DEPARTMENT OF ECONOMICS AND BUSINESS ECONOMICS
AARHUS UNIVERSITY

CREATES
Center for Research in Econometric Analysis of Time Series

The role of cointegration for optimal hedging with heteroscedastic error term

Lukasz Gatarek and Søren Johansen

CREATES Research Paper 2017-12

Department of Economics and Business Economics
Aarhus University
Fuglesangs Allé 4
DK-8210 Aarhus V
Denmark

Email: oekonomi@au.dk
Tel: +45 8716 5515

The role of cointegration for optimal hedging with heteroscedastic error term

Lukasz Gatarek[†]
University of Łódź

Søren Johansen[‡]
University of Copenhagen
and CREATES

March 5, 2017

Abstract

The role of cointegration is analysed for optimal hedging of an h-period portfolio. Prices are assumed to be generated by a cointegrated vector autoregressive model allowing for stationary martingale errors, satisfying a mixing condition and hence some heteroscedasticity. The risk of a portfolio is measured by the conditional variance of the h-period return given information at time t. If the price of an asset is nonstationary, the risk of keeping the asset for h periods diverges for large h. The $h-$period minimum variance hedging portfolio is derived, and it is shown that it approaches a cointegrating vector for large h, thereby giving a bounded risk. Taking the expected return into account, the portfolio that maximizes the Sharpe ratio is found, and it is shown that it also approaches a cointegration portfolio. For constant conditional volatility, the conditional variance can be estimated, using regression methods or the reduced rank regression method of cointegration. In case of conditional heteroscedasticity, however, only the expected conditional variance can be estimated without modelling the heteroscedasticity. The findings are illustrated with a data set of prices of two year forward contracts for electricity, which are hedged by forward contracts for fuel prices. The main conclusion of the paper is that for optimal hedging, one should exploit the cointegrating properties for long horizons, but for short horizons more weight should be put on the remaining dynamics.

Keywords: hedging, cointegration, minimum variance portfolio, maximum Sharpe ratio portfolio
JEL Classification: C22, C58, G11

[*]The first author is grateful to National Science Center Poland for funding with grant Preludium No. 2013/09/N/HS4/03751. The second author is grateful to CREATES - Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation. The authors would like to thank Katarina Juselius for expert advice on modelling the data.

[†]Chair of Econometric Models and Forecasts, Institute of Econometrics and Statistics, University of Łódź, ul. Rewolucji 1905, 90-214 Łódź, Poland. E-mail: gatarek@tlen.pl

[‡]Department of Economics, University of Copenhagen and CREATES, Department of Economics and Business, Aarhus University, DK-8000 Aarhus C. Correspondance: Department of Economics, University of Copenhagen, Øster Farimagsgade 5, building 26, 1353 Copenhagen K, Denmark. E-mail: soren.johansen@econ.ku.dk

1 Introduction, some notation and summary

1.1 Motivation for the problem investigated

The use of cointegration for analyzing financial data is well established over the last 20 years. The problem of price discovery is discussed by Hasbrouck (1995), Lehmann (2002), de Jong and Schotman (2010), and Grammig, Melvin, and Schlag (2005). Gatev, Goetzmann, and Rouwe (2006) study pairs trading, and continuous time models with a heteroscedastic error process are developed by Duan and Pliska (2004) and Nakajima and Ohashi (2011). Alexander (1999), and more recently Juhl, Kawaller, and Koch (2012), studied optimal hedging using cointegration. The idea of a minimum variance portfolio dates back to the seminal paper by Markowitz (1952) and has since been explored and extended in both the financial and econometric literature, see for instance Grinold and Kahn (1999).

In general, the hedging methods can be divided in two classes: static and dynamic methods. The static hedging techniques assume that the hedging portfolio is selected, given information available in period t, and remains unchanged during the entire holding period $t+1, \ldots, t+h$. This is opposed to the dynamic hedging methods which allows for rebalancing the portfolio during the holding period, but we are only concerned with static hedging.

This paper studies optimal hedging for an h-period investment. It is assumed that there are n assets with prices $y_t = (y_{1t}, \ldots, y_{nt})'$, and that the first asset is held for h periods, using the other assets to hedge the risk, as measured by conditional variance of returns $\Sigma_{t,h} = Var_t(y_{t+h} - y_t)$ given information at time t, that is y_s, $s = 1, \ldots, t$.

The cointegrated vector autoregressive model (CVAR) with a restricted constant term and an error term that allows for heteroscedasticity is assumed to describe the variation of the prices. This model allows for nonstationary prices with stationary linear combinations, that is cointegration.

The first set of results concerns the derivation of an expression for the risk, $\Sigma_{t,h}$, which depends on conditional volatility of the error term. Based on this expression, the optimal $h-$period hedging portfolio, which minimizes this risk is derived. The limit for $h \to \infty$ of the inverse risk matrix, $\Sigma_{t,h}^{-1}$, is found and used to show that the optimal portfolio approaches a variance minimal cointegrating portfolio, which has a bounded risk.

Thus for longer horizons we should choose the variance minimal cointegrating portfolio, which has a bounded risk, and for shorter horizons we should take conditional volatility into account.

The second set of results concerns estimation of risk, and the optimal h-period hedging portfolio based on data $y_t, t = 1, \ldots, T$. Under assumptions on the error term that allows for heteroscedasticity, we show two results. First we show that a regression of returns $y_{t+h} - y_t$ on information at time t gives a consistent estimator for Σ_h, and a similar result holds if the CVAR is estimated by reduced rank regression. Next it is shown that a regression of y_{1t} on the other prices and a constant gives a consistent estimator of the optimal limiting hedging portfolio.

The conclusion of this is, that if the conditional variance is used as risk measure, in the case of conditional volatility, this has to be modelled by a multivariate GARCH model, like the BEKK model, see for instance Engle and Kroner (1995) and Comte and Lieberman (2003), or a multivariate ARCH model like Li, Ling, and Wong (2001). The combined theory of cointegration and a model for heteroscedasticity is challenging. The obvious two-step procedure of first estimating the CVAR assuming i.i.d. Gaussian errors and then use the estimated residuals as input in a BEKK model has not been work out in details.

The well-known formula

$$Var(y_{t+h} - y_t) = E(Var_t(y_{t+h} - y_t)) + Var(E_t(y_{t+h} - y_t)),$$

shows that the choice between the conditional variance, $\Sigma_{h,t}$ and its expectation, Σ_h, does not involve the variation of the information y_t given at the time of investment.

If a consistent estimator of $\Sigma_{t,h} = Var_t(y_{t+h} - y_t)$ is needed, one has to model conditional volatility, but if the first term $\Sigma_h = E(Var_t(y_{t+h} - y_t))$ can be used, it can be estimated by the simple regression methods or from the CVAR.

The role of cointegration for hedging was analysed by Juhl, Kawaller, and Koch (2011). They considered a special case of the CVAR, and we want in this paper to generalize their results to a CVAR with more lags and more cointegrating relations and allow for a some degree of heteroscedasticity in the martingale error term.

Finally we analyze some daily data for futures of electricity prices, and compare the optimal hedging portfolio with the cointegrating portfolio. All proofs are given in the Appendix.

We conclude that cointegration plays an important role in hedging. It allows for the possibility that an h-period hedging portfolio has a risk that is bounded in the horizon h, as opposed to the unhedged risk. As important is the result that for moderate horizons, it is important not to use the cointegrating portfolio, but to use the optimal hedging portfolio which interpolates between the short and long-horizon cointegrating portfolio.

2 Optimal hedging in the CVAR with ergodic, mixing, martingale difference error terms

The results are formulated in Theorem 2 for the cointegrated VAR (CVAR) model with two lags

$$\Delta y_t = \alpha(\beta' y_{t-1} - \xi) + \Gamma_1 \Delta y_{t-1} + \varepsilon_t. \tag{1}$$

It is only a question of a more elaborate notation to handle the case of more lags using the companion form, see the proof of Theorem 2.

We formulate the assumptions on the parameters of the data generating process, see Johansen (1996, Theorem 4.2), and define the characteristic polynomial for the lag two model, $\Psi(z) = (1-z)I_n - \Pi z - \Gamma_1 z(1-z)$. In the following we define $a_\perp$, which for any $n \times m$ matrix, a, of rank $m < n$ is defined as an $n \times (n-m)$ matrix of rank $n-m$, for which $a' a_\perp = 0$.

Assumption 1 *The roots of* $\det \Psi(z) = 0$ *satisfy* $|z| > 1$ *or* $z = 1$, *and* $\Pi = \alpha\beta'$, *where* α *and* β *are* $n \times r$ *matrices of full rank* $r < n$. *The matrix* $\alpha'_\perp(I_n - \Gamma_1)\beta_\perp$ *has full rank, and*

$$C = \beta_\perp(\alpha'_\perp(I_n - \Gamma_1)\beta_\perp)^{-1}\alpha'_\perp. \tag{2}$$

Next we formulate the assumptions on the error term.

