Optimal hedging with the cointegrated vector autoregressive model

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

Søren Johansen[‡]
University of Copenhagen
and CREATES

May 16, 2015

Abstract

We analyse the role of cointegration for the problem of hedging an asset using other assets, when the prices are generated by a Cointegrated Vector Autoregressive model (CVAR). 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 ilustrate 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.

Keywords: hedging, cointegration, minimum variance portfolio, optimal 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.

1 Introduction

In this paper we consider the situation that there are given n tradable assets with prices $y_t = (y_{1t}, \ldots, y_{nt})'$, and we construct a *portfolio* as a linear combination $\eta = (\eta_1, \ldots, \eta_n)'$ of the assets with *value* at time t, $\eta' y_t = \sum_{i=1}^n \eta_i y_{it}$. A positive coefficient, $\eta_i > 0$, indicates that we buy η_i units of asset i, and a negative coefficient $\eta_j < 0$ means that we sell $|\eta_j|$ units of asset j. We use the terminology that we have taken a long position in asset i and a short position in asset j. We define the prediction variance $\Sigma_h = Var(y_{t+h}|\mathcal{I}_t)$, the variance of y_{t+h} given the information in the process up to time t, $\mathcal{I}_t = \sigma(y_s, s \leq t)$, and measure the *risk* of a portfolio, η, at time $t+h$ as $Var(\eta' y_{t+h}|\mathcal{I}_t) = \eta' \Sigma_h \eta$.

The simplest example of what we want to analyse is the case of two tradeable assets. If we hold one unit of asset 1, we have the portfolio $\eta = (1, 0)'$. The risk at time $t+h$ is Σ_{h11}, which, for nonstationary prices, will diverge with h. By selling β units of the second asset y_{2t}, we have the portfolio $\eta = (1, -\beta)'$, with risk at time $t+h$ given by $\Sigma_{h11} + \beta^2 \Sigma_{h22} - 2\beta \Sigma_{h21}$. This clearly minimized for $\beta = \Sigma_{h12}/\Sigma_{h22}$ giving the minimal risk $\Sigma_{h11} - \Sigma_{h12}^2/\Sigma_{h22}$ which is less than the unhedged risk of the first asset Σ_{h11}. We show that if there is cointegration among the prices, we can exploit this and show that for long horizons, it is a cointegrating relation that gives the best portfolios. In general we have more than two tradable assets, and we maintain throughout the idea that we have one unit of asset 1 and want to invest in $y_{2t}, \ldots, y_{nt}$ in order to offset the risk in asset one, as measured by the conditional variance, if we hold the portfolio for h periods. More precisely we want to choose a portfolio $\eta = (1, \eta_2, \ldots, \eta_n)'$ in such a way that $Var(\eta'(y_{t+h} - y_t)|\mathcal{I}_t) = \eta' \Sigma_h \eta$, is as small as possible. In this context we call asset one the *hedged asset* and the assets $(2, \ldots, n)$ the *hedging assets*. The coefficients $\eta_2, \ldots, \eta_n$ are called *hedging ratios* and η is the *hedging portfolio*. Finally we shall use the term *optimal hedging portfolio* or *minimum variance portfolio* for the portfolio minimizing the risk of $\eta'(y_{t+h} - y_t)$, among all portfolios normalized on $\eta_1 = 1$. Thus, hedging only considers risk and not the expected return of the investment. To discuss this problem, we define $\mu_h = E(y_{t+h} - y_t|\mathcal{I}_t)$, such that the *expected return* of η is $\eta' \mu_h = E(\eta'(y_{t+h} - y_t)|\mathcal{I}_t)$. To balance the expected return by the risk we consider the (squared) Sharpe ratio, Sharpe (1966), which takes into account both expected return and risk by considering $S_h^2 = (\eta' \mu_h)^2 / \eta' \Sigma_h \eta$. The portfolio maximizing the Sharpe ratio is called the *optimal Sharpe portfolio*. If we can normalize on the first coordinate we can use the optimal Sharp portfolio as a hedging portfolios, as we shall do in the analysis in Section 5.

The idea of minimum variance portfolio dates back to the seminal paper by Markowitz (1952) and has since been explored and extended in both 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 $t, \ldots, t+h$. This is opposed to the dynamic hedging methods which allows for rebalancing the portfolio during the holding period.

We are only concerned with static hedging, and our contribution is to analyse the properties of the optimal portfolios under the assumption that the asset prices are driven by a Cointegrated Vector Autoregressive model (CVAR). We start with a simple example of a cointegrating regression model, which relates the hedged asset to the hedging assets via a cointegrating relation, and the hedges are strongly exogenous and modelled by random walks. It must be pointed out that the assumption that the data is generated by a CVAR is not an assumption that holds for all assets and all frequencies of data. It has to be checked carefully using the available data. Cointegration is used in pairs trading, see

> "If the long and short components fluctuate with common nonstationary factors, then the prices of the component portfolios would be co-integrated and the pairs trading strategy would be expected to work." (Gatev, Goetzmann, Rouwe 2006, p. 801)

and the contribution in this paper is a framework and some results, that can be used if the assumptions of the CVAR are satisfied.

We then turn to the general CVAR, and find an expression for the optimal hedging portfolio, and the optimal Sharpe ratio portfolio as functions of the parameters of the model. There is no simple relation between the expected returns in the two situations, except when the assets are strongly exogenous, in which case the returns are the same.

Our main conclusion is that for large h both the optimal hedging portfolio and the optimal Sharpe portfolio converge to cointegrating relations, which we find explicitly and characterize as the minimum variance stationary portfolio normalized on $\eta_1 = 1$, and as the limit of the Sharpe optimal stationary portfolio respectively. If $r = 1$ they are equal when normalized on $\eta_1 = 1$. As an illustration of the results we analyse a set of data on prices of futures of electricity and fuels in the Netherlands.

Thus a conclusion is that cointegration plays an important role in hedging. It allows for the possibility that the hedging portfolio has a bounded risk, as opposed to the unhedged risk. More importantly, however, is that the results show 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 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 that are 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 analyse 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 p_t cointegrates with $coal$, gas, and CO_2 which are modelled as $n-1=3$ exogenous random walks, $y_{2t} \in \mathbb{R}^{n-1}$,

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

where $u_t = (u_{1t}, u'_{2t})'$ are independent identically distributed (i.i.d.) random errors with mean zero and variance split accordingly

$$\Psi = Var(u_t) = \begin{pmatrix} \Psi_{11} & \Psi_{12} \\ \Psi_{21} & \Psi_{22} \end{pmatrix}.$$

The stylized story is that 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. We define the expected return and the prediction variance h periods ahead

$$\mu_h = E(y_{t+h} - y_t|\mathcal{I}_t) = \begin{pmatrix} \mu_{h1} \\ \mu_{h2} \end{pmatrix},$$
$$\Sigma_h = Var(y_{t+h}|\mathcal{I}_t) = \begin{pmatrix} \Sigma_{h11} & \Sigma_{h12} \\ \Sigma_{h21} & \Sigma_{h22} \end{pmatrix}.$$

In fact the producer of electricity is doing the opposite, see section 5, by going short in electricity and long in the fuels, but that is just a question of a change of sign of the portfolio.

2.2 The hedging problem and its solution in cointegrating regression

We want to hedge one unit of the first asset by going short in the portfolio with value $\beta'_h y_{2t}$ and consider therefore the portfolio $\eta_h = (1, -\beta_h)'$ with value

$$\eta' y_t = y_{1t} - \beta'_h y_{2t}. \tag{2}$$

In portfolio hedging, a long position in asset one, is traditionally hedged with a short position in another set of assets. Thus the sign in front of the hedging ratios, β_h, indicates the market convention regarding hedging practice. The optimal portfolio is selected in period t and it is held up to period $t + h$.

We want to determine β_h to minimizes the risk measured by $\eta'\Sigma_h\eta = Var(\eta'_h y_{t+h}|\mathcal{I}_t)$, that is, we want to solve

$$\min_{\beta_h} Var(\eta'_h y_{t+h}|\mathcal{I}_t) = \min_{\beta_h}(\Sigma_{h11} + \beta'_h\Sigma_{h22}\beta_h - \beta'_h\Sigma_{h21} - \Sigma_{h12}\beta_h). \tag{3}$$

This is solved by the best linear predictor of $y_{1,t+h}$ given $y_{2,t+h}$ and $\mathcal{I}_t$, which is $\beta^*_h = \Sigma^{-1}_{h22}\Sigma_{h21}$. Therefore the optimal hedging portfolio becomes

$$\eta^*_h = \begin{pmatrix} 1 \\ -\Sigma^{-1}_{h22}\Sigma_{h21} \end{pmatrix}, \tag{4}$$

with expected return and risk

$$\eta^{*\prime}_h \mu_h = \mu_{h1} - \Sigma^{-1}_{h22}\Sigma_{h21}\mu_{h2},$$
$$\eta^{*\prime}_h \Sigma_h \eta^*_h = \Sigma_{h11} - \Sigma_{h12}\Sigma^{-1}_{h22}\Sigma_{h21}.$$

For the regression model (1) we can find explicit expressions for these quantities.

