,1], \mathbb{R}^3)$*, the* 3*-dimensional cross product of the Càdlàg-spaces on the interval* $[0,1]$*,*

$$W_{T,S_W}(\cdot) \Rightarrow_d W_{S_W}(\cdot),$$

where $\{W_{S_W}(s), s \in [0,1]\}$ *is a* 3*-dimensional Wiener process with limiting* 3*-dimensional covariance matrix*

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

with

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

$$\varepsilon_t = (I_n - \rho_1 W_1 - \rho_2 W_2 - \rho_3 W_3)^{-1} y_t,$$

Note that Assumption 1.8 is fulfilled if standard conditions on moments and serial dependence for multivariate functional central limit theorems apply. Since the norm of $(f_G(y_t, W_1, W_2, W_3)\lambda(\rho) + f_g(y_t, W_1, W_2, W_3)$ is bounded by the second-order cross moments of y_t, one typically needs finite fourth moments. Regarding serial dependence, e.g. the functional central limit theorem in Wooldridge and White (1988) requires near-epoch dependence with respect to a mixing process.

Theorem 1. *Under H_0 and Assumption 1, the suitably standardized estimator $(\hat{\rho}_j, j = 1, \ldots, T)$ converges against a Gaussian process, i.e. it holds for the process $W_{T,\Sigma}(s) = s\sqrt{T}(\hat{\rho}_{[sT]} - \rho_0), s \in [0,1]$, that*

$$W_{T,\Sigma}(\cdot) \Rightarrow_d W_\Sigma(\cdot),$$

where $\{W_\Sigma(s), s \in [0,1]\}$ is a 3-dimensional Wiener process with mean zero and covariance matrix $\Sigma = d_0^{-1} S_W d_0'^{-1}$, the asymptotic covariance matrix of $\hat{\rho}_j$.
Furthermore, Σ can be consistently estimated by an estimator $\hat{\Sigma}$.

Note that the matrix d_0 can be estimated by plug-in methods while estimation of S_W requires a kernel-based variance estimator. This leads to

$$\hat{d}_0 = G_T \cdot \begin{pmatrix} 1 & 0 & 0 & 2\hat{\rho}_{1,T} & 0 & 0 & \hat{\rho}_{2,T} & \hat{\rho}_{3,T} & 0 \\ 0 & 1 & 0 & 0 & 2\hat{\rho}_{2,T} & 0 & \hat{\rho}_{1,T} & 0 & \hat{\rho}_{3,T} \\ 0 & 0 & 1 & 0 & 0 & 2\hat{\rho}_{3,T} & 0 & \hat{\rho}_{1,T} & \hat{\rho}_{2,T} \end{pmatrix}$$

and

$$\begin{aligned} \hat{S}_W &= \frac{1}{T} \sum_{t=1}^{T} f(y_t, \hat{\rho}_T) f(y_t, \hat{\rho}_T)' \\ &+ \frac{1}{T} \sum_{i=1}^{T} k\left(\frac{i}{\gamma_T}\right) \sum_{t=1}^{T-i} [f(y_t, \hat{\rho}_T) f(y_{t+i}, \hat{\rho}_T)' + f(y_{t+i}, \hat{\rho}_T) f(y_t, \hat{\rho}_T)'] \end{aligned}$$

with

$$f(y_t, \hat{\rho}_T)' = \begin{pmatrix} \hat{\varepsilon}_t' W_1 \hat{\varepsilon}_t & \hat{\varepsilon}_t' W_2 \hat{\varepsilon}_t & \hat{\varepsilon}_t' W_3 \hat{\varepsilon}_t \end{pmatrix},$$

$$\hat{\varepsilon}_t = (I_n - \hat{\rho}_{1,T} W_1 - \hat{\rho}_{2,T} W_2 - \hat{\rho}_{3,T} W_3)^{-1} y_t,$$

the kernel function $k(\cdot)$ and the bandwidth γ_T.

