

Gatarek, Lukasz; Hoogerheide, Lennart; van Dijk, Herman K.

Working Paper

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

Tinbergen Institute Discussion Paper, No. 14-039/III

Provided in Cooperation with:

Tinbergen Institute, Amsterdam and Rotterdam

Suggested Citation: Gatarek, Lukasz; Hoogerheide, Lennart; van Dijk, Herman K. (2014) : Return and Risk of Pairs Trading using a Simulation-based Bayesian Procedure for Predicting Stable Ratios of Stock Prices, Tinbergen Institute Discussion Paper, No. 14-039/III, Tinbergen Institute, Amsterdam and Rotterdam

This Version is available at:

<https://hdl.handle.net/10419/98882>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

TI 2014-039/III
Tinbergen Institute Discussion Paper



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

Lukasz Gatarek¹

Lennart F. Hoogerheide²

Herman K. van Dijk^{1,2}

¹ *Erasmus School of Economics, Erasmus University Rotterdam, Tinbergen Institute;*

² *Faculty of Economics and Business Administration, VU University Amsterdam, Tinbergen Institute, the Netherlands.*

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

More TI discussion papers can be downloaded at <http://www.tinbergen.nl>

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Duisenberg school of finance is a collaboration of the Dutch financial sector and universities, with the ambition to support innovative research and offer top quality academic education in core areas of finance.

DSF research papers can be downloaded at: <http://www.dsf.nl/>

Duisenberg school of finance
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 8579

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

Lukasz T. Gatarek, Lennart Hoogerheide, and Herman K. van Dijk

March 2014

Abstract

In this paper we consider two cases of pairs trading strategies: a conditional statistical arbitrage method and an implicit statistical arbitrage method. We use a simulation-based Bayesian procedure for predicting stable ratios, defined in a cointegration model, of pairs of stock prices. We show the effect that using an encompassing 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 is done using stocks that are ingredients of the Dow Jones Composite Average index. The results show that the 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 classification: C11, C15, C32, C58, G17

1 Introduction

In this paper we consider statistical arbitrage strategies. These statistical arbitrage strategies presume that the patterns observed in the historical data are expected to be repeated in the future. As such the statistical arbitrage is a purely statistical approach designed to exploit market inefficiencies.

Khadani and Lo (2007) consider a specific strategy – first proposed by Lehmann (1990) and Lo and MacKinlay (1990) – 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 the 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, we buy it. Gatev et al. (2006) 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 deviation from the equilibrium in the cointegration model. This approach differs from Gatev et al. (2006), who implement a nonparametric framework. Gatev et al. (2006) 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. On the other hand, in the cointegration analysis that we perform the spread between two assets is modeled as the temporary deviation from the long-term equilibrium 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. Therefore, the pairs trading strategy is strongly dependent on the stability of the ratios of pairs of stocks.

The estimated and predicted spread are both computed from the estimated cointegration model. For the estimation of the cointegration model, we work with the Metropolis-Hastings (M-H) type of sampler derived under an encompassing prior. This sampling algorithm is derived by Kleibergen and Van Dijk (1998) for the Simultaneous Equations Model and by Kleibergen and Paap (2002) for the cointegration model. They specify a linear normalization to identify the parameters in the model. However, Strachan and Van Dijk (2004) point out the possible distortion 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. 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.

We find that 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 do not only consider the normal distribution, but we also extend our approach to the case of the Student's 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 the 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* (Burgess, 1999) between the prices $y_{t,1}$ and $y_{t,2}$ of two stocks, where the spread

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

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 (1999). 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

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

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 the case of $k = 0$ we only look at the sign of the expected change in the next day. 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 will 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 will use the posterior median, since the posterior distribution has Cauchy type tails in one of the model specifications that we compare, where these Cauchy type tails cause that the coefficients have no posterior means.

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

$$ISA(s_t) = -s_t. \tag{3}$$

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 will consider how the Bayesian analysis of the cointegration model (under linear or orthogonal normalization) answers all these questions. In order to answer the first and third questions we will use the posterior distribution (more precisely, the posterior median) of the parameters in the cointegration model. In order to answer the second question we will 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. (2006). In the strategy of Gatev et al. (2006) 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 significant 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 $\{Y_t\}_{t=1}^T$

$$Y_t = \Phi Y_{t-1} + \varepsilon_t, \tag{4}$$

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

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

where $\Pi' = \Phi - I_n$ (with I_n the $n \times n$ identity matrix) is the long-run multiplier matrix, see e.g. Johansen (1991) and Kleibergen and Paap (2002).

