

Article

Return and Risk of Pairs Trading Using a Simulation-Based Bayesian Procedure for Predicting Stable Ratios of Stock Prices

David Ardia ^{1,2}, Lukasz T. Gatarek ³, Lennart Hoogerheide ⁴ and Herman K. van Dijk ^{4,5,*}

¹ Institute of Financial Analysis, University of Neuchatel, Neuchatel, 2000, Switzerland; david.ardia@unine.ch

² Department of Finance, Insurance and Real Estate, Laval University, Quebec City, G1V 0A6, Canada

³ Institute of Econometrics and Statistics, Faculty of Economics and Sociology, University of Lodz, Lodz, 90-255, Poland; gatarek@tlen.pl

⁴ Department of Econometrics and Tinbergen Institute, Vrije Universiteit Amsterdam, Amsterdam, 1081 HV, The Netherlands; l.f.hoogerheide@vu.nl

⁵ Econometric Institute, Erasmus University Rotterdam, Rotterdam, 3062 PA, The Netherlands

* Correspondence: hkvandijk@ese.eur.nl; Tel.: +31-(0)10-4088900

Academic Editors: Nalan Baştürk, Francesco Ravazzolo and Roberto Casarin

Received date: 3 September 2015; Accepted: 28 January 2016; Published: 10 March 2016

Abstract: We investigate the direct connection between the uncertainty related to estimated stable ratios of stock prices and risk and return of two pairs trading strategies: a conditional statistical arbitrage method and an implicit arbitrage one. A simulation-based Bayesian procedure is introduced for predicting stable stock price ratios, defined in a cointegration model. Using this class of models and the proposed inferential technique, we are able to connect estimation and model uncertainty with risk and return of stock trading. In terms of methodology, we show the effect that using an encompassing prior, which is shown to be equivalent to a Jeffreys' prior, has under an orthogonal normalization for the selection of pairs of cointegrated stock prices and further, its effect for the estimation and prediction of the spread between cointegrated stock prices. We distinguish between models with a normal and Student t distribution since the latter typically provides a better description of daily changes of prices on financial markets. As an empirical application, stocks are used that are ingredients of the Dow Jones Composite Average index. The results show that normalization has little effect on the selection of pairs of cointegrated stocks on the basis of Bayes factors. However, the results stress the importance of the orthogonal normalization for the estimation and prediction of the spread—the deviation from the equilibrium relationship—which leads to better results in terms of profit per capital engagement and risk than using a standard linear normalization.

Keywords: Bayesian analysis; cointegration; linear normalization; orthogonal normalization; pairs trading; statistical arbitrage

JEL: C11; C15; C32; C58; G17

1. Introduction

In this paper we consider statistical arbitrage strategies. Such strategies presume that the patterns observed in the historical data are expected to be repeated in the future. That is, a statistical arbitrage is a purely advanced descriptive approach designed to exploit market inefficiencies.

Khandani and Lo [1] consider a specific strategy—first proposed by Lehmann [2] and Lo and MacKinlay [3]—that can be analyzed directly using individual equities returns. Given a collection of securities, they consider a long/short market-neutral equity strategy consisting of an equal dollar amount of long and short positions, where at each rebalancing interval, the long positions are made up of “losers” (underperforming stocks, relative to some market average) and the short positions are made up of “winners” (outperforming stocks, relative to the same market average). By buying yesterday’s losers and selling yesterday’s winners at each date, such a strategy actively bets on mean reversion across all stocks, profiting from reversals that occur within the rebalancing interval. For this reason, such strategies have been called “contrarian” trading strategies that benefit from market overreaction, *i.e.*, when underperformance is followed by positive returns and *vice-versa* for outperformance. The same key idea is the basis of pairs trading strategies, which constitute another form of statistical arbitrage strategies.

The idea of pairs trading relies on long-term equilibrium among a pair of stocks. If such an equilibrium exists, then it is presumed that a specific linear combination of prices reverts to zero. A trading rule can be set up to exploit the temporary deviations (spread) to generate profit. When the spread between two assets is positive it is sold; that is, the outperforming stock is shorted and the long position is opened in the underperforming stock. In the opposite case, when the spread is negative: one buys. Gatev *et al.* [4] investigate the performance of this arbitrage rule over a period of 40 years and they find huge empirical evidence in favor of it. It is fundamental for the pairs trading strategy to precisely estimate the current and expected spread among the stock prices.

In this paper we interpret spread as the temporary deviation from the equilibrium in a cointegration model. Equilibrium in a cointegration model is interpreted as time series behavior that is characterized by stable, or otherwise stated stationary, long-run relations to which actual series return after temporary deviations. This approach differs from Gatev *et al.* [4], who implement a nonparametric framework. These authors choose a matching partner for each stock by finding the security that minimizes the sum of squared deviations between the two normalized price series; pairs are thus formed by exhaustive matching between normalized daily prices, where price includes reinvested dividends. However, as argued above, in the cointegration analysis that we perform, the spread between two assets is modeled as the temporary deviation from the long-run stable relations among the time series of asset prices. This deviation is computed as a linear combination of stock prices, where the weights in the linear combination are given by the cointegrating vector. Long-run stability also implies that there exists a finite uncertainty in the predictability of stock prices that can be used in devising trading strategies. Therefore, pairs trading strategies are strongly dependent on the stability of ratios of pairs of stocks.

The estimated and predicted spreads are both computed from the estimated cointegration model. We introduce a simulation-based Bayesian estimation procedure that allows us to combine estimation and model uncertainty in a natural way with decision uncertainty associated with a decision process like a trading strategy. For the Bayesian estimation of the cointegration model, we work with a Metropolis-Hastings (M-H) type of sampler derived under an encompassing prior where we show that the encompassing prior is equivalent under certain conditions to the well-known Jeffreys’ or Information matrix prior. This sampling algorithm is derived by Kleibergen and Van Dijk [5] for the Simultaneous Equations Model and extended by Kleibergen and Paap [6] for the cointegration model. The latter authors specify a linear normalization to identify the parameters in the model. However, Strachan and Van Dijk [7] point at possible distortions of prior beliefs associated with the linear normalization. Moreover, in our application we find out that the distribution of the spread is particularly sensitive to the choice of normalization.

Therefore we make use of an alternative normalization, the orthogonal normalization, in order to identify the parameters in the cointegration model. Given that one is usually only interested in a linear combination of price series, this normalization is a natural one since it treats the variables in the series in a symmetric way. More details are given in Section 3. Hence, we implement the M-H

sampler for the cointegration model under this normalization. We compare the performance of the pairs trading strategy under the orthogonal normalization with the performance of the counterpart under the linear normalization and find that, for our set of data, the orthogonal normalization is highly favored over the linear normalization with respect to the profitability and risk of the trading strategies.

The results imply that within the statistical arbitrage approach of pairs trading based on the cointegration model, the normalization is not only a useful device easing the parameter identification but it primarily becomes an important part of the model.

To take into account the non-normality of the conditional distribution of daily returns, we extend our approach of using the normal distribution to the case of the Student- t distribution.

The outline of the paper is as follows. In Section 2 the conditional and implicit statistical arbitrage approaches are discussed. In Section 3 our Bayesian analysis of the cointegration model under the encompassing prior is explained. In Section 4 we consider an empirical application using stocks in the Dow Jones Composite Average index. Section 5 concludes. The appendices contain technical derivations and additional tables with detailed results from our empirical application.

2. Pairs Trading: Implicit and Conditional Statistical Arbitrage

Suppose that there exists a *statistical fair price relationship* [8] between the prices $y_{t,1}$ and $y_{t,2}$ of two stocks, where the spread

$$s_t = \beta_1 y_{t,1} + \beta_2 y_{t,2} \quad (1)$$

is the deviation from this statistical fair price relationship, or “statistical mispricing”, at the end of day t . In this paper we consider two types of trading strategies that are based upon the existence of such a long-run equilibrium relationship: conditional statistical arbitrage (CSA) and implicit statistical arbitrage (ISA), where we use the classification of Burgess [8]. We will implement these strategies in such a way that at the end of each day the holding is updated, after which the holding is kept constant for a day. In the CSA strategy the desired holding at the end of day t is given by

$$CSA(s_t, k) = \text{sign}(E(\Delta s_{t+1} | \mathcal{I}_t)) |E(\Delta s_{t+1} | \mathcal{I}_t)|^k, \quad (2)$$

where \mathcal{I}_t is the information set at the end of day t , and where we consider $k = 0$ and $k = 1$. A positive value of $CSA(s_t, k)$ means that we buy $CSA(s_t, k)$ spreads and a negative value of $CSA(s_t, k)$ means that we short $CSA(s_t, k)$ spreads. That is, if $\beta_1 > 0$ and $\beta_2 < 0$, then a positive value of $CSA(s_t, k)$ means that we buy $\beta_1 \times CSA(s_t, k)$ of stock 1 and short $(-\beta_2) \times CSA(s_t, k)$ of stock 2. For $k = 1$ the obvious intuition of the CSA strategy is that we want to invest more in periods with larger expected profits. In this way we consider the *accuracy* of the used method. In the case of $k = 0$ we only look at the sign of the expected change in the next day. In this way we consider the *directional accuracy* of the used method. Note that the expectation in (2) is taken over the distribution of Δs_{t+1} (given the information set \mathcal{I}_t and the ‘fixed’ values of β_1 and β_2). In the sequel of this paper, we use the posterior median to obtain estimates of model parameters, where the expectation in (2) will still be taken given these “fixed” estimated values. We use the posterior median, since the posterior distribution has Cauchy type tails in one of the model specifications that we investigate and these Cauchy type tails imply that the coefficients have no posterior means.