Assumption 2 *The innovations,* ε_t, *form an ergodic martingale difference sequence with respect to a filtration* $\mathcal{F}_t$, $t = \cdots -1, 0, 1, \ldots$, *satisfying for some* $\delta > 0$

$$E_t(\varepsilon_{t+1}) = E(\varepsilon_{t+1}|\mathcal{F}_t) = 0, \qquad E||\varepsilon_t||^{4+\delta} \le c < \infty, \tag{3}$$

and a mixing condition

$$E_t(\varepsilon_{t+h}\varepsilon_{t+h}') \xrightarrow{P} \Omega = E(\varepsilon_t\varepsilon_t') > 0,\ h \to \infty. \tag{4}$$

Note that the condition of constant conditional volatility, $E_t(\varepsilon_{t+1}\varepsilon_{t+1}') = \Omega$, implies (4) and if further the errors are $i.i.d.(0, \Omega)$ with $4+\delta$ moments, then also (3) is satisfied.

Suppose Assumptions 1 and 2 hold, then we find from the theory of the CVAR, that y_t is a nonstationary process, whereas Δy_t and $\beta' y_t$ are stationary. For the CVAR, we can find an expression for the conditional mean and variance of the h period return given information at time t, and therefore analyze analytically the role of cointegration for the optimal h period portfolio, in particular the limit behaviour for $h \to \infty$, but first we discuss optimal hedging.

2.1 Optimal hedging

Let $y_t = (y_{1t}, y_{2t}, \ldots, y_{nt})'$ denote prices of assets $1, 2, \ldots, n$, and let $\eta = (\eta_1, \ldots, \eta_n)'$ denote portfolio weights, such that $\eta' y_t$ is the price of the portfolio. We define conditional expected return $\mu_{t,h} = E_t(y_{t+h} - y_t)$ and conditional variance $\Sigma_{t,h} = Var_t(y_{t+h} - y_t)$ given information at time t and use the notation $\Sigma_h = E(\Sigma_{t,h})$. We formulate the optimization problem, see Markowitz (1952), as minimizing conditional variance of the return $\eta'(y_{t+h} - y_t)$, that is $\eta'\Sigma_{t,h}\eta$, under the constraint that $a'\eta = 1$, for some vector $a \in \mathbb{R}^n$. In particular, for $a = e_{n1} = (1, 0_{n-1}')'$ we find the optimal hedging portfolio, and for $a = \mu_{t,h}$ we find the optimal portfolio in the sense of Markowitz. This portfolio also maximizes (squared) Sharpe ratio $(\eta'\mu_{t,h})^2/\eta'\Sigma_{t,h}\eta$, see Theorem 4.

Using the notation Σ, minimization of $\eta'\Sigma\eta$ under the constraint $a'\eta = 1$, is solved by the Lagrange multiplier problem

$$\frac{\partial}{\partial\eta} : \eta'\Sigma\eta - 2\lambda(\eta' a - 1) = 0,$$

giving

$$\eta_{opt} = \Sigma^{-1}a/a'\Sigma^{-1}a, \tag{5}$$

with risk

$$\eta_{opt}'\Sigma\eta_{opt} = (a'\Sigma^{-1}a)^{-1}.$$

For hedging, $a = e_{n1}$. A different expression is found using

$$I_n = \Sigma\Sigma^{-1} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \begin{pmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{pmatrix},$$

where Σ_{22} is $(n-1)\times(n-1)$. It follows that $\Sigma_{21}\Sigma^{11} + \Sigma_{22}\Sigma^{21} = 0$, or

$$\Sigma^{21}/\Sigma^{11} = -\Sigma_{22}^{-1}\Sigma_{21}. \tag{6}$$

The optimal hedging portfolio is denoted η_{opt}, and is given by

$$\eta_{opt} = \Sigma^{-1}e_{n1}/e_{n1}'\Sigma^{-1}e_{n1} = \begin{pmatrix} \Sigma^{11} \\ \Sigma^{21} \end{pmatrix} /\Sigma^{11} = \begin{pmatrix} 1 \\ \Sigma^{21}/\Sigma^{11} \end{pmatrix} = \begin{pmatrix} 1 \\ -\Sigma_{22}^{-1}\Sigma_{21} \end{pmatrix}. \tag{7}$$

Thus, applying this to $\Sigma_{t,h}$, the price of the optimal $h-$period hedging portfolio, $\eta^*_{t,h}$, is given by two expressions

$$\eta^{*\prime}_{t,h} y_t = y_t' \Sigma_{t,h}^{-1} e_{n1} / e_{n1}' \Sigma_{t,h}^{-1} e_{n1} = y_{1t} - \Sigma_{t,h,12} \Sigma_{t,h,22}^{-1} (y_{2t}, \ldots, y_{nt})'. \tag{8}$$

It turns out that the formulation (5) is more convenient for asymptotic analysis, which consists of finding the limit of $\Sigma_{t,h}^{-1}$ for $h \to \infty$, see Lemma 1, whereas the second shows that the coefficient can be found as a regression coefficient, $\Sigma_{t,h,12}\Sigma_{t,h,22}^{-1}$.

2.2 Properties of the CVAR

Here some results for the solution, y_t, of equation (1) are collected. Expressions for conditional variance, $\Sigma_{t,h}$, its expectation $\Sigma_h = E(\Sigma_{t,h})$, and conditional mean return, $\mu_{t,h}$, are given, and the main limit result for the inverse conditional variance is proved using Lemma 1. To simplify the notation, we analyse a lag one model, and apply these results later to a lag two model in companion form.

Theorem 1 *Let Assumptions 1 and 2 be satisfied, and let $y_t \in \mathbb{R}^n, t = 1, \ldots, T$, be given by*

$$\Delta y_t = \alpha(\beta' y_{t-1} - \xi) + \varepsilon_t. \tag{9}$$

The solution satisfies for $\rho = I_r + \beta'\alpha$,

$$\beta' y_t = \sum_{i=0}^{\infty} \rho^i \beta' \varepsilon_{t-i} + \xi, \tag{10}$$

$$y_{t+h} - y_t = \sum_{i=0}^{h-1} \{C + \alpha(\beta'\alpha)^{-1}\rho^i\beta'\}\varepsilon_{t+h-i} + \alpha(\beta'\alpha)^{-1}(\rho^h - I_r)(\beta' y_t - \xi). \tag{11}$$

It follows that $y_{t+h} - y_t$ has conditional mean

$$\mu_{t,h} = E_t(y_{t+h} - y_t) = \alpha(\beta'\alpha)^{-1}(\rho^h - I_r)(\beta' y_t - \xi), \tag{12}$$

and conditional variance

$$\Sigma_{t,h} = Var_t(y_{t+h} - y_t) = \sum_{i=0}^{h-1} \{C + \alpha(\beta'\alpha)^{-1}\rho^i\beta'\} E_t \varepsilon_{t+h-i}\varepsilon_{t+h-i}' \{C' + \beta\rho'^i(\alpha'\beta)^{-1}\alpha'\}, \tag{13}$$

with expectation $\Sigma_h = E(\Sigma_{t,h})$ given by

$$\Sigma_h = \sum_{i=0}^{h-1} \{C + \alpha(\beta'\alpha)^{-1}\rho^i\beta'\}\Omega\{C' + \beta\rho'^i(\alpha'\beta)^{-1}\alpha'\}. \tag{14}$$

For $h \to \infty$, the limit behaviour of $\Sigma_{t,h}$ is given by

$$\beta'\Sigma_{t,h}\beta \xrightarrow{\mathrm{P}} Var(\beta' y_t),\ h^{-1}\beta_\perp'\Sigma_{t,h}\beta_\perp \xrightarrow{\mathrm{P}} \beta_\perp' C\Omega C'\beta_\perp,\ \beta_\perp'\Sigma_{t,h}\beta = O_P(1), \tag{15}$$

and it follows that

$$\Sigma_{t,h}^{-1} \xrightarrow{\mathrm{P}} \beta Var^{-1}(\beta' y_t)\beta'. \tag{16}$$

The same results hold for $\Sigma_h = E(\Sigma_{h,t})$,

$$\Sigma_h^{-1} \to \beta Var^{-1}(\beta' y_t)\beta'.$$

2.3 Optimal hedging in the CVAR

The main result for the hedging problem in the CVAR with ergodic, mixing, martingale difference sequence as error term is given for the two lag model. It is assumed that parameters and hence conditional mean and variance are known, and we return in Section 3 to the question of how to estimate these quantities based on data $y_{1t}, t = 1, \ldots, T$. Because the first coordinate of the portfolio has a special role, $\eta_1 = 1$, we introduce the notation

$$\Sigma_{t,h} = \begin{pmatrix} \Sigma_{t,h,11} & \Sigma_{t,h,12} \\ \Sigma_{t,h,21} & \Sigma_{t,h,22} \end{pmatrix}.$$

Theorem 2 *Let y_t be given by model (1), and let Assumption 1 and 2 hold. The optimal hedging portfolio, $\eta^*_{t,h}$, and its limit if $\beta' e_{n1} \neq 0$, are given by*

$$\eta^*_{t,h} = (1, -\Sigma_{t,h,12}\Sigma^{-1}_{t,h,22})' = \frac{\Sigma^{-1}_{t,h} e_{n1}}{e'_{n1}\Sigma^{-1}_{t,h} e_{n1}} \xrightarrow{P} \frac{\beta\Phi^{-1}_{\beta\beta}\beta' e_{n1}}{e'_{n1}\beta\Phi^{-1}_{\beta\beta}\beta' e_{n1}}, \tag{17}$$

where $\Phi_{\beta\beta} = Var(\beta' y_t)$. The limits for $h \to \infty$ of conditional mean return and risk of the optimal hedging portfolio are

$$\eta^{*\prime}_{t,h}\mu_{t,h} \xrightarrow{P} -\frac{e'_{n1}\beta\Phi^{-1}_{\beta\beta}(\beta' y_t - \xi)}{e'_{n1}\beta\Phi^{-1}_{\beta\beta}\beta' e_{n1}}, \tag{18}$$

$$\eta^{*\prime}_{t,h}\Sigma_{t,h}\eta^*_{t,h} \xrightarrow{P} (e'_{n1}\beta\Phi^{-1}_{\beta\beta}\beta' e_{n1})^{-1}. \tag{19}$$

Finally, if $e'_{n1}\beta \neq 0$ and $e'_{n1}\beta_\perp \neq 0$, the fraction of explained variation, $R^2_{t,h}$, satisfies

$$1 - R^2_{t,h} = \frac{(e'_{n1}\Sigma^{-1}_{t,h} e_{n1})^{-1}}{e'_{n1}\Sigma_{t,h} e_{n1}} \xrightarrow{P} 0, \text{ for } h \to \infty. \tag{20}$$

The limit results are the same, if we replace $\Sigma_{t,h}$ by its expectation $\Sigma_h = E(\Sigma_{t,h})$.