If Π is a zero matrix, the series Y_t contains n unit roots. If the matrix Π has full rank, the univariate series in Y_t are stationary. 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 the 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 $\beta\alpha = \beta BB^{-1}\alpha$ 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. The first normalization is the most commonly used normalization, the linear normalization, 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 second normalization is the orthogonal normalization, where we have

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

In this paper we will 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 $r = 1$:

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

¹We 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.

with spread

$$s_t = \beta_1 y_{t,1} + \beta_2 y_{t,2} \quad (10)$$

and

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

so that

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

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 $\beta_1 = 1$:

$$\beta = \begin{pmatrix} 1 \\ \beta_2 \end{pmatrix}, \quad (13)$$

whereas under the orthogonal normalization we have

$$\beta' \beta = \beta_1^2 + \beta_2^2 = 1, \quad (14)$$

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

$$\beta_2 = \sqrt{1 - \beta_1^2}. \quad (15)$$

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 Encompassing 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 the 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 $\beta'\beta_{\perp} \equiv 0$, $\beta'_{\perp}\beta \equiv I_{n-r}$, $\alpha_{\perp}\alpha' \equiv 0$ 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 (16) 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 (16) 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

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

The matrices in decomposition (16) in terms of the blocks of U , S and V are given by

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

Under the orthogonal normalization, the matrices are partitioned as

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

and the following relations hold:

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

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 (2002) 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 we derive a computationally more convenient expression for this Jacobian under the linear normalization. In Appendix C the Jacobian is derived under the orthogonal normalization of β . Gatarek et al. (2010) 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.

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 (2002), 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 (2002), 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 (2002) 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 C 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).

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 (1993).

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 use the method of Kleibergen and Paap (2002) who evaluate the Bayes factor as the Savage-Dickey density ratio, see Dickey (1971) and Verdinelli and Wasserman (1995). 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 (1999) 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 observe 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 January 1, 2009 until December 31, 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% ($c = 0.001$).

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 $(50 + \xi/2)\%$ percentile and the $(50 - \xi/2)\%$ percentile of the posterior distribution of the current spread have the same sign, where we consider the cases of $\xi = 20\%$, 30% , 40% , 50% or 60% . The case of $\xi = 60\%$ 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 $\xi = 0\%$ 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

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 January 1, 2009 until June 30, 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.

months of 2009). Then the average daily capital engagement is given by:

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

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

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:

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

$$Profitability_{ISA} \equiv \frac{\text{Cumulative Return of } ISA \text{ at time } T}{ADCE_{ISA}} \quad (23)$$

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 D. 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 $k = 0$ is compared with the counterpart under $k = 1$. 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 ξ . 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. 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 ($k = 0$)						
		ξ						
		0%	20%	30%	40%	50%	60%	
		linear normalization						
normal		12.52	17.08	19.72	26.94	31.32	47.37	
Student's t		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 t		60.27	66.41	67.67	81.78	91.53	114.05	
		ISA	CSA ($k = 1$)					
			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 t	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 t	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 ($k = 0$)						
		ξ						
		0%	20%	30%	40%	50%	60%	
		linear normalization						
normal		0.45	0.43	0.42	0.42	0.42	0.35	
Student's t		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 t		0.39	0.40	0.40	0.39	0.36	0.33	
		ISA	CSA ($k = 1$)					
			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 t	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 t	0.46	0.47	0.47	0.45	0.42	0.39	0.35	

5 Conclusions and further research

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 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.

If we consider 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 more important to specify the distribution of the innovations more 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. (2012) to perform the posterior simulation for the cointegration model with errors obeying a finite mixture distribution.

