Optimal hedging with the cointegrated vector autoregressive model allowing for heteroscedastic errors

Lukasz Gatarek[*,†]
Institute of Econometrics and Statistics,

Søren Johansen[‡,§]
University of Copenhagen
and CREATES

August 31, 2016

Abstract

We analyse the role of cointegration for hedging an asset using other assets, when the prices are generated by a cointegrated vector autoregressive model allowing for stationary martingale errors. We first note that if the price of the asset is nonstationary, the risk of keeping the asset diverges. We then derive the minimum variance hedging portfolio as a function of the holding period, h, and show that it approaches a cointegrating relation for large h, thereby giving a serious reduction in the risk. We then take into account the expected return and find the portfolio that maximizes the Sharpe ratio. We show that it also approaches a cointegration portfolio, with weights depending on the price of the portfolio. We illustrate the finding with a data set of electricity prices which are hedged by 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 remaining part of the dynamics.

We then analyse the situation with some heteroscedasticity, and find the same results provided one applies the average conditional variance of the return to measure the risk.

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.

[†]Chair of Econometric Models and Forecasts, Institute of Econometrics and Statistics, University of Lodz, ul. Rewolucji 1905, 90-214 Lodz, 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. E-mail: soren.johansen@econ.ku.dk. Correspondance: Department of Economics, University of Copenhagen, Øster Farimagsgade 5, building 26, 1353 Copenhagen K, Denmark.

[§]The authours would like to thank Katarina Juselius for expert advice on modelling the data.

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) studied optimal hedging using cointegration, see also Juhl, Kawaller, Koch (2012). 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.

Under the assumption of stationary martingale errors and constant volatility, we derive the optimal hedging ratios for horizon h, which can determined by a regression of h period returns $y_{t+h} - y_t$ on information at time t and a constant. By analyzing the limit of the relevant product moments, we discuss the role of cointegration for the optimal hedging portfolio, when prices are assumed to follow a cointegrated vector autoregressive model (CVAR) with conditionally heteroscedastic error terms. The role of cointegration for hedging was analysed by Juhl, Kawaller, and Koch (1011). They considered a special case of the CVAR and we want in this paper to generalize their results obtained to a CVAR with more lags and more cointegrating relations and allow for a some degree of heteroscedasticity in the martingale error term.

1.2 Content of the paper

We start with a simple example of a cointegrating regression model, which relates the hedged asset to the hedging assets via a cointegrating relation. In this example the hedging assets are assumed strongly exogenous and modelled by random walks to facilitate the derivations. We then turn to the general CVAR with an error term which is a stationary martingale difference sequence with constant volatility,. We find expressions for conditional mean and variance of the h period return, and use these to analyze the role of cointegration.

For a fixed horizon portfolio we note that because of the nonstationarity of prices, the risk will in general diverge as a function of the holding period. In the presence of cointegration, however, the optimal portfolio has a bounded risk. We find that for long horizons the optimal hedging portfolio approaches a cointegrating portfolio, whereas for shorter horizons the short-run dynamics and the error variance has to be taken into account.

When we allow for a more general nonstationary martingale difference error term with heteroscedasticity, the results are different, if we measure risk by conditional variance. The assumption that the average volatility converges implies that one can only derive similar results by using the average conditional variance as a measure of risk.

Finally we analyze some daily data for futures of electricity prices, and show how the risk of the optimal hedging ratios change with h and compare the optimal hedging portfolio with the cointegrating portfolio.

1.3 Main conclusion

Our main conclusion is that in the case of stationary martingale difference errors with constant volatility, both the optimal hedging portfolio and the maximum Sharpe ratio portfolio converge to cointegrating relations for large h, which we find explicitly and characterize as the minimum variance cointegrating portfolio normalized on $\eta_1 = 1$, and as the limit of the Sharpe optimal cointegrating portfolio respectively.

We conclude that cointegration plays an important role in hedging. It allows for the possibility that the 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. If we allow for some degree of volatility, the same results can not be proved, unless we replace the conditional variance of the returns by the conditional variance average over shorter horizons. All proofs are given in the Appendix.

2 A simple example of hedging cointegrated variables

This section analyses a simple model, where the hedged asset is cointegrated with the hedging assets, modelled as random walks. We compare the optimal hedging portfolio with the unhedged position in the first asset, and show how we find a substantial reduction in risk, due to the nonstationarity of the asset prices.

2.1 The cointegrating regression model

We first consider a simple model for the variables in the example in Section 5. This model is too simple to describe the data, which we analyze in Section 5, and is used here only because the derivations are simpler in this case. Thus, p_t is the price of a future on electricity and there are three "fuels", *coal*, *gas* and the price of CO_2 permits collected in y_{2t}. We consider a cointegrating regression model, where the endogenous variable $y_{1t} = p_t$ cointegrates with *coal*, *gas*, and CO_2, modelled as exogenous random walks,

$$\begin{aligned} y_{1t} &= \gamma' y_{2t} + u_{1t}, \\ y_{2t} &= y_{2,t-1} + u_{2t}, \end{aligned} \tag{1}$$

where $u_t = (u_{1t}, u_{2t}')'$ are here assumed independent identically distributed random errors with mean zero and variance split accordingly $\Psi = (\Psi_{ij}, i, j = 1, 2)$. We hold one unit of electricity, and want to hedge by going short in the fuels in the hope of reducing the risk associated with the prices.

For the regression model (1) is it easy to estimate the optimal h period portfolio, by regression of $y_{1,t+h} - y_t$ on $y_{2,t+h} - y_{2t}$, or based on the product moment for the stationary process $y_{t+h} - y_t$

$$\hat{\Sigma}_h = (T-h)^{-1} \sum_{t=1}^{T-h} (y_{t+h} - y_t)(y_{t+h} - y_t)'. \tag{2}$$

The idea of the paper is to find an expression for the expectation or probability limit of $\hat{\Sigma}_h$, in order to analyse the role of cointegration.

We find from model equations (1), that y_{2t} is a random walk in $n-1$ dimensions. This is used to find a representation of $y_{1,t+h}$ and $y_{2,t+h}$ as functions of y_{2t} and the errors

$$y_{2,t+h} - y_{2t} = u_{2,t+1} + \cdots + u_{2,t+h}, \tag{3}$$

$$y_{1,t+h} - y_{1t} = \gamma' y_{2,t+h} - y_{1t} + u_{1,t+h} = -(y_{1,t} - \gamma' y_{2t}) + \gamma' \sum_{i=1}^{h} u_{2,t+i} + u_{1,t+h}. \tag{4}$$

This representation is now used to find the conditional expected return and conditional variance of $y_{t+h} - y_t$ given y_t,

$$\mu_h = E_t(y_{t+h} - y_t) = \begin{pmatrix} -(y_{1t} - \gamma' y_{2t}) \\ 0 \end{pmatrix}, \tag{5}$$

$$\Sigma_h = V_t(y_{t+h} - y_t) = \begin{pmatrix} h\gamma'\Psi_{22}\gamma + \gamma'\Psi_{21} + \Psi_{12}\gamma + \Psi_{11} & h\gamma'\Psi_{22} + \Psi_{12} \\ h\Psi_{22}\gamma + \Psi_{21} & h\Psi_{22} \end{pmatrix}. \tag{6}$$