In the ISA strategy the desired holding at the end of day t is given by:

$$ISA(s_t) = -s_t. \quad (3)$$

A positive value of $ISA(s_t)$ means that we buy $ISA(s_t)$ spreads and a negative value of $ISA(s_t)$ means that we short $ISA(s_t)$ spreads. Or equivalently, a negative value of s_t means that we buy $-s_t$ spreads and a positive value of s_t means that we short s_t spreads. That is, if $\beta_1 > 0$ and $\beta_2 < 0$, then a positive value of $ISA(s_t)$ means that we buy $\beta_1 \times ISA(s_t) = \beta_1 \times (-s_t)$ of stock 1 and short

$(-\beta_2) \times (-ISA(s_t)) = (-\beta_2) \times s_t$ of stock 2. In the sequel of this paper, we will substitute the posterior medians of β_1 and β_2 to obtain an estimate of the spread in (1).

The CSA and ISA strategies raise several questions. First, how do we define such long-run equilibrium relationships? How are the coefficients β_1 and β_2 estimated? Second, how do we find pairs of stocks that satisfy such a long-run equilibrium relationship? Third, how do we estimate how the stock prices adjust towards their long-run equilibrium relationship? In the next section, we consider how our Bayesian analysis of the cointegration model (under linear or orthogonal normalization) provides answers to all these questions. In order to answer the first and third questions we use the posterior distribution (more precisely, the posterior median) of the parameters in the cointegration model. In order to answer the second question we compute the Bayes factor of a model with a cointegration relationship *versus* a model without a cointegration relationship for a large number of pairs of stocks.

At this point, we stress why we make use of the CSA and ISA strategies, rather than the approach of Gatev *et al.* [4]. In the strategy of Gatev *et al.* [4] a holding is taken as soon as it is found that a pair of prices has substantially diverged. After that, the holding remains constant until the prices have completely converged to the equilibrium relationship. A disadvantage of that trading strategy is that there is not much trading going on (*i.e.*, in most periods there is no trading at all), which makes it more difficult to investigate the difference in quality between different models given a finite period, or equivalently a very long period may be required to be able to find substantially credible differences in trading results between models.

3. Bayesian Analysis of the Cointegration Model Under Linear and Orthogonal Normalization

Consider a vector autoregressive model of order 1 (VAR(1)) for an n -dimensional vector of time series $\{Y_t\}_{t=1}^T$

$$Y_t = \Phi Y_{t-1} + \varepsilon_t, \quad (4)$$

ε_t is an independent n -dimensional vector normal process with zero mean and $n \times n$ positive definite symmetric (PDS) covariance matrix Σ . We will consider two alternative distributions for ε_t : a multivariate normal distribution and a multivariate Student's t distribution. Φ is an $n \times n$ matrix with autoregressive coefficients. The initial values in Y_1 are assumed fixed. The VAR model in (4) can be written in error correction form

$$\Delta Y_t = \Pi' Y_{t-1} + \varepsilon_t, \quad (5)$$

where $\Pi' = \Phi - I_n$ (with I_n the $n \times n$ identity matrix) is the long-run multiplier matrix, see *e.g.*, Johansen [9] and Kleibergen and Paap [6].

If Π is a zero matrix, the series Y_t contains n unit roots and there is no opportunity for long term predictability with finite uncertainty. If the matrix Π has full rank, the univariate series in Y_t are stationary and long-run equilibrium relations are assumed to hold. Cointegration appears if the rank Π equals r with $0 < r < n$. The matrix Π' can be written as the outer product of two full rank $n \times r$ matrices α' and β :

$$\Pi' = \alpha' \beta'.$$

The matrix β contains the cointegration vectors, which reflect the stationary long-run (equilibrium) relations between the univariate series in Y_t ; that is, each element of $\beta' Y_t$ can be interpreted as a temporary deviation from a long-run (equilibrium) relations. The matrix α contains the adjustment parameters, which indicate the speed of adjustment to the long-run (equilibrium) relations.

To save on notation, we write (5) in matrix notation

$$\Delta Y = Y_{-1} \Pi + \varepsilon$$

with $(T - 1) \times n$ matrices $\Delta Y = (\Delta Y_2, \dots, \Delta Y_T)'$, $Y_{-1} = (Y_1, \dots, Y_{T-1})'$ and $\varepsilon = (\varepsilon_2, \dots, \varepsilon_T)$.

Under the cointegration restriction $\Pi = \beta\alpha$, this model is given by: ¹

$$\Delta Y = Y_{-1}\beta\alpha + \varepsilon. \tag{6}$$

The individual parameters in $\beta\alpha$ are non-identified as $\beta\alpha = \beta B B^{-1}\alpha$ for any nonsingular $r \times r$ matrix B . That is, postmultiplying β by an invertible matrix B and premultiplying α by its inverse leaves the matrix $\beta\alpha$ unchanged. Therefore, $r \times r$ identification restrictions are required to identify the elements of β and α , so that these become estimable. In this paper we will consider two different normalization restrictions for identification purposes. The first normalization is the linear normalization, which is commonly used, where we have

$$\beta = \begin{pmatrix} I_r \\ \beta_2 \end{pmatrix}. \tag{7}$$

That is, the $r \times r$ elements of the first r rows must form an identity matrix. The intuition behind this normalization is that for the case of two series it is assumed that the second series has an effect on the first series that is similar to the case of the linear regression model, where one measures the effect of a right-hand side explanatory variable on a left-hand side dependent variable. The second normalization is the orthogonal normalization, where we have

$$\beta'\beta = I_r. \tag{8}$$

Here the interpretation is that the two series are treated as symmetrically effective and only the linear combination matters. This normalization and interpretation comes natural for a set of time series of different, symmetrically treated prices, where one is mainly interested in stable linear combinations.

In this paper we consider the case of $n = 2$ time series (of stock prices) in $Y_t = (y_{t,1}, y_{t,2})$, where the rank of Π is equal to $r = 1$:

$$\begin{pmatrix} \Delta y_{t,1} \\ \Delta y_{t,2} \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} (\beta_1 y_{t-1,1} + \beta_2 y_{t-1,2}) + \begin{pmatrix} \varepsilon_{t,1} \\ \varepsilon_{t,2} \end{pmatrix} \tag{9}$$

with spread

$$s_t = \beta_1 y_{t,1} + \beta_2 y_{t,2} \tag{10}$$

and

$$\begin{pmatrix} E(\Delta y_{t+1,1} | \mathcal{I}_t) \\ E(\Delta y_{t+1,2} | \mathcal{I}_t) \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} (\beta_1 y_{t,1} + \beta_2 y_{t,2}) = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} s_t, \tag{11}$$

so that

$$E(\Delta s_{t+1} | \mathcal{I}_t) = \beta_1 E(\Delta y_{t+1,1}) + \beta_2 E(\Delta y_{t+1,2}) = (\alpha_1 \beta_1 + \alpha_2 \beta_2) s_t. \tag{12}$$

From (10) and (12) it is clear that our ISA trading strategy depends on β_1 and β_2 , whereas our CSA trading strategy also depends on α_1 and α_2 .

Under the linear normalization we have $\beta_1 = 1$:

$$\beta = \begin{pmatrix} 1 \\ \beta_2 \end{pmatrix}, \tag{13}$$

¹ We also considered models with a constant term inside the cointegration relationship and/or drift terms in the model equation. The inclusion of such terms did not change the conclusions of our paper.

whereas under the orthogonal normalization we have

$$\beta' \beta = \beta_1^2 + \beta_2^2 = 1, \tag{14}$$

which is (under the further identification restriction $\beta_2 \geq 0$) equivalent with

$$\beta_2 = \sqrt{1 - \beta_1^2}. \tag{15}$$

Since the adjustment coefficients α_1 and α_2 may be close to 0, there may be substantial uncertainty about the equilibrium relationship. The linear normalization allows β_2 to take values in $(-\infty, \infty)$, whereas the orthogonal normalization allows β_1 to take values in $[-1, 1]$ and β_2 in $[0, 1]$. One may argue that the spread under the linear normalization is just a re-scaled version of the spread under the orthogonal normalization (where the spread under the linear normalization would result by dividing the spread under the orthogonal normalization by β_1). However, we will consider a moving window, where the parameters will be updated every day, so that the re-scaling factor is not constant over time. Therefore, the profit/loss of the ISA strategy under the linear normalization is not just a re-scaled version of the profit/loss of the ISA strategy under the orthogonal normalization. Further, we estimate the parameters using their posterior median, where the posterior median of β_2 under the linear normalization will typically differ from the ratio of the posterior medians of β_2 and β_1 under the orthogonal normalization. The profit/loss of the strategies under the linear normalization may be much affected by a small number of days at which the β_2 is estimated very large (in an absolute sense), whereas under the orthogonal normalization the profit/loss may be more evenly affected by the different days, as (the estimates of) β_1 and β_2 can not ‘escape’ to extreme values outside $[-1, 1] \times [0, 1]$.

3.1. The Encompassing and Jeffreys’ Framework for Prior Specification and Posterior Simulation

As mentioned above, we consider the case of $n = 2$ time series (of stock prices), where the rank of $\Pi = \beta\alpha$ is equal to $r = 1$. That is, the matrix Π needs to satisfy a reduced rank restriction. A natural way to specify a prior for α and β is given by the encompassing framework, in which one first specifies a prior on Π without imposing a reduced rank restriction and then obtains the prior in our model as the conditional prior of Π given that the rank of Π is equal to 1.