Similar to Lee et al. (2003), the test statistic is the maximum over a weighted quadratic form of $(\hat{\rho}_j - \hat{\rho}_T)$ and is defined as

$$Q_T = \max_{1 \leq j \leq T} \frac{j^2}{T} (\hat{\rho}_j - \hat{\rho}_T)' \hat{\Sigma}^{-1} (\hat{\rho}_j - \hat{\rho}_T).$$

The central asymptotic result is

Theorem 2. *Under H_0 and Assumption 1,*

$$Q_T \to_d \sup_{s \in [0,1]} \sum_{i=1}^{3} B_i^2(s),$$

where $\{B_i(s), s \in [0,1]\}, i = 1, 2, 3,$ are independent standard Brownian Bridges.

There is an explicit form of the distribution function of the limit random variable in Theorem 2; some relevant critical values, which are provided in Kiefer (1959), p. 438, are 2.623 for $\alpha = 0.90$, 3.053 for $\alpha = 0.95$ and 4.004 for $\alpha = 0.99$.

We run a small simulation study to examine the size and power of our test in small samples. For both, we use the weighting matrices described above for the special case of Euro Stoxx 50 returns in the composition of January 2010 for the period from 2003 until 2009 using the partition into countries and industrial branches from Table 1.

– Table 1 here –

The weighting matrices (with $n = 50$) are constructed as described above with the additional characteristic that the nonzero entries consist of the stock weights in the Euro Stoxx 50 (with respect to the row-standardization) in order to consider the market capitalization. In the simulation study, there are always the parameters $\rho_1 = 0.5, \rho_2 = 0.2, \rho_1 = 0.1$ in the first part of the sample. In the power study, we change the parameters after one half of the sample. We always use several values of T, let the components of the ε_t be identically standard normally distributed, use 1000 replications, a nominal level of $\alpha = 5\%$

and for the long-run variance estimation the Bartlett kernel with bandwidth $\log(T)$. In a second experiment, we also consider the case $n = 100$ by building (row-standardized) diagonal block matrices of the form

$$\begin{pmatrix} W_i & \mathbf{0} \\ \mathbf{0} & W_i \end{pmatrix}, i = 1, 2, 3.$$

The results can be found in Table 2 and 3. We see that the test asymptotically keeps its size and has considerable power even for small values of T and small shifts of the correlations. For smaller T and $n = 50$, the empirical size is considerably larger than the nominal size. In general, the test detects increasing dependencies much better than decreasing dependencies; for $T = 100$ and $n = 50$, the test cannot detect decreasing dependencies in our setup. For larger n, in general, the empirical size is closer to the nominal size and the power is higher. This is plausible because the sums in G_j and g_j consist of quadratic forms and basically the amount of summmands is doubled if n is doubled.

– Table 2 here –

– Table 3 here –

Analogously, one could obtain constancy tests for just one or two of the three parameters by extracting the relevant parts of the 3-dimensional vector $(\hat{\rho}_j - \hat{\rho}_T)$ and the relevant parts of the 3×3 covariance matrix Σ. Then,

$$Q_T^* \to_d \sup_{s \in [0,1]} \sum_{i=1}^{k} B_i^2(s),$$

for $k \in \{1, 2\}$ and the "reduced" test statistic Q_T^*.

If one were for example interested in the constancy of the general dependence parameter

ρ_g, one would have

$$Q_T^* = \max_{1\le j\le T} \frac{j^2}{T}(\hat{\rho}_{1,j} - \hat{\rho}_{1,T})'\hat{\Sigma}_{11}^{-1}(\hat{\rho}_{1,j} - \hat{\rho}_{1,T})$$

with Σ_{11} the entry in the first row and first column of Σ. Then, $Q_T^* \to_d \sup_{s\in[0,1]} B_1^2(s)$, where $\{B_1(s), s \in [0,1]\}$ is a standard Brownian Bridge.

Relevant critical values for this case are also provided in Kiefer (1959), p. 438.

3. Local power

In this section, we analyze local power properties of our fluctuation test. We formulate the sequence of local alternatives in terms of the moment conditions, i.e. we have

$$H_1 : \mathsf{E}[f_G(\rho_t, y_t, W_1, W_2, W_3)\lambda(\rho_0) + f_g(\rho_t, y_t, W_1, W_2, W_3)] = \frac{1}{\sqrt{T}} h\left(\frac{t}{T}\right), t = 1, \ldots, T, \tag{3.1}$$

where $h(s) = (h_1(s), h_2(s), h_3(s))$ is a bounded 3-dimensional function that can be approximated by step functions in each component such that

$$\sup_{s\in[0,1]} \sup_{i\in\{1,2,3\}} \left| \int_0^s h_i(u)du - s\int_0^1 h_i(u)du \right| > 0.$$