The best linear predictor of $y_{1,t+h}$ given $(y_{2,t+h}, y_t)$ is

$$\gamma_h^* = \Sigma_{h22}^{-1}\Sigma_{h21} = (h\Psi_{22})^{-1}(h\Psi_{22}\gamma + \Psi_{21}) = \gamma + h^{-1}\Psi_{22}^{-1}\Psi_{21}.$$

and the optimal hedging portfolio becomes

$$\eta_h^* = \begin{pmatrix} 1 \\ -\Sigma_{h22}^{-1}\Sigma_{h21} \end{pmatrix}. \tag{7}$$

The conditional expected return and risk are

$$\eta_h^{*\prime}\mu_h = \mu_{h1} - \Sigma_{h22}^{-1}\Sigma_{h21}\mu_{h2}, \tag{8}$$

$$\eta_h^{*\prime}\Sigma_h^{-1}\eta_h^* = \Sigma_{h11} - \Sigma_{h12}\Sigma_{h22}^{-1}\Sigma_{h21}. \tag{9}$$

In order to interpret the consequences of these results, note that holding the first asset for h periods leads to a diverging risk for $h \to \infty$,

$$Var_t(y_{1,t+h}) = \Psi_{11} + h\gamma'\Psi_{22}\gamma + \Psi_{12}\gamma + \gamma'\Psi_{21} \to \infty,$$

because y_{1t} is driven by the nonstationary random walk y_{2t}. If we use the optimal hedging portfolio, however, we find the increasing but converging risk

$$Var_t(\eta_h^{*\prime} y_{1,t+h}) = \Psi_{11} - h^{-1}\Psi_{12}\Psi_{22}^{-1}\Psi_{21} \to \Psi_{11}.$$

Thus for large h, one obtains a substantial reduction in risk by hedging.

The conditional expected return of holding the first asset is the same as the conditional expected return of the hedged asset, so in this case it is enough to compare the risks.

Two assets modelled by correlated random walks are substitutes. In the extreme case that two assets are fully correlated, having only one of them as hedging asset, is enough for an optimal portfolio. The expression for the optimal risk $\Psi_{11} - h^{-1}\Psi_{12}\Psi_{22}^{-1}\Psi_{21}$ shows that the more hedging assets are used, the smaller is the risk.

3 Optimal hedging in the CVAR with stationary martingale difference error terms

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

$$\Delta y_t = \alpha(\beta' y_{t-1} - \xi) + \Xi \Delta y_{t-1} + \varepsilon_t. \tag{10}$$

It is only a question of a more elaborate notation to handle the case of more lags using the companion form.

We formulate the assumptions on the data generating process, see Johansen (1996), and define the characteristic polynomial for the lag two model, $\Psi(z) = (1-z)I_n - \Pi z - \Xi z(1-z)$, see (10). For any $n \times m$ matrix, a, of rank $m < n$ we denote by $a_\perp$ an $n \times (n-m)$ matrix of rank $n - m$ for which $a' a_\perp = 0$.

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

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

Next we formulate the assumptions on the stationary error term, but discuss a weaker assumption in Section 4.

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

$$E_t(\varepsilon_{t+1}) = E(\varepsilon_{t+1}|\mathcal{F}_t) = 0, \qquad E|\varepsilon_t|^4 \le c < \infty, \tag{11}$$

and with constant volatility

$$E_t(\varepsilon_{t+1}\varepsilon'_{t+1}|\mathcal{F}_t) = \Omega > 0. \tag{12}$$

With this assumption we find that y_t is a nonstationary process, whereas Δy_t and $\beta' y_t$ are stationary. Moreover 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.

It is convenient to formulate the optimal hedging problem in the same way as Markowitz (1952) formulated the optimal portfolio choice, as a constrained optimization problem. We define $\mu_{t,h} = E_t(y_{t+h} - y_t)$ and $\Sigma_h = Var_t(y_{t+h})$ and want to minimize the conditional variance given information at time t of $\eta' y_{t+h}$, that is $\eta'\Sigma_h\eta$, 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 $= \mu_{t,h}$ we find the optimal portfolio in the sense of Markowitz, which also maximizes the Sharpe ratio.

The solution is easily found by solving the Lagrange multiplier problem

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

giving

$$\eta_{opt} = \Sigma_h^{-1} a / a' \Sigma_h^{-1} a, \tag{13}$$

with risk

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

For the case $a = e_{n1}$, we can find a different expression using

$$I_n = \Sigma_h \Sigma_h^{-1} = \begin{pmatrix} \Sigma_{h11} & \Sigma_{h12} \\ \Sigma_{h21} & \Sigma_{h22} \end{pmatrix} \begin{pmatrix} \Sigma_h^{11} & \Sigma_h^{12} \\ \Sigma_h^{21} & \Sigma_h^{22} \end{pmatrix},$$

where Σ_{h22} is $(n-1) \times (n-1)$, such that $\Sigma_{h21}\Sigma_h^{11} + \Sigma_{h22}\Sigma_h^{21} = 0$, or

$$\Sigma_{h22}^{-1}\Sigma_{h21} = -\Sigma_h^{21}/\Sigma_h^{11}.$$

In this case we denote the optimal portfolio η_h^*, and find

$$\eta_h^* = \Sigma_h^{-1} e_{n1} / e_{n1}' \Sigma_h^{-1} e_{n1} = \begin{pmatrix} \Sigma_h^{11} \\ \Sigma_h^{21} \end{pmatrix} / \Sigma_h^{11} = \begin{pmatrix} 1 \\ \Sigma_h^{21}/\Sigma_h^{11} \end{pmatrix} = \begin{pmatrix} 1 \\ -\Sigma_{h22}^{-1}\Sigma_{h21} \end{pmatrix}. \tag{14}$$

Thus, the optimal choice is found as a population regression of $y_{1,t+h} - y_{1,t}$ on $(y_{2,t+h}, \ldots, y_{n,t+h})$ correcting for information at time t and a constant. It turns out that the formulation (13) is more convenient for the asymptotic analysis, which reduces to finding the limit of Σ_h^{-1} for $h \to \infty$.

We shall use the following elementary lemma for the asymptotic analysis.

Lemma 1 *Let $\Theta_h = h\beta_\perp \Phi_h \beta_\perp' + \Psi_h$ for Θ_h and Φ_h positive definite symmetric possibly stochastic matrices, and assume that $\Psi_h = O_P(1)$, $\beta' \Psi_h \beta \xrightarrow{P} \Gamma > 0$, and $\Phi_h \xrightarrow{D} \Phi > 0$ a.s., then*

$$\Theta_h^{-1} \xrightarrow{P} \beta \Gamma^{-1} \beta'.$$

We first find a representation of the process for the lag one model, and use that to calculate the conditional variance, Σ_h, and conditional mean return, $\mu_{t,h}$, and their limits.