As singular values are generalized eigenvalues of non-symmetric matrices, they are a natural way to represent the rank of a matrix. Using singular values we can artificially construct the full rank specification of Π via an auxiliary parameter given by the $(n - r) \times (n - r)$ matrix λ ; *i.e.*, λ is a scalar in our case with $n = 2$ and $r = 1$. The reduced rank matrix $\beta\alpha$ is extended into the full rank specification:

$$\Pi = \beta\alpha + \beta_{\perp} \lambda \alpha'_{\perp}, \tag{16}$$

where β_{\perp} and α'_{\perp} are $n \times (n - r)$ matrices that are specified such that $\beta' \beta_{\perp} \equiv 0$, $\beta'_{\perp} \beta_{\perp} \equiv I_{n-r}$, $\alpha_{\perp} \alpha' \equiv 0$ and $\alpha_{\perp} \alpha'_{\perp} \equiv I_{n-r}$. The full rank specification encompasses the reduced rank case given by $\lambda = 0$. In this framework the probability $p(\lambda = 0|Y)$ can be interpreted as a measure quantifying the likelihood of reduced rank. The specification in (26) is obtained using the singular value decomposition $\Pi = USV'$ of Π , where the $n \times n$ matrices U and V are orthogonal such that $U'U = I_n$ and $V'V = I_n$ and the $n \times n$ matrix S is diagonal and has the singular values of Π on its diagonal in a decreasing order.

To derive the elements of equation (26) in terms of parameters Π we partition Π according to the specifics of the chosen normalization. Under the linear normalization, we partition the matrices U , S and V as follows

$$U = \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix}, S = \begin{pmatrix} S_1 & 0 \\ 0 & S_2 \end{pmatrix}, \text{ and } V' = \begin{pmatrix} V'_{11} & V'_{21} \\ V'_{12} & V'_{22} \end{pmatrix}.$$

The matrices in decomposition (26) in terms of the blocks of U , S and V are given by

$$\begin{aligned}\alpha &= U_{11}S_1 \begin{pmatrix} V'_{11} & V'_{21} \end{pmatrix}, \alpha_{\perp} = (V_{22}V'_{22})^{1/2}V_{22}^{-1'} \begin{pmatrix} V'_{12} & V'_{22} \end{pmatrix}, \\ \beta_2 &= U_{21}U_{11}^{-1}, \beta_{\perp} = \begin{pmatrix} U_{12} \\ U_{22} \end{pmatrix} U_{22}^{-1}(U_{22}U'_{22})^{1/2}, \\ \lambda &= (U_{22}U'_{22})^{-1/2}U_{22}S_2V'_{22}(V_{22}V'_{22})^{-1/2}.\end{aligned}$$

Under the orthogonal normalization, the matrices are partitioned as

$$U = \begin{pmatrix} U_1 & U_2 \end{pmatrix}, S = \begin{pmatrix} S_1 & 0 \\ 0 & S_2 \end{pmatrix} \text{ and } V' = \begin{pmatrix} V'_1 \\ V'_2 \end{pmatrix},$$

and the following relations hold:

$$\begin{aligned}\alpha &= S_1V'_1, \alpha_{\perp} = V'_2 \\ \beta &= U_1, \beta_{\perp} = U_2, \\ \lambda &= S_2.\end{aligned}$$

Under the orthogonal normalization λ is directly equal to S_2 , whereas under the linear normalization it is just a rotation of S_2 . In both cases restriction $\lambda = 0$ is equivalent with restricting the $n - r$ smallest singular values of Π to 0.

The prior on (α, β) is equal to the conditional prior of the parameters (α, β, λ) given that $\lambda = 0$, which is proportional to the joint prior for (α, β, λ) evaluated at $\lambda = 0$:

$$p(\alpha, \beta) \propto p(\alpha, \beta, \lambda)|_{\lambda=0} \propto p(\Pi(\alpha, \beta, \lambda))|_{\lambda=0} |J(\Pi, (\alpha, \beta, \lambda))| |_{\lambda=0}, \quad (17)$$

where $|_{\lambda=0}$ stands for evaluated in $\lambda = 0$, where $J(\Pi, (\alpha, \beta, \lambda))$ denotes the Jacobian of the transformation from Π to (α, β, λ) . Kleibergen and Paap [6] derive the closed form expression for the determinant of the Jacobian $|J(\Pi, (\alpha, \beta, \lambda))|$ for the general case of n variables and reduced rank r under the linear normalization. In Appendix B the Jacobian is derived under the orthogonal normalization of β .

Baştürk *et al.* [10] prove that under certain conditions the encompassing prior is equivalent to Jeffreys' prior in the cointegration model with normally distributed innovations, irrespective of the normalization applied. We emphasize this equivalence, since the use of the information matrix or Jeffreys' prior is more well-known than the encompassing approach. Since the information matrix prior may yield certain desirable properties of the posterior, we conclude that an encompassing approach may also serve this purpose.

In a similar fashion, the posterior of (α, β) is equal to the conditional posterior of the parameters (α, β, λ) given that $\lambda = 0$, which is proportional to the joint posterior for (α, β, λ) evaluated at $\lambda = 0$:

$$\begin{aligned}p(\alpha, \beta|Y) &= p(\alpha, \beta|\lambda = 0, Y) \\ &\propto p(\alpha, \beta, \lambda|Y)|_{\lambda=0} = p(\Pi(\alpha, \beta, \lambda|Y))|_{\lambda=0} |J(\Pi, (\alpha, \beta, \lambda))| |_{\lambda=0},\end{aligned} \quad (18)$$

where the detailed expression for $p(\Pi(\alpha, \beta, \lambda)|Y)$ is given by Kleibergen and Paap [6], and where

$$p(\alpha, \beta, \lambda|Y) = p(\Pi(\alpha, \beta, \lambda)|Y)|_{\Pi=\beta\alpha+\beta_{\perp}\lambda\alpha_{\perp}} |J(\Pi, (\alpha, \beta, \lambda))|. \quad (19)$$

For Bayesian estimation of the cointegration model we need an algorithm to sample from the posterior density in (18). However this posterior densities does not belong to any known

class of distributions, see Kleibergen and Paap [6], and as such can not be sampled directly. The idea of the Metropolis-Hastings (M-H) algorithm is to generate draws from the target density by constructing a Markov chain of which the distribution converges to the target distribution, using draws from a candidate density and an acceptance-rejection scheme. Kleibergen and Paap [6] present the M-H algorithm to sample from (18) for the cointegration model with normally distributed disturbances under the linear normalization. In this algorithm (19) is used to form a candidate density. The general outline of this sampling algorithm is presented in Appendix A. Appendix B presents the approach to evaluate the acceptance-rejection weights under the orthogonal normalization. The posteriors of the coefficients under the linear normalization have Cauchy type tails, so that there exist no posterior means for the coefficients. Therefore, we estimate the coefficients using the posterior median (which we do under both normalizations to keep the comparison between the normalizations as fair as possible).

Given that the time series considered have a non-normal shape, we also consider the model under a multivariate Student's t distribution for the innovations ε_t . Then the M-H algorithms are straightforwardly extended, see Geweke [11].

Since we make use of the independence-chain Metropolis-Hastings algorithm, the simulation of candidate draws and the evaluation of the importance weights (to be used in the probability of accepting the candidate draw) can be easily performed in a parallel fashion. This would enormously increase the speed of our computations. Only the final step of the method, the actual acceptance or rejection of candidate draws, can not be performed in a parallel fashion. But this step takes relatively very little computing time. As an alternative, one can make use of importance sampling, where the whole method can be performed in a parallel fashion.

3.2. Bayes Factors

We evaluate the Bayes factor of rank 1 *versus* rank 2 and the Bayes factor of rank 0 *versus* rank 2. The Bayes factor of rank 1 *versus* rank 0 is obviously given by the ratio of these Bayes factors. For the evaluation of these Bayes factors we extend the method of Kleibergen and Paap [6] who evaluate the Bayes factor as the Savage-Dickey density ratio, see Dickey [12] and Verdinelli and Wasserman [13] to the case of orthogonal normalization. The Bayes factor for the restricted model with $\lambda = 0$ (where Π has rank 0 or 1) *versus* the unrestricted model with unrestricted λ (where Π has rank 2) equals the ratio of the marginal posterior density of λ , and the marginal prior density of λ , both evaluated in $\lambda = 0$. However, in the case of our diffuse prior specification this Bayes factor for rank reduction is not defined, as the marginal prior density of λ is improper.

Therefore, we follow Chao and Phillips [14] who use as prior height $(2\pi)^{-(2nr-r^2)/2}$ to construct their posterior information criterium (PIC). We assume equal prior probabilities $\frac{1}{3}$ for the rank 0, 1 or 2, so that the Bayes factor is equal to the posterior odds, the ratio of posterior model probabilities. For pairs of stock prices we will mostly observe that the estimated posterior model probability is highest for rank 0, the case of two random walk processes without cointegration. Only for a small fraction of pairs, we will observed that the estimated posterior model probability is highest for rank 1, the case of two cointegrated random walk processes.

4. Empirical Application

The CSA and ISA strategies are applied to components of the Dow Jones Composite Average index. We work with daily closing prices recorded over the period of one year, from 1 January 2009 until 31 December 2009. We consider the 65 stocks with the highest liquidity. First, we identify cointegrated pairs based on the estimated posterior probability of cointegration (*i.e.*, Π having rank 1) computed for the first half year of the data. That is, among the $\frac{65 \times 64}{2} = 2080$ pairs we select the 10 pairs with the highest Bayes factor of rank 1 *versus* rank 0 (where these Bayes factors are larger than 1) for both the linear and orthogonal normalization. The 10 pairs are identical for both normalizations; these pairs are given by Table 1. Second, those pairs are used in the CSA and ISA

trading strategies during the last 6 months of 2009. We use a rolling window, where the parameter estimates are updated at the end of each trading day, after which the positions are updated and kept constant until the end of the next trading day. We will analyze the profits from these trading strategies, where we take into account the common level of transaction costs of 0.1% ($c = 0.001$).