Theorem 1 *Let $y_t \in \mathbb{R}^n, t = 1, \ldots, T$, be given by the regression model (1), and let η_h^* be the optimal hedging portfolio for horizon h, see (4).*

1. The expected return and risk are

$$\mu_h = \begin{pmatrix} -(y_{1t} - \beta' y_{2t}) \\ y_{2t} \end{pmatrix} \tag{5}$$

$$\Sigma_h = \begin{pmatrix} h\beta'\Psi_{22}\beta + \beta'\Psi_{21} + \Psi_{12}\beta + \Psi_{11} & h\beta'\Psi_{22} + \Psi_{12} \\ h\Psi_{22}\beta + \Psi_{21} & h\Psi_{22} \end{pmatrix} \tag{6}$$

2. The optimal hedging portfolio is

$$\eta_h^* = \begin{pmatrix} 1 \\ -(\beta + \Psi_{22}^{-1}\Psi_{21})h^{-1} - \beta(1 - h^{-1}) \end{pmatrix}, \tag{7}$$

which has expected return and risk

$$\eta_h^{*\prime}\mu_h = -(y_{1t} - \beta' y_{2t}), \tag{8}$$

$$\eta_h^{*\prime}\Sigma_h\eta_h^* = \Psi_{11} - h^{-1}\Psi_{12}\Psi_{22}^{-1}\Psi_{21}. \tag{9}$$

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

$$Var(y_{1,t+h}|\mathcal{I}_t) = \Psi_{11} + h\beta'\Psi_{22}\beta + \Psi_{12}\beta + \beta'\Psi_{21} \to \infty,$$

whereas using the optimal hedging portfolio, we find the increasing but converging risk

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

Thus for large h one obtains a substantial reduction in risk by hedging. Even for $h = 1$, the risk associated with not hedging is $\Psi_{11} + \beta'\Psi_{22}\beta + \Psi_{12}\beta + \beta'\Psi_{21}$, which is larger than the optimal risk when hedging: $\Psi_{11} - \Psi_{12}\Psi_{22}^{-1}\Psi_{21}$.

The expected return of holding the first asset is the same as the 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 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

The analysis of the model, where the hedging assets are exogenous, is now generalized to the general cointegration model, see Johansen (1996), where we use the error correction formulation, which allows general adjustment coefficients as well as a constant term in the cointegrating space, allowing for nonzero expectation of the cointegrating relations. In the case of $n - 1$ exogenous random walks and $r = 1$, see the regression (1), the optimizing portfolio approaches the only cointegrating vector. In the general case where $r \geq 1$, we show that the optimal portfolio converge to η_{stat}^*, the minimum variance stationary portfolio normalized on the first asset.

This result is formulated in Theorem 4 for the cointegrated VAR model with two lags

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

It is only a question of a more elaborate notation to handle the case of more lags. For a lag k model, we can express the model as a lag one model for the stacked process $\tilde{y}_t = (y_t', \dots y_{t-k+1}',)'$, using the companion form, see Johansen (1996) and Hansen (2006). The portfolios we investigate, however, have the form $\tilde{\eta}'\tilde{y}_t = (\eta', 0_{n(k-1)}')\tilde{y}_t = \eta' y_t$, where $0_{n(k-1)} = (0, \dots, 0)' \in \mathbb{R}^{n(k-1)}$. Thus we are not optimizing over all linear combinations of $\tilde{y}_t$, but only linear combinations of the first n coordinates. This requires a slightly modified form of the optimal portfolio.

The prediction variance $\Sigma_h = Var(y_{t+h}|\mathcal{I}_t)$ can be calculated recursively from the estimated parameters, as in Lütkepohl (2005, pp. 259-260), by defining the matrices $\Phi_0 = I_n$ and $\Phi_1 = I_n + \alpha\gamma' + \Phi$ and $\Phi_i = \Phi_{i-1}(I_n + \alpha\gamma' + \Phi) - \Phi_{i-2}\Phi$, $i = 2, 3, \dots$ Then the variance is given by

$$\Sigma_h = \Sigma_{h-1} + \Phi_{h-1}\Omega\Phi_{h-1}',$$

but we need a more explicit expression for the detailed analysis below. The first result is formulated for the lag one model to simplyfy the notation.

Theorem 2 *Let $y_t \in \mathbb{R}^n, t = 1, \dots, T$ be given by*

$$\Delta y_t = \alpha(\gamma' y_{t-1} - \xi) + \varepsilon_t, \tag{11}$$

where ε_t are i.i.d. $(0, \Omega)$ and α and γ are $n \times r$ matrices. We assume the usual $I(1)$ conditions, see Johansen (1996, Theorem 4.2). This implies that the eigenvalues of $\rho = I_r + \gamma'\alpha$ have absolute value less than 1, such that $\gamma' y_t - \xi$ is stationary with mean zero. We define $\gamma_\perp$ as an $n \times (n - r)$ matrix of rank $n - r$, such that $\gamma'\gamma_\perp = 0$, and similarly for $\alpha_\perp$, and use them to construct $C = \gamma_\perp(\alpha_\perp'\gamma_\perp)^{-1}\alpha_\perp'$. We then find the conditional mean and variance

$$\mu_h = E(y_{t+h} - y_t|\mathcal{I}_t) = \alpha(\gamma'\alpha)^{-1}(\rho^h - 1)(\gamma' y_t - \xi), \tag{12}$$

$$\Sigma_h = Var(y_{t+h} - y_t|\mathcal{I}_t) = \sum_{i=0}^{h-1}[C + \alpha(\gamma'\alpha)^{-1}\rho^i\gamma']\Omega[C' + \gamma\rho'^i(\alpha'\gamma)^{-1}\alpha']. \tag{13}$$

It follows that the optimal hedging portfolio is

$$\eta_h^* = \begin{pmatrix} 1 \\ -\Sigma_{h22}^{-1}\Sigma_{h21} \end{pmatrix}. \tag{14}$$

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 $\gamma_1' y_t = y_{1t} + \beta_1' y_{2t}$. 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

$$\gamma = (\gamma_1, \gamma_2) = \begin{pmatrix} 1 & 0 \\ \beta_1 & \beta_2 \end{pmatrix}, \tag{15}$$

for $\gamma_1 \in \mathbb{R}^n$ and $\gamma_2 \in \mathbb{R}^{n\times(r-1)}$. We use the notation for mean and variance of the stationary variables $\gamma' y_t$

$$\xi = E(\gamma' y_t) = E\begin{pmatrix} y_{1t} + \beta_1' y_{2t} \\ \beta_2' y_{2t} \end{pmatrix} = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}, \tag{16}$$

$$\Gamma = Var(\gamma' y_t) = Var\begin{pmatrix} y_{1t} + \beta_1' y_{2t} \\ \beta_2' y_{2t} \end{pmatrix} = \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{pmatrix}. \tag{17}$$

If the portfolio is chosen as a cointegrating relation, we find the optimal portfolio in the next Theorem.

Theorem 3 *Under the assumptions of Theorem 2 and if the cointegrating relations are normalized as in (15), then the variance of a stationary portfolio $\eta' y_t = y_{1t} - \beta' y_{2t}$ is minimized for $\beta \in \mathbb{R}^{n-1}$ by the optimal hedging portfolio*

$$\eta^*_{stat} = \begin{pmatrix} 1 \\ \beta_1 - \beta_2 \Gamma_{22}^{-1} \Gamma_{21} \end{pmatrix} = \gamma \begin{pmatrix} 1 \\ -\Gamma_{22}^{-1}\Gamma_{21} \end{pmatrix}, \tag{18}$$

with expected return and risk

$$E(\eta^{*\prime}_{stat} y_t) = \xi_1 - \Gamma_{12}\Gamma_{22}^{-1}\xi_2, \tag{19}$$

$$Var(\eta^{*\prime}_{stat} y_t) = \Gamma_{11} - \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{21}. \tag{20}$$

The coefficient $\beta^*_{stat} = \beta_1 - \beta_2\Gamma_{22}^{-1}\Gamma_{21}$ is the probability limit of the estimated coefficient in a regression of y_{1t} on y_{2t}.

Note that with the parametrization (15), the parameter β_1 is not identified, because we could choose the parameters,

$$\gamma_\kappa = \gamma\kappa = \begin{pmatrix} 1 & 0 \\ \beta_1 & \beta_2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \kappa_1 & \kappa_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \beta_1 + \beta_2\kappa_1 & \beta_2\kappa_2 \end{pmatrix},$$

and $\alpha_\kappa = \alpha\kappa'^{-1}$ for which $\alpha\gamma' = \alpha_\kappa \gamma_\kappa'$, and that would not change the cointegrating space and therefore not the model (11), as long as κ has full rank r. The result in (18), however, is invariant to this choice of parametrization, because if γ_κ were the cointegrating relations, then using the expression in (18), we would find

$$\eta^*_\kappa = \begin{pmatrix} 1 \\ \beta_1 + \beta_2\kappa_1 - \beta_2\kappa_2(\kappa_2'\Gamma_{22}\kappa_2)^{-1}\kappa_2'(\Gamma_{21} + \Gamma_{22}\kappa_1) \end{pmatrix} = \begin{pmatrix} 1 \\ \beta_1 - \beta_2\Gamma_{22}^{-1}\Gamma_{21} \end{pmatrix}.$$

Thus, the result (18) does not depend on the parametrization of the cointegrating space.

We next formulate the main result for the hedging problem in the CVAR.