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

$$\Delta y_t = \alpha(\beta' y_{t-1} - \xi) + \varepsilon_t. \tag{15}$$

The process $y_{t+h} - y_t$ has the representation as function of errors, $\varepsilon_{t+1}, \ldots, \varepsilon_{t+h}$, and initial values, y_t,

$$y_{t+h} - y_t = \sum_{i=0}^{h-1} \{C + \alpha(\beta'\alpha)^{-1}\rho^i\beta'\}\varepsilon_{t+h-i} + \alpha(\beta'\alpha)^{-1}(\rho^h - I_r)(\beta' y_t - \xi). \tag{16}$$

The conditional mean and variance are therefore

$$\mu_{t,h} = E_t(y_{t+h} - y_t) = \alpha(\beta'\alpha)^{-1}(\rho^h - I_r)(\beta' y_t - \xi), \tag{17}$$

$$\Sigma_h = Var_t(y_{t+h} - y_t) = \sum_{i=0}^{h-1} [C + \alpha(\beta'\alpha)^{-1}\rho^i\beta']\Omega[C' + \beta\rho'^i(\alpha'\beta)^{-1}\alpha']. \tag{18}$$

Finally for $h \to \infty$

$$\mu_{t,h} \to -\alpha(\beta'\alpha)^{-1}(\beta' y_t - \xi), \tag{19}$$
$$\Sigma_h^{-1} \to \beta(Var(\beta' y_t))^{-1}\beta'. \tag{20}$$

In order to formulate the main result it is convinient to normalize the cointegrating vectors. Because we are interested in hedging the first asset and investigate the influence of cointegration, we assume that there exists a cointegrating relation of the form $y_{1t} + \gamma_1' y_{2t}$. By taking linear combinations of the cointegrating relations, we can eliminate the first asset from the remaining relations and assume, without loss of generality, that

$$\beta = \begin{pmatrix} 1 & 0 \\ \gamma_1 & \gamma_2 \end{pmatrix}, \tag{21}$$

for $\gamma_1 \in \mathbb{R}^{n-1}$ and $\gamma_2 \in \mathbb{R}^{(n-1)\times(r-1)}$.

We next formulate the main result for the hedging problem in the CVAR with stationary martingale difference sequence as error term and two lags.

Theorem 2 *Let y_t be given by the model (10) and let Assumption 1 and 2 hold.*

For $h = 1$, we find $\Sigma_1 = E_t\varepsilon_{t+1}\varepsilon_{t+1}' = \Omega$, $\mu_{t,1} = \alpha(\beta' y_t - \xi) + \Xi\Delta y_t$ and the optimal hedging portfolio is $\eta_1^{*\prime} = (1, -\Omega_{12}\Omega_{22}^{-1})$, which has conditional mean return and risk

$$\eta_1^{*\prime}\mu_{t,1} = (1, -\Omega_{12}\Omega_{22}^{-1})(\alpha(\beta' y_t - \xi) + \Xi\Delta y_t), \tag{22}$$
$$\eta_1^{*\prime}\Sigma_1\eta_1^* = \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}. \tag{23}$$

For $h \to \infty$, the limiting optimal hedging portfolio is

$$\eta_h^* = \frac{\Sigma_h^{-1} e_{n1}}{e_{n1}'\Sigma_h^{-1} e_{n1}} \to \frac{\beta\Gamma^{-1}\beta' e_{n1}}{e_{n1}'\beta\Gamma^{-1}\beta' e_{n1}} = \beta\begin{pmatrix} 1 \\ -\Gamma_{22}^{-1}\Gamma_{21} \end{pmatrix} = \begin{pmatrix} 1 \\ \gamma_1 - \gamma_2\Gamma_{22}^{-1}\Gamma_{21} \end{pmatrix}, \tag{24}$$

where $\Gamma = Var(\beta' y_t)$. The limits for $h \to \infty$ of the conditional mean return and risk are

$$\eta_h^{*\prime}\mu_{t,h} \to -(1, -\Gamma_{12}\Gamma_{22}^{-1})(\beta' y_t - \xi), \tag{25}$$
$$\eta_h^{*\prime}\Sigma_h\eta_h^* \to \Gamma_{11}^{-1} = \Gamma_{11} - \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{21}. \tag{26}$$

Finally, the fraction of explained variation is given by

$$R_h^2 = \Sigma_{11h}^{-1}\Sigma_{12h}\Sigma_{22h}^{-1}\Sigma_{21h} \to 1. \tag{27}$$

The interpretation of these results is the following. For $h = 1$, the optimal portfolio depends only on the conditional error variance $E_t\varepsilon_{t+1}\varepsilon_{t+1}' = \Omega$, assumed constant, and cointegration plays no role. The minimal variance is $\Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21} < \Omega_{11}$, where the latter is the variance of the unhedged asset.

For any h the risk of the optimal portfolio is

$$(\Sigma_h^{11})^{-1} = \Sigma_{h11} - \Sigma_{h21}\Sigma_{h22}^{-1}\Sigma_{h21} < \Sigma_{h11}, \tag{28}$$

where Σ_{h11} is the risk of the unhedged portfolio, which diverges to infinity for large h, if the price of asset one is nonstationary, whereas the risk of the optimal portfolio stays bounded, so a lot is gained by hedging. We also see that $R_h^2 \longrightarrow 1$, 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 $\Sigma_{h12}\Sigma_{h22}^{-1}\Sigma_{h21} > 0$, see (28), and the mean return is changed, but there is no simple comparison between the mean returns for $h = 1$ and the limit for $h \to \infty$. We note that the limit of the optimal portfolio is a cointegrating portfolio, that is, a linear combination of the columns of β, see (24). We next analyse cointegrating portfolios.

Theorem 3 *Under Assumptions 1 and 2, the optimal cointegrating hedging portfolio and its limit are*

$$\eta^*_{h,coint} = \frac{\beta(\beta'\Sigma_h\beta)^{-1}\beta' e_{r1}}{e'_{r1}\beta(\beta'\Sigma_h\beta)^{-1}\beta' e_{r1}} \to \beta \begin{pmatrix} 1 \\ -\Gamma_{22}^{-1}\Gamma_{21} \end{pmatrix}. \tag{29}$$

Note that the limit is the same as in (24) and therefore the limits of the conditional return $\eta^{*\prime}_{h,coint}\mu_{t,h}$ and conditional variance $\eta^{*\prime}_{h,coint}\Sigma_h\eta^{*\prime}_{h,coint}$ are given in (25) and (26).

Note that with the parametrization (21), the parameter γ_1 is not identified, because we could choose the parameters,

$$\beta_\kappa = \beta\kappa = \begin{pmatrix} 1 & 0 \\ \gamma_1 & \gamma_2 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ \kappa_1 & \kappa_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \gamma_1 + \gamma_2\kappa_1 & \gamma_2\kappa_2 \end{pmatrix},$$

and $\alpha_\kappa = \alpha\kappa'^{-1}$ where κ is $r \times r$ of full rank. Then $\alpha\beta' = \alpha_\kappa\beta'_\kappa$, and that would not change the cointegrating space and therefore not the model (15).

The result in (24), however, is invariant to this choice of parametrization, because it depends only on $\beta(Var(\beta' y_t))^{-1}\beta'$, which is invariant under the transformation $\beta \to \beta\kappa$ for nay full rank matrix κ.

3.1 Optimizing the Sharpe ratio for the CVAR

We define the (squared) Sharpe ratio after h periods as

$$S_h(\eta) = \frac{[E_t\{\eta'(y_{t+h} - y_t)\}]^2}{Var_t(\eta'(y_{t+h} - y_t))} = \frac{(\eta'\mu_{t,h})^2}{\eta'\Sigma_h\eta}. \tag{30}$$

Theorem 4 *Under Assumptions 1 and 2, the portfolio which maximizes the Sharpe ratio after h periods and its limit are given, up to a constant factor, by*

$$\breve{\eta}_h = \Sigma_h^{-1}\mu_{t,h} \to -\beta\Gamma^{-1}(\beta' y_t - \xi). \tag{31}$$

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

$$\breve{\eta}_{h,coint} = \beta(\beta'\Sigma_h\beta)^{-1}\beta'\mu_{t,h} \to -\beta\Gamma^{-1}(\beta' y_t - \xi). \tag{32}$$

3.2 Regression results

The analysis above shows how to determine the optimal portfolios if the parameters and hence the conditional mean and variance are known. In practice one would have to estimate the parameters. This can be done by either using the Gaussian maximum likelihood estimators of the CVAR or by simply running suitable regressions. We next give the asymptotic properties of the product moments involved in these regressions.