The limit expression in (17) for $\eta^*_{t,h}$ shows explicitly how cointegrating vectors, β, should be combined by their variance $\Phi_{\beta\beta} = Var(\beta' y_t)$ to give the optimal hedging portfolio for large horizons.

Another expression for the optimal portfolio, $\eta^*_{t,h}$, can be found by normalizing the cointegrating relations such that

$$\beta = \begin{pmatrix} 1 & 0'_{r-1} \\ \beta_1 & \beta_2 \end{pmatrix}.$$

In this case

$$\Phi_{\beta\beta} = \begin{pmatrix} Var(y_{1t} + \beta'_1 y_{2t}) & Cov(y_{1t} + \beta'_1 y_{2t}, \beta'_2 y_{2t}) \\ Cov(\beta'_2 y_{2t}, y_{1t} + \beta'_1 y_{2t}) & Var(\beta'_2 y_{2t}) \end{pmatrix} = \begin{pmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{21} & \Phi_{22} \end{pmatrix},$$

say, and the limit of the optimal portfolio can be expressed as

$$\begin{aligned} \frac{\beta\Phi^{-1}_{\beta\beta}\beta' e_{n1}}{e'_{n1}\beta\Phi^{-1}_{\beta\beta}\beta' e_{n1}} &= \begin{pmatrix} 1 & 0'_{r-1} \\ \beta_1 & \beta_2 \end{pmatrix} \Phi^{-1}_{\beta\beta} \begin{pmatrix} 1 \\ 0_{r-1} \end{pmatrix} / \Phi^{11} \\ &= \begin{pmatrix} 1 & 0'_{r-1} \\ \beta_1 & \beta_2 \end{pmatrix} \begin{pmatrix} 1 \\ -\Phi^{-1}_{22}\Phi_{21} \end{pmatrix} = \begin{pmatrix} 1 \\ \beta_1 - \beta_2\Phi^{-1}_{22}\Phi_{21} \end{pmatrix}, \end{aligned}$$

see (7).

The results in Theorem 2 can be interpreted as follows. For $h = 1$, the optimal hedging portfolio depends only on the conditional error variance is $\Sigma_{t,1} = Var_t(\Delta y_{t+1}) = E_t\varepsilon_{t+1}\varepsilon'_{t+1}$, and cointegration plays no role. The minimal conditional variance is $\Sigma_{t,1,11} - \Sigma_{t,1,12}\Sigma^{-1}_{t,1,22}\Sigma_{t,1,21} < \Sigma_{t,1,11}$, where $\Sigma_{t,1,11}$ is the risk of asset one.

For any h, the risk of the optimal portfolio is

$$(e'_{n1}\Sigma^{-1}_{t,h}e_{n1})^{-1} = (\Sigma^{11}_{t,h})^{-1} = \Sigma_{t,h,11} - \Sigma_{t,h,21}\Sigma^{-1}_{t,h,22}\Sigma_{t,h,21} < \Sigma_{t,h,11} = e'_{n1}\Sigma_{t,h}e_{n1}, \tag{21}$$

where $\Sigma_{t,h,11}$ is the risk of asset one, which diverges to infinity for large h, if the price of asset one is nonstationary, that is $e'_{n1}\beta_\perp \neq 0$, whereas the risk of the optimal portfolio stays bounded, if $e'_{n1}\beta \neq 0$, so a lot is gained by hedging. In this case $R^2_{t,h} \xrightarrow{P} 1$, see Juhl, Kawaller, and Koch (2012, p. 838), for a discussion of $R^2 > 0.8$ as a necessary condition to qualify for hedge accounting treatment. By the optimal hedging portfolio, the risk is reduced by $\Sigma_{t,h,12}\Sigma^{-1}_{t,h,22}\Sigma_{t,h,21} > 0$, see (21), and the conditional mean return is changed, but there is no simple comparison between the conditional mean returns for $h = 1$, and the limit for $h \to \infty$.

Note that the optimal portfolio converges to a cointegrating portfolio, that is, a linear combination of the columns of β, see (17). The result in (17) is invariant to the choice of normalization, because it depends only on $\beta(Var(\beta' y_t))^{-1}\beta'$, which is invariant under the transformation $\beta \to \beta\kappa$ for any full rank $r \times r$ matrix κ.

A different interpretation of the limit, as an optimal cointegrating portfolio of the form $\eta = \beta\psi$, where $\psi \in \mathbb{R}^r$, is given next.

Theorem 3 *Under Assumptions 1 and 2, and if $\beta' e_{n1} \neq 0$, the optimal cointegrating hedging portfolio, $\beta\psi^*_{t,h}$, and its limit are*

$$\beta\psi^*_{t,h} = \frac{\beta(\beta'\Sigma_{t,h}\beta)^{-1}\beta' e_{n1}}{e'_{n1}\beta(\beta'\Sigma_{t,h}\beta)^{-1}\beta' e_{n1}} \xrightarrow{P} \frac{\beta\Phi^{-1}_{\beta\beta}\beta' e_{n1}}{e'_{n1}\beta\Phi^{-1}_{\beta\beta}\beta' e_{n1}}, \textit{ for } h \to \infty. \tag{22}$$

Note that the limit is the same as in (17), and therefore the limits of the conditional expected return $\psi^{*\prime}_{t,h}\beta'\mu_{t,h}$ and conditional variance $\psi^{*\prime}_{t,h}\beta'\Sigma_{t,h}\beta\psi^*_{t,h}$ are given in (18) and (19).

The optimization of the squared h-period Sharpe ratio, see Sharpe (1966), defined by

$$S_h(\eta) = \frac{[E_t\{\eta'(y_{t+h} - y_t)\}]^2}{Var_t(\eta'(y_{t+h} - y_t))} = \frac{(\eta'\mu_{t,h})^2}{\eta'\Sigma_{t,h}\eta}, \tag{23}$$

is analysed in the next Theorem.

Theorem 4 *Under Assumptions 1 and 2, the portfolio which maximizes the Sharpe ratio after h periods, $\eta^\dagger_{t,h}$, and its limit are given, up to a constant factor, by*

$$\eta^\dagger_{t,h} = \Sigma^{-1}_{t,h}\mu_{t,h} \xrightarrow{P} -\beta\Phi^{-1}_{\beta\beta}(\beta' y_t - \xi). \tag{24}$$

The maximizing cointegrating portfolio and its limit are given up to a constant factor by

$$\beta\psi^\dagger_{t,h} = \beta(\beta'\Sigma_{t,h}\beta)^{-1}\beta'\mu_{t,h} \xrightarrow{P} -\beta\Phi^{-1}_{\beta\beta}(\beta' y_t - \xi). \tag{25}$$

3 Estimation results

The results above show how to determine optimal portfolios, if parameters and hence conditional mean and variance are known. In practice, one would have to estimate parameters and conditional mean and variance of returns. The parameters $\theta = (\alpha, \beta, \Gamma_1, \xi)$ and Ω can be estimated using the Gaussian quasi-likelihood, which assumes i.i.d. $N(0, \Omega)$ errors, that is by reduced rank regression, see Anderson (1951). In this case the asymptotic properties of $\hat{\theta}$, $\hat{\Omega}$, and rank test, under Assumptions 1 and 2, are the roughly same as for i.i.d. errors, see Theorem 5. This means in particular, that a consistent estimator of the limiting optimal cointegrating hedging portfolio, see (17), can be found. For the conditional variance $\Sigma_{t,h}$, only its expectation Σ_h, can be estimated. For constant conditional volatility, $\Sigma_{t,h}$ is the same as Σ_h, but this is not the case if there is heteroscedasticity.

By a regression of the returns $y_{1,t+h} - y_{1,t}$ on returns of the remaining assets and $(y_t, y_{t-1}, 1)$, it is shown that the corresponding estimator $\hat{\Sigma}_{t,h}^{reg}$ is consistent for Σ_h, whereas a regression of $y_{1,t}$ on the other assets and a constant gives a consistent estimator of the limiting optimal cointegrating portfolio.