Theorem 4 *Let y_t be given by the CVAR (10) and assume usual $I(1)$ conditions, see Johansen (1996, Theorem 4.2), such that y_t is $I(1)$ and $\gamma' y_t$ is stationary with mean ξ.*

1. If $h = 1$, we find $\Sigma_1 = \Omega$, $\mu_1 = \alpha(\gamma' y_{t-1} - \xi) + \Phi\Delta y_{t-1}$ and the optimal hedging portfolio is $\eta_1^{*\prime} = (1, -\Omega_{12}\Omega_{22}^{-1})$, which has mean return and risk

$$\eta_1^{*\prime}\mu_1 = (1, -\Omega_{12}\Omega_{22}^{-1})(\alpha(\gamma' y_t - \xi) + \Phi\Delta y_t), \tag{21}$$

$$\eta_1^{*\prime}\Sigma_1\eta_1^* = \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}. \tag{22}$$

2. If $h \to \infty$, we find

$$\eta_h^* \to \eta_{stat}^* = \gamma \begin{pmatrix} 1 \\ -\Gamma_{22}^{-1}\Gamma_{21} \end{pmatrix}, \tag{23}$$

and the limits of mean return and risk are

$$\eta_h^{*\prime}\mu_h \to -(1, -\Gamma_{12}\Gamma_{22}^{-1})(\gamma' y_t - \xi) = -(\eta_{stat}^{*\prime} y_t - E(\eta_{stat}^{*\prime} y_t)), \tag{24}$$

$$\eta_h^{*\prime}\Sigma_h\eta_h^* \to \Gamma_{11} - \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{21} = Var(\eta_{stat}^{*\prime} y_t). \tag{25}$$

The interpretation of these results is the following. For $h = 1$, the optimal portfolio depends only on the error variance Ω, and cointegration plays no role. The minimal variance is $\Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21} < \Omega_{11}$, which is the variance of the unhedged asset. For $h \to \infty$ we find that the limit of the optimal portfolio is the cointegrating relation, which we would estimate by regression of y_{1t} on y_{2t}, that is η_{stat}^*. For any h we find the risk of the optimal portfolio is

$$\Sigma_{h11} - \Sigma_{h12}\Sigma_{h22}^{-1}\Sigma_{h21} < \Sigma_{h11},$$

which is the risk of the unhedged portfolio, which diverges to infinity if the price of asset one is nonstationary, whereas the risk of the optimal portfolio stay bounded, so a lot is gained by hedging. The mean return of the unhedged asset and the optimal hedging portfolio are, for $e_{n1} = (1, 0_{2n-1}')' \in \mathbb{R}^{2n}$,

$$E(y_{1,t+h} - y_{1t}|\mathcal{I}_t) = e_{n1}'\tilde{\alpha}(\tilde{\gamma}'\tilde{\alpha})^{-1}(\tilde{\rho}^h - 1)(\tilde{\gamma}'\tilde{y}_t - \tilde{\xi}),$$
$$\eta_h^{*\prime}\mu_h = (1, -\Sigma_{h12}\Sigma_{h22}^{-1}, 0_n')\tilde{\alpha}(\tilde{\gamma}'\tilde{\alpha})^{-1}(\tilde{\rho}^h - 1)(\tilde{\gamma}'\tilde{y}_t - \tilde{\xi}).$$

Thus the risk is reduced by $\Sigma_{h12}\Sigma_{h22}^{-1}\Sigma_{h21} > 0$, and the mean return is changed, but not necessarily increased, by

$$-\Sigma_{h12}\Sigma_{h22}^{-1}e_{n2}'\tilde{\alpha}(\tilde{\gamma}'\tilde{\alpha})^{-1}(\tilde{\rho}^h - 1)(\tilde{\gamma}'\tilde{y}_t - \tilde{\xi}),$$

where $e_{n2}' = (0_{n-1}, I_{n-1}, 0_{(n-1)\times n})$. Note that

$$\begin{pmatrix} e_{n1}' \\ e_{n2}' \end{pmatrix} = (I_n, 0_{n\times n})$$

If in particular $e_{n2}'\tilde{\alpha} = (\alpha_2, \Phi_2) = 0$, then y_{2t} is strongly exogenous, as in model (1) and the mean return is not changed, and only risk need to be taken into account.

In the next section we analyse the balance between expected return and risk using the Sharpe ratio.

4 Optimizing the Sharpe ratio for the CVAR

We first derive the well know result for the portfolio optimizing the Sharpe ratio, see for instance Gourieroux and Jasiak (2001 pp. 74–76). We define the (squared) Sharpe ratio after h periods as

$$S_h(\eta) = \frac{[E\{\eta'(y_{t+h} - y_t)|\mathcal{I}_t\}]^2}{Var(\eta'(y_{t+h} - y_t)|\mathcal{I}_t)} = \frac{(\eta'\mu_h)^2}{\eta'\Sigma_h\eta}. \tag{26}$$

Theorem 5 *The portfolio which maximizes the Sharpe ratio after h periods is given, up to a constant factor, by*

$$\bar{\eta}_h = \Sigma_h^{-1}\mu_h, \tag{27}$$

and the maximal value is

$$S_h(\bar{\eta}_h) = \mu_h'\Sigma_h^{-1}\mu_h. \tag{28}$$

The optimal stationary portfolio $\eta_{h,stat} = \gamma(\gamma'\Sigma_h\gamma)^{-1}\gamma'\mu_h$, *satisfies for* $h \to \infty$

$$\bar{\eta}_{h,stat} \to -\gamma\Gamma^{-1}(\gamma' y_t - \xi) = \bar{\eta}_{stat}, \tag{29}$$

say.

Note that the expected return of the portfolio optimizing the Sharpe ratio is equal to the risk and given by

$$E\{\bar{\eta}'(y_{t+h} - y_t)|\mathcal{I}_t\} = \bar{\eta}'\mu_h = \mu_h'\Sigma_h^{-1}\mu_h = Var\{\bar{\eta}'(y_{t+h} - y_t)|\mathcal{I}_t\} > 0.$$

Thus, the mean and variance of the optimal portfolio are equal to the maximized value of the squared Sharpe ratio and the positive expected return is positive.

In the following we analyse (27) and (28) further for assets that are driven by the cointegration model with two lags, in order to investigate the role of the cointegrating relations.

Theorem 6 *Under the assumption of Theorem 4, we find*

1. *For* $h = 1$, *the optimal Sharpe portfolio and its expected risk are*

$$\bar{\eta}_1 = \Omega^{-1}(\alpha(\gamma' y_t - \xi) + \Phi\Delta y_t), \tag{30}$$

$$\bar{\eta}_1'\mu_1 = \bar{\eta}_1'\Sigma_1\bar{\eta}_1 = \mu_1'\Sigma_1^{-1}\mu_1 = (\alpha(\gamma' y_t - \xi) + \Phi\Delta y_t)'\Omega^{-1}(\alpha(\gamma' y_t - \xi) + \Phi\Delta y_t). \tag{31}$$

2. *For* $h \to \infty$, *the optimal Sharpe portfolio and its expected risk satisfy*

$$\bar{\eta}_h \to -\gamma\Gamma^{-1}(\gamma' y_t - \xi) = \bar{\eta}_{stat}, \tag{32}$$

$$\bar{\eta}_h'\mu_h = \bar{\eta}_h'\Sigma_h\bar{\eta}_h = \mu_h'\Sigma_h^{-1}\mu_h \to (\gamma' y_t - \xi)'\Gamma^{-1}(\gamma' y_t - \xi). \tag{33}$$

Note that the optimal Sharpe portfolio for $h = 1$, is a combination of the columns of the inverse error variance Ω with weights depending on the expected return after one period, $\mu_1 = E(\Delta y_{t+1}|\mathcal{I}_t)$. For large h the optimal portfolio approaches a cointegrating relation with weights determined by the inverse variance of the cointegrating relations, and the disequilibrium error $\gamma' y_t - \xi$ at the time of investment.

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. 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 next h periods. 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 price 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

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

$\varepsilon_t, t = 1, \ldots, T$ 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. The measurements are taken on the first trading day for each month January 2006 to April 2015, a total of 112 observations. 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 magnitude of roots of the companion matrix when $r = 1$. One finds as expected three unit roots, and the remaining are well within the unit disc.