Theorem 5 *Let $y_t, t = 1, \ldots, T$, be generated by (10) and assume that Assumptions 1 and 2 hold.*

Let $(y_{t+h} - y_t|1, y_t, y_{t-1})$ be the residual of a regression of $y_{t+h} - y_t$ on $(1, y_t, y_{t-1})$, then

$$S_{hT} = (T-h)^{-1} \sum_{t=1}^{T-h} (y_{t+h} - y_t|1, y_t, y_{t-1})(y_{t+h} - y_t|1, y_t, y_{t-1})' \overset{a.s.}{\to} \Sigma_h, \ T \to \infty. \tag{33}$$

Hence $S_{hT}^{-1} e_{n1}/e_{n1}' S_{hT}^{-1} e_{n1}$ is a consistent estimator of the optimal h period hedging portfolio.

Let similarly $(y_t|1) = y_t - \bar{y}_T$, and

$$S_T = T^{-1} \sum_{t=1}^{T} (y_t|1)(y_t|1)', \tag{34}$$

then

$$\beta' S_T \beta \overset{a.s.}{\to} \Gamma, \ \text{and} \ S_T^{-1} \overset{P}{\to} \beta \Gamma^{-1} \beta'. \tag{35}$$

Hence $S_T^{-1} e_{n1}/e_{n1}' S_T^{-1} e_{n1}$ is a consistent estimator of the limiting $(h \to \infty)$ optimal hedging portfolio.

4 Optimal hedging in the CVAR with martingale difference heteroscedastic errors

We now formulate a more general set of assumptions for the error term which allows for heteroscedasticity, and replace Assumption 2 by the weaker Assumption 3.

Assumption 3 *The innovations, ε_t, form a martingale difference sequence with respect to a filtration $\mathcal{F}_t$,$t = \ldots, -1, 0, 1, \ldots,$ and satisfy*

$$E_t(\varepsilon_{t+1}) = 0, \qquad E|\varepsilon_t|^4 \leq c < \infty, \tag{36}$$

and for $\Omega_t = E_{t-1}\varepsilon_t\varepsilon_t'$ we assume

$$T^{-1} \sum_{t=1}^{T} \Omega_{t+1} \overset{a.s.}{\to} \Omega > 0. \tag{37}$$

Thus the volatility Ω_t is not constant, but the time verage converges almost surely. Under Assumption 3, y_t is nonstationary, but both Δy_t and $\beta' y_t$ are nonstationary too, due to the variation of the volatility Ω_t, and $Var_t(\beta' y_{t+h})$ does not necessarily converge for $h \to \infty$. Thus the role of cointegration is not so simple for heteroscedastic errors.

We first note that under Assumption 3 we can define the general linear process $z_t = \sum_{i=0}^{\infty} \phi_i' \varepsilon_{t-i}$, for coefficients which satisfy $\sum_{i=0}^{\infty} |\phi_i| < \infty$. This has autocovariance function

$$\gamma_j = Cov(z_t, z_{t+j}) = E\sum_{i=0}^{\infty} \phi_i' \Omega_{t-i} \phi_{i+j},$$

and by averaging for $1 \leq j \leq h$ we get

$$\bar{\gamma}_h = h^{-1} \sum_{j=1}^{h} Cov(z_t, z_{t+j}) \to \sum_{i=0}^{\infty} \phi_i' \Omega \phi_{i+j}, \text{ for } h \to \infty. \tag{38}$$

Thus Ω from (37) does not in general play a role in the autocovariance function, only in the limit of the average autocovariance function.

These processes are also studied by Hannan and Heyde (1972, Theorem 1), who proved that if Assumption 3 holds[1] and $\sum_{i=0}^{\infty} |\phi_i'| < \infty$, then for $j \geq 0$,

$$\bar{z}_T = T^{-1} \sum_{t=1}^{T} z_t \overset{a.s.}{\to} 0, \qquad \hat{\gamma}_j = T^{-1} \sum_{t=1}^{T} (z_t - \bar{z}_T)(z_{t+j} - \bar{z}_T)' \overset{P}{\to} \sum_{i=0}^{\infty} \phi_i' \Omega \phi_{i+j} = \lim_{h\to\infty} \bar{\gamma}_h. \tag{39}$$

Thus, the emprical autocovariance function of z_t does not converge to the theoretical autocovariance function, but to a limit of an average of the theoretical autocovariance function, see (38).

Therefore the role of cointegration is not the same as for homogeneous conditional variances. To discuss this we define

$$\mu_{t,h} = E_t(y_{t+h} - y_t), \qquad \Sigma_{t,h} = Var_t(y_{t+h} - y_t), \qquad \overline{\Sigma}_{t,h} = h^{-1} \sum_{j=1}^{h} \Sigma_{t,j}.$$

Theorem 6 *Let y_t be given by model (10) and let Assumption 1 and 3 hold.*
Then for some $\bar{\Gamma} > 0$,

$$\beta' \overline{\Sigma}_{t,h} \beta \overset{P}{\to} \beta \overline{\Gamma} \beta' \text{ and } \overline{\Sigma}_{t,h}^{-1} \overset{P}{\to} \beta \overline{\Gamma}^{-1} \beta'. \tag{40}$$

The portfolio minimizing $\eta' \overline{\Sigma}_{t,h} \eta$ under the constraint $\eta' e_1 = 1$ and its limit are given by

$$\overline{\Sigma}_{t,h}^{-1} e_1 / e_1' \overline{\Sigma}_{t,h}^{-1} e_1 \overset{P}{\to} \beta \overline{\Gamma} \beta' e_1 / e_1' \beta \overline{\Gamma} \beta' e_1, \text{ for } h \to \infty. \tag{41}$$

Thus, if we use the average conditional variance, $\overline{\Sigma}_{t,h}$, to measure risk, instead of the conditional variance itself, $\Sigma_{t,h}$, the results of Theorem 2 hold.

We next give an analysis of the regression approach outlined in Section 3.2 and prove the analogue of Theorem 5.

---

[1] The condition (6) of Hannan and Heide (1972) assumes that there is a random variable X with $E|X|^2 < \infty$, such that $P(|\varepsilon_t| \geq u) \leq cP(|X| \geq u)$ for all $t, u \geq 0$ and some c. This is implied by our condition $E(|\varepsilon_t|^4) \leq c < \infty$, using a random variable with distribution function $F(x) = (1 - c/x^4)^+$.

Theorem 7 *Let $y_t, t = 1, \ldots, T$, be generated by (10) and let Assumptions 1 and 3 hold. Let $(y_{t+h} - y_t | 1, y_t, y_{t-1})$ be the residual of a regression of $y_{t+h} - y_t$ on $(1, y_t, y_{t-1})$, then the residual sum of squares, see (33), satisfies*

$$S_{hT} \overset{P}{\to} \overline{\Sigma}_{t,h}, \ T \to \infty. \tag{42}$$

Hence the portfolio $S_{hT}^{-1} e_{n1} / e_{n1}' S_{hT}^{-1} e_{n1}$ is a consistent estimator of the optimal portfolio if we use the average variance, $\overline{\Sigma}_{t,h}$, as risk measure.

Let S_T be the residual sum of squares of a regression of y_t on a constant, see (34), then

$$\beta' S_T \beta \overset{P}{\to} \overline{\Gamma} \text{ and } S_T^{-1} \overset{P}{\to} \beta \overline{\Gamma}^{-1} \beta'. \tag{43}$$

Hence $S_T^{-1} e_{n1} / e_{n1}' S_T^{-1} e_{n1}$ is a consistent estimator of the limiting $(h \to \infty)$ optimal hedging portfolio, if we use the average variance $\overline{\Sigma}_{t,h}$ as risk measure.