3.1 Estimation of parameters and Σ_h

Asymptotic properties of the estimated parameters based on quasi-likelihood, which assumes Gaussian i.i.d. errors, have been analysed under Assumptions 1 and 2 by Cavaliere, Rahbek and Taylor (2010) and Boswijk, Cavaliere, Rahbek and Taylor (2016), and we formulate their results in the next theorem.

Theorem 5 *Let $y_t, t = 1, \ldots, T$, be generated by (1) and assume that Assumptions 1 and 2 hold. Consider the estimators $\hat{\theta} = (\hat{\alpha}, \hat{\beta}, \hat{\Gamma}_1, \hat{\xi})$ and $\hat{\Omega}$, derived from the Gaussian quasi-likelihood that assumes i.i.d. $N(0, \Omega)$ errors, that is,*

$$\varepsilon_t(\theta) = \Delta y_t - \alpha\beta' y_{t-1} - \Gamma_1 \Delta y_{t-1} + \alpha\xi, \tag{26}$$

$$-2 \log L(\theta, \Omega) = T \log \det \Omega + tr\{\Omega^{-1} \sum_{t=1}^{T} \varepsilon_t(\theta)\varepsilon_t(\theta)'\}. \tag{27}$$

Asymptotic properties of $T(\hat{\beta} - \beta), T^{1/2}(\hat{\xi} - \xi)$ and rank test, are the same as asymptotic properties, when errors are i.i.d. But $T^{1/2}(\hat{\alpha} - \alpha, \hat{\Gamma}_1 - \Gamma_1)$ is asymptotically Gaussian with a variance that involves the fourth moments of the error term.

Thus, Gaussian quasi-likelihood (27) can be used to estimate parameters consistently. In particular the optimal limiting cointegrating portfolio given in (17), whereas the conditional variance (13) contains $E_t \varepsilon_{t+h-1} \varepsilon_{t+h-1}'$, which cannot be estimated without modelling the error term explicitly, only its expectation Ω can be estimated.

Theorem 6 *Let $y_t, t = 1, \ldots, T$, be generated by (1) and let Assumptions 1 and 2 hold. Consider the process y_t conditional on initial values y_0, y_{-1}.*

Let $(y_{t+h} - y_t | 1, y_t, y_{t-1})$ be the residual of y_{t+h} regressed on $(y_t, y_{t-1}, 1)$, $t = 1, \ldots, T - h$. For fixed h and $T \to \infty$,

$$S_{h,T} = (T-h)^{-1} \sum_{t=1}^{T-h} (y_{t+h} - y_t | 1, y_t, y_{t-1})(y_{t+h} - y_t | 1, y_t, y_{t-1})' \xrightarrow{P} \Sigma_h. \tag{28}$$

If parameters are estimated by reduced rank regression, one obtains similarly for fixed (h, t) *and* $T \to \infty$,

$$\hat{\Sigma}_{h,t} \xrightarrow{\mathrm{P}} \Sigma_h. \tag{29}$$

Thus, estimating $\Sigma_{t,h}$ using the Gaussian quasi-likelihood for i.i.d. innovations, or running a regression of $y_{t+h} - y_t$ on $1, y_t, y_{t-1}$, implies in both cases that $\hat{\Sigma}_{t,h} \xrightarrow{P} \Sigma_h$ for $T \to \infty$. This has the implication that the optimal h-period portfolio, can only be estimated, if we use Σ_h as risk measure.

Next we prove that the limiting optimal h-period portfolio can be estimated by a regression of y_{1t} on $y_{2t}, \ldots, y_{nt}$, see Theorem 7

Theorem 7 *Let* $y_t, t = 1, \ldots, T$, *be generated by (1) and assume that Assumptions 1 and 2 hold and let initial values* y_0, y_{-1} *be fixed. Let* $(y_t|1) = y_t - \bar{y}_T$, $t = 1, \ldots, T$, *and*

$$S_T = T^{-1} \sum_{t=1}^{T} (y_t|1)(y_t|1)'. \tag{30}$$

Then

$$\beta' S_T \beta \xrightarrow{\mathrm{P}} Var(\beta' y_t), \text{ and } S_T^{-1} \xrightarrow{\mathrm{P}} \beta Var(\beta' y_t)^{-1} \beta', \; T \to \infty. \tag{31}$$

Hence, a regression of y_{1t} *on* $(y_{2t}, \ldots, y_{nt})$, *and a constant, for* $t = 1, \ldots, T$, *gives a consistent estimator of the limiting* $(h \to \infty)$ *optimal hedging portfolio (17).*

4 Empirical example

Consider the situation that a producer of electricity enters an agreement to deliver to customers two years from today one MWh of electricity every day of the year. Therefore she/he sells to the customers, today at the price p_t, the right to having delivered one MWh of electricity in two years, that is, a two year forward contract in electricity. The seller is worried about the risk due to changing fuel prices and decides to hedge these risks by buying two year futures in the price of fuels. The problem is which amounts, the hedge ratios, should be bought of the futures to hedge optimally, the risk due to the variation of fuel prices, as measured by conditional variance. Note that instead of holding the first asset, we are selling it and buying the hedging assets, but that is just a matter of a change of sign. A detailed analysis of some aspects of the electricity market in Europe, using cointegration analysis, can be found in Bosco, Parisio, Pelagatti, and Baldi (2010) and Mohammadi (2009).

Above a theory for this situation has been developed, under the assumption that a constant parameter model describes the data well, and for which we can assume that the model parameters remain fixed in the entire period. The model describes a cointegration relation between electricity and fuels. This theory is applied to a set of data, and it is shown how in this particular case, the optimal risk change with h.

We take Dutch electricity prices for trades for two year ahead forward contracts for electricity, p_t, and two year futures prices for $coal_t$, gas_t and CO_{2t} (CO_2 is the European Emission

Allowances for carbon dioxide) which are main determinants of the price of electricity, denoted fuels below. The data is from Datastream. The variables $y_t = (p_t, coal_t, gas_t, CO_{2t})'$ are modelled using a cointegration model with two lags of the form

$$\Delta y_t = \alpha(\beta' y_{t-1} - \xi) + \Gamma_1 \Delta y_{t-1} + \varepsilon_t, \tag{32}$$

and we estimate it, using the Gaussian likelihood assuming that ε_t, $t = 1, \ldots, T$, are independent identically distributed $N_4(0, \Omega)$. Note that in order to interpret a cointegrating relation as a portfolio, the prices, not the log prices are modelled. The analysis is summarized as follows.

Time series of the data are presented in Figure 1 and consists of daily observations for 2009. A CVAR with two lags is fitted to the data, and a few dummy variables are needed to account for outliers at observations $(10, 25, 55, 63, 117)$, using the software CATS in RATS, Dennis (2006).

A model with two lags is a reasonable description of the data, and we first test for the number of cointegrating relations. The test for rank is given in Table 2. The test that there is no cointegration, $r = 0$, has a $p-$value 0.028 and is rejected in favor of the hypothesis that $r = 1$, with $p-$value 0.534. This model is estimated and one finds, see Table 2, the three unit roots imposed and that the remaining are well within the unit circle. There is, however, a problem with conditional volatility, as is seen from the test for ARCH in Table 2.

	ARCH	p-value
$\Delta elec$	11.591	0.003
$coal$	14.923	0.001
gas	3.584	0.167
CO_2	9.880	0.007

Table 1: The test for ARCH effects for the individual variables indicate that there is conditional heteroscedasticity

There are now two possibilities, one is to model the conditional volatility with a multivariate GARCH model, like the BEKK model. The theory for this combination of CVAR and BEKK has yet to be worked out. Even the two-step procedure of first estimating the CVAR, using reduced rank regression, and then analysing the estimated residuals by BEKK model, is challenging. We have instead formulated general assumptions on the error term, to see how far one can get with the usual cointegration analysis based on reduced rank regression. It was a conclusion of the above analysis that if we are willing to use the expected conditional variance Σ_h as risk measure, we can estimate the optimal hedging portfolio.

The estimated cointegrating relation and the adjustment coefficients are

$$\begin{aligned}
\beta' y &= elec. - \underset{[t=-0.071]}{0.006}\ coal\ \underset{[t=-9.408]}{-1.143}\ gas - \underset{[t=-9.678]}{1.100}\ CO_2 - \underset{[t=-3.541]}{13.792} \\
\alpha' &= (\underset{[t=-4.495]}{-0.162}, \underset{[t=-1.244]}{-0.069}, \underset{[t=-0.970]}{-0.020}, \underset{[t=-0.483]}{-0.011}),
\end{aligned}$$

It is seen that the coefficient to $coal$ is not significant ($t = -0.071$), and that electricity is adjusting to the cointegrating relation with coefficient -0.162, ($t = -4.495$).