References

- Burgess, A. N., 1999. A computational methodology for modelling the dynamics of statistical arbitrage. PhD thesis, University of London, London Business School.
- Chao, J. C., Phillips, P. C. B., 1999. Model selection in partially nonstationary vector autoregressive processes with reduced rank structure. *Journal of Econometrics* 91, 227–271.
- Chen, M. H., 1994. Importance-weighted marginal Bayesian posterior density estimation. *Journal of the American Statistical Association* 89, 818–824.
- Dickey, J., 1971. The weighted likelihood ratio, linear hypothesis on normal location parameters. *The Annals of Mathematical Statistics* 42, 204–223.
- Gatarek, L., Hoogerheide, L. F., Kleijn, R., Van Dijk, H. K., 2010. Prior ignorance, normalization and reduced rank probabilities in cointegration models.
- Gatev, E., Goetzmann, W. N., Rouwenhorst, K. G., 2006. Pairs trading: Performance of a relative-value arbitrage rule. *The Review of Financial Studies* 19, 797–827.
- Geweke, J., 1993. Bayesian treatment of independent Student-t linear model. *Journal of Applied Econometrics* 8, 19–40.
- Hoogerheide, L. F., Opschoor, A., Van Dijk, H. K., 2012. A class of adaptive importance sampling weighted EM algorithms for efficient and robust posterior and predictive simulation. *Journal of Econometrics* 171 (2), 101–120.
- Johansen, S., 1991. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* 59, 1551–1580.
- Khadani, A. E., Lo, A. W., 2007. What happened to the quants in August 2007? Working Paper, Massachusetts Institute of Technology Sloan School of Management, Cambridge.
- Kleibergen, F. R., Paap, R., 2002. Priors, posteriors and Bayes factors for a Bayesian analysis of cointegration. *Journal of Econometrics* 111, 223–249.

- Kleibergen, F. R., Van Dijk, H. K., 1994. On the shape of the likelihood/posterior in cointegration models. *Econometric Theory* 10, 514–551.
- Kleibergen, F. R., Van Dijk, H. K., 1998. Bayesian simultaneous equation analysis using reduced rank structures. *Econometric Theory* 14, 701–743.
- Lehmann, B., 1990. Fads, martingales and market efficiency. *Quarterly Journal of Economics* 105, 1–28.
- Lo, A. W., MacKinlay, A. C., 1990. When are contrarian profits due to stock market overreaction? *Review of Financial Studies* 3, 175–206.
- Magnus, J. R., Neudecker, H., 1999. *Matrix differential calculus with application in statistics and econometric*. Wiley, (2. ed.).
- Strachan, R. W., Van Dijk, H. K., 2004. Valuing structure, model uncertainty and model averaging in vector autoregressive processes. *Econometric Institute Report EI 2004-23*, Erasmus University Rotterdam.
- Verdinelli, I., Wasserman, L., 1995. Computing Bayes factors using a generalization of the Savage-Dickey density ratio. *Journal of the American Statistical Association* 90, 614–618.
- Zellner, A., 1971. *An introduction to Bayesian inference in econometrics*. Wiley, New York.

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 (2002). 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 (1994).

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

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

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:

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

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 (1971) 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 (2002) adopt the approach suggested by Chen (1994). 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 (2002). In Appendix C 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 $\min\left(\frac{w(\alpha^{i+1}, \beta^{i+1}, \lambda^{i+1}, \Sigma^{i+1})}{w(\alpha^i, \beta^i, \lambda^i, \Sigma^i)}, 1\right)$.

B Jacobian of the transformation from Π to (α, β, Σ) under linear normalization

This discussion is based on Kleibergen and Paap (2002). They derive the Jacobian transformation under the linear normalization.

It is convenient to derive the Jacobian in two steps based on (26)

$$\Pi = \beta\alpha + \beta_{\perp}\lambda\alpha_{\perp} = \beta\alpha_1\theta + \beta_{\perp}\lambda\theta_{\perp},$$

where $\alpha = \begin{pmatrix} \alpha_1 & \alpha_2 \end{pmatrix}$, $\theta = \alpha_1^{-1}\alpha$, and $\theta = \begin{pmatrix} I_r & -\theta_2 \end{pmatrix}$. First the derivatives of Π with respect to $(\alpha_1, \theta_2, \lambda, \beta_2)$ are computed, and then in the final step, with respect to $(\alpha, \lambda, \beta_2)$.

