SFB 823

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

Discussion Paper

Dominik Wied

Nr. 30/2011

SFB 823

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

Dominik Wied*

TU Dortmund

This Version: August 30, 2011

*Fakultät Statistik, CDI-Gebäude, TU Dortmund, 44221 Dortmund. Email: wied@statistik.tu-dortmund.de, Phone: +49 231 755 3869.

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 allows for superior risk forecasts in portfolio management.

JEL Classification: C13, C31, C32, C51

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

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 parameter 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 fills this gap by proposing a 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, Lee et al. (2003) propose a general framework for parameter constancy tests in time series models, Busetti and Harvey (2011), Krämer and van Kampen (2011) and van Kampen and Wied (2011) propose a copula constancy test and Wied et al. (2011) propose a non-parametric test for constant Pearson correlation. However, as far as the author knows, a special test for spatial parameters 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 $t = 1, \ldots, T$, 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:

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

$\rho_{1,t}$ denotes the general dependence parameter, $\rho_{2,t}$ the parameter of dependence inside branches and $\rho_{3,t}$ the local dependence parameter. W_1, W_2 and W_3 are weighting matrices which are specified in Arnold et al. (2011).

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

$$H_0 : \forall t \in \{1, \ldots, T\} : \rho_t = \rho_0 \text{ vs. } H_1 : \exists t \in \{1, \ldots, T-1\} : \rho_t \neq \rho_{t+1}$$

for a constant $\rho_0 \in \mathbb{R}^3$.

We propose a CUSUM-type test, i.e. we first estimate ρ_t with a GMM-estimator successively from the data and then compare the estimates with the estimate from the whole data set. To be more precise, let $\hat{\rho}_t := h(y_1, \ldots, y_t)$ be the estimator for ρ_t, then the test statistic is some suitably transformed version of the process

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

The vector ρ_t is estimated by a GMM-estimator. The estimator based on the first j observations is defined as

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

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

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

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

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

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

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,

$$\mathsf{E}[f_G(\rho_0, y_t, W_1, W_2, W_3)\lambda(\rho) + f_g(\rho_t, y_0, W_1, W_2, W_3)] =: \Gamma\lambda + \gamma = 0,$$

see Arnold et al. (2011), 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. These assumptions are also needed for the derivation of the asymptotic null distribution.

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 $W_{S_W}(s)$ *is a* 3*-dimensional Wiener process with limiting* 3*-dimensional covariance matrix* S_W *as specified in Arnold et al. (2011).*

The spatial weighting 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. The n variances may depend on additional parameters, respectively, but for estimating ρ_j, the specific structure of ε_t is irrelevant. After estimating ρ_j, the variances can be estimated in a two-stage procedure, see Arnold et al. (2011). The model does not include any explanatory variables and in the context of daily stock returns, the zero mean assumption is plausible, see also Aue et al. (2009). Due to the structure of innovations, the spatial autoregressive model can be denoted as $SAR(3,0)$.

Note that Assumption 1.3 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 e.g. Wooldridge and White (1988) requires near-epoch dependence with respect to a mixing process.

Lemma 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 $W_\Sigma(\cdot)$ 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, see e.g. de Jong and Davidson (2000) and Arnold et al. (2011). This leads to

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

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

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

The central asymptotic result is

Theorem 1. *Under H_0 and Assumption 1,*

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

where $B_1(s), \ldots, B_3(s), s \in [0,1]$, are independent standard Brownian Bridges.

There is an explicit form of the distribution function of the limit random variable in Theorem 1, see Aue et al. (2009), p. 4051. 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 from Arnold et al. (2011), i.e. $n = 50$ and 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 ε_t be NID, use 1000 replications, a nominal level of $\alpha = 5\%$ and for the long-run variance estimation the Barlett kernel with bandwidth $\log(T)$.

The results can be found in Table 1. 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.

– Table 1 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^*. 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

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

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 $\rho_t =: (\rho_{1,t}, \rho_{2,t}, \rho_{3,})'$ 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 $W_{S_W}(s)$ is a 3-dimensional Wiener process with limiting 3-dimensional covariance matrix S_W which is written down in Arnold et al. (2011).

The following two results are then corollaries of Theorem 1 and Theorem 1 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 $W_\Sigma(\cdot)$ 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 $B_1(s), \ldots, B_3(s), s \in [0,1]$ *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 (see Arnold et al., 2011 for a detailed description of the data) 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,

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

with the 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

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

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 grateful to Matthias Arnold and Walter Krämer for helpful comments.

References

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

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

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

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, 149–163.

Busetti, F. and A. Harvey (2011): "When is a copula constant? A test for changing relationships," *Journal of Financial Econometrics*, 9(1), 106–131.

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

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 financial assets," *submitted for publication.*

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, 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.

Krämer, W. and M. van Kampen (2011): "A simple nonparametric test for structural change in joint tail probabilities," *Economics Letters*, 110, 245–247.

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.

Ploberger, W., W. Krämer, and K. Kontrus (1989): "A new test for structural stability in the linear regression model," *Journal of Econometrics*, 40, 307–318.

van Kampen, M. and D. Wied (2011): "A nonparametric constancy test for copulas under mixing conditions," *submitted for publication.*

Wied, D., W. Krämer, and H. Dehling (2011): "Testing for a change in correlation at an unknown point in time using an extended functional delta method," *Econometric Theory, forthcoming.*

Wooldridge, J. and H. White (1988): "Some invariance principles and central limit theorems for dependent heterogeneous processes," *Econometric Theory*, 4, 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 multivariate mean value theorem 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

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

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 $f(\rho_0, \hat\rho_j)$ converges to $2\lambda^{(1)'}(\rho_0)s\Gamma's\Gamma\lambda^{(1)}(\rho_0)$ and $s\sqrt{T}(\hat\rho_{[sT]} - \rho_0)$ converges in distribution to the stochastic 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 $W_{S_W}(s)$ with covariance matrix S_W. This means that $s\sqrt{T}(\hat{\rho}_{[sT]} - \rho_0)$ 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

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

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

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

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

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

where C is a constant and $W^*(s)$ 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 1

By Theorem 1, $W_T(\cdot) \Rightarrow_d W(\cdot)$, so that the process

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

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 $B(s)$, 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

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

yields the convergence

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

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

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

■

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 1. ■

Table 1: Empirical size and empirical power

Parameter vector in the second half	$T = 200$	$T = 250$	$T = 300$	$T = 400$
$(0.5, 0.2, 0.1)$	0.096	0.083	0.065	0.050
$(0.55, 0.2, 0.1)$	0.741	0.822	0.899	0.962
$(0.5, 0.15, 0.1)$	0.168	0.319	0.463	0.704
$(0.45, 0.15, 0.05)$	0.468	0.779	0.961	1
$(0.55, 0.25, 0.15)$	1	1	1	1

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

(a) Window length 100 days (b) Window length 200 days