A typical example for h might for example be a step function which jumps from 0 to $h_0 \neq 0$ in the point $z_0 = \frac{1}{2}$ in each component. Formally, we deal with triangular arrays in this setup because the distribution of the y_t changes with T, but we stick to the former notation for ease of exposition. The local alternatives (3.1) are equivalent to changes in the spatial correlation parameters. Note that $\lambda(\rho_0)$ does not change with t or T; the changes in ρ_t only affect the expectations of $f_G(\rho_t, y_t, W_1, W_2, W_3)$ and $f_g(\rho_t, y_t, W_1, W_2, W_3)$. To ensure the local alternative to converge properly against the null hypothesis, we impose

Assumption 2. *For $T \to \infty$ and fixed $t = 1, \ldots, T$, $\mathsf{E}[f_G(\rho_t, y_t, W_1, W_2, W_3)] \to \Gamma$ and $\mathsf{E}[f_g(\rho_t, y_t, W_1, W_2, W_3)] \to \gamma$.*

With this assumption, it is also ensured that $\rho_t =: (\rho_{1,t}, \rho_{2,t}, \rho_{3,t})'$ converges to ρ_0.

For the derivation of asymptotic properties, slight modifications of the previous assumptions are necessary. Assumption 1.1 and 1.8 are (in this order) replaced by

Assumption 3. *1. The sequence $(y_t, t \in \mathbb{Z})$ has zero mean and is ergodic.*

2. The process $\left(f_G(y_t, W_1, W_2, W_3)\lambda(\rho) + f_g(y_t, W_1, W_2, W_3) - \frac{1}{\sqrt{T}} h\left(\frac{t}{T}\right), t = 1, \ldots, T\right)$ fulfills a functional central limit theorem, i.e. it holds for the process

$$W_{T,S_W}(s) := \frac{1}{\sqrt{T}} \sum_{t=1}^{[sT]} \left[f_G(y_t, W_1, W_2, W_3)\lambda(\rho_0) + f_g(y_t, W_1, W_2, W_3) - \frac{1}{\sqrt{T}} h\left(\frac{t}{T}\right) \right],$$

$s \in [0, 1]$, that

$$W_{T,S_W}(\cdot) \Rightarrow_d W_{S_W}(\cdot),$$

where $\{W_{S_W}(s), s \in [0, 1]\}$ is a 3-dimensional Wiener process with limiting 3-dimensional covariance matrix S_W from 1.8.

Note that the comments on Assumption 1.8 apply in a similar way on 3.2.

The following two results are then corollaries of Theorem 1 and Theorem 2 as they can be obtained with similar proofs.

Corollary 1. *Under the sequence of local alternatives, Assumptions 1.2 - 1.7, 2 and 3, the suitably standardized estimator $(\hat{\rho}_t, t = 1, \ldots, T)$ converges against a Gaussian process, i.e. it holds for the process $W_{T,\Sigma}(s) = s\sqrt{T}(\hat{\rho}_{[sT]} - \rho_0), s \in [0, 1]$, that*

$$W_{T,\Sigma}(\cdot) \Rightarrow_d W_\Sigma(\cdot) + D(\cdot),$$

where $\{W_\Sigma(s), s \in [0, 1]\}$ is a 3-dimensional Wiener process with mean zero and covariance matrix Σ, and

$$D(s) = (D_1(s), D_2(s), D_3(s)) = -d_0^{-1} \int_0^s h(u)du.$$

Furthermore, Σ *can be consistently estimated by an estimator* $\hat{\Sigma}$.

Corollary 2. *Under the sequence of local alternatives, Assumptions 1.2 - 1.7, 2 and 3,*

$$Q_T \to_d \sup_{s\in[0,1]} \sum_{i=1}^{3} [B_i(s) + \Sigma^{-1/2}(D_i(s) - sD_i(1))]^2,$$

where $\{B_i(s), s \in [0,1]\}, i = 1,2,3,$ *are independent standard Brownian Bridges.*

Corollary 2 gives us two different information: First, for a given alternative, the Corollary provides a detailed description of the test's behavior and enables the applicant to approximate the rejection probability for fixed T. Second, the rejection probability becomes arbitrarily large for sufficient large shifts in the alternatives.