$$\frac{\partial \text{vec}(\Pi)}{\partial \text{vec}(\alpha_1)'} = \theta' \otimes \beta$$

$$\frac{\partial \text{vec}(\Pi)}{\partial \text{vec}(\theta_2)'} = \begin{pmatrix} 0 \\ -I_{n-r} \end{pmatrix} \otimes \beta\alpha_1 + (I_n \otimes \beta_{\perp}\lambda) \frac{\partial \text{vec}(\theta_{\perp})}{\partial \text{vec}(\theta_2)'}$$

$$\frac{\partial \text{vec}(\Pi)}{\partial \text{vec}(\lambda)'} = \theta'_{\perp} \otimes \beta_{\perp}$$

$$\frac{\partial \text{vec}(\Pi)}{\partial \text{vec}(\beta_2)'} = \theta'\alpha_1' \otimes \begin{pmatrix} 0 \\ -I_{n-r} \end{pmatrix} + (\theta'_{\perp}\lambda' \otimes I_n) \frac{\partial \text{vec}(\beta_{\perp})}{\partial \text{vec}(\beta_2)'}$$

with

$$\frac{\partial \text{vec}(\theta_{\perp})}{\partial \text{vec}(\theta_2)'} = (I_n \otimes \theta_n^{-\frac{1}{2}}) \frac{\partial \text{vec} \begin{pmatrix} \theta_2' & I_{n-r} \end{pmatrix}}{\partial \text{vec}(\theta_2)'} + \left(\begin{pmatrix} \theta_2 \\ I_{n-r} \end{pmatrix} \otimes I_{n-r} \right) \frac{\partial \text{vec}(\theta_n^{-\frac{1}{2}})}{\partial \text{vec}(\theta_n^{\frac{1}{2}})'} \frac{\partial \text{vec}(\theta_n^{\frac{1}{2}})}{\partial \text{vec}(\theta_n)'} \frac{\partial \text{vec}(\theta_n)}{\partial \text{vec}(\theta_2)'}$$

$$\frac{\partial \text{vec}(\beta_{\perp})}{\partial \text{vec}(\beta_2)'} = (\beta_n^{-\frac{1}{2}} \otimes I_n) \frac{\partial \text{vec} \left(\begin{pmatrix} \beta_2 & I_{n-r} \end{pmatrix}' \right)}{\partial \text{vec}(\beta_2)'} + (I_{n-r} \otimes \begin{pmatrix} \beta_2' \\ I_{n-r} \end{pmatrix}) \frac{\partial \text{vec}(\beta_n^{-\frac{1}{2}})}{\partial \text{vec}(\beta_n^{\frac{1}{2}})'} \frac{\partial \text{vec}(\beta_n^{\frac{1}{2}})}{\partial \text{vec}(\beta_n)'} \frac{\partial \text{vec}(\beta_n)}{\partial \text{vec}(\beta_2)'}$$

where $\theta_n = I_{n-r} + \theta_2'\theta_2$ and $\beta_n = I_{n-r} + \beta_2\beta_2'$. Also, $\theta_{\perp} = \theta_n^{-\frac{1}{2}} \begin{pmatrix} \theta_2' & I_{n-r} \end{pmatrix}$ and $\beta_{\perp} = \begin{pmatrix} \beta_2 & I_{n-r} \end{pmatrix} \beta_n^{-\frac{1}{2}}$.

$$\begin{aligned}
\frac{\partial \text{vec} \begin{pmatrix} \theta'_2 & I_{n-r} \end{pmatrix}}{\partial \text{vec}(\theta_2)'} &= \begin{pmatrix} I_r \\ 0 \end{pmatrix} \otimes I_{n-r} K_{n-r,r} \\
\frac{\partial \text{vec}(\theta_n^{-\frac{1}{2}})}{\partial \text{vec}(\theta_n^{\frac{1}{2}})'} &= -\theta_n^{-\frac{1}{2}'} \otimes \theta_n^{-\frac{1}{2}} \\
\frac{\partial \text{vec}(\theta_n^{\frac{1}{2}})}{\partial \text{vec}(\theta_n)'} &= (\theta_n^{\frac{1}{2}'} \otimes I_{n-r} + I_{n-r} \otimes \theta_n^{\frac{1}{2}})^{-1} \\
\frac{\partial \text{vec}(\theta_n)}{\partial \text{vec}(\theta_2)'} &= I_{n-r} \otimes \theta'_2 + (\theta'_2 \otimes I_{n-r}) K_{n-r,r} \\
\frac{\partial \text{vec} \begin{pmatrix} \beta_2 & I_{n-r} \end{pmatrix}'}{\partial \text{vec}(\beta_2)'} &= (I_{n-r} \otimes \begin{pmatrix} I_r \\ 0 \end{pmatrix}) K_{n-r,r} \\
\frac{\partial \text{vec}(\beta_n^{-\frac{1}{2}})}{\partial \text{vec}(\beta_n^{\frac{1}{2}})'} &= -\beta_n^{-\frac{1}{2}'} \otimes \beta_n^{-\frac{1}{2}} \\
\frac{\partial \text{vec}(\beta_n^{\frac{1}{2}})}{\partial \text{vec}(\beta_n)'} &= (\beta_n^{\frac{1}{2}'} \otimes I_{n-r} + I_{n-r} \otimes \beta_n^{\frac{1}{2}})^{-1} \\
\frac{\partial \text{vec}(\beta_n)}{\partial \text{vec}(\beta_2)'} &= \beta_2 \otimes I_{n-r} + (I_{n-r} \otimes \beta'_2) K_{n-r,r}
\end{aligned}$$