5 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, in the sense of smallest variance, the risk due to the variation of fuel prices. 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 we have developed a theory for this situation under the assumption that we have a constant parameter model, which describes the data well and for which we can assume that the model parameters remain fixed in the entire period. The model describes the cointegration relation between electricity and the fuels. We now want to apply this theory to a set of data, and show how in this particular case, the optimal hedge ratios and its 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. We model these variables $y_t = (p_t, coal_t, gas_t, CO_{2t})'$ using a cointegration model with two lags of the form

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

where ε_t, $t = 1, \ldots, T$, are independent identically distributed $(0, \Omega)$. Note that in order to interpret a cointegrating relation as a portfolio, we model the prices, not the log prices. We summarize the analysis as follows.

The time series of the data are presented in Figure 1 and consists of daily observations for 2009. We fit a CVAR with two lags and we need a few dummy variables to account for outliers at observations $(62, 24, 54, 116)$

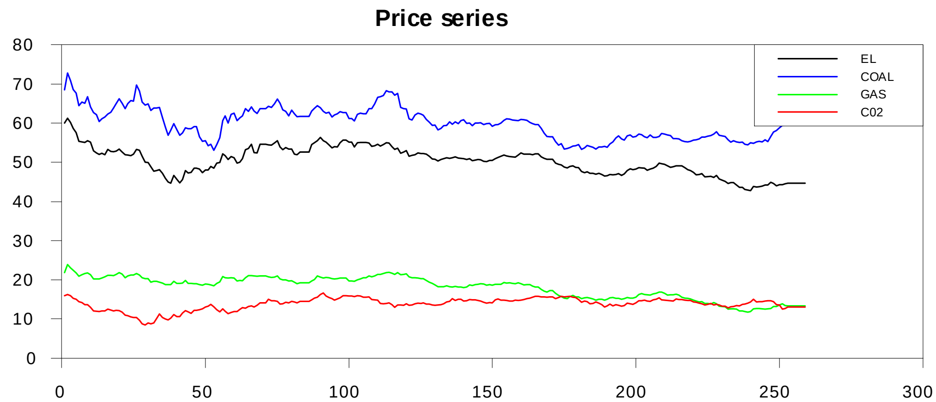

Figure 1: The daily prices of a two year forward contract for delivery of electricity and the prices of *coal*, *gas* and *CO2* permits

Test for cointegrating rank				
r	Eig.Value	Trace	Frac95	P-Value
0	0.136	62.475	53.945	0.006
1	0.055	24.899	35.070	0.411
2	0.028	10.473	20.164	0.600
3	0.012	3.153	9.142	0.562
8 abs(roots) of companion matrix for $r = 1$				
$1, 1, 1, 0.87, 0.25, 0.069, 0.066, 0.066$				

Table 1: The tests for rank indicate that $r = 0$ can be rejected (p-value 0.006), and that $r = 1$ looks acceptable (p-value 0.411). The absolute value of the roots of the companion matrix are three imposed unit roots for $r = 1$ and the next largest is 0.72.

We estimate the model using the Gaussian maximum likelihood procedure, Johansen (1988), and the calculations are performed using the software CATS in RATS, Dennis (2006).

We find that 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 1 together with the absolute value of the roots of the companion matrix when $r = 1$. One finds therefore three unit roots, and the remaining roots are well within the unit disc.

The cointegrating relation is given below together with the adjustment coefficients. It is seen that the coefficient to *coal* and the constant term are not significant, and that all variables are adjusting to the cointegrating relations except the price of CO2 permits with a t-value of -0.852.

$$\beta' y = elec. - \underset{[t=-0.434]}{0.033}\ coal \underset{[t=-5.306]}{-0.987}\ gas - \underset{[t=-8.158]}{1.145}\ CO2 + \underset{[t=0.691]}{0.004}$$
$$\alpha' = (\underset{[t=-5.147]}{-0.194}, \underset{[t=-2.100]}{-0.117}, \underset{[t=-2.399]}{-0.048}, \underset{[t=-0.852]}{-0.020}),$$

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

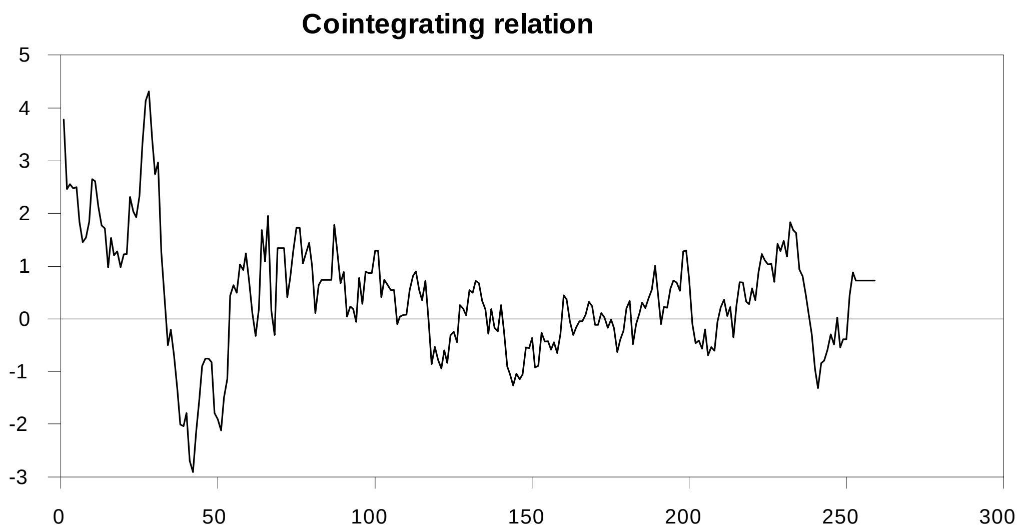

Figure 2: $\beta' y = elec. - \underset{[t=-0.434]}{0.033}\ coal\ \underset{[t=-5.306]}{-0.987}\ gas - \underset{[t=-8.158]}{1.145}\ CO2 + \underset{[t=0.691]}{0.004}$

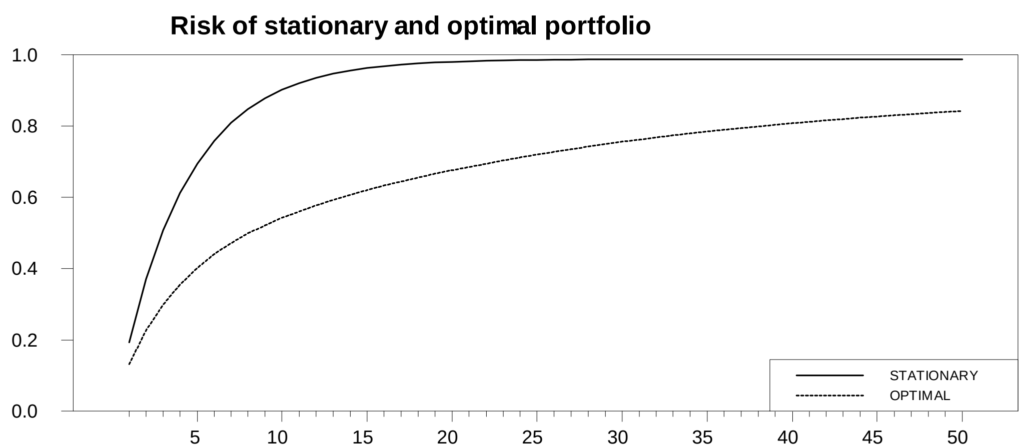

Figure 3: We plot the risk of the stationary portfolio $Var_t(\beta' y_{t+h}) = \beta' \Sigma_h \beta$, (——) which converges to $\Gamma = 0.987$ with an exponential rate, and the optimal risk $Var_t(\eta_h^{*\prime} y_{t+h}) = \eta_h^{*\prime} \Sigma_h \eta_h^*$, $(\cdots\cdots)$ which converges to Γ like h^{-1}. The unhedged risk (not plotted) for asset one is $\Sigma_{h11} \approx 0.35 + 0.58(h-1)$, which goes from 0.35 to 13.73 for $h = 24$.

cointegrating relation as a hedging portfolio has a much greater risk than the optimal hedging portfolio. The unhedged risk grows linearly from 0.35 $(h = 1)$ to 13.73 $(h = 24)$, whereas the optimally hedged risk grows from 0.13 $(h = 1)$ but stays below the limit $\Gamma = 0.987$.