4. Application to risk management

We investigate the utility of the test for structural breaks by comparing the accuracy of predicted Values at Risk (VaR) for Euro Stoxx 50 members in the time period 2003-2009 using the spatial model with and without taking structural changes into account. Arnold et al. (2011) compare the spatial model without taking structural changes into account with a factor model and the sample covariance matrix with the same data set and demonstrate with this example that the spatial model can lead to more accurate risk forecasts. The present paper shows that the accuracy can be further improved by considering structural changes.

Replacing the unknown parameters by their estimates in the formula for the covariance matrix of y_t,

$$\begin{aligned} \mathsf{Cov}(y_t) &= (I_n - \rho_1 W_1 - \rho_2 W_2 - \rho_3 W_3)^{-1} \cdot diag\{\sigma_1^2, \ldots, \sigma_n^2\} \\ &\quad \cdot (I_n - \rho_1 W_1' - \rho_2 W_2' - \rho_3 W_3')^{-1} \\ &=: V, \end{aligned} \tag{4.1}$$

yields an estimate $\hat{V}_{spat}$ for the stock returns' covariance matrix V. Arnold et al. (2011)

estimate the parameters with a rolling window of 100 days. We will compare this method by taking structural changes into account; our procedure is as follows: Basically, we also use a rolling window, but at each time point, we perform the parameter constancy test on a significance level of 0.01% to decide if the parameters have been constant in the last 100 days. If the test does not reject, we use the 100 days for estimation. If the test rejects, we only use the data after the change point (with the constraint that we always use at least the past 10 days for estimation). The idea is that an estimator for the correlation parameters based on all data cannot produce reasonable results if the true parameters change. Of course, in practice the true change point is unknown, but we estimate it by

$$\mathsf{argmax}_{1 \le j \le T} \frac{j^2}{T} (\hat{\rho}_j - \hat{\rho}_T)' \hat{\Sigma}^{-1} (\hat{\rho}_j - \hat{\rho}_T)$$

which is a common and intuitive estimator in change point analysis, see e.g. Inclán and Tiao (1994), Galeano and Wied (2011) and the references therein. The procedure is repeated for a rolling window of 200 days, respectively.

By performing several tests, the nominal significance level might not be attained, but we do not discuss this issue here as we just use the test's decisions in an explorative way.

Each of the methods 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 two different ways of estimating the covariance matrix can thus be compared in the following way: The covariance matrix 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 both models in the same way so that the comparison of the models would remain the same.
For each α and each of the two models, we thus get daily updated estimated VaR. We compare these with the realized portfolio returns of the following day. For a convincing 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 1 shows the results for $\alpha \in (0, 0.05)$ and for the window lengths 100 days and 200 days.

- Figure 1 here -

Indeed, the spatial model with taking structural changes into account seems to be slightly more adequate to estimate risk than the other approach. It is basically either closer to or has equal distance to α. Consider e.g. the estimated VaR for $\alpha = 0.04$ for a sample period of 200 days. For the spatial model with structural breaks, portfolio returns fall below $\widehat{VaR}_\alpha$ in 5.6% of all days, whereas this happens slightly more frequently for the other approach (6.2%). This pattern can be found for almost all values of α considered here.
We conclude that accounting for structural breaks can lead to more accurate risk forecasts and might thus be relevant in practice.

Acknowledgements: Financial support by Deutsche Forschungsgemeinschaft (SFB 823, project A1) is gratefully acknowledged. I am very grateful to Matthias Arnold, Walter Krämer, an anonymous associate editor and an anonymous referee for helpful comments.

References

Amann, H. and J. Escher (2006): *Analysis II*, Birkhäuser Verlag, Basel, 2. ed.

Arnold, M., S. Stahlberg, and D. Wied (2011): "Modelling different kinds of spatial dependence in stock returns," *Empirical Economics, forthcoming.*

Asgharian, H., W. Hess, and L. Liu (2011): "A Spatial Analysis of International Stock Market Linkages," *working paper.*

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

Billingsley, P. (1968): *Convergence of probability measures*, Wiley, New York.

Brown, R., J. Durbin, and J. Evans (1975): "Techniques for Testing the Constancy of Regression Relationships over Time," *Journal of the Royal Statistical Society B*, 37(2), 149–163.