Table 1. Ten pairs of stocks with highest Bayes factor of model with Π having rank 1 (cointegration) versus model with Π having rank 0 (two random walks) under both the linear normalization and the orthogonal normalization (among stocks in the Dow Jones Composite Average index, using daily closing prices recorded over the period of 1 January 2009 until 30 June 2009).

AA	-	OSG	:	ALCOA Inc.	-	Overseas Shipholding Group, Inc.
CNP	-	OSG	:	CenterPoint Energy, Inc.	-	Overseas Shipholding Group, Inc.
DUK	-	IBM	:	Duke Energy Corp.	-	International Business Machines Corp.
DUK	-	OSG	:	Duke Energy Corp.	-	Overseas Shipholding Group, Inc.
MO	-	UPS	:	Altria Group, Inc.	-	United Parcel Service, Inc.
NI	-	NSC	:	NiSource, Inc.	-	Norfolk Southern Corp.
NI	-	OSG	:	NiSource, Inc.	-	Overseas Shipholding Group, Inc.
NI	-	R	:	NiSource, Inc.	-	Ryder System, Inc.
NI	-	UNP	:	NiSource, Inc.	-	Union Pacific Corp.
NI	-	UTX	:	NiSource, Inc.	-	United Technologies Corp.

Next to the “standard” CSA approach described before, we will also perform a more cautious, more conservative CSA strategy that takes into account parameter uncertainty. Here we only take a position if we are more certain about the sign of the current spread (and hence the sign of the expected change of the spread, which is the opposite sign). We only take a position if the $(50 + \zeta/2)\%$ percentile and the $(50 - \zeta/2)\%$ percentile of the posterior distribution of the current spread have the same sign, where we consider the cases of $\zeta = 20\%, 30\%, 40\%, 50\%$ or 60% . The case of $\zeta = 60\%$ is the most cautious strategy, where the sign of the posterior 20% and 80% percentiles of the spread must be the same. Note that for $\zeta = 0\%$ this strategy reduced to the original CSA strategy.

In order to evaluate the CSA and ISA strategies in the cointegration models under the linear and orthogonal normalization and under a normal and Student’s t distribution for the innovations, we compute two measures. First of all, the strategy can not be evaluated in terms of the percentage return on initial capital investment, as we are not only buying stocks but we are also shorting stocks. Suppose that we perform our strategies for T consecutive trading days (where in our case T is the number of trading days in the last 6 months of 2009). Then the average daily capital engagement is given by:

$$ADCE_{CSA(k,x\%)} \equiv \frac{1}{T} \sum_{t=0}^{T-1} |CSA(s_t, k, x\%)| (|\hat{\beta}_{t,1}|y_{t,1} + |\hat{\beta}_{t,2}|y_{t,2}), \tag{20}$$

$$ADCE_{ISA} \equiv \frac{1}{T} \sum_{t=0}^{T-1} |ISA(s_t)| (|\hat{\beta}_{t,1}|y_{t,1} + |\hat{\beta}_{t,2}|y_{t,2}), \tag{21}$$

where $\hat{\beta}_{t,1}$ and $\hat{\beta}_{t,2}$ are the posterior medians of β_1 and β_2 computed at the end of the t -th day in the trading period. That is, $\hat{\beta}_{t=0,1}$ and $\hat{\beta}_{t=0,2}$ are computed at the end of the last trading day before the trading period. Our first performance measure is a profitability measure that is given by the total return of the strategy divided by the average daily capital engagement:

$$Profitability_{CSA(k,x\%)} = \frac{\text{Cumulative Return of } CSA(k, x\%) \text{ at time } T}{ADCE_{CSA(k,x\%)}} \tag{22}$$

$$Profitability_{ISA} = \frac{\text{Cumulative Return of } ISA \text{ at time } T}{ADCE_{ISA}} \tag{23}$$

Our second performance measure concerns the risk of the strategies. In order to estimate risk we use paths of cumulative return. When the cumulative return at time $t + 1$ is often lower than the cumulative return at time t , then a strategy can be considered risky. On the other hand, if the cumulative return is growing or remains steady over most periods the strategy can be considered as having low risk. In the latter case the signals generated by the trading rule are accurate and yield (mostly) profit. We define our measure as:

$$Risk \equiv \frac{\#\{\text{Cum. return at time } t + 1 < \text{Cum. return at time } t\}}{\#\{\text{Cum. return at time } t + 1 \neq \text{Cum. return at time } t\}} \quad (24)$$

for each strategy.

Table 2 presents the average of the *Profitability* measure in (22)–(23) over the ten selected (cointegrated) pairs of stocks. Detailed results, for every pair of stocks, are presented in Appendix C. The hypothesis that the normalization plays an important role in pairs trading strategies is confirmed by the empirical findings. Table 2 confirms that the orthogonal normalization substantially outperforms the linear counterpart, irrespectively of the assumed distribution for the innovations. It is particularly pronounced when the profitability of the CSA strategies under $k = 0$ (directional accuracy) is compared with the counterpart under $k = 1$ (accuracy). For the linear normalization the increase from $k = 0$ to $k = 1$ is linked with substantial decrease in profitability. It means that predictions of the change of the spread under the linear normalization are relatively poor compared with the orthogonal case. For $k = 1$ where not only the direction but also the size of predicted change of the spread play an important role, the linear normalization performs relatively poorly. On the contrary, the orthogonal normalization shows an appreciable increase in profitability for $k = 1$ compared to $k = 0$.

As expected, the *Risk* measure for the CSA strategies in Table 3 decreases for more cautious, more conservative strategies with larger values of ζ . The *Risk* measure under the linear normalization is similar or worse than the counterpart under the orthogonal normalization. Obviously, this only means that the percentage of trading days with a decrease of the cumulative return is similar. The size of these decreases may be larger. In further research, we will take a closer look at the riskiness of the alternative strategies in the different models.

Now we compare the performance between the normal distribution and the Student's t distribution. The *Risk* measure is slightly better under the Student's t distribution. For the *Profitability* measure the difference (in favor of the Student's t distribution) seems somewhat more clearly present. However, it should be noted that the difference between the orthogonal and linear normalizations is much larger than the difference between the Student's t and the normal distribution. The normalization is clearly the key factor for the profitability of the trading strategies for our set of data. One possible reason for this result is that the profit/loss of the strategies under the linear normalization may be much affected by a small number of days at which the β_2 is estimated very large (in an absolute sense), whereas under the orthogonal normalization the profit/loss may be more evenly affected by the different days, as (the estimates of) β_1 and β_2 can not "escape" to extreme values far outside $[-1, 1] \times [0, 1]$. The latter may happen in the case of the linear normalization if the adjustment coefficients α_1 and α_2 are close to 0. The latter may be found for certain pairs of empirical time series of stock prices (in certain periods), where the error correction may be rather slow. In future research, we will take a closer look at the reasons for the substantial difference in profitability between the normalizations.

Table 2. Performance evaluation measure *Profitability* in (22)–(23) (in %). Average over 10 pairs of assets.

		CSA ($k = 0$, Directional Accuracy)					
		ξ					
		0%	20%	30%	40%	50%	60%
		linear normalization					
normal		12.52	17.08	19.72	26.94	31.32	47.37
Student's <i>t</i>		19.87	24.23	30.11	32.9	38.74	61.03
		orthogonal normalization					
normal		57.10	64.8	63.61	76.56	83.49	112.2
Student's <i>t</i>		60.27	66.41	67.67	81.78	91.53	114.05
		ISA	CSA ($k = 1$, accuracy)				
		ξ					
		0%	20%	30%	40%	50%	60%
		linear normalization					
normal	15.42	4.80	5.73	6.07	6.98	8.82	11.89
Student's <i>t</i>	16.82	7.81	9.77	12.57	13.53	15.32	20.34
		orthogonal normalization					
normal	46.11	67.78	83.03	90.24	104.62	118.09	160.68
Student's <i>t</i>	46.32	68.38	83.75	91.37	106.04	113.93	155.67

Table 3. Risk evaluation measure *Risk* in (24). Average over 10 pairs of assets.

		CSA ($k = 0$, Directional Accuracy)					
		ξ					
		0%	20%	30%	40%	50%	60%
		linear normalization					
normal		0.45	0.43	0.42	0.42	0.42	0.35
Student's <i>t</i>		0.44	0.42	0.41	0.41	0.40	0.32
		orthogonal normalization					
normal		0.40	0.40	0.42	0.41	0.40	0.34
Student's <i>t</i>		0.39	0.40	0.40	0.39	0.36	0.33
		ISA	CSA ($k = 1$, accuracy)				
		ξ					
		0%	20%	30%	40%	50%	60%
		linear normalization					
normal	0.47	0.51	0.48	0.45	0.45	0.44	0.35
Student's <i>t</i>	0.48	0.50	0.47	0.46	0.46	0.44	0.34
		orthogonal normalization					
normal	0.45	0.48	0.46	0.48	0.46	0.45	0.37
Student's <i>t</i>	0.46	0.47	0.47	0.45	0.42	0.39	0.35

5. Conclusions and Topics for Further Research

In this paper we explored the connection between the statistical wellknown cointegration model and decision strategies for the selection of pairs trading of stocks using a simulation-based Bayesian procedure. We considered two cases of pairs trading strategies: a conditional statistical arbitrage method and an implicit statistical arbitrage method. We used a simulation-based Bayesian procedure for predicting stable ratios, defined in a cointegration model, of pairs of stock prices. We showed the effect that using an encompassing or Jeffreys' prior under an orthogonal normalization has for the selection of pairs of cointegrated stock prices and for the estimation and prediction of the spread between cointegrated stock prices and its uncertainty. An empirical application was done using stocks that are ingredients of the Dow Jones Composite Average index. The results showed that the normalization has little effect on the selection of pairs of cointegrated stocks on the basis of Bayes factors. However, the results stressed the importance of the orthogonal normalization for the

estimation and prediction of the spread, which leads to better results in terms of profit per capital engagement and risk than using a standard linear normalization.