where $K_{m,n}$ is a commutation matrix such that $K_{m,n} \text{vec}(A) = \text{vec}(A')$ for any matrix A of dimension $m \times n$.

B.1 Jacobian with λ unrestricted

This discussion goes beyond Kleibergen and Paap (2002). We derive some convenient computational reductions compared to Kleibergen and Paap (2002). We start with (26) again in order to obtain:

$$\begin{aligned}
d\Pi &= d(\beta\alpha) + d(\beta_\perp \lambda \alpha_\perp) \\
&= \beta(d\alpha) + (d\beta)\alpha + \beta_\perp \lambda(d\alpha_\perp) + \beta_\perp(d\lambda)\alpha_\perp + (d\beta_\perp)\lambda\alpha_\perp.
\end{aligned} \tag{36}$$

Hence,

$$\begin{aligned}
\frac{\partial \text{vec}(\Pi)}{\partial \text{vec}(\alpha)'} &= I_n \otimes \beta + (I_n \otimes \beta_\perp \lambda) \frac{\partial \text{vec}(\alpha_\perp)}{\partial \text{vec}(\alpha)'} \\
\frac{\partial \text{vec}(\Pi)}{\partial \text{vec}(\beta_2)'} &= \alpha' \otimes \begin{pmatrix} 0 \\ -I_{n-r} \end{pmatrix} + (\alpha'_\perp \lambda' \otimes I_n) \frac{\partial \text{vec}(\beta_\perp)}{\partial \text{vec}(\beta_2)'} \\
\frac{\partial \text{vec}(\Pi)}{\partial \text{vec}(\lambda)'} &= \alpha'_\perp \otimes \beta_\perp.
\end{aligned}$$

For the derivation of the Jacobian we thus need $\frac{\partial \text{vec}(\alpha_\perp)}{\partial \text{vec}(\alpha)'}$ and $\frac{\partial \text{vec}(\beta_\perp)}{\partial \text{vec}(\beta_2)'}$. We shall derive the derivative $\frac{\partial \text{vec}(\beta_\perp)}{\partial \text{vec}(\beta)'} using the implicit function theorem. The other derivatives can be obtained from it. By definition, it holds$

$$\beta'_\perp \beta = 0_{n-r \times r}. \quad (37)$$

Also, we have restricted β_\perp to be semi-orthogonal, that is

$$\beta'_\perp \beta_\perp = I_{n-r}. \quad (38)$$

Define the function

$$f(\beta_\perp; \beta) = \begin{pmatrix} \beta'_\perp \beta & \beta'_\perp \beta_\perp - I_{n-r} \end{pmatrix} = 0_{n-r \times n}.$$

Then a sufficient condition for the implicit function theorem is that the matrix $\frac{\partial \text{vec} f(\beta_\perp; \beta)}{\partial \text{vec}(\beta_\perp)'}$ is non-singular. We find that

$$\left| \frac{\partial \text{vec}(f(\beta_\perp; \beta))}{\partial \text{vec}(\beta_\perp)'} \right| = \left| \frac{\frac{\partial \text{vec}(\beta'_\perp \beta)}{\partial \text{vec}(\beta_\perp)'}}{\frac{\partial \text{vec}(\beta'_\perp \beta_\perp - I_{n-r})}{\partial \text{vec}(\beta_\perp)'}} \right| \quad (39)$$

$$= \left| \frac{(\beta' \otimes I_{n-r}) K_{n, n-r}}{I_{n-r} \otimes \beta'_\perp + (\beta'_\perp \otimes I_{n-r}) K_{n, n-r}} \right| \quad (40)$$