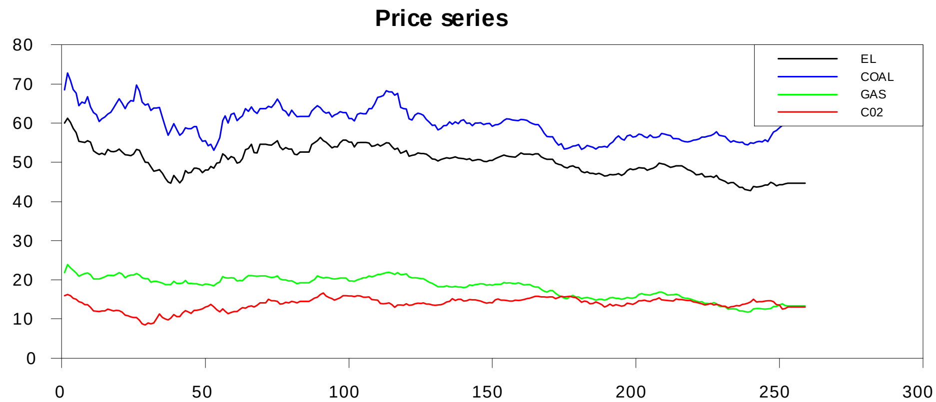

Figure 1: The daily prices of a two year forward contract for delivery of electricity and the prices of $coal, gas$ and CO_2 permits

Test for cointegrating rank				
r	Eig.Value	Trace	Frac95	P-Value
0	0.124	56.476	53.945	0.028
1	0.048	22.978	35.070	0.534
2	0.030	10.367	20.164	0.610
3	0.010	2.608	9.142	0.661
8 abs(roots) of companion matrix for $r = 1$				
$1, 1, 1, 0.864, 0.261, 0.068, 0.068, 0.049$				

Table 2: The tests for rank indicate that $r = 0$ can be rejected (p-value 0.028), and that $r = 1$ looks acceptable (p-value 0.534). The absolute value of the roots of the companion matrix consists of three imposed unit roots for $r = 1$, and the next largest is 0.864.

The estimated cointegrating relation is plotted in Figure 2, and the risk of the optimal portfolio compared to the stationary portfolio is given in Figure 3. Note that using the cointegrating relation as a hedging portfolio has a much greater risk than the optimal hedging portfolio. The unhedged, not shown, risk grows linearly from 0.35 ($h = 1$) to 13.73 ($h = 24$), whereas the optimally hedged risk grows from 0.13 ($h = 1$) and stays below the limit $\Phi = 0.987$.

5 Conclusion

The role of cointegration for hedging is analysed for an asset held for h periods. Prices are assumed generated by a cointegrated vector autoregressive model allowing for stationary and mixing martingale errors. The risk of a portfolio is measured by conditional variance of returns given information at time t. For nonstationary prices, the risk of keeping an asset for h periods diverges for large h. An expression is derived for the minimum variance hedging portfolio as a function of the holding period, h, and it is shown that it approaches a cointegrating vector for large h, thereby giving a bounded risk. Taking into account expected

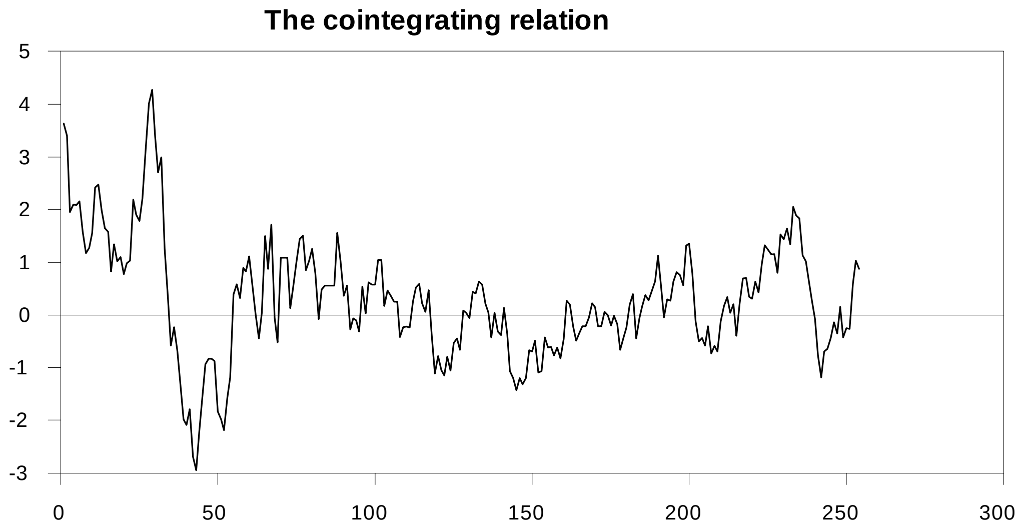

Figure 2: $\beta' y = elec. - \underset{[t=-0.071]}{0.006}\ coal\ \underset{[t=-9.408]}{-1.143}\ gas - \underset{[t=-9.678]}{1.100}\ CO2 - \underset{[t=-3.541]}{13.792}$

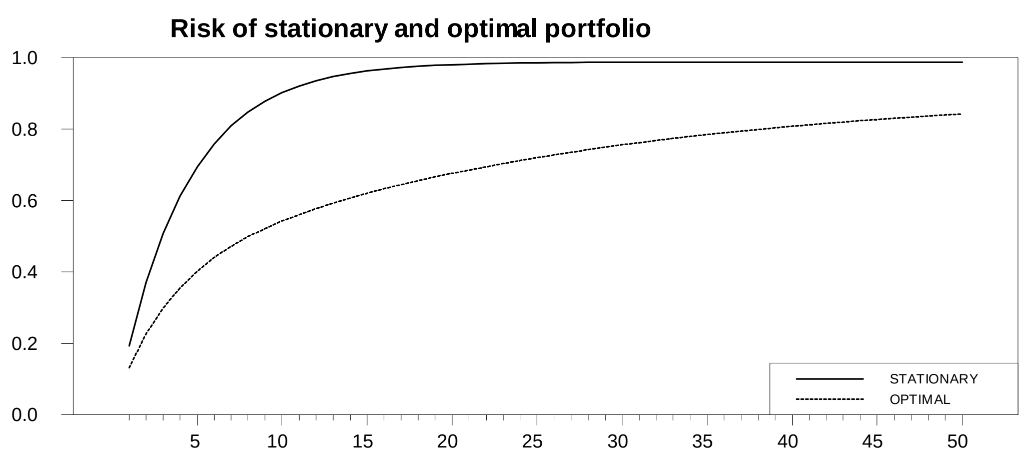

Figure 3: The risk of the cointegrating portfolio $Var_t(\beta' y_{t+h}) = \beta' \Sigma_h \beta$, (——) and the optimal risk $Var_t(\eta_h^{*\prime} y_{t+h}) = \eta_h^{*\prime} \Sigma_h \eta_h^*$, $(\cdots\cdots)$ are plotted. They both converge to $\Phi = 0.987$, with exponential an hyperbolic rates respectively. The unhedged risk (not plotted) for asset one is $\Sigma_{h11} \approx 0.35 + 0.58(h-1)$, which goes from 0.35 ($h = 1$) to 13.73 ($h = 24$).

return, the portfolio that maximizes the Sharpe ratio is derived. It is shown that this also approaches a cointegration portfolio, with weights depending on the price of the portfolio. These results are derived when parameters and hence the conditional variance is known.

If parameters have to be estimated, it is shown that, under Assumption 2, we can estimate the expected conditional variance, using either regression methods, or the reduced rank regression method of cointegration. The conditional variances, however, can only be estimated if heteroscedasticity is modelled. The results are illustrated with an analysis of a data set of prices of forward contracts on electricity prices, which are hedged by forward contracts on fuel prices. The main conclusion of the paper is that for optimal hedging, the cointegrating properties for long horizons should be exploited, but for short horizons more weight should be put on the remaining part of the dynamics.

References

Alexander, C., 1999, Optimal hedging using cointegration. *Philosophical Transactions: Mathematical Physical and Engineering Sciences*, Vol 357 No 1758 *Mathematics and Finance* 2039–2058.

Anderson, T.W., 1951, Estimating linear restrictions on regression coefficients for multivariate normal distributions, *Annals of Mathematical Statistics* 22, 327–351.

Bosco, B., Parisio, L., Pelagatti, M. and F. Baldi, 2010, Long-run relations in European electricity prices, *Journal of Applied Econometrics* 25, 805–832.

Boswijk, P. G., Cavalliere, A. Rahbek, and A.M.R. Taylor, 2016, Inference on co-integration parameters in heteroscedastic vector autoregressions. *Journal of Econometrics*, 192, 64-85.

Cavaliere, G., A.C. Rahbek and A.M.R. Taylor, 2010, Cointegration rank testing under conditional heteroskedasticity. *Econometric Theory* 26, 1719–1760.

Comte, F. and O. Lieberman, 2003, Asymptotic Theory for Multivariate GARCH Processes. *Journal of Multivariate Analysis* 84, 61-84.

de Jong, F. and P.C. Schotman, 2010, Price discovery in fragmented markets. *Journal of Financial Econometrics* Vol. 8, No. 1, 1–28.

Dennis, J. 2006, CATS in RATS, *Cointegration Analysis of Time Series,* Estima, Evanston.

Duan J-C., and S. Pliska, 2004, Option valuations with co-integrated asset prices. *Journal of Economics Dynamics & Control* 28 727-754.

Engle, R.F. and K.F. Kroner, 1995, Multivariate Simultaneous Generalized ARCH, *Econometric Theory* 11, 122-150.

Gatev, E., Goetzmann, W.N., and K.G. Rouwe, 2006, Pairs trading: performance of a relative-value arbitrage rule. *The Review of Financial Studies* 19(3) 797–827.