6 Conclusion

We have analyzed the role of cointegration for hedging under the assumption that asset prices are driven by a CVAR with stationary martingale difference errors with constant volatility.

We have found the optimal hedging portfolio and maximum Sharpe ratio portfolio and compared them with the unhedged portfolio for horizon h.

We find that, due to the nonstationarity of the asset prices, there is a substantial gain in risk by hedging, especially for longer horizons. There is no simple comparison between the expected return of the hedged and unhedged portfolio. Thus the main advantage of hedging is the reduction of the risk. The minimum variance optimal hedging portfolio does not take into account the expected return, and we therefore also analyze the maximum Sharpe ratio portfolio, which balances the expected return and risk.

For long horizons, the optimal portfolio in both cases approaches a cointegrating relation, which we find explicitly together with a formula for the expected return and risk.

If we allow for some degree of volatility, the same results can not be proved, unless we replace the conditional variance of the returns by the average conditional variance as a measure for risk.

References

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

Baba, Y., R.F. Engle, D. Kraft, and K. Kroner, 1990, Multivariate simultaneous generalized ARCH. Unpublished manuscript, University of California, San Diego

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

Brown, B.M., 1971, Martingale central limit theorems. *Annals of Mathematical Statistics* 42, 59–66.

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

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.

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.

Gourieroux, C., and J. Jasiak, 2001, *Financial Econometrics*, Princeton University Press.

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.

Hannan, E.J. and Heyde, C.C. 1972, On limit theorems of quadratic functions of discrete time series. *The Annals of Mathematical Statistics* 43(6), 2058–2066.

Hansen, P.R. 2006, Granger's representation theorem: A closed-form expression for I(1) processes. *The Econometrics Journal*, 8(1), 23–38.

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.

7 Appendix

Proof of Lemma 1. We use $\bar\beta_\perp = \beta_\perp(\beta_\perp'\beta_\perp)^{-1}$and find

$$\begin{aligned}
\Theta_h^{-1} &= (\beta, h^{-1/2}\bar\beta_\perp)\begin{pmatrix} \beta'\Theta_h\beta & h^{-1/2}\beta'\Theta_h\bar\beta_\perp \\ h^{-1/2}\bar\beta_\perp'\Theta_h\beta & h^{-1}\bar\beta_\perp'\Theta_h\bar\beta_\perp \end{pmatrix}^{-1}(\beta, h^{-1/2}\bar\beta_\perp)' \\
&= (\beta, h^{-1/2}\bar\beta_\perp)\begin{pmatrix} \beta'\Psi_h\beta & h^{-1/2}\beta'\Psi_h\bar\beta_\perp \\ h^{-1/2}\bar\beta_\perp'\Psi_h\beta & \Phi_h + h^{-1}\bar\beta_\perp'\Psi_h\bar\beta_\perp \end{pmatrix}^{-1}(\beta, h^{-1/2}\bar\beta_\perp)' \\
&\xrightarrow{P} (\beta, 0)\begin{pmatrix} \Gamma & 0 \\ 0 & \Phi \end{pmatrix}^{-1}(\beta, 0)' = \beta\Gamma^{-1}\beta'.
\end{aligned}$$

■

Proof of Theorem 1. For model (15) it is seen that for $\rho = I_r + \beta'\alpha$,

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

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

We combine these results using the identity

$$I_n = \beta_\perp(\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp' + \alpha(\beta'\alpha)^{-1}\beta' = C + \alpha(\beta'\alpha)^{-1}\beta', \tag{44}$$

and find

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

which reduces to (16) using (44). We then find the conditional mean (17) and conditional variance (18). The result (19) follows immediately and (20) follows from Lemma 1. ■

Proof of Theorem 2. *Proof of (22) and (23):* We find from equation (10) that

$$\mu_{t,1} = E_t(\Delta y_{t+1}) = \alpha(\beta' y_t - \xi) + \Xi \Delta y_t, \quad \Sigma_1 = \Omega,$$

and the optimal hedging portfolio is $\eta_1^* = (1, -\Omega_{12}\Omega_{22}^{-1})$, see (14), which proves (22) and (23).

The companion form: The model with two lags (10) can be expressed in companion form as

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

which we formulate as a lag one model for the stacked process $\tilde{y}_t = (y_t', y_{t-1}')'$ and errors $\tilde{\varepsilon}_t = (\varepsilon_t', 0_n')'$

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

where we use the notation

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

We then find for $C = \beta_\perp(\alpha_\perp'(I_n - \Xi)\beta_\perp)^{-1}\alpha_\perp'$, see Assumption 1, the derived parameters

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

Assumption 1 implies that $\tilde{\rho}^h$ converges exponentially to zero for $h \to \infty$.

The results (17) and (18) hold for the process $\tilde{y}_t$ by adding a tilde on all parameters,

$$\begin{aligned} \tilde{\Sigma}_h &= \tilde{C}\sum_{i=0}^{h-1}\tilde{\Omega}\tilde{C}' + \tilde{\alpha}(\tilde{\beta}'\tilde{\alpha})^{-1}(\sum_{i=0}^{h-1}\tilde{\rho}^i\tilde{\beta}'\tilde{\Omega}\tilde{\beta}\tilde{\rho}'^i)(\tilde{\alpha}'\tilde{\beta})^{-1}\tilde{\alpha}' \qquad (45) \\ &\quad + \tilde{C}\sum_{i=0}^{h-1}\tilde{\Omega}\tilde{\beta}\tilde{\rho}'^i(\tilde{\alpha}'\tilde{\beta})^{-1}\tilde{\alpha}' + \tilde{\alpha}(\tilde{\beta}'\tilde{\alpha})^{-1}\sum_{i=0}^{h-1}\tilde{\rho}^i\tilde{\beta}'\tilde{\Omega}\tilde{C}', \\ \tilde{\mu}_{t,h} &= \tilde{\alpha}(\tilde{\beta}'\tilde{\alpha})^{-1}(\tilde{\rho}^h - I_{r+n})(\tilde{\beta}'\tilde{y}_t - \tilde{\xi}), \qquad (46) \end{aligned}$$

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 $\Sigma_h = (I_n, 0_{n\times n})\tilde{\Sigma}_h(I_n, 0_{n\times n})'$.