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

Galeano, P. and D. Wied (2011): "Multiple change point detection in the correlation structure of random variables," *arXiv:1206.5367v1.*

Inclán, C. and G. Tiao (1994): "Use of cumulative sums of squares for retrospective detection of change of variance," *Journal of the American Statistical Association*, 89(427), 913–923.

KIEFER, J. (1959): “K-sample analogues of the Kolmogorov-Smirnov and Cramer-V. Mises tests,” *The Annals of Mathematical Statistics*, 30(2), 420–447.

LEE, S., J. HA, O. NA, AND S. NA (2003): “The Cusum Test for Parameter Change in Time Series Models,” *Scandinavian Journal of Statistics*, 30(4), 781–796.

MANLY, B. AND D. MACKENZIE (2000): “A cumulative sum type of method for environmental monitoring,” *Environmetrics*, 11(2), 151–166.

——— (2003): “CUSUM environmental monitoring in time and space,” *Environmental and Ecological Statistics*, 10, 231–247.

PLOBERGER, W., W. KRÄMER, AND K. KONTRUS (1989): “A new test for structural stability in the linear regression model,” *Journal of Econometrics*, 40(2), 307–318.

SOWELL, F. (1996): “Optimal Tests for Parameter Instability in the Generalized Method of Moments Framework,” *Econometrica*, 64(5), 1085–1107.

WIED, D., W. KRÄMER, AND H. DEHLING (2012): “Testing for a change in correlation at an unknown point in time using an extended functional delta method,” *Econometric Theory*, 28(3), 570–589.

WOOLDRIDGE, J. AND H. WHITE (1988): “Some invariance principles and central limit theorems for dependent heterogeneous processes,” *Econometric Theory*, 4(2), 210–230.

A. PROOFS

Proof of Theorem 1

The proof bases on a Taylor expansion on the 3-dimensional derivative $D\Psi_j(\rho) := \frac{\partial \Psi_j(\rho)}{\partial \rho}$ of the target function $\Psi_j(\rho) := (G_j\lambda(\rho)+g_j)'(G_j\lambda(\rho)+g_j)$ using the fact that $D\Psi_j(\hat{\rho}_j) = 0$ due to the smoothness of $\Psi_j(\rho)$. It holds

$$D\Psi_j(\rho) = 2\lambda^{(1)'}(\rho)G_j'(G_j\lambda(\rho) + g_j).$$

With the mean value theorem for vector-valued functions in integral form (see Amann and Escher, 2006, Theorem 3.10) we have

$$
\begin{aligned}
D\Psi_j(\hat{\rho}_j) = 0 = D\Psi_j(\rho_0) + \int_0^1 \left[D^2\Psi_j(\rho_0 + t(\hat{\rho}_j - \rho_0))dt\right](\hat{\rho}_j - \rho_0) \\
\Leftrightarrow (\hat{\rho}_j - \rho_0) = -\left\{\int_0^1 \left[D^2\Psi_j(\rho_0 + t(\hat{\rho}_j - \rho_0))dt\right]\right\}^{-1} D\Psi_j(\rho_0) =: f(\rho_0, \hat{\rho}_j) D\Psi_j(\rho_0)
\end{aligned}
$$

with $D^2\Psi_j(\bar{\rho}) = 2\lambda^{(1)'}(\bar{\rho})G_j'G_j\lambda^{(1)}(\bar{\rho}) + o_{\mathsf{P}}(1)$ for any $\bar{\rho}$ between ρ_0 and $\hat{\rho}_j$.
It follows, for $s \in [0,1]$,

$$
\begin{aligned}
W_{T,S_W}(s) &= s\sqrt{T}(\hat{\rho}_{[sT]} - \rho_0) \\
&= -sf(\rho_0, \hat{\rho}_{[sT]})2\lambda^{(1)'}(\rho_0)G'_{[sT]}\sqrt{T}(G_{[sT]}\lambda(\rho_0) + g_{[sT]}) \\
&= -sf(\rho_0, \hat{\rho}_{[sT]})2\lambda^{(1)'}(\rho_0)G'_{[sT]} \\
&\quad \sqrt{T}\frac{1}{T}\sum_{t=1}^{[sT]}[f_G(y_t, W_1, W_2, W_3)\lambda(\rho_0) + f_g(y_t, W_1, W_2, W_3)].
\end{aligned}
$$