An important issue for future research is to investigate the robustness of our empirical results. Here we list three topics of further research. First, there may exist a sensitivity of the results for the specific data. We already indicated that our results may be sensitive for some particular data in a sub period. Second and more generally, if one considers the percentiles of the predictive distribution for the future spread during the trading strategy, taking into account the uncertainty in future innovations, then it is important to specify the distribution of the innovations carefully. In future research, we will consider a finite mixture of Gaussian distributions for the innovations. However, the algorithm for the posterior simulation will not be a straightforward extension of the algorithm under the normal distribution, which was the case for the Student's t distribution. We will extend the partial and permutation-augmented MitISEM (Mixture of t by Importance Sampling weighted Expectation Maximization) approaches of Hoogerheide *et al.* [15] to perform the posterior simulation for the cointegration model with errors obeying a finite mixture distribution. A third point is the introduction of a learning strategy by the decision maker. Fourth, the economic performance of econometric predictions can be evaluated using a utility based metric to obtain a certainty equivalence of strategies. This penalizes the excess variation in predictions perceived as "risk" of the strategy. See West *et al.* [16]. Fifth, the methods can be compared with a strategy where one invests 50% in the risk-free rate and 50% in a risky asset, which is quite a successful and robust strategy, see Marquering and Verbeek [17]. Sixth, the performance of the methods can be compared with alternative Bayesian cointegration approaches, see Furmston *et al.* [18] and Bracegirdle and Barber [19].

Author Contributions: The authors contributed equally to the paper.

Funding: This research was funded by the National Science Center, Poland, grant number 2013/09/N/HS4/03751.

Acknowledgments: The authors are indebted to two anonymous reviewers and a guest-editor for very helpful comments on an earlier version of this paper.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A: Posterior Distribution and Sampling Algorithm Under Encompassing Prior

In this appendix we make use of two models, where we follow the terminology of Kleibergen and Paap [6]. First, the linear error correction (LEC) model is given by

$$\Delta Y_t = \Pi' Y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \Sigma), \quad (25)$$

in which we impose no rank reduction restriction on

$$\Pi = \beta\alpha + \beta_{\perp}\lambda\alpha_{\perp}. \quad (26)$$

We will use the posterior of $(\alpha, \beta, \lambda, \Sigma)$ in this LEC model as a candidate distribution in a Metropolis-Hastings algorithm.

Second, the cointegration model is given by

$$\Delta Y_t = \Pi' Y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \Sigma), \quad (27)$$

in which we impose the rank reduction restriction

$$\Pi = \beta\alpha. \quad (28)$$

The posterior of (α, β, Σ) in this cointegration model is the actual target distribution that we are interested in. Straightforwardly using a Gibbs sampler by simulating α and β from their full

conditional posteriors is not possible due to their difficult dependence structure, see Kleibergen and Van Dijk [20].

The encompassing prior assumes that rank restriction on Π is expressed explicitly using the decomposition (26), where not only the prior in the LEC model but also the posterior of $\alpha, \beta, \lambda | \Sigma, Y$ in the LEC model satisfies the transformation of random variables defined by (26) such that

$$p_{LEC}(\alpha, \beta, \lambda | \Sigma, Y) = p_{LEC}(\Pi | \Sigma, Y) |_{\Pi = \beta\alpha + \beta_{\perp}\lambda\alpha_{\perp}} |J(\Pi, (\alpha, \beta, \lambda))| \tag{29}$$

where $J(\Pi, (\alpha, \beta, \lambda))$ denotes the Jacobian of the transformation from Π to (α, β, λ) .

The conditional posterior $\alpha, \beta | \lambda, \Sigma, Y$ in the LEC model, which is obviously proportional to the joint, can be evaluated in $\lambda = 0$ to obtain the posterior of $\alpha, \beta | \Sigma, Y$ in our cointegration model with rank reduction:

$$\begin{aligned} p(\alpha, \beta | \Sigma, Y) &= p_{LEC}(\alpha, \beta | \lambda, \Sigma, Y) |_{\lambda=0} \\ &\propto p_{LEC}(\alpha, \beta, \lambda | \Sigma, Y) |_{\lambda=0} \\ &= p_{LEC}(\Pi | \Sigma, Y) |_{\Pi = \beta\alpha} |J(\Pi, (\alpha, \beta, \lambda))| |_{\lambda=0}, \end{aligned} \tag{30}$$

since our model results from the linear error correction model by imposing $\lambda = 0$. Hence, we can consider the rank reduction as a parameter realization $\lambda = 0$.

Sampling Algorithm Based on Diffuse Prior on Π and Nesting

We specify a diffuse prior on the parameters Π and Σ in the LEC model: $p_{LEC}(\Sigma) \propto |\Sigma|^{-\frac{n+1}{2}}$ and $p_{LEC}(\Pi | \Sigma) \propto 1$ (where n is the dimension of Y_t). The conditional posterior of Σ given Π in the LEC model is a matrix-variate normal distribution. The marginal posterior of Σ in the LEC model is an inverted Wishart distribution. See Zellner [21] for a discussion of Bayesian analysis in the linear model. The decomposition in (26) allows us to obtain a (joint) draw of α and β (and λ) from a draw of Π . The dependencies between α and β are fully taken into account by determining α and β simultaneously.

This poses the problem that our posteriors of interest in the cointegration model, $p(\alpha, \beta, \Sigma | Y)$ and $p(\alpha, \beta | Y, \Sigma)$, do not involve λ while it is sampled in the posterior draws for the LEC model. Kleibergen and Paap [6] adopt the approach suggested by Chen [22]. For simulating from the posterior $p(\alpha, \beta | \Sigma, Y)$ it is first extended with an artificial extra parameter λ whose density we denote by $g(\lambda | \alpha, \beta, \Sigma, Y)$. We use a Metropolis-Hastings (M-H) sampling algorithm for simulating from the joint density

$$p_g(\alpha, \beta, \lambda, \Sigma | Y) = g(\lambda | \alpha, \beta, \Sigma, Y) p(\alpha, \beta, \Sigma | Y). \tag{31}$$

The posterior $p(\alpha, \beta, \lambda | \Sigma, Y)$ from (29) is used as the candidate generating density. When $p_g(\alpha, \beta, \lambda, \Sigma, Y)$ is marginalized with respect to λ in order to remove the artificial parameter λ , the resultant distribution is $p(\alpha, \beta, \Sigma | Y)$. The simulated values of α, β, Σ (discarding λ) therefore are a sample from $p(\alpha, \beta, \Sigma | Y)$.

The choice of $g(\lambda | \alpha, \beta, \Sigma, Y)$ leads to the weight function $w(\alpha, \beta, \lambda, \Sigma)$ for use in the M-H algorithm. The acceptance probability in the M-H depends on a weight function which is the ratio of the target density (31) and the candidate generating density (29),

$$\begin{aligned} w(\alpha, \beta, \lambda, \Sigma) &= \frac{p_g(\alpha, \beta, \lambda, \Sigma | Y)}{p_{LEC}(\alpha, \beta, \lambda, \Sigma | Y)} \\ &= \frac{g(\lambda | \alpha, \beta, \Sigma, Y) p(\alpha, \beta | \Sigma, Y)}{p_{LEC}(\alpha, \beta, \lambda, \Sigma | Y)} \\ &= \frac{g(\lambda | \alpha, \beta, \Sigma, Y) \exp(-\frac{1}{2}(\text{tr}(\beta\alpha - \hat{\Pi})'(\beta\alpha - \hat{\Pi})))}{\exp(-\frac{1}{2}(\text{tr}(\beta\alpha + \beta_{\perp}\lambda\alpha_{\perp} - \hat{\Pi})'(\beta\alpha + \beta_{\perp}\lambda\alpha_{\perp} - \hat{\Pi})))} \frac{|J|_{\lambda=0}}{|J|} \end{aligned} \tag{32}$$

where $\hat{\Pi}$ is the OLS estimator.