One can simplify the coefficients in α and β and find that there is a stationary relation between *electricity*, *gas* and CO_2 without a constant, and that only CO_2 is significantly adjusting to the disequilibrium error. The restrictions are tested by a χ^2 test with 5 degrees of freedom. There are three restrictions on α and one on β and one on the constant:

$$\beta' y = p_t - \underset{[t=-17.20]}{1.459} gas - \underset{[t=-10.95]}{1.500} CO_2,$$

$$\alpha' = (0, \;\; 0, \;\; 0, \;\; \underset{[t=3.86]}{0.163}),$$

$$\text{LR test} = 6.833 \;\sim \chi^2(5) \; [p - \text{value} = 0.23].$$

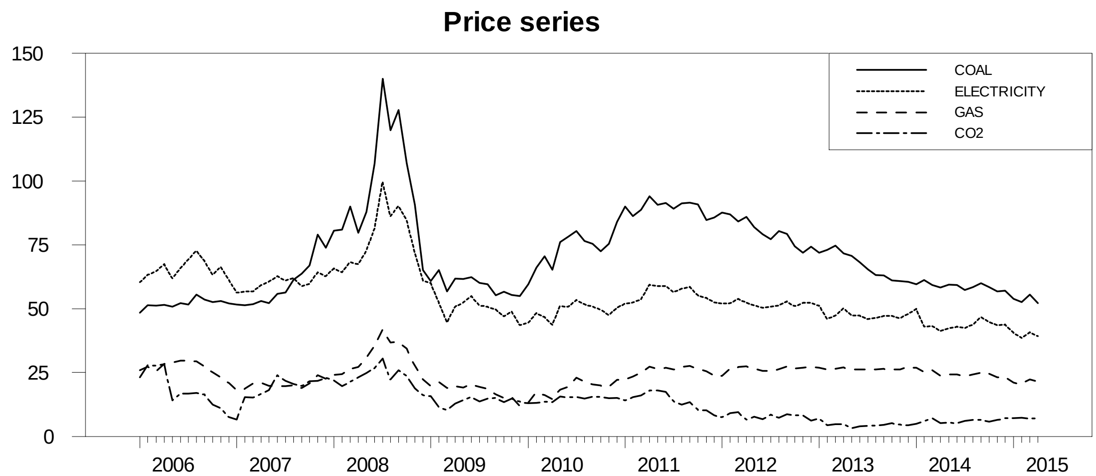

Figure 1: The four monthly series from January 2006 to April 2015

Test for cointegrating rank				
r	Eig.Value	Trace	Frac95	P-Value
0	0.248	60.135	53.945	0.012
1	0.123	28.785	35.070	0.210
2	0.097	14.391	20.164	0.270
3	0.028	3.151	9.142	0.562
8 abs(roots) of companion matrix for $r = 1$				
$1, 1, 1, 0.72, 0.24, 0.16, 0.08, 0.01$				

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

The estimated cointegrating relation is plotted in Figure 2 and the optimal β_h^*, is plotted in Figure 3, and the risk of the optimal portfolio compared to the stationary portfolio is given in Figure 4. Note that using the cointegrating relation as a hedging portfolio has a much greater risk than the optimal hedging portfolio. The unhedged risk grows linearly from 15.14 ($h = 1$) to 420.74 ($h = 24$), whereas the optimally hedged risk grows from 3.00 ($h = 1$) but stays below the limit $\Gamma = 20.49$.

We have illustrated the findings with some plots in Figure 5. The example has the special feature that $r = 1$, so we get some simplification. Because $\bar{\eta}_{h1} \to -\gamma' y_t/\Gamma$,we find that the optimal portfolios $\bar{\eta}_h/\bar{\eta}_{h1}$ and η^* converge to γ. The corresponding expected returns $\bar{\eta}_h'\mu_h/\bar{\eta}_{h1}$ and $\eta_h^{*\prime}\mu_h$ converge to $-\gamma' y_t$ and the risks $\bar{\eta}_h'\Sigma_h\bar{\eta}_h/\bar{\eta}_{h1}^2$ and $\eta^{*\prime}\Sigma_h\eta^*$ converge to $\Gamma = 20.49$.

In Figure 5 *panel a* and *b* we have chosen $t = 2006 : 2$ and plotted, in *panel a*, the risk and expected return of the optimal hedging portfolio. They converge towards their limits, $\eta_h^{*\prime}\Sigma_h\eta_h^* \to \Gamma = 20.49$ and $\eta_h^{*\prime}\mu_h \to -\gamma' y_t = 20.26$, see (24-25). The same holds for the optimal Sharpe portfolio in *panel b*, see (31-32), when normalized by $\bar{\eta}_{h1}$, the coefficient to p_t, and it has for general h a higher return and a larger risk.

In *panel c* and *panel d* we have chosen $t = 2010 : 2$, where $\gamma' y_t = 2.57$ has the opposite sign, and we plot the same curves. We note that again the optimal Sharpe risk is larger

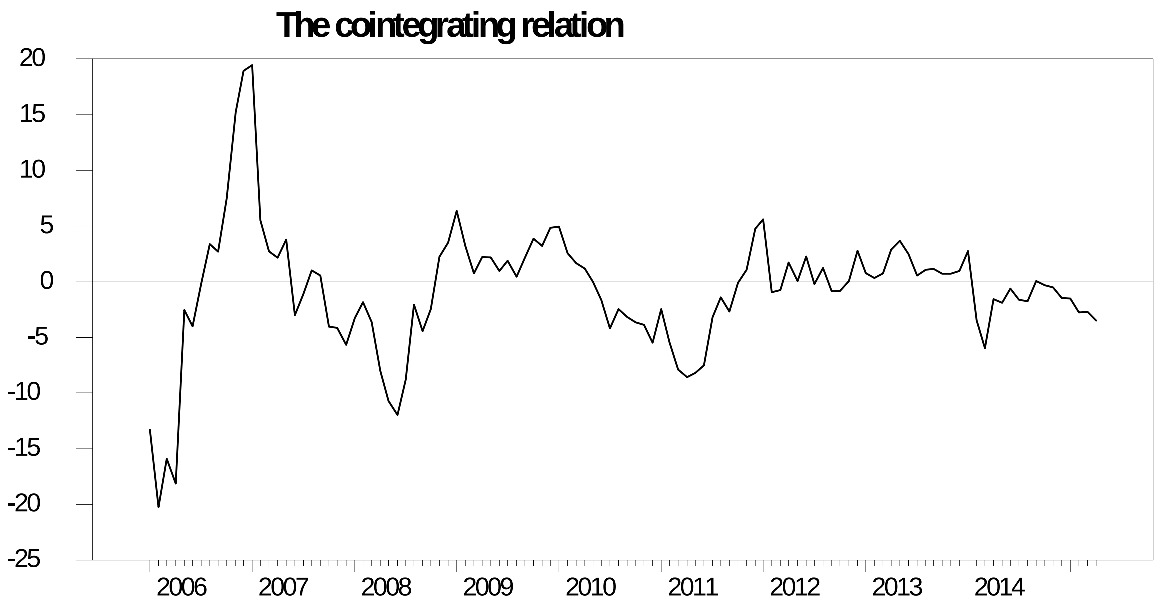

Figure 2: The cointegrating relation $p_t - 1.459 gas_t - 1.550 CO_{2t}$ only shows significant coeffcients for gas and CO_2. The adjustment coefficents to the changes in $p, coal, gas, CO_2$ are $\alpha' = (0, 0, 0, 0.163)$, so that only CO_2 is adjusting to disequilibrium.

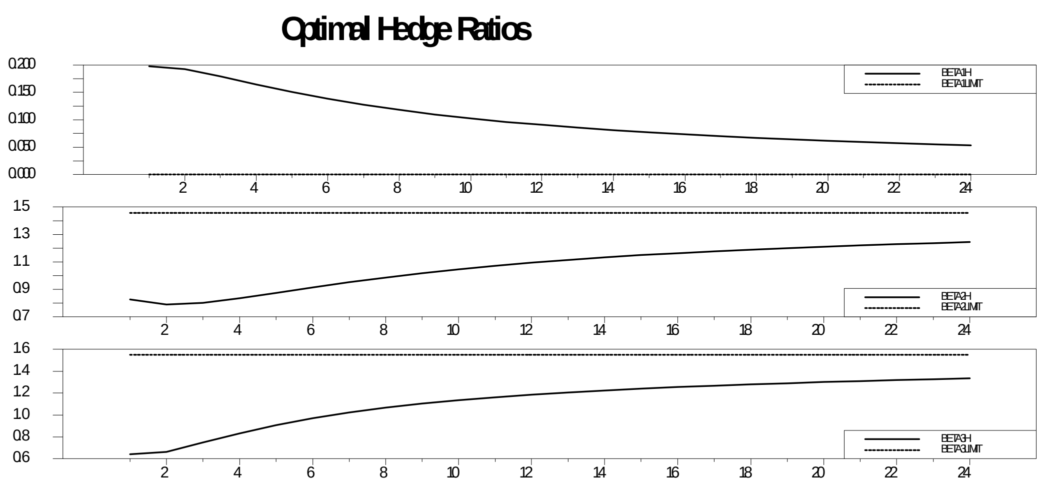

Figure 3: For $h = 1, \ldots, 24$, we plot the optimal hedge ratios $\beta_h^* = (\beta_{1h}, \beta_{2h}, \beta_{3h})'$ for the estimated model. It is seen how the hedge ratios converge to the coefficients of the cointegrating relation $(\beta_1, \beta_2, \beta_3) = (0, 1.459, 1.550)$ given as the dotted lines.