$$= \left| \begin{pmatrix} (\beta' \otimes I_{n-r}) K_{n, n-r} \\ I_{n-r} \otimes \beta'_\perp + (\beta'_\perp \otimes I_{n-r}) K_{n, n-r} \end{pmatrix} \begin{pmatrix} (\beta' \otimes I_{n-r}) K_{n, n-r} \\ I_{n-r} \otimes \beta'_\perp + (\beta'_\perp \otimes I_{n-r}) K_{n, n-r} \end{pmatrix}' \right|^{\frac{1}{2}} \quad (41)$$

$$= \left| \begin{matrix} \beta' \beta \otimes I_{n-r} & 0 \\ 0 & 4N_{n-r} \end{matrix} \right| = 0, \quad (42)$$

because

$$\begin{aligned} & (I_{n-r} \otimes \beta'_\perp + (\beta'_\perp \otimes I_{n-r}) K_{n, n-r}) (I_{n-r} \otimes \beta'_\perp + (\beta'_\perp \otimes I_{n-r}) K_{n, n-r})' = \\ & 2I_{(n-r)^2} + (I_{n-r} \otimes \beta'_\perp) K_{n-r, n} (\beta_\perp \otimes I_{n-r}) + (\beta'_\perp \otimes I_{n-r}) K_{n, n-r} (I_{n-r} \otimes \beta'_\perp) = \\ & 2I_{(n-r)^2} + (I_{n-r} \otimes \beta'_\perp) (I_{n-r} \otimes \beta_\perp) K_{n-r} + (\beta'_\perp \otimes I_{n-r}) (\beta'_\perp \otimes I_{n-r}) K_{n-r} = \\ & \qquad \qquad \qquad 4N_{n-r}, \end{aligned}$$

and $N_{n-r} = \frac{1}{2}(I_{(n-r)^2} + K_{n-r})$ is a matrix with rank $\frac{1}{2}(n-r)(n-r+1)$, see e.g. Magnus and Neudecker (1999). The Jacobian (39) thus has a rank deficiency of $(n-r)n - \frac{1}{2}(n-r)$

$r)(n - r + 1) = \frac{1}{2}(n - r)(n - r - 1)$. There are currently not enough restrictions to uniquely identify the derivative. For instance, for any orthogonal matrix C of dimension $(n - r) \times (n - r)$ with $C'C = I_{n-r}$ the matrix $\beta_{\perp}C$ also satisfies (37) and (38).

One might be tempted to conclude that restrictions (37) and (38) taken together suffice for identification of $\frac{\partial \text{vec}(\beta_{\perp})}{\partial \text{vec}(\beta)'}'$, since $n \times (n - r)$ restrictions are needed. However, restriction (38) represents at most $\frac{1}{2}(n - r)(n - r + 1)$ unique restrictions, since both left and right hand sides are symmetric matrices. Additional restrictions are needed.

The additional restrictions can be chosen to be consistent with the explicit expressions for β_{\perp} and α'_{\perp} from Kleibergen and Paap (2002) by partitioning β_{\perp} as $\begin{pmatrix} \beta_{\perp 1} \\ \beta_{\perp 2} \end{pmatrix}$, with $\beta_{\perp 1} : r \times (n - r)$ and $\beta_{\perp 2} : (n - r) \times (n - r)$ such that $\beta_{\perp 2} = \begin{pmatrix} 0_{n-r \times r} & I_{n-r} \end{pmatrix} \beta_{\perp}$, and requiring $\beta_{\perp 2}$ to be symmetric, i.e.

$$\beta_{\perp 2} = \beta'_{\perp 2}. \quad (43)$$

It is convenient to combine the restrictions (38) and (43) into a single restriction such that the non-singularity condition for the implicit function theorem can be made to hold. Therefore, we replace (38) by the sum of (38) and (43) such that

$$\beta'_{\perp} \beta_{\perp} + \begin{pmatrix} 0_{n-r \times r} & I_{n-r} \end{pmatrix} \beta_{\perp} - \beta'_{\perp} \begin{pmatrix} 0_{r \times n-r} \\ I_{n-r} \end{pmatrix} = I_{n-r}. \quad (44)$$

Note that restriction (44) implies both (38) and (43). This is easily seen by adding (44) and its transpose together and dividing the result by two, which reduces to (38).