The exponentiated trace expressions in numerator and denominator are related to each other by

$$\begin{aligned} & \text{tr}((\beta\alpha + \beta_{\perp}\lambda\alpha_{\perp} - \hat{\Pi})'(\beta\alpha + \beta_{\perp}\lambda\alpha_{\perp} - \hat{\Pi})) \\ &= \text{tr}((\beta\alpha - \hat{\Pi})'(\beta\alpha - \hat{\Pi})) + \text{tr}((\lambda - \beta_{\perp}'\hat{\Pi}\alpha_{\perp}')'(\lambda - \beta_{\perp}'\hat{\Pi}\alpha_{\perp}')) \\ & \quad + \text{tr}((\beta_{\perp}'\hat{\Pi}\alpha_{\perp}')'\beta_{\perp}'\hat{\Pi}\alpha_{\perp}') \\ &= \text{tr}((\beta\alpha - \hat{\Pi})'(\beta\alpha - \hat{\Pi})) + \text{tr}((\lambda - \tilde{\lambda})'(\lambda - \tilde{\lambda})) + \text{tr}(\tilde{\lambda}'\tilde{\lambda}) \end{aligned} \tag{33}$$

where $\tilde{\lambda} = \beta_{\perp}'\hat{\Pi}\alpha_{\perp}'$. A sensible choice for the density function $g(\lambda|\alpha, \beta, \Sigma, Y)$ thus turns out to be

$$g(\lambda|\alpha, \beta, \Sigma, Y) \propto \exp\left(-\frac{1}{2}\text{tr}((\lambda - \tilde{\lambda})'(\lambda - \tilde{\lambda}))\right). \tag{34}$$

Using this choice of $g(\lambda|\alpha, \beta, \Sigma, Y)$ the weight function reduces to

$$w(\alpha, \beta, \lambda, \Sigma) \propto \exp\left(-\frac{1}{2}\text{tr}(\tilde{\lambda}'\tilde{\lambda})\right) \frac{|J(\Pi, (\alpha, \beta, \lambda))|_{\lambda=0}}{|J(\Pi, (\alpha, \beta, \lambda))|}. \tag{35}$$

For the determinant of the Jacobian $|J(\Pi, (\alpha, \beta, \lambda))|$ under the linear normalization we refer to the appendix of Kleibergen and Paap [6]. In Appendix B we derive the determinant of the Jacobian under the orthogonal normalization.

The steps required in the sampling algorithm are,

1. Draw Σ^{i+1} from $p_{LEC}(\Sigma|Y)$.
2. Draw Π^{i+1} from $p_{LEC}(\Pi|\Sigma, Y)$.
3. Compute $\alpha^{i+1}, \beta^{i+1}, \lambda^{i+1}$ from Π^{i+1} using the singular value decomposition.
4. Accept $\Sigma^{i+1}, \alpha^{i+1}$ and β^{i+1} with probability $\min\left(\frac{w(\alpha^{i+1}, \beta^{i+1}, \lambda^{i+1}, \Sigma^{i+1})}{w(\alpha^i, \beta^i, \lambda^i, \Sigma^i)}, 1\right)$.

Appendix B: Jacobian of the Transformation from Π to (α, β, Σ) under Orthogonal Normalization

Under the orthogonal normalization the components in decomposition (26) can be computed from Π using the singular value decomposition with matrices U, S and V partitioned according to

$$\Pi = USV' = \begin{pmatrix} U_1 & U_2 \end{pmatrix} \begin{pmatrix} S_1 & 0 \\ 0 & S_2 \end{pmatrix} \begin{pmatrix} V_1' \\ V_2' \end{pmatrix}, \tag{36}$$

where $U = \begin{pmatrix} U_1 & U_2 \end{pmatrix}$ and $V = \begin{pmatrix} V_1 & V_2 \end{pmatrix}$ are orthonormal matrices. U_1 and V_1 are $p \times r$, U_2 and V_2 are $p \times (p - r)$, and S_1 and S_2 are diagonal $r \times r$ and $(p - r) \times (p - r)$. Then for orthogonal normalization the following relation hold: $\beta = U_1, \alpha = S_1 V_1', \beta_{\perp} = U_2, \alpha_{\perp} = V_2'$, and $\lambda = S_2$.

β_1 is identified uniquely by β_2 .² So, it suffices to derive $J(\Pi, (\alpha, \beta_2, \lambda))$

$$J(\Pi, (\beta_2, \alpha, \lambda)) = \begin{pmatrix} \frac{\partial \text{vec}\Pi}{\partial (\text{vec}\beta_2)'} & \frac{\partial \text{vec}\Pi}{\partial (\text{vec}\alpha)'} & \frac{\partial \text{vec}\Pi}{\partial (\text{vec}\lambda)'} \end{pmatrix}. \tag{37}$$

The expression for $\frac{\partial \text{vec}\Pi}{\partial (\text{vec}\alpha)'}$ is given by

² In general we can add restrictions to the normalization restriction $\beta'\beta = I_r$ in order to uniquely identify the $r \times r$ matrix β_1 given the $(n - r) \times r$ matrix β_2 , where $\beta = (\beta_1' \beta_2')'$. For example, if we assume that $n = 2, r = 1$, then we can add the restriction $\beta_1 \geq 0$, so that only $\beta_1 = \sqrt{1 - \beta_2^2}$ satisfies $\beta_1^2 + \beta_2^2 = 1$. Note that without that restriction $\beta_1 \geq 0$ we could have $\beta_1 = \sqrt{1 - \beta_2^2}$ or $\beta_1 = -\sqrt{1 - \beta_2^2}$, so that in that case β_1 would not be uniquely identified by β_2 .

$$\frac{\partial \text{vec}\Pi}{\partial (\text{vec}\alpha)'} = (I_p \otimes \beta) + \frac{\partial \text{vec}\Pi}{\partial (\text{vec}\alpha_\perp)'} \frac{\partial \text{vec}\alpha_\perp}{\partial (\text{vec}\alpha)'} \tag{38}$$

and

$$\frac{\partial \text{vec}\Pi}{\partial (\text{vec}\alpha_\perp)'} = (I \otimes \beta_\perp \lambda). \tag{39}$$

If we assume $c = \begin{pmatrix} I_r & 0 \end{pmatrix}'$ and $c_\perp = \begin{pmatrix} 0 & I_{p-r} \end{pmatrix}$, we have $\alpha_\perp = c_\perp' (I_r - \alpha'(\alpha c)'^{-1} c')$ and

$$\frac{\partial \text{vec}\alpha_\perp}{\partial (\text{vec}\alpha)'} = \left((c(\alpha c)^{-1} \otimes (c(\alpha c)^{-1} \alpha c_\perp)') - (c(\alpha c)^{-1} \otimes c_\perp') \right) K_{r,p}, \tag{40}$$

so that

$$\frac{\partial \text{vec}\Pi}{\partial (\text{vec}\alpha)'} = (I_p \otimes \beta) + \left(c(\alpha c)^{-1} \otimes \beta_\perp \lambda c_\perp' ((c(\alpha c)^{-1} \alpha)' - I_p) \right) K_{r,p}. \tag{41}$$

Then for β_2 we obtain

$$\begin{aligned} \frac{\partial \text{vec}\Pi}{\partial (\text{vec}\beta_2)'} &= \frac{\partial \text{vec}\Pi}{\partial (\text{vec}\beta)'} \frac{\partial \text{vec}\beta}{\partial (\text{vec}\beta_1)'} \frac{\partial \text{vec}\beta_1}{\partial (\text{vec}\beta_2)'} + \frac{\partial \text{vec}\Pi}{\partial (\text{vec}\beta_\perp)'} \frac{\text{vec}\beta_\perp}{\partial (\text{vec}\beta)'} \frac{\partial \text{vec}\beta}{\partial (\text{vec}\beta_1)'} \frac{\partial \text{vec}\beta_1}{\partial (\text{vec}\beta_2)'} = \\ &(\alpha' \otimes I_n) \left[I_r \otimes \begin{pmatrix} I_r \\ 0_{(n-r) \times r} \end{pmatrix} \right] \frac{\partial \text{vec}\beta_1}{\partial (\text{vec}\beta_2)'} + (\alpha_\perp' \lambda' \otimes I_n) \frac{\text{vec}\beta_\perp}{\partial (\text{vec}\beta)'} \left[I_r \otimes \begin{pmatrix} I_r \\ 0_{(n-r) \times r} \end{pmatrix} \right] \frac{\partial \text{vec}\beta_1}{\partial (\text{vec}\beta_2)'} \end{aligned}$$

The formula for $\frac{\partial \text{vec}\beta_1}{\partial (\text{vec}\beta_2)'}$ is derived based on the orthogonal normalization condition. We have:

$$\begin{aligned} I_r &= \beta_1' \beta_1 + \beta_2' \beta_2 \\ 0 &= d(\beta_1' \beta_1) + d(\beta_2' \beta_2) \\ 0 &= (\beta_1 \otimes I_r) d\text{vec}\beta_1' + (I_r \otimes \beta_1') d\text{vec}\beta_1 \\ &\quad + (\beta_2 \otimes I_r) d\text{vec}\beta_2' + (I_r \otimes \beta_2') d\text{vec}\beta_2 \end{aligned}$$

As $K_r \text{vec}(A) = \text{vec}(A')$, see we have

$$\begin{aligned} 0 &= K_r (I_r \otimes \beta_1') d\text{vec}\beta_1 + (I_r \otimes \beta_1') d\text{vec}\beta_1 + K_r (I_r \otimes \beta_2') d\text{vec}\beta_2 + (I_r \otimes \beta_2') d\text{vec}\beta_2 \\ &= (K_r + I_r) ((I_r \otimes \beta_1') d\text{vec}\beta_1) + (K_r + I_r) ((I_r \otimes \beta_2') d\text{vec}\beta_2) \\ &= 2N_r ((I_r \otimes \beta_1') d\text{vec}\beta_1) + 2N_r ((I_r \otimes \beta_2') d\text{vec}\beta_2), \end{aligned} \tag{42}$$

where $N_r = \frac{1}{2}(I_r + K_r)$. Thus we obtain

$$0 = 2N_r (I_r \otimes \beta_1') d\text{vec}\beta_1 + 2N_r (I_r \otimes \beta_2') d\text{vec}\beta_2 \tag{43}$$

and

$$\frac{\partial \text{vec}\beta_1}{\partial (\text{vec}\beta_2)'} = -(N_r (I_r \otimes \beta_1'))^{-1} (N_r (I_r \otimes \beta_2')). \tag{44}$$

Further, because $\beta = U_1$ and $\beta_\perp = U_2$ we can derive $\frac{\text{vec}\beta_\perp}{\partial (\text{vec}\beta)'}$ based on the orthomorphic transformation between $U = \begin{pmatrix} U_1 & U_2 \end{pmatrix}$ and \tilde{X} , where $\tilde{X} = (I + U)^{-1}(I - U)$ and $U = (I + \tilde{X})^{-1}(I - \tilde{X})$. We find that

$$\frac{\text{vec}\beta_{\perp}}{\partial(\text{vec}\beta)'} = \frac{\partial\text{vec}\beta_{\perp}}{\partial(\text{vec}\tilde{X})'} \frac{\partial\text{vec}\tilde{X}}{\partial(\text{vec}\beta)'} \tag{45}$$

$$= - \left(I \otimes \begin{pmatrix} 0 \\ I_{p-r} \end{pmatrix} \right) \left((I_p + U)' \otimes (I_p + \tilde{X})^{-1} \right) \times \left[- \left(I \otimes \begin{pmatrix} I_r \\ 0 \end{pmatrix} \right) \left((I_p + \tilde{X})' \otimes (I_p + U)^{-1} \right) \right]$$

where $U_1 = U \begin{pmatrix} I_r & 0 \end{pmatrix}'$ and $U_2 = U \begin{pmatrix} 0 & I_{p-r} \end{pmatrix}'$ and $d((I + U)^{-1}(I - U)) = -(I + U)^{-1}dU(I + U)^{-1}(I - U) - (I + U)^{-1}dU$.