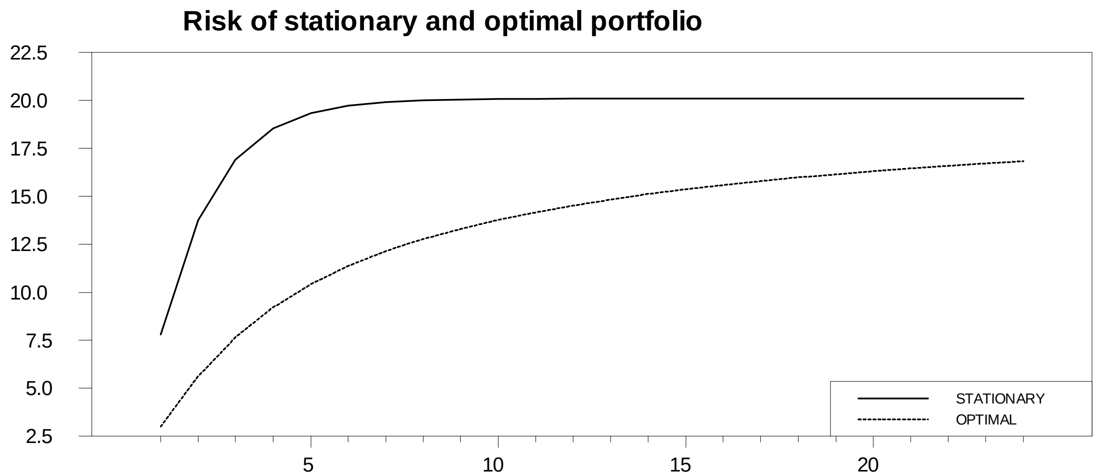

Figure 4: We plot the risk of the stationary portfolio $Var(\gamma' y_{t+h}|\mathcal{I}_t) = \gamma'\Sigma_h\gamma$, (——) which converges to Γ with an exponential rate, and the optimal risk $Var(\eta_h^{*\prime} y_{t+h}|\mathcal{I}_t) = \eta_h^{*\prime}\Sigma_h\eta_h^*$, $(\cdots\cdots)$ which converges to $\Gamma = 20.49$ like h^{-1}. The unhedged risk (not plotted) for asset one is $\Sigma_{h11} \approx 15.14 + 16.9(h-1)$, which goes from 15.14 to 420.74 for $h = 24$.

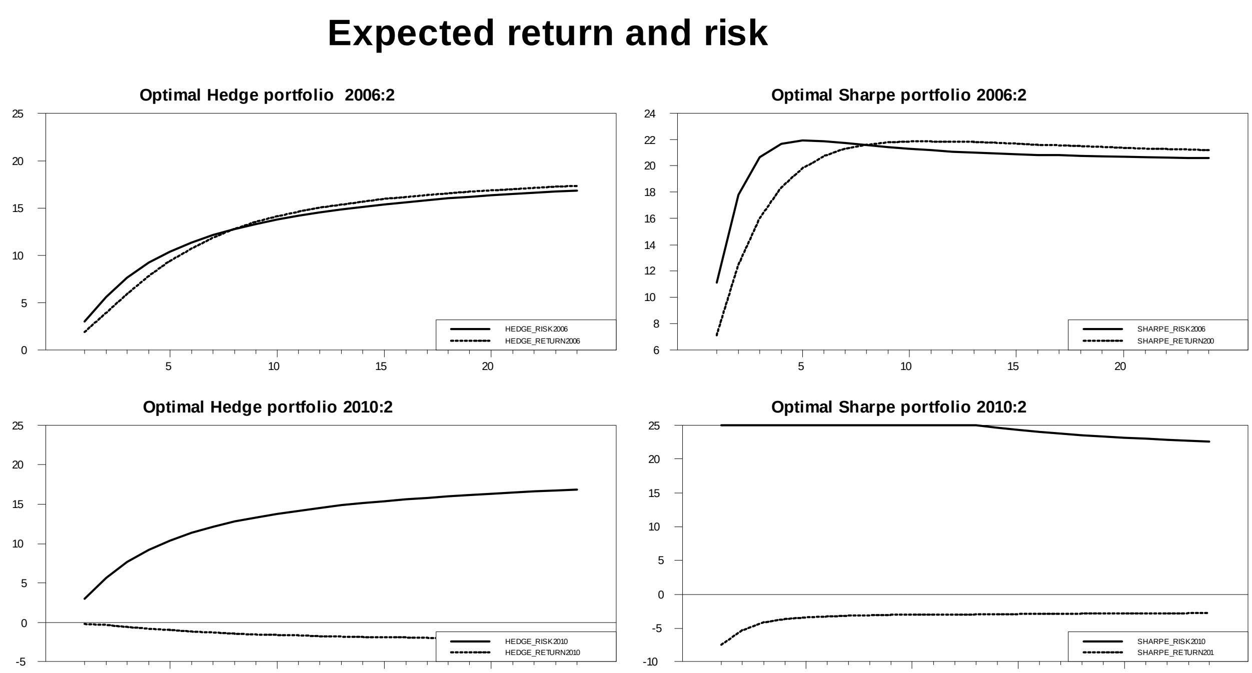

Figure 5: The expected return and risk of the optimal hedging portfolio and the optimal Sharpe portfolio for two different dates, plotted for horizon $h = 1, \ldots, 24$.

than the optimal hedging risk, but the expected returns are both negative now. The optimal Sharpe risk is larger than 25 for $h \leq 14$, and therefore truncated in the plot.

The conclusion of this is, that if we want to buy one unit of electricity and hedge using the fuels, then, if we start on February 2006 (where $\gamma' y_t = -20.26$), we can expect a positive return which converges to 20.26, and we can use the optimal Sharpe portfolio which has a higher expected return and a higher risk but the same limit. Thus, if the stationary relation takes a negative value at the time of investment, it pays to invest.

If, however, we start in February 2010 (where $\gamma' y_t = 2.57$) the optimal hedging portfolio has a much smaller risk than the unhedged portfolio, but we can expect a negative return as the price paid for getting rid of risk. The risk of the optimal Sharpe portfolio is larger than 25 for $h \leq 14$.

If, however, we want to go short in electricity, as the producer in the example above, then we have to change the sign of the portfolio, which leaves the risk the same but changes the sign of the expected return. Thus starting in February 2006 will imply a negative expected risk for the electricity producer and it would be better to start February 2010.

In summary. If $\gamma' y_t < 0$ it pays to go long in electricity, and if $\gamma' y_t > 0$ it pays to go short. Thus the electricity producer should sell the future in electricity in a month where $p_t - 1.459 gas_t - 1.550 CO_{2t} > 0$.

6 Conclusion

We have analysed the role of cointegration for hedging under the assumption that asset prices are driven by a CVAR. We have found the optimal hedging portfolio and optimal Sharpe ratio portfolio and compare 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, except in the special situation of strongly exogenous hedging assets. Thus the main advantage of hedging is the reduction of the risk. The minimum variance optimal portfolio does not take into account the expected return, and we therefore also analyse the optimal Sharpe 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 the first asset enters the optimal Sharpe portfolio with a positive coefficient, we can choose the portfolio as a hedging portfolio by normalizing on y_{1t}, and we have found a good balance between mean return and risk for the hedging problem. If, however, the coefficient to y_{1t} is negative, the corresponding hedging portfolio, normalized on y_1, will give the largest negative expected return in relation to its risk.

References

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

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

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.

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

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

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.)

Lütkepohl, H. 2005, New Introduction to Multiple Time Series Analysis, Springer, New York.

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.

Sharpe, W.F., 1966, Mutual Fund Performance, Journal of Business, 39 (S1), 119–138.

7 Appendix

Proof of Theorem 1. We find from model equation (1), that y_{2t} is a random walk in $n-1$ dimensions and that can be used to find $y_{1,t+h}$ and $y_{2,t+h}$ as function of y_{1t}, y_{2t} and the errors

$$y_{2,t+h} = y_{2t} + u_{2,t+1} + \cdots + u_{2,t+h},$$
$$y_{1,t+h} = \beta' y_{2,t+h} + u_{1,t+h} = \beta' y_{2t} + \beta' \sum_{i=0}^{h-1} u_{2,t+h-i} + u_{1,t+h}.$$

We find the expected return and prediction variance in (5) and (6). The best linear predictor is $\beta_h^* = (h\Psi_{22})^{-1}(h\Psi_{22}\beta + \Psi_{21}) = \beta + h^{-1}\Psi_{22}^{-1}\Psi_{21}$. We note in particular that for $h = 1$, $\beta_1^* = \beta + \Psi_{22}^{-1}\Psi_{21}$ and $\beta_h^* \to \beta$, $h \to \infty$, and that we can write $\beta + h^{-1}\Psi_{22}^{-1}\Psi_{21} = \beta_1^* h^{-1} + (1-h^{-1})\beta$, which proves (7), (8) and (9). ■

Proof of Theorem 2. For model (10) the cointegrating relation $\gamma' y_t$ is an $r-$dimensional $AR(1)$ process with autoregressive $r \times r$ parameter $\rho = I_r + \gamma'\alpha$, and $\gamma' y_t$ is given by the equation

$$\gamma' y_t - \xi = \rho(\gamma' y_{t-1} - \xi) + \gamma'\varepsilon_t.$$

By forward recursion from $i = t+1, \ldots, t+h$, we find that $\alpha'_\perp y_t$ is a random walk, and that

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

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

We combine these results using the identity

$$I_n = \gamma_\perp(\alpha_\perp'\gamma_\perp)^{-1}\alpha_\perp' + \alpha(\gamma'\alpha)^{-1}\gamma' = C + \alpha(\gamma'\alpha)^{-1}\gamma'.$$

This gives

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

From this we can find the conditional mean (12) and variance (13), and the optimal hedging portfolio, using (14). ■

Proof of Theorem 3. A cointegrating vector $\eta' = (1, -\beta')'$ is a linear combination of the vectors in γ, see (15), and therefore there exists a vector $(1, \kappa')' \in \mathbb{R}^r$ such that

$$\eta = \begin{pmatrix} 1 \\ -\beta \end{pmatrix} = \gamma \begin{pmatrix} 1 \\ \kappa \end{pmatrix} = \begin{pmatrix} 1 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} 0 \\ \beta_2 \end{pmatrix}\kappa = \begin{pmatrix} 1 \\ \beta_1 + \beta_2\kappa \end{pmatrix},$$

that is, $\eta' y_t = y_{1t} + \beta_1' y_{2t} + \kappa'\beta_2' y_{2t}$. The variance of this is $\Gamma_{11} + \kappa'\Gamma_{21} + \Gamma_{12}\kappa + \kappa'\Gamma_{22}\kappa$, which is minimized for

$$\kappa^* = -\Gamma_{22}^{-1}\Gamma_{21},$$

giving the optimal cointegrating portfolio (18) with mean and variance as given in (19) and (20).