Let $\epsilon > 0$ arbitrary and $s \geq \epsilon$. With Assumption 1 and a standard argmin argument, $\hat{\rho}_{[sT]}$ converges uniformly to ρ_0 (see Arnold et al., 2011) so that the process $f(\rho_0, \hat{\rho}_{[sT]})$ converges to the process $\left[2\lambda^{(1)'}(\rho_0)s\Gamma' s\Gamma\lambda^{(1)}(\rho_0)\right]^{-1}$ and the process $s\sqrt{T}(\hat{\rho}_{[sT]} - \rho_0)$ converges in distribution to the process

$$
\begin{aligned}
&- s[2\lambda^{(1)'}(\rho_0)s\Gamma' s\Gamma\lambda^{(1)}(\rho_0)]^{-1}2\lambda^{(1)'}(\rho_0)s\Gamma' W_{S_W}(s) \\
=&- [\lambda^{(1)'}(\rho_0)\Gamma'\Gamma\lambda^{(1)}(\rho_0)]^{-1}\lambda^{(1)'}(\rho_0)\Gamma' W_{S_W}(s)
\end{aligned}
$$

with the 3-dimensional Wiener process $\{W_{S_W}(s), s \in [0,1]\}$ with covariance matrix S_W. This means that the process $s\sqrt{T}(\hat{\rho}_{[sT]} - \rho_0)$ converges to a 3-dimensional Wiener process with covariance matrix $d_0^{-1}S_W d_0'^{-1}$ and $d_0 = \Gamma\lambda^{(1)}$.

Now, let

$$W^{\epsilon}_{T,S_W}(s) = \begin{cases} W_{T,S_W}(s), & s \geq \epsilon \\ 0 & s < \epsilon \end{cases},$$

$$W^{\epsilon}(s) = \begin{cases} W_{\Sigma}(s), & s \geq \epsilon \\ 0 & s < \epsilon \end{cases}.$$

The previous calculations imply that

$$W^{\epsilon}_{T,S_W}(\cdot) \Rightarrow_d W^{\epsilon}(\cdot)$$

in $D([0,1],\mathbb{R}^3)$ and also

$$W^{\epsilon}(\cdot) \Rightarrow_d W_{\Sigma}(\cdot)$$

for rational $\epsilon \to 0$ in $D([0,1],\mathbb{R}^3)$.

The convergence of $W_{T,S_W}(\cdot)$ in $D([0,1],\mathbb{R}^3)$ follows with Theorem 4.2 in Billingsley (1968) if we can show that

$$\lim_{\epsilon\to 0}\limsup_{T\to\infty} \mathsf{P}(\sup_{s\in[0,1]} |W^{\epsilon}_{T,S_W}(s) - W_{T,S_W}(s)| \geq \eta) = \lim_{\epsilon\to 0}\limsup_{T\to\infty} \mathsf{P}(\sup_{s\in[0,\epsilon]} |W_{T,S_W}(s)| \geq \eta) = 0$$

for all $\eta > 0$.

For this, note that

$$\limsup_{T\to\infty} \mathsf{P}(\sup_{s\in[0,\epsilon]} |W_{T,S_W}(s)| \geq \eta) \leq \mathsf{P}(\sup_{s\in[0,\epsilon]} C|W^*(s)| \geq \eta)$$

where C is a constant and $\{W^*(s), s \in [0,\epsilon]\}$ is a Brownian Motion. This sum becomes arbitrarily small for $\epsilon \to 0$ and so the limit result is proved.