Grammig, J., J. M. Melvin, and C. Schlag, 2005, Internationally cross-listed stock prices during overlapping trading hours: price discovery and exchange rate effects. *Journal of Empirical Finance* 12 139–164.

Grinold, R., and R. Kahn, 1999, *Active Portfolio Management: A Quantitative Approach for Producing Superior Returns and Controlling Risk,* McGraw-Hill, New York.

Hasbrouck, J. 1995, One security, many markets: determining the contributions to price discovery. *The Journal of Finance* 50, 1175–1199.

Johansen, S. 1988, Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* 12, 231–254.

Johansen, S. 1996, *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press, Oxford (2. ed.)

Juhl, T., I.G., Kawaller, and P.D. Koch, 2012, The effect of the hedge horizon on optimal hedge size and effectivenes when prices are cointegrated. *The Journal of Futures Markets* 32(9), 837-876.

Lehmann, B.N. 2002, Some desiderata for the measurement of price discovery across markets. *Journal of Financial Markets* 5, 259-276.

Markowitz, H., 1952, Portfolio selection. *Journal of Finance*, 7(1), 77–90.

Mohammadi, H., 2009, Electricity prices and fuel costs: long-run relations and short-run dynamics, *Energy Economics*, 31, 503–509.

Nakajima L., and K. Ohashi, 2011, A cointegrated commodity pricing model. *The Journal of Futures Markets* Vol 32 11, 995-1033.

Sharpe, W.F., 1966, Mutual fund performance. *Journal of Business*, 39 (S1), 119–138.

W.K. Li, Shiqing Ling, and H. Wong, 2001, Estimation for partially nonstationary multivariate autoregressive models with conditional heteroscedasticity, *Biometrika*, 88, 1135-1152.

6 Appendix

The following elementary lemma is used for asymptotic analysis of $\Sigma_{t,h}^{-1}$.

Lemma 1 *Let for $m = 1, 2, \ldots,$ $\Theta_m > 0$ be a sequence of symmetric stochastic $k \times k$ matrices, and let γ be $k \times q$ of rank $q < k$.*

If

$$\gamma'\Theta_m\gamma \xrightarrow{P} \Phi > 0, \tag{33}$$

$$\gamma_\perp'\Theta_m\gamma_\perp \xrightarrow{D} \Xi > 0, \tag{34}$$

$$\gamma'\Theta_m\gamma_\perp = O_P(1), \tag{35}$$

then

$$\Theta_m^{-1} \xrightarrow{P} \gamma\Phi^{-1}\gamma', \text{ for } m \to \infty.$$

Proof of Lemma 1. Let $B_m = (\gamma, m^{-1/2}\gamma_\perp)$, then

$$\Theta_m^{-1} = B_m\{B_m'\Theta_m B_m\}^{-1}B_m' = (\gamma, m^{-1/2}\gamma_\perp)\begin{pmatrix} \gamma'\Theta_m\gamma & m^{-1/2}\gamma'\Theta_m\gamma_\perp \\ m^{-1/2}\gamma_\perp'\Theta_m\gamma & m^{-1}\gamma_\perp'\Theta_m\gamma_\perp \end{pmatrix}^{-1}(\gamma, m^{-1/2}\gamma_\perp)'$$

$$\xrightarrow{D} (\gamma, 0_{k\times(k-q)})\begin{pmatrix} \Phi & 0_{q\times(k-q)} \\ 0_{(k-q)\times q} & \Xi \end{pmatrix}^{-1}(\gamma, 0_{k\times(k-q)})' = \gamma\Phi^{-1}\gamma',$$

because of the assumptions (33), (33), and (33). ∎

Proof of Theorem 1. The representation (10) follows from

$$\beta' y_t = \rho\beta' y_{t-1} - \beta'\alpha\xi + \beta'\varepsilon_t$$

by backward elimination which gives y_{t+h} as function of y_t and later innovations

$$\alpha'_{\perp} y_{t+h} = \alpha'_{\perp} y_t + \alpha'_{\perp} \sum_{i=0}^{h-1} \varepsilon_{t+h-i},$$

$$\beta' y_{t+h} - \xi = \rho(\beta' y_{t+h-1} - \xi) + \beta' \varepsilon_{t+h} = \cdots = \rho^h(\beta' y_t - \xi) + \sum_{i=0}^{h-1} \rho^i \beta' \varepsilon_{t+h-i}.$$

These results are combined using the identity

$$I_n = \beta_{\perp}(\alpha'_{\perp}\beta_{\perp})^{-1}\alpha'_{\perp} + \alpha(\beta'\alpha)^{-1}\beta' = C + \alpha(\beta'\alpha)^{-1}\beta', \tag{36}$$

to give

$$\begin{aligned} y_{t+h} - y_t &= C y_{t+h} + \alpha(\beta'\alpha)^{-1}\beta' y_{t+h} - y_t \\ &= C\sum_{i=0}^{h-1} \varepsilon_{t+h-i} + C y_t - y_t + \alpha(\beta'\alpha)^{-1}\{\xi + \rho^h(\beta' y_t - \xi) + \sum_{i=0}^{h-1} \rho^i \beta' \varepsilon_{t+h-i}\}, \end{aligned} \tag{37}$$

which reduces to (11) using (36). We then find (12) and (13). It follows from Assumption 1 that the eigenvalues of ρ are less than one in absolute value, so that $\rho^h \to 0, h \to \infty$. Moreover, from (37) it follows that

$$\begin{aligned} \beta'\Sigma_{t,h}\beta &= \sum_{i=0}^{h-1} \rho^i \beta' E_t \varepsilon_{t+h-i}\varepsilon'_{t+h-i}\beta\rho'^i \xrightarrow{\mathrm{P}} \sum_{i=0}^{\infty} \rho^i \beta'\Omega\beta\rho'^i = Var(\beta' y_t), \\ h^{-1}\beta'_{\perp}\Sigma_{t,h}\beta_{\perp} &= \beta'_{\perp} C h^{-1} \sum_{i=0}^{h-1} E_t \varepsilon_{t+h-i}\varepsilon'_{t+h-i} C'\beta_{\perp} + O_P(1) \xrightarrow{\mathrm{P}} \beta'_{\perp} C\Omega C'\beta_{\perp}, \\ \beta'_{\perp}\Sigma_{t,h}\beta &= O_P(1). \end{aligned}$$

Note that the limits do not depend on t because of the mixing condition (4). Thus (15) holds and using Lemma 1, also (16) follows. The same proof can be applied to show the result for Σ_h. ■

Proof. *Proof of (17):* We introduce the companion form of the lag two model (1), which we formulate as a lag one model for the stacked process $\tilde{y}_t = (y'_t, y'_{t-1})'$ with errors $\tilde{\varepsilon}_t = (\varepsilon'_t, 0'_n)'$

$$\begin{pmatrix} \Delta y_t \\ \Delta y_{t-1} \end{pmatrix} = \begin{pmatrix} \alpha & \Gamma_1 \\ 0_{n\times r} & I_n \end{pmatrix} \left\{ \begin{pmatrix} \beta & I_n \\ 0_{n\times r} & -I_n \end{pmatrix}' \begin{pmatrix} y_{t-1} \\ y_{t-2} \end{pmatrix} - \begin{pmatrix} \xi \\ 0_n \end{pmatrix} \right\} + \begin{pmatrix} \varepsilon_t \\ 0_n \end{pmatrix},$$

or

$$\Delta\tilde{y}_t = \tilde{\alpha}(\tilde{\beta}'\tilde{y}_{t-1} - \tilde{\xi}) + \tilde{\varepsilon}_t.$$

where

$$\tilde{\alpha} = \begin{pmatrix} \alpha & \Gamma_1 \\ 0_{n\times r} & I_n \end{pmatrix}, \ \tilde{\beta} = \begin{pmatrix} \beta & I_n \\ 0_{n\times r} & -I_n \end{pmatrix}, \ \tilde{\xi} = \begin{pmatrix} \xi \\ 0_n \end{pmatrix}, \ \tilde{\Omega} = \begin{pmatrix} \Omega & 0_{n\times n} \\ 0_{n\times n} & 0_{n\times n} \end{pmatrix}.$$

Let $C = \beta_\perp(\alpha'_\perp(I_n - \Gamma_1)\beta_\perp)^{-1}\alpha'_\perp$, see (2) in Assumption 1, then the derived parameters are

$$\tilde{\alpha}_\perp = \begin{pmatrix} \alpha_\perp \\ -\Gamma'_1\alpha_\perp \end{pmatrix}, \; \tilde{\beta}_\perp = \begin{pmatrix} \beta_\perp \\ \beta_\perp \end{pmatrix}, \; \tilde{C} = \begin{pmatrix} C & -\Gamma_1 C \\ C & -\Gamma_1 C \end{pmatrix}, \; \tilde{\rho} = \begin{pmatrix} I_r + \beta'\alpha & \beta'\Gamma_1 \\ \alpha & \Gamma_1 \end{pmatrix}.$$

It follows that

$$(I_r, 0_{n\times n})\tilde{C}\sum_{i=0}^{h-1} E_t(\tilde{\varepsilon}_{t+h-i}\tilde{\varepsilon}'_{t+h-i})\tilde{C}'(I_n, 0_{n\times n})' = C\sum_{i=0}^{h-1} E_t(\varepsilon_{t+h-i}\varepsilon'_{t+h-i})C'. \tag{38}$$