Regressing y_{1t} on y_{2t} we find $\hat\beta$, which satisfies

$$\hat\beta - \beta_{stat}^* = (\sum_{t=1}^n y_{2t}y_{2t}')^{-1}\sum_{t=1}^n y_{2t}(y_{1t} - y_{2t}'\beta_{stat}^*).$$

We then analyse the matrices by pre and post multiplying by $B_T = (T^{-1/2}\beta_2, T^{-1}\beta_{2\perp})$ and find, using the rules that product moments of $I(1)$ variables are $O_P(T^2)$ and product moments of an $I(1)$ variable and an $I(0)$ variable is $O_P(T)$, that

$$B_T'\sum_{t=1}^n y_{2t}y_{2t}'B_T = O_P(1), \quad T^{-1}\sum_{t=1}^n \beta_{2\perp}' y_{2t}(y_{1t} - y_{2t}'\beta_{stat}^*) = O_P(1).$$

For details see Johansen (1996). Next we apply the law of large numbers for stationary (ergodic) processes and find using the definition of $\beta_{stat}^* = \beta_1 - \beta_2\Gamma_{22}^{-1}\Gamma_{21}$ that

$$T^{-1}\sum_{t=1}^n \beta_2' y_{2t}(y_{1t} - y_{2t}'\beta_{stat}^*) = T^{-1}\sum_{t=1}^n \beta_2' y_{2t}(y_{1t} - y_{2t}'\beta_1 + y_{2t}'\beta_2\Gamma_{22}^{-1}\Gamma_{21}) \xrightarrow{P} \Gamma_{21} - \Gamma_{22}\Gamma_{22}^{-1}\Gamma_{21} = 0.$$

This implies that because $\beta_2' y_{2t}(y_{1t} - y_{2t}'\beta_{stat}^*)$ is a stationary mean zero process,

$$\sum_{t=1}^n \beta_2' y_{2t}(y_{1t} - y_{2t}'\beta_{stat}^*) = O_P(T^{1/2}).$$

This means that, using $B_T^{-1} = (T^{-1/2}\bar{\beta}_2, T^{-1}\bar{\beta}_{2\perp})'$,

$$\begin{pmatrix} T^{1/2}\bar{\beta}_2'(\hat{\beta} - \beta^*_{stat}) \\ T\bar{\beta}_{2\perp}'(\hat{\beta} - \beta^*_{stat}) \end{pmatrix} = B_T^{-1}(\hat{\beta} - \beta^*_{stat}) = (B_T' \sum_{t=1}^n y_{2t}y_{2t}'B_T)^{-1}B_T' \sum_{t=1}^n y_{2t}(y_{1t} - y_{2t}'\beta^*_{stat}) = O_P(1),$$

which implies that $\hat{\beta} \xrightarrow{P} \beta^*_{stat}$ for $n \to \infty$. ∎

Proof of Theorem 4. *Proof of 1:* We find from equation (10) that

$$\mu_1 = E(\Delta y_t|\mathcal{I}_t) = \alpha(\gamma' y_t - \xi) + \Phi\Delta y_t, \quad \Sigma_1 = Var(\Delta y_t|\mathcal{I}_t) = \Omega, \quad \beta_1^* = \Omega_{22}^{-1}\Omega_{21},$$

which proves (21) and (22).

Proof of 2: The model with two lags (10) can be expressed in companion form as

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

We express that as the lag one model

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

where

$$\tilde{y}_t = \begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix}, \tilde{\alpha} = \begin{pmatrix} \alpha & \Phi \\ 0_{n\times r} & I_n \end{pmatrix}, \tilde{\gamma} = \begin{pmatrix} \gamma & I_n \\ 0_{n\times r} & -I_n \end{pmatrix},$$
$$\tilde{\xi} = \begin{pmatrix} \xi \\ 0_n \end{pmatrix}, \quad \tilde{\varepsilon}_t = \begin{pmatrix} \varepsilon_t \\ 0_n \end{pmatrix}, \quad \tilde{\Omega} = \begin{pmatrix} \Omega & 0 \\ 0 & 0 \end{pmatrix}.$$

We then find for $C = \gamma_\perp(\alpha_\perp'(I_n - \Phi)\gamma_\perp)^{-1}\alpha_\perp'$ the derived parameters,

$$\tilde{\alpha}_\perp = \begin{pmatrix} \alpha_\perp \\ -\Phi'\alpha_\perp \end{pmatrix}, \tilde{\gamma}_\perp = \begin{pmatrix} \gamma_\perp \\ \gamma_\perp \end{pmatrix}, \tilde{C} = \begin{pmatrix} C & -\Phi C \\ C & -\Phi C \end{pmatrix}, \quad \tilde{\rho} = \begin{pmatrix} I_r + \gamma'\alpha & \gamma'\Phi \\ \alpha & \Phi \end{pmatrix}.$$

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

$$\tilde{\Sigma}_h = h\tilde{C}\tilde{\Omega}\tilde{C}' + \tilde{\alpha}(\tilde{\gamma}'\tilde{\alpha})^{-1}(\sum_{i=0}^{h-1}\tilde{\rho}^i\tilde{\gamma}'\tilde{\Omega}\tilde{\gamma}\tilde{\rho}'^i)(\tilde{\alpha},\tilde{\gamma})^{-1}\tilde{\alpha}' \qquad (34)$$
$$+ \tilde{C}\tilde{\Omega}\tilde{\gamma}(\sum_{i=0}^{h-1}\tilde{\rho}'^i)(\tilde{\alpha},\tilde{\gamma})^{-1}\tilde{\alpha}' + \tilde{\alpha}(\tilde{\gamma}'\tilde{\alpha})^{-1}(\sum_{i=0}^{h-1}\tilde{\rho}^i)\tilde{\gamma}'\tilde{\Omega}\tilde{C}'.$$

We note that for $h \to \infty$,

$$\tilde{\gamma}'\tilde{\Sigma}_h\tilde{\gamma} = \sum_{i=0}^{h-1}\tilde{\rho}^i\tilde{\gamma}'\tilde{\Omega}\tilde{\gamma}\tilde{\rho}'^i \to \tilde{\Gamma} = Var(\tilde{\gamma}'\tilde{y}_t),$$
$$\sum_{i=0}^{h-1}\tilde{\rho}^i \to -(\tilde{\rho} - I_{r+n})^{-1} = -(\tilde{\gamma}'\tilde{\alpha})^{-1},$$
$$\tilde{\rho}^h \to 0,$$

and all converge exponentially fast, because the $I(1)$ condition implies that the absolute roots of the companion form are bounded by 1. We therefore replace all three by their limits in the limit argument below. We introduce the matrices

$$\tilde{\Theta} = (\tilde{\gamma}'\tilde{\alpha})^{-1}\tilde{\gamma}'(I_n, 0_{n\times n})'\Omega\alpha_\perp(\alpha_\perp'(I_n - \Phi)\gamma_\perp)^{-1},$$
$$\Upsilon = (\alpha_\perp'(I_n - \Phi)\gamma_\perp)^{-1}\alpha_\perp'\Omega\alpha_\perp(\gamma_\perp'(I_n - \Phi')\alpha_\perp)^{-1},$$

and find

$$(I_n, 0_{n\times n})\tilde{C}\tilde{\Omega}\tilde{C}'(I_n, 0_{n\times n})' = C\Omega C' = \gamma_\perp\Upsilon\gamma_\perp',$$
$$(I_n, 0_{n\times n})\tilde{C}\tilde{\Omega}\tilde{\gamma}(\tilde{\alpha}'\tilde{\gamma})^{-2}\tilde{\alpha}' = C\Omega(I_n, 0_{n\times n})\tilde{\gamma}(\tilde{\alpha}'\tilde{\gamma})^{-2}\tilde{\alpha}' = \gamma_\perp\tilde{\Theta}'(\tilde{\alpha}'\tilde{\gamma})^{-1}\tilde{\alpha}'.$$