All entries of the limiting covariance matrix can be estimated consistently by plug-in-methods and kernel-based estimators. ■

Proof of Theorem 2

By Theorem 1, $W_{T,S_W}(\cdot) \Rightarrow_d W_\Sigma(\cdot)$, so that the process

$$B_T(s) := s\sqrt{T}(\hat{\rho}_{[sT]} - \hat{\rho}_T) = s\sqrt{T}(\hat{\rho}_{[sT]} - \rho_0) - s\sqrt{T}(\hat{\rho}_T - \rho_0)$$

converges weakly to the process $B_\Sigma(s) := W_\Sigma(s) - sW_\Sigma(1)$ which is a k-dimensional Brownian Bridge with covariance matrix Σ. With Slutzky's theorem, the process $B_T^*(s) := \Sigma^{-1/2} s\sqrt{T}(\hat{\rho}_{[sT]} - \hat{\rho}_T)$ converges weakly to $\{B(s), s \in [0,1]\}$, a k-dimensional standard Brownian Bridge, i.e. a k-dimensional vector whose components are independent one-dimensional standard Brownian Bridges. With the consistent estimator $\hat{\Sigma}$ from Theorem 1, the process $B_T^{**}(s) := \hat{\Sigma}^{-1/2} s\sqrt{T}(\hat{\rho}_{[sT]} - \hat{\rho}_T)$ converges to the same limit. An application of the Continuous Mapping Theorem with the function

$$f : D([0,1], \mathbb{R}^3) \to \mathbb{R}$$
$$(x_1(\cdot), x_2(\cdot), x_3(\cdot)) \to \sup_{s \in [0,1]} \sum_{i=1}^{3} x_i(s)^2$$

yields the convergence

$$\sup_{s \in [0,1]} (B_T^{**}(s))' B_T^{**}(s) \to_d \sup_{s \in [0,1]} (B(s))' B(s).$$

Using the identity $j \in \{1, \dots, T\} \leftrightarrow [sT]$ with $s \in [0,1]$ we directly see that

$$\sup_{s \in [0,1]} (B_T^{**}(s))' B_T^{**}(s) = Q_T.$$

■

Proof of Corollary 1

The proof straightforwardly follows the arguments of the proof of Theorem 1 with the additional argument

$$\sup_{z\in[0,1]} \left| \frac{1}{T} \sum_{t=1}^{[sT]} h\left(\frac{t}{T}\right) \right| \to_{T\to\infty} \int_0^z h(u)du$$

from e.g. Ploberger et al. (1989). ■

Proof of Corollary 2

The proof straightforwardly follows the arguments of the proof of Theorem 2. ■

Table 1: Partitioning of Euro Stoxx 50 members into branches and countries

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
Small countries	Aegon, Anheuser Busch, Arcelor, CRH, ING, Nokia, 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

Table 2: Empirical size and empirical power, $n = 50$

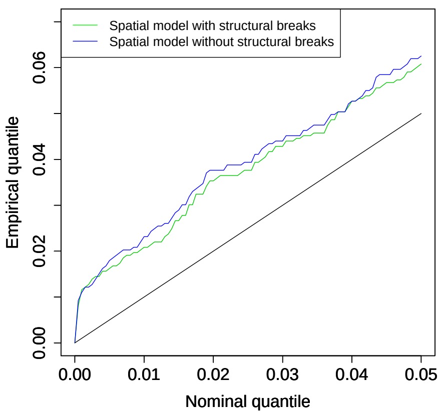
Parameter vector in the second half	$T = 100$	$T = 200$	$T = 300$	$T = 400$
$(0.50, 0.20, 0.10)$	0.165	0.096	0.065	0.050
$(0.55, 0.20, 0.10)$	0.566	0.741	0.899	0.962
$(0.55, 0.25, 0.15)$	1.000	1.000	1.000	1.000
$(0.50, 0.15, 0.10)$	0.109	0.168	0.463	0.704
$(0.45, 0.15, 0.05)$	0.062	0.468	0.961	1.000
$(0.40, 0.10, 0.00)$	0.152	0.929	1.000	1.000

Table 3: Empirical size and empirical power, $n = 100$

Parameter vector in the second half	$T = 100$	$T = 200$	$T = 300$	$T = 400$
$(0.50, 0.20, 0.10)$	0.059	0.047	0.039	0.036
$(0.55, 0.20, 0.10)$	0.714	0.947	0.992	1.000
$(0.55, 0.25, 0.15)$	1.000	1.000	1.000	1.000
$(0.50, 0.15, 0.10)$	0.069	0.514	0.888	0.989
$(0.45, 0.15, 0.05)$	0.113	0.955	1.000	1.000
$(0.40, 0.10, 0.00)$	0.506	1.000	1.000	1.000

Figure 1: Estimated VaR for the spatial model with and without structural breaks

(a) Window length 100 days



(b) Window length 200 days