For $\frac{\partial\text{vec}\Pi}{\partial(\text{vec}\lambda)'}$ we obtain

$$\frac{\partial\text{vec}\Pi}{\partial(\text{vec}\lambda)'} = \alpha'_{\perp} \otimes \beta_{\perp}. \tag{46}$$

Appendix C: Tables for Ten Pairs of Stocks

Table 4. Performance evaluation measure *Profitability* in (22)–(23), which is the ratio of cumulative income to the average daily capital absorption (in %), under a normal distribution for the innovations. The meaning of the ten pairs of stocks is indicated in Table 1. Outperforming normalization in boldface.

	<i>k</i> = 0 (Directional Accuracy)						ISA	<i>k</i> = 1 (Accuracy)					
	CSA							CSA					
	0%	20%	30%	40%	50%	60%		0%	20%	30%	40%	50%	60%
linear normalization of β													
AA-OSG	19.5	25.2	15.5	23.2	25.5	40.1	22.5	4.2	2.1	3.5	2.1	7	5.2
DUK-IBM	7.3	17	16.2	22.9	16.7	74.3	6.5	0.9	1.8	1.5	2.5	2.3	7.5
DUK-OSG	-3	10.5	15.9	47.7	46.2	82.7	24.9	2.9	4.3	4.9	9.2	11.4	16.1
NI-NSC	24.9	35.9	40.4	33.5	33.9	38.8	7.1	13.2	19.1	25	25.8	33.8	37.4
NI-OSG	11.2	19.9	22.3	34.1	47.9	70.6	33.2	3.4	4.8	5.7	7.7	10	14.7
CNP-OSG	14.8	24.8	32	29.7	23	37.3	10.3	1.6	2.1	2.7	2.8	3.5	8.8
MO-UPS	-5.8	-5.9	10.1	23.3	30.5	16.6	19.8	1.4	1.5	3	4.3	5.2	4.6
NI-R	39.2	32.9	22.9	14.4	20	13.4	24.1	17.8	15.2	8.3	4.7	3.8	3.3
NI-UNP	11.2	9.1	8.3	18.6	31.6	44	6.4	0.9	1.3	1.6	2.8	4.7	6.9
NI-UTX	12.9	9.5	9.4	18.3	32.1	48.6	6.5	1.1	1.5	1.9	3	4.7	7.7
orthogonal normalization of β													
AA-OSG	108.7	138.6	145.2	155.2	167.4	181.2	133.3	229	293.3	324.6	358.6	397.1	428.2
DUK-IBM	23.6	36.3	32.6	60.7	82.3	364	14.9	19.7	34.7	36.4	65.9	112.9	626.1
DUK-OSG	80.9	85.7	85.3	124.3	136.5	59.7	62.2	77.5	96.5	100.7	130.7	174.4	12.1
NI-NSC	46.4	54.8	59.8	81.2	81.5	80.9	17.5	32.3	44.4	56.4	82.1	95.1	132.8
NI-OSG	67.1	100.3	94.2	113.3	160.4	150.2	65.5	74	96.6	108.9	130.2	164.4	103.5
CNP-OSG	25.5	21.4	16.9	17.4	6.4	9.2	8.8	3.4	2.7	1.8	2.8	1.1	2.1
MO-UPS	0.2	4.4	2.8	15.6	32	66.7	17.9	1	1.1	1.3	2.8	4.6	7.5
NI-R	23.7	12.9	2.5	-10.2	-9.4	12.4	17.2	9.7	7.7	5.5	1	1.5	2.6
NI-UNP	14.2	7.5	1.5	16.8	30.7	21.6	5.8	2	1	1.1	2.7	4.7	3.8
NI-UTX	180.7	186.1	195.3	191.3	147.1	176.1	118	229.2	252.3	265.7	269.4	225.1	288.1

Table 5. Performance evaluation measure *Profitability* in (22)–(23), which is the ratio of cumulative income to the average daily capital absorption (in %), under a Student’s *t*-distribution for the innovations. The meaning of the ten pairs of stocks is indicated in Table 1.

	<i>k</i> = 0 (Directional Accuracy)						ISA	<i>k</i> = 1 (Accuracy)					
	CSA							CSA					
	0%	20%	30%	40%	50%	60%		0%	20%	30%	40%	50%	60%
linear normalization of β													
AA-OSG	24.5	31.1	18.8	24.7	28.8	41.6	24.6	2.5	3.2	3	3.5	4	6.4
DUK-IBM	10.5	16.5	15.4	25.3	17.8	95.4	7	1.3	2.4	2.1	3.2	3.3	14.9
DUK-OSG	13.8	27	35.8	45.2	47.7	82.6	24.7	2.8	4.5	5.1	7.1	8	11.7
NI-NSC	41.5	53.1	73.2	71.3	67.6	79.8	17.8	34.8	48.2	70.4	87.7	93.8	106.3
NI-OSG	29	35.7	42.8	51	77	86.8	34.1	4.8	6.9	8	9.9	13.7	18
CNP-OSG	17.1	24.3	34.2	17.7	26.4	61.2	10.3	1.5	2.2	3	1.7	3.1	11.6
MO-UPS	9.9	7.9	20.8	39.8	44	48.7	20.5	2.3	2.2	3.7	5.6	6.1	5.6
NI-R	32.4	28.1	24.6	4.2	11.5	14.8	23.9	17.4	15.9	16.2	1.5	2.6	3.3
NI-UNP	4.9	8.3	7.4	18.6	31.5	50.2	6.5	0.9	1.7	1.7	2.4	4.3	8.2
NI-UTX	19.7	17.2	16.8	23	25.2	29.8	6.6	4.5	3.9	2.9	2.7	3	3.5
orthogonal normalization of β													
AA-OSG	118.3	150.9	170.1	182.5	192.7	213.2	141.1	245.1	314	351.8	391.9	427.4	466.3
DUK-IBM	25.9	29.7	26.6	63.1	98.4	350.3	14.9	19.8	30.4	35	65.3	97.7	608.5
DUK-OSG	78.4	84.2	96.7	115.3	148.5	101.9	60.5	73.7	90.5	102.5	121.3	133.5	16.4
NI-NSC	43.5	54.6	58.2	69.9	53	16.1	12.8	23.4	41	41.5	54.8	25.2	7.8
NI-OSG	55.5	77	74.3	101.5	152.1	119.6	54.8	53.6	69.9	80.8	114	145.1	100.2
CNP-OSG	29.7	26.9	18.9	20.8	15.4	6.6	9.7	5.8	7.3	2.5	3.3	2.9	2.9
MO-UPS	11.1	4.9	2.3	15.4	30.7	68.9	18.5	1.7	1.5	1.5	2.9	4.6	7.7
NI-R	33.4	22.1	13.8	2.6	-12.4	1.7	16.9	9.3	6.7	5	0.7	1.3	1.8
NI-UNP	16	15	9.2	35.4	40.4	48.8	6.5	2.9	3.4	3.4	6.9	7.8	11.3
NI-UTX	190.9	198.8	206.6	211.3	196.5	213.4	127.5	248.5	272.8	289.7	299.3	293.8	333.8

Table 6. Risk evaluation measure *Risk* in (24), fraction of trading days with decreases in cumulative income, under a normal distribution for the innovations. The meaning of the ten pairs of stocks is indicated in Table 1. Outperforming normalization in boldface.