Proof of (24): It is seen from (45) that $\Sigma_h = hC\Omega C' + \Psi_h$, for some Ψ_h which is convergent. We use the identity $\beta'(I_n, 0_{n\times n}) = (I_r, 0_{r\times n})\tilde{\beta}'$ to find

$$\begin{aligned} Var_t(\beta' y_{t+h}) &= \beta'\Psi_h\beta = (I_r, 0_{r\times n})\sum_{i=0}^{h-1}\tilde{\rho}^i\tilde{\beta}'\tilde{\Omega}\tilde{\beta}\tilde{\rho}'^i(I_r, 0_{r\times n})' \\ &\to (I_r, 0_{r\times n})\sum_{i=0}^{\infty}\tilde{\rho}^i\tilde{\beta}'\tilde{\Omega}\tilde{\beta}\tilde{\rho}'^i(I_r, 0_{r\times n})' = (I_r, 0_{r\times n})Var\begin{pmatrix} \beta' y_t \\ \Delta y_t \end{pmatrix} = Var(\beta' y_t) = \Gamma. \end{aligned}$$

It follows from Lemma 1 that

$$\Sigma_h^{-1} \to \beta\Gamma^{-1}\beta',$$

such that the optimal hedging portfolio as given in (13) has limit

$$\eta_h^* = \Sigma_h^{-1} e_{n1}/e_{n1}'\Sigma_h^{-1} e_{n1} \to \beta\Gamma^{-1}\beta' e_{n1}/e_{n1}'\beta\Gamma^{-1}\beta' e_{n1}, \text{ for } h \to \infty.$$

Using the normalization (21) we find $e_{n1}'\beta = (1, 0_{r-1}')$, $\Gamma^{-1}\beta' e_{n1} = (\Gamma^{11}, \Gamma^{12})'$ and $e_{n1\perp}'\beta\Gamma^{-1}\beta' e_{n1} = \gamma_1\Gamma^{11} + \gamma_2\Gamma^{21}$, such that

$$e_{n1\perp}'\beta\Gamma^{-1}\beta' e_{n1}/e_{n1}'\beta\Gamma^{-1}\beta' e_{n1} = \gamma_1 + \gamma_2\Gamma^{21}(\Gamma^{11})^{-1} = \gamma_1 - \gamma_2\Gamma_{22}^{-1}\Gamma_{21},$$

see (14), and therefore we find (24).

The proof of (25), (26) and (27): Using $\beta'(I_n, 0_{n\times n}) = (I_r, 0_{r\times n})\tilde{\beta}'$, the limits are

$$\eta_h^{*\prime}\mu_{t,h} \to (1, \Gamma^{12}/\Gamma^{11})(\beta' y_t - \xi) = (1, -\Gamma_{12}\Gamma_{22}^{-1})(\beta' y_t - \xi),$$
$$\eta_h^{*\prime}\Sigma_h^{-1}\eta_h^* = 1/e_1'\Sigma_h^{-1} e_1 \to (\Gamma^{11})^{-1} = \Gamma_{11} - \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{21}.$$

■

Proof of Theorem 3. A cointegrating portfolio has the form $\beta\kappa$ for some $\kappa \in \mathbb{R}^r$. The conditional variance of $\beta' y_{t+h}$ is $\beta'\Sigma_h\beta$, and hence the optimal cointegrating portfolio is

$$\frac{\beta(\beta'\Sigma_h\beta)^{-1}\beta' e_{1n}}{e_{1n}'\beta(\beta'\Sigma_h\beta)^{-1}\beta' e_{1n}} \to \frac{\beta\Gamma^{-1}\beta' e_{1n}}{e_{1n}'\beta\Gamma^{-1}\beta' e_{1n}},$$

which is the same as in (24). Hence the results follow. ■

Proof of Theorem 4. Maximizing the Sharpe ratio is equivalent to minimizing the variance $\eta'\Sigma_h\eta$ subject to the constraint $\eta'\mu_{t,h} = 1$, and the optimizing portfolio and its limit are therefore given by any portfolio proportional to

$$\Sigma_h^{-1}\mu_{t,h} \to -\beta\Gamma^{-1}\beta'(I_n, 0_{n\times n})\tilde{\alpha}(\tilde{\beta}'\tilde{\alpha})^{-1}(\tilde{\beta}'\tilde{y}_t - \tilde{\xi}).$$

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

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

If we restrict the portfolio to be stationary $\eta = \beta\kappa$, $\kappa \in \mathbb{R}^r$, we find

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

such that for the optimal $\breve{\kappa}$ we find

$$\breve{\eta}_{h,coint} = \beta\breve{\kappa} = \beta(\beta'\Sigma_h\beta)^{-1}\beta'\mu_{t,h}$$
$$\to -\beta\Gamma^{-1}\beta'(I_n, 0_{n\times n})\tilde{\alpha}(\tilde{\beta}'\tilde{\alpha})^{-1}(\tilde{\beta}'\tilde{y}_t - \tilde{\xi}) = -\beta\Gamma^{-1}(\beta' y_t - \xi).$$

■

Proof of Theorem 5. *Proof of (33)*: We find from the representation (16) applied to the stacked process $\tilde{y}_t$, multiplying by $(I_n, 0_{n\times n})$, that

$$y_{t+h} - y_t = \sum_{i=0}^{h-1} C\varepsilon_{t+h-i} + (I_n, 0_{n\times n})\tilde{\alpha}(\tilde{\beta}'\tilde{\alpha})^{-1}\sum_{i=0}^{h-1}\tilde{\rho}^i\tilde{\beta}'\tilde{\varepsilon}_{t+h-i} \tag{47}$$
$$+ (I_n, 0_{n\times n})\tilde{\alpha}(\tilde{\beta}'\tilde{\alpha})^{-1}(\tilde{\rho}^h - I_r)(\tilde{\beta}'\tilde{y}_t - \tilde{\xi}).$$

It is seen that regressing on y_t, y_{t-1}, and a constant, eliminates the last term, and we note that the first two terms only depend on $\varepsilon_{t+1}, \ldots, \varepsilon_{t+h}$, which are uncorrelated with y_t, y_{t-1}, because ε_t form a martingale difference sequence. Thus, correcting for y_t, y_{t-1} and a constant, we find the residuals $(y_{t+h} - y_t|1, y_t, y_{t-1})$ and their sum of squares has the same limit as the sum of squares of the first two terms.

The behavior of such product moments under stationary martinale difference assumptions with constant velocity for ε_t, were studied by Hannan and Heyde (1972, Theorem 1). They considered a general linear process of the form $z_t = \sum_{i=0}^{\infty}\phi_i'\varepsilon_{t-i}$, for coefficients which satisfy $\sum_{i=0}^{\infty}|\phi_i'| < \infty$ and ε_t which satisfy Assumption 2. They define the empirical and theoretical autocovariance functions for $j \geq 0$

$$\hat{\gamma}_j = T^{-1}\sum_{t=1}^{T-j}(z_t - \bar{z})(z_{t+j} - \bar{z})', \tag{48}$$

$$\gamma_j = \sum_{i=0}^{\infty}\phi_i'\Omega\phi_{i+j} = Cov(z_t, z_{t+j}). \tag{49}$$

They then prove that

$$\bar{z}_T \overset{a.s.}{\rightarrow} 0, \text{ and } \hat{\gamma}_j \overset{a.s.}{\rightarrow} \gamma_j. \tag{50}$$

We apply these results to the processes $\varepsilon_{t+1}, \ldots, \varepsilon_{t+h}$ in (47) under Assumption 2, and find that S_{hT} from (33) converges almost surely to Σ_h.