For $\mu_h = (I_n, 0_{n\times n})\tilde{\mu}_h$ and $\Sigma_h = (I_n, 0_{n\times n})\tilde{\Sigma}_h(I_n, 0_{n\times n})'$ we therefore get

$$\mu_h = (I_n, 0_{n\times n})\tilde{\alpha}(\tilde{\gamma}'\tilde{\alpha})^{-1}(\tilde{\rho}^h - I_{r+n})(\tilde{\gamma}'\tilde{y}_t - \tilde{\xi}), \tag{35}$$
$$\begin{aligned}\Sigma_h = {} & h\gamma_\perp\Upsilon\gamma_\perp' + (I_n, 0_{n\times n})\tilde{\alpha}(\tilde{\gamma}'\tilde{\alpha})^{-1}\tilde{\Gamma}(\tilde{\alpha}, \tilde{\gamma})^{-1}\tilde{\alpha}'(I_n, 0_{n\times n})' \\ & - \gamma_\perp\tilde{\Theta}'(\tilde{\alpha}'\tilde{\gamma})^{-1}\tilde{\alpha}'(I_n, 0_{n\times n})' - (I_n, 0_{n\times n})\tilde{\alpha}(\tilde{\gamma}'\tilde{\alpha})^{-1}\tilde{\Theta}\gamma_\perp'.\end{aligned} \tag{36}$$

Next we introduce the notation

$$e_{n1}' = (1, 0_{n-1}', 0_n'),$$
$$\tilde{\alpha}_1' = e_{n1}'\tilde{\alpha}(\tilde{\gamma}'\tilde{\alpha})^{-1},$$

for the first unit vector in $\mathbb{R}^{2n}$, and the first row of the $2n \times (r+n)$ matrix $\tilde{\alpha}(\tilde{\gamma}'\tilde{\alpha})^{-1}$. We also need the next $n-1$ rows of the matrix $\tilde{\alpha}(\tilde{\gamma}'\tilde{\alpha})^{-1}$, and define

$$e_{n2}' = (0_{n-1}, I_{n-1}, 0_{(n-1)\times n}),$$
$$\tilde{\alpha}_2' = e_{n2}'\tilde{\alpha}(\tilde{\gamma}'\tilde{\alpha})^{-1},$$

such that e_{n2}' consists of the $n-1$ unit vectors in $\mathbb{R}^{2n}$, which picks out the rows $2, \ldots, n$ of $\tilde{\alpha}(\tilde{\gamma}'\tilde{\alpha})^{-1}$.

We use below the simplifying relations

$$\tilde{\alpha}_1' + \beta_1'\tilde{\alpha}_2 = (e_{n1}' + \beta_1'e_{n2}')\tilde{\alpha}(\tilde{\gamma}'\tilde{\alpha})^{-1} = \tilde{\gamma}_1'\tilde{\alpha}(\tilde{\gamma}'\tilde{\alpha})^{-1} = (1, 0_{r-1+n}') = e_{r1}', \tag{37}$$
$$\beta_2'\tilde{\alpha}_2' = \beta_2'e_{n2}'\tilde{\alpha}(\tilde{\gamma}'\tilde{\alpha})^{-1} = \tilde{\gamma}_2'\tilde{\alpha}(\tilde{\gamma}'\tilde{\alpha})^{-1} = (0_{r-1}, I_{r-1}, 0_{(n-1)\times n}) = e_{r2}', \tag{38}$$

say. We next want to derive expressions for $\Sigma_{h22}, \Sigma_{h21}$, and Σ_{h11} and note that

$$\gamma = \begin{pmatrix} 1 & 0 \\ \beta_1 & \beta_2 \end{pmatrix}, \gamma_\perp = \begin{pmatrix} -\beta_1'\beta_{2\perp} \\ \beta_{2\perp} \end{pmatrix},$$

which implies that from (36) we find the expressions

$$\Sigma_{h22} = h\beta_{2\perp}\Upsilon\beta_{2\perp}' + \tilde{\alpha}_2'\tilde{\Gamma}\tilde{\alpha}_2 - \beta_{2\perp}\tilde{\Theta}'\tilde{\alpha}_2 - \tilde{\alpha}_2'\tilde{\Theta}\beta_{2\perp}',$$
$$\Sigma_{h21} = -n\beta_{2\perp}\Upsilon\beta_{2\perp}'\beta_1 + \tilde{\alpha}_2'\tilde{\Gamma}\tilde{\alpha}_1 - \beta_{2\perp}\tilde{\Theta}'\tilde{\alpha}_1 + \tilde{\alpha}_2'\tilde{\Theta}\beta_{2\perp}'\beta_1,$$
$$\Sigma_{h11} = n\beta_1'\beta_{2\perp}\Upsilon\beta_{2\perp}'\beta_1 + \tilde{\alpha}_1'\tilde{\Gamma}\tilde{\alpha}_1 + \beta_1'\beta_{2\perp}\tilde{\Theta}'\tilde{\alpha}_1 + \tilde{\alpha}_1'\tilde{\Theta}\beta_{2\perp}'\beta_1.$$

We see from (36) that Σ_{h22} tends to infinity, and in order to analyse Σ_{h22}, its inverse, and the limit of the best linear predictor, $\beta_h^* = \Sigma_{h22}^{-1}\Sigma_{h21}$, we introduce the normalizing matrices

$$A_h = (\beta_2, h^{-1}\bar{\beta}_{2\perp}),\ A = (\beta_2, \bar{\beta}_{2\perp}),$$

where $\bar{\beta}_{2\perp} = \beta_{2\perp}(\beta_{2\perp}'\beta_{2\perp})^{-1}$, such that $\beta_{2\perp}'\bar{\beta}_{2\perp} = I_{r-1}$, and find

$$\beta_h^* = \Sigma_{h22}^{-1}\Sigma_{h21} = A(A_h'\Sigma_{h22}A)^{-1}A_h'\Sigma_{h21}.$$

Using $\beta_{2\perp}'\bar{\beta}_{2\perp} = I_{r-1}$ and $\beta_2'\tilde{\alpha}_2'\tilde{\Gamma}\tilde{\alpha}_2\beta_2 = e_{r2}'\tilde{\Gamma}e_{r2} = Var(\beta_2'y_{2t}) = \Gamma_{22}$, see (17), we find

$$A_h'\Sigma_{h22}A = \begin{pmatrix} \Gamma_{22} & e_{r2}'(\tilde{\Gamma}\tilde{\alpha}_2\bar{\beta}_{2\perp} - \tilde{\Theta}) \\ 0 & \Upsilon \end{pmatrix} + O(h^{-1}),$$

$$A_h'\Sigma_{h21} = \begin{pmatrix} e_{r2}'(\tilde{\Gamma}\tilde{\alpha}_1 + \tilde{\Theta}\beta_{2\perp}'\beta_1) \\ -\Upsilon\beta_{2\perp}'\beta_1 \end{pmatrix} + O(h^{-1}).$$

Hence for $h \to \infty$,

$$\begin{aligned}
&A(A_h'\Sigma_{h22}A)^{-1}A_h'\Sigma_{h21} \\
&\to (\beta_2, \bar{\beta}_{2\perp})\begin{pmatrix} \Gamma_{22}^{-1} & -\Gamma_{22}^{-1}e_{r2}'(\tilde{\Gamma}\tilde{\alpha}_2\bar{\beta}_{2\perp} - \tilde{\Theta})\Upsilon^{-1} \\ 0 & \Upsilon^{-1} \end{pmatrix}\begin{pmatrix} e_{r2}'(\tilde{\Gamma}\tilde{\alpha}_1 + \tilde{\Theta}\beta_{2\perp}'\beta_1) \\ -\Upsilon\beta_{2\perp}'\beta_1 \end{pmatrix} \\
&= \beta_2\Gamma_{22}^{-1}e_{r2}'(\tilde{\Gamma}\tilde{\alpha}_1 + \tilde{\Theta}\beta_{2\perp}'\beta_1 + (\tilde{\Gamma}\tilde{\alpha}_2\bar{\beta}_{2\perp} - \tilde{\Theta})\beta_{2\perp}'\beta_1) - \bar{\beta}_{2\perp}\beta_{2\perp}'\beta_1 \\
&= \beta_2\Gamma_{22}^{-1}e_{r2}'\tilde{\Gamma}(\tilde{\alpha}_1 + \tilde{\alpha}_2\bar{\beta}_{2\perp}\beta_{2\perp}'\beta_1) - \bar{\beta}_{2\perp}\beta_{2\perp}'\beta_1.
\end{aligned}$$

In this expression we find, using (37) and (38),

$$\tilde{\alpha}_1 + \tilde{\alpha}_2\bar{\beta}_{2\perp}\beta_{2\perp}'\beta_1 = \tilde{\alpha}_1 + \tilde{\alpha}_2\beta_1 - \tilde{\alpha}_2\beta_2\bar{\beta}_2'\beta_1 = e_{n1} - e_{n2}\bar{\beta}_2'\beta_1.$$

Inserting this we find the limit for $h \to \infty$,

$$\begin{aligned}
A(A_h'\Sigma_{h22}A)^{-1}A_h'\Sigma_{h21} &\to \beta_2\Gamma_{22}^{-1}e_{n2}'\tilde{\Gamma}(e_{n1} - e_{n2}\bar{\beta}_2'\beta_1) - \bar{\beta}_{2\perp}\beta_{2\perp}'\beta_1 \\
&= \beta_2\Gamma_{22}^{-1}\Gamma_{21} - \beta_2\Gamma_{22}^{-1}\Gamma_{22}\bar{\beta}_2'\beta_1 - \bar{\beta}_{2\perp}\beta_{2\perp}'\beta_1 = \beta_2\Gamma_{22}^{-1}\Gamma_{21} - \beta_1,
\end{aligned}$$

because $\beta_2\bar{\beta}_2'\beta_1 + \bar{\beta}_{2\perp}\beta_{2\perp}'\beta_1 = \beta_1$. This proves (23).