The results (12) and (13) hold for the process $\tilde{y}_t$ by adding a tilde on all parameters,

$$\begin{aligned}
\tilde{\Sigma}_{t,h} &= \tilde{C}\sum_{i=0}^{h-1} E_t(\tilde{\varepsilon}_{t+h-i}\tilde{\varepsilon}'_{t+h-i})\tilde{C}' + \tilde{\alpha}(\tilde{\beta}'\tilde{\alpha})^{-1}(\sum_{i=0}^{h-1}\tilde{\rho}^i\tilde{\beta}'E_t(\tilde{\varepsilon}_{t+h-i}\tilde{\varepsilon}'_{t+h-i})\tilde{\beta}\tilde{\rho}'^i)(\tilde{\alpha}'\tilde{\beta})^{-1}\tilde{\alpha}' \qquad (39)\\
&\quad + \tilde{C}\sum_{i=0}^{h-1} E_t(\tilde{\varepsilon}_{t+h-i}\tilde{\varepsilon}'_{t+h-i})\tilde{\beta}\tilde{\rho}'^i(\tilde{\alpha}'\tilde{\beta})^{-1}\tilde{\alpha}' + \tilde{\alpha}(\tilde{\beta}'\tilde{\alpha})^{-1}\sum_{i=0}^{h-1}\tilde{\rho}^i\tilde{\beta}'E_t(\tilde{\varepsilon}_{t+h-i}\tilde{\varepsilon}'_{t+h-i})\tilde{C}',\\
\tilde{\mu}_{t,h} &= \tilde{\alpha}(\tilde{\beta}'\tilde{\alpha})^{-1}(\tilde{\rho}^h - I_{r+n})(\tilde{\beta}'\tilde{y}_t - \tilde{\xi}). \qquad (40)
\end{aligned}$$

The conditional mean and variance of $y_{t+h} - y_t$ are then $\mu_{t,h} = (I_n, 0_{n\times n})\tilde{\mu}_{t,h}$ and

$$\begin{aligned}
\Sigma_{t,h} &= (I_n, 0_{n\times n})\tilde{\Sigma}_{t,h}(I_n, 0_{n\times n})'\\
&= C\sum_{i=0}^{h-1} E_t\tilde{\varepsilon}_{t+h-i}\tilde{\varepsilon}'_{t+h-i}C' + (I_n, 0_{n\times n})\tilde{\alpha}(\tilde{\beta}'\tilde{\alpha})^{-1}(\sum_{i=0}^{h-1}\tilde{\rho}^i\tilde{\beta}'E_t\tilde{\varepsilon}_{t+h-i}\tilde{\varepsilon}'_{t+h-i}\tilde{\beta}\tilde{\rho}'^i)(\tilde{\alpha}'\tilde{\beta})^{-1}\tilde{\alpha}'(I_n, 0_{n\times n})'\\
&+ C\sum_{i=0}^{h-1} E_t\tilde{\varepsilon}_{t+h-i}\tilde{\varepsilon}'_{t+h-i}(I_n, 0_{n\times n})\tilde{\beta}\tilde{\rho}'^i(\tilde{\alpha}'\tilde{\beta})^{-1}\tilde{\alpha}' + \tilde{\alpha}(\tilde{\beta}'\tilde{\alpha})^{-1}\sum_{i=0}^{h-1}\tilde{\rho}^i\tilde{\beta}'(I_n, 0_{n\times n})'E_t\tilde{\varepsilon}_{t+h-i}\tilde{\varepsilon}'_{t+h-i}C'.
\end{aligned}$$

It is seen that conditions (33)–(35) from Lemma 1 are satisfied, and this implies that $\Sigma_{t,h}^{-1} \xrightarrow{P} \beta Var^{-1}(\beta' y_t)\beta'$, such that the optimal hedging portfolio, as given in (5), has limit

$$\eta^*_{t,h} = \Sigma_{t,h}^{-1}e_{n1}/e'_{n1}\Sigma_{t,h}^{-1}e_{n1} \xrightarrow{P} \beta\Phi_{\beta\beta}^{-1}\beta'e_{n1}/e'_{n1}\beta\Phi_{\beta\beta}^{-1}\beta'e_{n1}, \text{ for } h \to \infty.$$

The proof of (18), (19) and (20): It follows from (40), using $\beta'(I_n, 0_{n\times n}) = (I_r, 0_{r\times n})\tilde{\beta}'$, that

$$\begin{aligned}
\beta'\mu_{t,h} &= \beta'(I_n, 0_{n\times n})(\tilde{\beta}'\tilde{\alpha})^{-1}(\tilde{\rho}^h - I_{r+n})(\tilde{\beta}'\tilde{y}_t - \tilde{\xi})\\
&= (I_r, 0_{r\times n})(\tilde{\rho}^h - I_{r+n})(\tilde{\beta}'\tilde{y}_t - \tilde{\xi}) \xrightarrow{P} -(I_r, 0_{r\times n})(\tilde{\beta}'\tilde{y}_t - \tilde{\xi}) = -(\beta' y_t - \xi),
\end{aligned}$$

such that

$$\begin{aligned}
\eta^*_{t,h} = \Sigma_{t,h}^{-1}e_{n1}/e'_{n1}\Sigma_{t,h}^{-1}e_{n1} &\xrightarrow{P} -(e'_{n1}\beta\Phi_{\beta\beta}^{-1}\beta'e_{n1})^{-1}e'_{n1}\beta\Phi_{\beta\beta}^{-1}(\beta' y_t - \xi),\\
\eta^{*\prime}_{t,h}\Sigma_{t,h}^{-1}\eta^*_{t,h} &= (e'_{n1}\Sigma_{t,h}^{-1}e_{n1})^{-1} \xrightarrow{P} e'_{n1}\beta\Phi_{\beta\beta}^{-1}\beta'e_{n1}.
\end{aligned}$$

Finally, to prove (20), note that

$$e'_{n1}\Sigma_{t,h}^{-1}e_{n1} \xrightarrow{P} e'_{n1}\beta\Phi_{\beta\beta}^{-1}\beta'e_{n1} > 0 \text{ if } \beta'e_{n1} \neq 0,$$

and

$$e_{n1}'\Sigma_{t,h}e_{n1} = e_{n1}'C\sum_{i=0}^{h-1}E_t\varepsilon_{t+h-i}\varepsilon_{t+h-i}'C'e_{n1} + O_P(1) \xrightarrow{P} \infty, \text{ if } e_{n1}'\beta_\perp \neq 0.$$

Thus, if both conditions are satisfied,

$$1 - R_{t,n}^2 = \frac{(e_{n1}'\Sigma_{t,h}^{-1}e_{n1})^{-1}}{e_{n1}'\Sigma_{t,h}e_{n1}} \xrightarrow{P} 0.$$

■

Proof of Theorem 3. A cointegrating portfolio has the form $\beta\psi$ for some $\psi \in \mathbb{R}^r$. The conditional variance of $\psi'\beta' y_{t+h}$ is $\psi'\beta'\Sigma_{t,h}\beta\psi$, which under the constraint $e_{n1}'\beta\psi = 1$ is optimized by (5) for $a' = e_{n1}'\beta$, and we get

$$\psi_{t,h}^* = (\beta'\Sigma_{t,h}\beta)^{-1}\beta' e_{n1}/(e_{n1}'\beta(\beta'\Sigma_{t,h}\beta)^{-1}\beta' e_{n1})^{-1} \xrightarrow{P} \Phi_{\beta\beta}^{-1}\beta' e_{n1}/(e_{n1}'\beta\Phi_{\beta\beta}^{-1}\beta' e_{n1})^{-1},$$

such that the limit of $\beta\psi_{t,h}^*$ is (22). ■

Proof of Theorem 4. Maximizing the Sharpe ratio is equivalent to minimizing the variance $\eta'\Sigma_{t,h}\eta$ subject to the constraint $\eta'\mu_{t,h}$ equal to a constant, and the optimizing portfolio and its limit are therefore given by any portfolio proportional to

$$\eta_{t,h}^\dagger = \Sigma_{t,h}^{-1}\mu_{t,h} \xrightarrow{P} -\beta\Phi_{\beta\beta}^{-1}\beta'(I_n, 0_{n\times n})\tilde{\alpha}(\tilde{\beta}'\tilde{\alpha})^{-1}(\tilde{\beta}'\tilde{y}_t - \tilde{\xi}).$$

Using $\beta'(I_n, 0_{n\times n}) = (I_r, 0_{r\times n})\tilde{\beta}'$, the limit becomes

$$-\beta\Phi_{\beta\beta}^{-1}(I_r, 0_{r\times n})(\tilde{\beta}'\tilde{y}_t - \tilde{\xi}) = -\beta\Phi_{\beta\beta}^{-1}(\beta' y_t - \xi).$$