Proof of (35): From (47) we find the representation

$$y_t - y_0 = C\sum_{i=0}^{t-1}\varepsilon_{t-i} + (I_n, 0_{n\times n})\tilde{\alpha}(\tilde{\beta}'\tilde{\alpha})^{-1}(\sum_{i=0}^{t-1}\tilde{\rho}^i\tilde{\beta}'\tilde{\varepsilon}_{t-i}) \tag{51}$$
$$+ (I_n, 0_{n\times n})\tilde{\alpha}(\tilde{\beta}'\tilde{\alpha})^{-1}(\tilde{\rho}^t - I_r)(\tilde{\beta}'\tilde{y}_0 - \tilde{\xi}).$$

Regressing on a constant the last term vanishes for $T \to \infty$. The first term is a martingale and $T^{-1/2}\sum_{t=1}^{[T.]}\varepsilon_t \overset{D}{\rightarrow} W(.)$, where W is Brownian motion, see Brown (1971, Theorem 3). This implies that

$$T^{-1}S_T = T^{-1}C(\sum_{i=1}^{t}\varepsilon_i|1)(\sum_{i=1}^{t}\varepsilon_i|1)'C' + o_P(1)$$
$$\overset{D}{\rightarrow} C\int_0^1(W(u) - \bar{W})(W(u) - \bar{W})'duC',$$

where $\bar{W} = \int_0^1 W(u)du$. In the second term we can replace $(\sum_{i=0}^{t-1} \tilde{\rho}^i \tilde{\beta}' \tilde{\varepsilon}_{t-i}|1)$ by $\sum_{i=0}^{\infty} \tilde{\rho}^i \tilde{\beta}' \tilde{\varepsilon}_{t-i}$, and find from (50) that $\beta' S_T \beta \overset{a.s.}{\to} Var(\beta' y_t) = \Gamma$. Finally we find that $\beta' S_T \bar{\beta}_\perp = O_P(1)$ and we therefore get from Lemma 1 that $S_T^{-1} \overset{P}{\to} \beta \Gamma^{-1} \beta'$. ∎

Proof of Theorem 6. *Proof of (42):* We find from (47), using $\beta'(I_n, 0_{n\times n}) = (I_r, 0_{rn})\tilde{\beta}'$, that

$$\beta' \Sigma_{t,j} \beta = (I_r, 0_{r\times n}) E_t \sum_{i=0}^{j-1} \tilde{\rho}^i \tilde{\beta}' \tilde{\Omega}_{t+j-i} \tilde{\beta} \tilde{\rho}^{i\prime} (I_r, 0_{r\times n})'.$$

Taking average for $j = 1, \ldots, h$ we find

$$\begin{aligned} \beta' \overline{\Sigma}_{t,h} \beta &= (I_r, 0_{r\times n}) E_t \sum_{i=0}^{h-1} \tilde{\rho}^i \tilde{\beta}' [h^{-1} \sum_{j=i+1}^{h} \begin{pmatrix} \Omega_{t+j-i} & 0_{n\times n} \\ 0_{n\times n} & 0_{n\times n} \end{pmatrix}] \tilde{\beta} \tilde{\rho}^{i\prime} (I_r, 0_{r\times n})' \\ &\overset{P}{\to} (I_r, 0_{r\times n}) \sum_{i=0}^{\infty} \tilde{\rho}^i \tilde{\beta}' \begin{pmatrix} \Omega & 0_{n\times n} \\ 0_{n\times n} & 0_{n\times n} \end{pmatrix} \tilde{\beta} \tilde{\rho}^{i\prime} (I_r, 0_{r\times n})' = \overline{\Gamma}, \end{aligned}$$

say. We next consider $h^{-1} \bar{\beta}_\perp' \overline{\Sigma}_{t,h} \bar{\beta}_\perp$ and find from (47) that the first term is dominating. Thus

$$h^{-1} \bar{\beta}_\perp' \overline{\Sigma}_{t,h} \bar{\beta}_\perp = \bar{\beta}_\perp' C E_t \{h^{-2} \sum_{j=1}^{h} j [j^{-1} \sum_{i=0}^{j-1} \Omega_{t+j-i}]\} C' \bar{\beta}_\perp + o_P(1) \overset{P}{\to} \frac{1}{2} \bar{\beta}_\perp' C \Omega C' \bar{\beta}_\perp.$$

Finally by the same methods we find that $\beta' \overline{\Sigma}_{t,h} \bar{\beta}_\perp = O_P(1)$ because the dominating term vanishes by multiplication by β and the remaining are bounded because $\tilde{\rho}^i$ is decreasing exponentially. Finally we find from Lemma 1 that $\overline{\Sigma}_{t,h}^{-1} \overset{P}{\to} \beta \overline{\Gamma}^{-1} \beta'$. ∎

Proof of Theorem 7. *Proof of (42):* From the representation (47) we find that only the first two terms are relevant and we consider the different terms separately. They only involve the errors $\varepsilon_{t+1}, \ldots, \varepsilon_{t+h}$ and by applying the result of Hannan and Heyde (1972) given in (39) we find (42).

Proof of (43): We apply the representation (51). In order to apply the result (39) we define the linear process $z_t = \sum_{i=0}^{\infty} \tilde{\rho}^i \tilde{\beta}' \tilde{\varepsilon}_{t-i} = \sum_{i=0}^{t-1} \tilde{\rho}^i \tilde{\beta}' \tilde{\varepsilon}_{t-i} + \sum_{i=t}^{\infty} \tilde{\rho}^i \tilde{\beta}' \tilde{\varepsilon}_{t-i}$, where the remainder is evaluated as

$$E| \sum_{i=t}^{\infty} \tilde{\rho}^i \tilde{\beta}' \tilde{\varepsilon}_{t-i}| \le c \sum_{i=t}^{\infty} \tilde{\rho}^i \le c|\tilde{\rho}|^t \to 0,$$

so we shall neglect it. Then from (39) we find

$$\begin{aligned} \beta' S_T \beta &= T^{-1} \sum_{t=1}^{T} (I_r, 0_{r\times n}) \sum_{i=0}^{t-1} \tilde{\rho}^i \tilde{\beta}' \tilde{\varepsilon}_{t-i} \sum_{j=0}^{t-1} \tilde{\varepsilon}_{t-j}' \tilde{\beta} \tilde{\rho}^{j\prime} (I_r, 0_{r\times n})' \\ &= (I_r, 0_{r\times n}) T^{-1} \sum_{t=1}^{T} z_t z_t' (I_r, 0_{r\times n})' + o_P(1) \overset{P}{\to} (I_r, 0_{r\times n}) \sum_{i=0}^{\infty} \tilde{\rho}^i \tilde{\beta}' \tilde{\Omega} \tilde{\beta} \tilde{\rho}^{i\prime} (I_r, 0_{r\times n})' = \overline{\Gamma}, \end{aligned}$$

say. We next consider

$$T^{-1}\bar{\beta}_{\perp}' S_T \bar{\beta}_{\perp} = \bar{\beta}_{\perp}' C T^{-1} \sum_{t=1}^{T} (T^{-1/2} \sum_{i=0}^{t-1} \varepsilon_{t-i}|1)(T^{-1/2} \sum_{j=0}^{t-1} \varepsilon_{t-j}'|1)' C' \bar{\beta}_{\perp}$$
$$\xrightarrow{D} \bar{\beta}_{\perp}' C \int_0^1 (W(u) - \bar{W})(W(u) - \bar{W})' du C' \bar{\beta}_{\perp}.$$

The same methods give $\beta' S_T \bar{\beta}_{\perp} = O_P(1)$ and we can apply Lemma 1to show that $S_T^{-1} \xrightarrow{P} \beta \overline{\Gamma}^{-1} \beta'$. ■