	<i>k</i> = 0 (Directional Accuracy)						ISA	<i>k</i> = 1 (Accuracy)					
	CSA							CSA					
	0%	20%	30%	40%	50%	60%		0%	20%	30%	40%	50%	60%
linear normalization of β													
AA-OSG	0.42	0.43	0.42	0.46	0.41	0.39	0.48	0.49	0.48	0.43	0.46	0.42	0.39
DUK-IBM	0.46	0.44	0.46	0.43	0.36	0	0.49	0.55	0.55	0.52	0.5	0.5	0
DUK-OSG	0.55	0.48	0.49	0.41	0.44	0.36	0.5	0.58	0.5	0.52	0.43	0.44	0.36
NI-NSC	0.4	0.38	0.37	0.43	0.53	0.54	0.48	0.5	0.42	0.4	0.48	0.53	0.54
NI-OSG	0.46	0.43	0.43	0.4	0.4	0.36	0.42	0.48	0.44	0.44	0.41	0.4	0.36
CNP-OSG	0.42	0.34	0.29	0.3	0.29	0.33	0.44	0.48	0.39	0.34	0.33	0.29	0.33
MO-UPS	0.52	0.52	0.47	0.42	0.39	0.42	0.5	0.58	0.58	0.53	0.46	0.43	0.46
NI-R	0.38	0.39	0.42	0.45	0.45	0.45	0.49	0.42	0.44	0.47	0.49	0.47	0.47
NI-UNP	0.41	0.45	0.43	0.45	0.42	0.33	0.44	0.46	0.49	0.43	0.45	0.42	0.33
NI-UTX	0.42	0.47	0.44	0.48	0.46	0.33	0.45	0.5	0.49	0.44	0.48	0.46	0.33
orthogonal distribution of β													
AA-OSG	0.45	0.42	0.42	0.4	0.39	0.37	0.4	0.51	0.48	0.49	0.45	0.45	0.4
DUK-IBM	0.46	0.42	0.46	0.4	0.36	0	0.49	0.54	0.48	0.52	0.43	0.36	0
DUK-OSG	0.4	0.44	0.44	0.38	0.42	0.36	0.4	0.49	0.52	0.51	0.45	0.48	0.36
NI-NSC	0.36	0.36	0.35	0.35	0.39	0.46	0.44	0.44	0.42	0.39	0.39	0.43	0.54
NI-OSG	0.39	0.32	0.36	0.36	0.31	0.27	0.37	0.44	0.39	0.41	0.39	0.34	0.23
CNP-OSG	0.4	0.39	0.4	0.42	0.42	0.4	0.45	0.44	0.41	0.42	0.45	0.42	0.4
MO-UPS	0.52	0.52	0.54	0.49	0.44	0.31	0.53	0.58	0.58	0.6	0.56	0.5	0.34
NI-R	0.45	0.46	0.5	0.54	0.57	0.51	0.5	0.49	0.52	0.56	0.61	0.63	0.56
NI-UNP	0.44	0.46	0.54	0.5	0.46	0.47	0.46	0.52	0.52	0.59	0.57	0.54	0.53
NI-UTX	0.22	0.25	0.22	0.23	0.26	0.29	0.41	0.35	0.34	0.32	0.32	0.35	0.38

Table 7. Risk evaluation measure *Risk* in (24), fraction of trading days with decreases in cumulative income, under a Student's *t* distribution for the innovations. The meaning of the ten pairs of stocks is indicated in Table 1.

	<i>k</i> = 0 (Directional Accuracy)						ISA	<i>k</i> = 1 (Accuracy)					
	CSA							CSA					
	0%	20%	30%	40%	50%	60%		0%	20%	30%	40%	50%	60%
linear normalization of β													
AA-OSG	0.45	0.46	0.49	0.46	0.42	0.37	0.48	0.48	0.48	0.49	0.46	0.42	0.37
DUK-IBM	0.45	0.48	0.49	0.46	0.43	0	0.5	0.54	0.55	0.53	0.5	0.43	0
DUK-OSG	0.51	0.46	0.46	0.38	0.44	0.36	0.49	0.54	0.47	0.48	0.41	0.44	0.36
NI-NSC	0.37	0.32	0.27	0.39	0.41	0.4	0.48	0.45	0.38	0.36	0.43	0.47	0.47
NI-OSG	0.41	0.39	0.39	0.38	0.32	0.33	0.44	0.45	0.4	0.4	0.39	0.34	0.36
CNP-OSG	0.45	0.38	0.33	0.34	0.29	0.25	0.46	0.5	0.42	0.39	0.42	0.33	0.25
MO-UPS	0.49	0.49	0.47	0.41	0.42	0.42	0.49	0.53	0.55	0.53	0.48	0.51	0.46
NI-R	0.42	0.41	0.44	0.49	0.5	0.47	0.48	0.45	0.44	0.48	0.53	0.52	0.49
NI-UNP	0.45	0.45	0.4	0.4	0.38	0.25	0.45	0.54	0.51	0.46	0.47	0.46	0.25
NI-UTX	0.43	0.44	0.45	0.43	0.41	0.4	0.5	0.49	0.48	0.48	0.47	0.44	0.44
orthogonal normalization of β													
AA-OSG	0.46	0.41	0.39	0.37	0.35	0.34	0.42	0.49	0.43	0.42	0.38	0.37	0.36
DUK-IBM	0.42	0.43	0.44	0.34	0.23	0	0.45	0.51	0.52	0.52	0.38	0.23	0
DUK-OSG	0.39	0.41	0.4	0.37	0.27	0.18	0.4	0.51	0.52	0.51	0.46	0.37	0.27
NI-NSC	0.29	0.32	0.32	0.35	0.4	0.5	0.43	0.39	0.37	0.32	0.31	0.35	0.42
NI-OSG	0.44	0.4	0.38	0.39	0.32	0.38	0.42	0.5	0.45	0.44	0.43	0.35	0.43
CNP-OSG	0.37	0.38	0.35	0.38	0.36	0.4	0.44	0.43	0.42	0.38	0.38	0.36	0.4
MO-UPS	0.52	0.53	0.54	0.49	0.43	0.3	0.5	0.55	0.55	0.56	0.51	0.43	0.3
NI-R	0.42	0.45	0.49	0.53	0.57	0.55	0.5	0.5	0.5	0.51	0.56	0.59	0.55
NI-UNP	0.43	0.47	0.48	0.4	0.44	0.4	0.47	0.46	0.47	0.48	0.4	0.44	0.4
NI-UTX	0.24	0.25	0.24	0.25	0.25	0.24	0.49	0.38	0.4	0.37	0.37	0.38	0.36

References

1. Khandani, A.E.; Lo, A.W. What happened to the quants in August 2007? *J. Invest. Manag.* **2007**, *5*, 29–78.
2. Lehmann, B. Fads, martingales and market efficiency. *Q. J. Econ.* **1990**, *105*, 1–28.
3. Lo, A.W.; MacKinlay, A.C. When are contrarian profits due to stock market overreaction? *Rev. Financ. Stud.* **1990**, *3*, 175–206.
4. Gatev, E.; Goetzmann, W.N.; Rouwenhorst, K.G. Pairs trading: Performance of a relative-value arbitrage rule. *Rev. Financ. Stud.* **2006**, *19*, 797–827.
5. Kleibergen, F.R.; van Dijk, H.K. Bayesian simultaneous equation analysis using reduced rank structures. *Econom. Theory* **1998**, *14*, 701–743.
6. Kleibergen, F.R.; Paap, R. Priors, posteriors and Bayes factors for a Bayesian analysis of cointegration. *J. Econom.* **2002**, *111*, 223–249.
7. Strachan, R.W.; van Dijk, H.K. *Valuing Structure, Model Uncertainty and Model Averaging in Vector Autoregressive Processes*; Econometric Institute Report EI 2004-23; Erasmus University Rotterdam: Rotterdam, the Netherlands, 2004.
8. Burgess, A.N. A Computational Methodology for Modelling the Dynamics of Statistical Arbitrage. Ph.D. Thesis, University of London, London Business School, London, UK, 1999.
9. Johansen, S. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* **1991**, *59*, 1551–1580.
10. Baştürk, N.; Hoogerheide, L.F.; Kleijn, R.; van Dijk, H.K. Corresponding author: H.K. van Dijk, Department of Econometrics and Tinbergen Institute, Vrije Universiteit Amsterdam and Econometric Institute, Erasmus University Rotterdam. Prior ignorance, likelihood shape and posterior existence in a cointegration model. Unpublished working paper, 2015.
11. Geweke, J. Bayesian treatment of independent Student-*t* linear model. *J. Appl. Econom.* **1993**, *8*, 19–40.
12. Dickey, J. The weighted likelihood ratio, linear hypothesis on normal location parameters. *Ann. Math. Stat.* **1971**, *42*, 204–223.
13. Verdinelli, I.; Wasserman, L. Computing Bayes factors using a generalization of the Savage-Dickey density ratio. *J. Am. Stat. Assoc.* **1995**, *90*, 614–618.
14. Chao, J.C.; Phillips, P.C.B. Model selection in partially nonstationary vector autoregressive processes with reduced rank structure. *J. Econom.* **1999**, *91*, 227–271.

15. Hoogerheide, L.F.; Opschoor, A.; van Dijk, H.K. A class of adaptive importance sampling weighted EM algorithms for efficient and robust posterior and predictive simulation. *J. Econom.* **2012**, *171*, 101–120.
16. West, K.D.; Edison, H.J.; Cho, D. A utility-based comparison of some models of exchange rate volatility. *J. Int. Econ.* **1993**, *35*, 23–45.
17. Marquering, W.; Verbeek, M. The economic value of predicting stock index returns and volatility. *J. Financ. Q. Anal.* **2004**, *39*, 407–429.
18. Furmston, T.; Hailes, S.; Morton, A.J. A Bayesian Residual-Based Test for Cointegration. 2013. Available online: <http://arxiv.org/abs/1311.0524> (accessed on 21 February 2016).
19. Bracegirdle, C.; Barber, D. Bayesian Conditional Cointegration. 2012. Available online: <http://arxiv.org/abs/1206.6459> (accessed on 21 February 2016).
20. Kleibergen, F.R.; van Dijk, H.K. On the shape of the likelihood/posterior in cointegration models. *Econom. Theory* **1994**, *10*, 514–551.
21. Zellner, A. *An Introduction to Bayesian Inference in Econometrics*; Wiley: New York, NY, USA, 1971.
22. Chen, M.H. Importance-weighted marginal Bayesian posterior density estimation. *J. Am. Stat. Assoc.* **1994**, *89*, 818–824.



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).