Restricting the portfolio to a cointegrating portfolio, $\eta = \beta\psi$, $\psi \in \mathbb{R}^r$, it holds that

$$\frac{(\eta'\mu_{t,h})^2}{\eta'\Sigma_{t,h}\eta} = \frac{(\psi'\beta'\mu_{t,h})^2}{\psi'\beta'\Sigma_{t,h}\beta\psi},$$

such that the optimal η is proportional to $\beta\psi^\dagger$

$$\beta\psi^\dagger = \beta(\beta'\Sigma_{t,h}\beta)^{-1}\beta'\mu_{t,h} \xrightarrow{P} -\beta\Phi_{\beta\beta}^{-1}\beta'(I_n, 0_{n\times n})\tilde{\alpha}(\tilde{\beta}'\tilde{\alpha})^{-1}(\tilde{\beta}'\tilde{y}_t - \tilde{\xi}) = -\beta\Phi_{\beta\beta}^{-1}(\beta' y_t - \xi).$$

■

Proof of Theorem 6. *Proof of (28)*: We find from representation (11), applied to $\tilde{y}_t$, multiplying by $(I_n, 0_{n\times n})$, that

$$\begin{aligned} y_{t+h} - y_t = {} & C\sum_{i=0}^{h-1}\varepsilon_{t+h-i} + (I_n, 0_{n\times n})\tilde{\alpha}(\tilde{\beta}'\tilde{\alpha})^{-1}\sum_{i=0}^{h-1}\tilde{\rho}^i\tilde{\beta}'\tilde{\varepsilon}_{t+h-i} \\ & + (I_n, 0_{n\times n})\tilde{\alpha}(\tilde{\beta}'\tilde{\alpha})^{-1}(\tilde{\rho}^h - I_r)(\tilde{\beta}'\tilde{y}_t - \tilde{\xi}) = z_t + u_t, \end{aligned} \tag{41}$$

say, where u_t is the last term and z_t the sum of the first two. It is seen that regressing on $(\tilde{y}_t', 1) = (y_t, y_{t-1}, 1)$, or equivalently on $(y_t - y_{t-1}, \beta' y_t, 1, \beta_\perp' y_t)$, eliminates u_t. Note that

z_t is uncorrelated with $\Delta y_t, \beta' y_t, \beta'_\perp y_t$, which only depend on $\varepsilon_t, \ldots, \varepsilon_1$, because ε_t form a martingale difference sequence. Therefore, correcting $y_{t+h} - y_t$ for the constant and the stationary processes $\Delta y_t, \beta' y_t$, has no effect asymptotically. Correcting the $I(0)$ process z_t for the $I(1)$ process $\beta'_\perp y_t$ has no effect asymptotically, and it follows by the law of large numbers applied to the ergodic process $z_t z_t'$, that

$$S_{hT} = T^{-1} \sum_{t=1}^{T} z_t z_t' + o_P(1) \xrightarrow{\text{P}} E(Var_t(y_{t+h} - y_t)) = \Sigma_h,$$

where Σ_h is given by (14).

Proof of (29): The CVAR parameters $\tilde{\alpha}, \tilde{\beta}, \tilde{\xi}, \tilde{\Omega}$ can be estimated consistently using the quasi-likelihood assuming i.i.d. errors, as was shown by Boswijk *et al.* (2016), see Theorem 5. Inserting these in (14) shows (29). ∎

Proof of Theorem 7. The representation

$$\begin{aligned} y_t - y_0 = C \sum_{i=0}^{t-1} \varepsilon_{t-i} + (I_n, 0_{n\times n}) \tilde{\alpha} (\tilde{\beta}' \tilde{\alpha})^{-1} (\sum_{i=0}^{t-1} \tilde{\rho}^i \tilde{\beta}' \tilde{\varepsilon}_{t-i}) \\ + (I_n, 0_{n\times n}) \tilde{\alpha} (\tilde{\beta}' \tilde{\alpha})^{-1} (\tilde{\rho}^t - I_r)(\tilde{\beta}' \tilde{y}_0 - \tilde{\xi}), \end{aligned} \tag{42}$$

follows from (41). Regressing on a constant, the last term vanishes for $T \to \infty$, and will be ignored, and for the second term correcting for the constant is asymptotically negligible. Thus the important part of the representation is

$$(y_t|1) = C(\sum_{i=0}^{t-1} \varepsilon_{t-i}|1) + (I_n, 0_{n\times n}) \tilde{\alpha} (\tilde{\beta}' \tilde{\alpha})^{-1} \sum_{i=0}^{t-1} \tilde{\rho}^i \tilde{\beta}' \tilde{\varepsilon}_{t-i}.$$

The conditions of Lemma 1 can now be checked. Using Boswijk *et al.* (2016), it is seen that

$$\begin{aligned} &T^{-1} \sum_{t=1}^{[Tu]} \varepsilon_t \xrightarrow{\text{D}} W(u), \\ &T^{-1} \sum_{t=1}^{T} \{\sum_{i=0}^{t-1} \tilde{\rho}^i \tilde{\beta}' \tilde{\varepsilon}_{t-i}\}\{\sum_{i=0}^{t-1} \tilde{\rho}^i \tilde{\beta}' \tilde{\varepsilon}_{t-i}\}' \xrightarrow{\text{P}} \sum_{i=0}^{\infty} \tilde{\rho}^i \tilde{\beta}' \tilde{\Omega} \tilde{\beta} \tilde{\rho}'^i. \end{aligned}$$

Here $W(u)$ is a Brownian motion with variance Ω. It follows, using $\beta'(I_n, 0_{n\times n}) = (I_r, 0_{r\times n})\tilde{\beta}'$, that

$$\begin{aligned} T^{-1}\beta'_\perp S_T \beta_\perp &\xrightarrow{\text{D}} \beta'_\perp C\{\int_0^1 (W(u)|1)(W(u)|1)' du\} C' \beta_\perp, \\ \beta' S_T \beta &= (I_r, 0_{r\times n}) T^{-1} \sum_{t=1}^{T} \{\sum_{i=0}^{t-1} \tilde{\rho}^i \tilde{\beta}' \tilde{\varepsilon}_{t-i}\}\{\sum_{i=0}^{t-1} \tilde{\rho}^i \tilde{\beta}' \tilde{\varepsilon}_{t-i}\}' (I_r, 0_{r\times n})' \\ &\xrightarrow{\text{P}} (I_r, 0_{r\times n})(\sum_{i=0}^{\infty} \tilde{\rho}^i \tilde{\beta}' \tilde{\Omega} \tilde{\beta} \tilde{\rho}'^i)(I_r, 0_{r\times n})' = Var(\beta' y_t). \end{aligned}$$

Finally, $\beta' S_T \beta_\perp = O_P(1)$, and applying Lemma 1 it follows that $S_T^{-1} \xrightarrow{\text{P}} \beta(Var(\beta' y_t))^{-1} \beta'$ for $T \to \infty$. ∎

Research Papers
2016

CREATES
Center for Research in Econometric Analysis of Time Series

2016-28: Kim Christensen, Roel Oomen and Roberto Renò: The Drift Burst Hypothesis

2016-29: Hossein Asgharian, Charlotte Christiansen, Rangan Gupta and Ai Jun Hou: Effects of Economic Policy Uncertainty Shocks on the Long-Run US-UK Stock Market Correlation

2016-30: Morten Ørregaard Nielsen and Sergei S. Shibaev: Forecasting daily political opinion polls using the fractionally cointegrated VAR model

2016-31: Carlos Vladimir Rodríguez-Caballero: Panel Data with Cross-Sectional Dependence Characterized by a Multi-Level Factor Structure

2016-32: Lasse Bork, Stig V. Møller and Thomas Q. Pedersen: A New Index of Housing Sentiment

2016-33: Joachim Lebovits and Mark Podolskij: Estimation of the global regularity of a multifractional Brownian motion

2017-01: Nektarios Aslanidis, Charlotte Christiansen and Andrea Cipollini: Predicting Bond Betas using Macro-Finance Variables

2017-02: Giuseppe Cavaliere, Morten Ørregaard Nielsen and Robert Taylor: Quasi-Maximum Likelihood Estimation and Bootstrap Inference in Fractional Time Series Models with Heteroskedasticity of Unknown Form

2017-03: Peter Exterkate and Oskar Knapik: A regime-switching stochastic volatility model for forecasting electricity prices

2017-04: Timo Teräsvirta: Sir Clive Granger's contributions to nonlinear time series and econometrics

2017-05: Matthew T. Holt and Timo Teräsvirta: Global Hemispheric Temperatures and Co–Shifting: A Vector Shifting–Mean Autoregressive Analysis

2017-06: Tobias Basse, Robinson Kruse and Christoph Wegener: The Walking Debt Crisis

2017-07: Oskar Knapik: Modeling and forecasting electricity price jumps in the Nord Pool power market

2017-08: Malene Kallestrup-Lamb and Carsten P.T. Rosenskjold: Insight into the Female Longevity Puzzle: Using Register Data to Analyse Mortality and Cause of Death Behaviour Across Socio-economic Groups

2017-09: Thomas Quistgaard Pedersen and Erik Christian Montes Schütte: Testing for Explosive Bubbles in the Presence of Autocorrelated Innovations

2017-10: Jeroen V.K. Rombouts, Lars Stentoft and Francesco Violante: Dynamics of Variance Risk Premia, Investors' Sentiment and Return Predictability

2017-11: Søren Johansen and Morten Nyboe Tabor: Cointegration between trends and their estimators in state space models and CVAR models

2017-12: Lukasz Gatarek and Søren Johansen: The role of cointegration for optimal hedging with heteroscedastic error term