Next we find for $h \to \infty$ the limiting expected return using $\gamma'(I_n, 0_{n\times n}) = (I_r, 0_{n\times n})\tilde{\gamma}'$

$$\begin{aligned}
\eta^{*\prime}\mu_h &= (1, -\Sigma_{h12}\Sigma_{h22}^{-1}, 0_{n\times n})\tilde{\alpha}(\tilde{\gamma}'\tilde{\alpha})^{-1}(\tilde{\rho}^h - I_{r+n})(\tilde{\gamma}'\tilde{y}_t - \tilde{\xi}) \qquad (39) \\
&\to -(1, \beta_1' - \Gamma_{12}\Gamma_{22}^{-1}\beta_2', 0_{n\times n})\tilde{\alpha}(\tilde{\gamma}'\tilde{\alpha})^{-1}(\tilde{\gamma}'\tilde{y}_t - \tilde{\xi}) \\
&= -(1, -\Gamma_{12}\Gamma_{22}^{-1})(I_r, 0_{r\times n})\tilde{\gamma}'\tilde{\alpha}(\tilde{\gamma}'\tilde{\alpha})^{-1}(\tilde{\gamma}'\tilde{y}_t - \tilde{\xi}) \\
&= -(1, -\Gamma_{12}\Gamma_{22}^{-1})(\gamma' y_t - \xi),
\end{aligned}$$

which proves (24).

From the above we find that the optimal portfolio satisfies

$$\eta_h^* = \eta_{stat}^* + r_h,$$

where $r_h = O(h^{-1})$, such that the minimal variance is

$$\eta_h^{*\prime}\Sigma_h\eta_h^* = (\eta_{stat}^* + r_h)'\Sigma_h(\eta_{stat}^* + r_h) = \eta_{stat}^{*\prime}\Sigma_h\eta_{stat}^* + 2r_h'\Sigma_h\eta_{stat}^* + r_h'\Sigma_h r_h.$$

Here the first term, using $\gamma'(I_n, 0_{n\times n}) = (I_r, 0_{n\times n})\tilde{\gamma}'$, is

$$\begin{aligned}\eta_{stat}^{*\prime}\Sigma_h\eta_{stat}^* &= (1, -\Gamma_{12}\Gamma_{22}^{-1})\gamma(I_n, 0_{n\times n})\tilde{\alpha}(\tilde{\gamma}'\tilde{\alpha})^{-1}\tilde{\Gamma}(\tilde{\alpha},\tilde{\gamma})^{-1}\tilde{\alpha}'(I_n, 0_{n\times n})\gamma(1, -\Gamma_{12}\Gamma_{22}^{-1})' \\ &= (1, -\Gamma_{12}\Gamma_{22}^{-1})\begin{pmatrix}\Gamma_{11} & \Gamma_{12}\\ \Gamma_{21} & \Gamma_{22}\end{pmatrix}(1, -\Gamma_{12}\Gamma_{22}^{-1})' = \Gamma_{11} - \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{21},\end{aligned}$$

and the remaining terms are $O(h^{-1})$ because $\eta_{coint}'\Sigma_h = O(h^{-1})$, which proves (25). ■

Proof of Theorem 5. Introducing $a = \Sigma_h^{1/2}\eta$ and $b = \Sigma_h^{-1/2}\mu_h$, we find $\eta'\mu_h = a'b$ and

$$\sup_\eta \frac{(\eta'\mu_h)^2}{\eta'\Sigma_h\eta} = b'b\sup_a \frac{(a'b)^2}{(a'a)(b'b)} \leq b'b = \mu_h'\Sigma_h^{-1}\mu_h,$$

by the Cauchy-Schwarz inequality, where equality holds for $a = cb$ or $\eta = c\Sigma_h^{-1}\mu_h$. This proves (27) and (28). See also Gourieroux and Jasiak (2001 pp. 74–76). If we restrict $\eta = \gamma\kappa$, $\kappa \in \mathbb{R}^r$, then

$$\frac{(\eta'\mu_h)^2}{\eta'\Sigma_h\eta} = \frac{(\kappa'\gamma'\mu_h)^2}{\kappa'\gamma'\Sigma_h\gamma\kappa},$$

such that for the optimal $\bar{\kappa}$ we find $\bar{\eta}_{h,stat} = \gamma\bar{\kappa} = \gamma(\gamma'\Sigma_h\gamma)^{-1}\gamma'\mu_h$. To find the limit we note that from (35) we find for $h \to \infty$,

$$\begin{aligned}\gamma'\mu_h &= \gamma'(I_n, 0_{n\times n})\tilde{\alpha}(\tilde{\gamma}'\tilde{\alpha})^{-1}(\tilde{\rho}^h - I_{r+n})(\tilde{\gamma}'\tilde{y}_t - \tilde{\xi}) \\ &= (I_r, 0_{n\times n})\tilde{\gamma}'\tilde{\alpha}(\tilde{\gamma}'\tilde{\alpha})^{-1}(\tilde{\rho}^h - I_{r+n})(\tilde{\gamma}'\tilde{y}_t - \tilde{\xi}) \\ &\to -(I_r, 0_{n\times n})(\tilde{\gamma}'\tilde{y}_t - \tilde{\xi}) = -(I_r, 0_{n\times n})\begin{pmatrix}\gamma'y_t - \xi \\ y_t - y_{t-1}\end{pmatrix} = -(\gamma'y_t - \xi).\end{aligned}$$

From (36) we find similarly

$$\gamma'\Sigma_h\gamma = \gamma'(I_n, 0_{n\times n})\tilde{\alpha}(\tilde{\gamma}'\tilde{\alpha})^{-1}\tilde{\Gamma}(\tilde{\alpha},\tilde{\gamma})^{-1}\tilde{\alpha}'(I_n, 0_{n\times n})'\gamma \to (I_r, 0_{n\times n})\tilde{\Gamma}(I_r, 0_{n\times n})' = \Gamma.$$

Thus

$$\bar{\eta}_{h,stat} = \gamma(\gamma'\Sigma_h\gamma)^{-1}\gamma'\mu_h \to -\gamma\Gamma^{-1}(\gamma'y_t - \xi) = \bar{\eta}_{stat}, \quad \text{for } h \to \infty.$$

■

Proof of Theorem 6. *Proof of 1*: For $h = 1$ we get from the model equations (10), that $\mu_1 = E(\Delta y_{t+1}|\mathcal{I}_t) = \alpha(\gamma'y_t - \xi) + \Phi\Delta y_t$ and $Var(\Delta y_{t+1}|\mathcal{I}_t) = \Omega$, which shows (30) and (31).

Proof of 2: From Theorem 5 we find the optimal Sharpe portfolio as

$$\bar{\eta}_h = \Sigma_h^{-1}\mu_h = [(I_n, 0)\tilde{\Sigma}_h(I_n, 0)']^{-1}(I_n, 0)\tilde{\mu}_h,$$

and the expressions (35) and (36) are valid for $\tilde{\mu}_h$ and $\tilde{\Sigma}_h$.

We introduce the matrices $B = (\gamma, \bar{\gamma}_\perp)$ and $B_h = (\gamma, h^{-1}\bar{\gamma}_\perp)$, $(\bar{\gamma}_\perp = (\gamma_\perp'\gamma_\perp)^{-1}\gamma_\perp)$ and find using $\gamma'(I_n, 0_{n\times n}) = (I_r, 0_{r\times n})\tilde{\gamma}'$ that from (36)

$$\begin{aligned}\Sigma_h^{-1} &= ((I_n,0)\tilde{\Sigma}_h(I_n,0)')^{-1} = B(B_h'(I_n,0)\tilde{\Sigma}_h(I_n,0)'B)^{-1}B_h' \\ &\to (\gamma, \bar{\gamma}_\perp)\begin{pmatrix} \Gamma & \gamma'(I_n,0)\tilde{\Sigma}_h(I_n,0)'\bar{\gamma}_\perp \\ 0 & \Upsilon \end{pmatrix}^{-1}\begin{pmatrix} \gamma' \\ 0 \end{pmatrix} = \gamma\Gamma^{-1}\gamma'.\end{aligned}$$

Similarly we find from (35) that for $h \to \infty$,

$$\bar{\eta}_h = \Sigma_h^{-1}\mu_h \to -\gamma\Gamma^{-1}\gamma'(I_n, 0_{n\times n})\tilde{\alpha}(\tilde{\gamma}'\tilde{\alpha})^{-1}(\tilde{\gamma}'\tilde{y}_t - \tilde{\xi}) = -\gamma\Gamma^{-1}(\gamma' y_t - \xi) = \bar{\eta}_{stat}.$$

Finally we find the limit of the optimal variance:

$$\tilde{\mu}_h'(I_n,0)'[(I_n,0)\tilde{\Sigma}_h(I_n,0)']^{-1}(I_n,0)\tilde{\mu}_h = \bar{\eta}_h'(I_n,0)\tilde{\mu}_h \to (\gamma' y_t - \xi)'\Gamma^{-1}(\gamma' y_t - \xi).$$

■