Define the function

$$g(\beta_{\perp}; \beta) = \begin{pmatrix} \beta'_{\perp} \beta & \beta'_{\perp} \beta_{\perp} + \begin{pmatrix} 0 & I \end{pmatrix} \beta_{\perp} - \beta'_{\perp} \begin{pmatrix} 0 \\ I \end{pmatrix} - I \end{pmatrix},$$

then

$$\begin{aligned}
\left| \frac{\partial \text{vec}(g(\beta_\perp; \beta))}{\partial \text{vec}(\beta_\perp)'} \right| &= \left| \frac{\frac{\partial \text{vec}(\beta'_\perp \beta)}{\partial \text{vec}(\beta_\perp)'}}{\partial \text{vec}(\beta'_\perp \beta_\perp + \begin{pmatrix} 0 & I \end{pmatrix} \beta_\perp - \beta'_\perp \begin{pmatrix} 0 \\ I \end{pmatrix} - I)} \right| \\
&= \left| \frac{(\beta' \otimes I_{n-r}) K_{n,n-r}}{I_{n-r} \otimes \beta'_\perp + (\beta'_\perp \otimes I_{n-r}) K_{n,n-r} + I_{n-r} \otimes \begin{pmatrix} 0 & I \end{pmatrix} - \left(\begin{pmatrix} 0 \\ I \end{pmatrix} \otimes I_{n-r} \right) K_{n,n-r}} \right| \\
&= \left| \begin{pmatrix} (\beta' \otimes I_{n-r}) K_{n,n-r} \\ I_{n-r} \otimes \beta'_\perp + (\beta'_\perp \otimes I_{n-r}) K_{n,n-r} \end{pmatrix} \begin{pmatrix} (\beta' \otimes I_{n-r}) K_{n,n-r} \\ I_{n-r} \otimes \beta'_\perp + (\beta'_\perp \otimes I_{n-r}) K_{n,n-r} \end{pmatrix}' \right|^{\frac{1}{2}} \\
&= \left| \begin{array}{cc} \beta' \beta \otimes I_{n-r} & (\beta'_\perp \otimes I)(I - K) \\ (I - K')(\beta_2 \otimes I) & 4N_{n-r} \end{array} \right| \neq 0.
\end{aligned}$$

Thus, we find that the matrix $\frac{\partial \text{vec}(g(\beta_\perp; \beta))}{\partial \text{vec}(\beta_\perp)'}$ is non-singular, which is what was required.

The additional condition imposed on β and β_\perp via (44) can be easily met by a transformation of a given matrix β_\perp that satisfies only (38). To that end we first define $\beta^+ = (\beta' \beta)^{-1} \beta'$. Now assume that there exists β_\perp^* that satisfies condition (38), i.e. $\beta_\perp^{*'} \beta = 0$. Then a transformation of this matrix, given by $\beta_\perp = \beta_\perp^* \beta_{\perp 2}^{*'} + [\beta_{\perp 2}^* (\beta_\perp^{*'} \beta_\perp^*)^{-1} \beta_{\perp 2}^{*'}]^{\frac{1}{2}}$ will satisfy both (38) and (44).

Condition (44) has implications on derivations of $\frac{\partial \text{vec}(\beta_\perp)}{\partial \text{vec}(\beta)'}$, $\frac{\partial \text{vec}(\beta_\perp)}{\partial \text{vec}(\beta_2)'}$ and $\frac{\partial \text{vec}(\alpha_\perp)}{\partial \text{vec}(\alpha)'}$. We derive $\frac{\partial \text{vec}(\beta_\perp)}{\partial \text{vec}(\beta)'}$ as it provides the workhorse for finding $\frac{\partial \text{vec}(\beta_\perp)}{\partial \text{vec}(\beta_2)'}$ and $\frac{\partial \text{vec}(\alpha_\perp)}{\partial \text{vec}(\alpha)'}$.

We denote by $d\beta_\perp$ the effect of the additional condition imposed by (38) and (44) together, compared to (38) only. Thus, if a given β_\perp satisfies condition (38), then $\beta_\perp + d\beta_\perp$ satisfies both (38) and (44). We substitute $\beta_\perp + d\beta_\perp$ for β_\perp in (44) using the convention that products of differentials vanish. Then,

$$\beta'_\perp \beta_\perp + \beta'_\perp d\beta_\perp + d\beta'_\perp \beta_\perp + \begin{pmatrix} 0 & I \end{pmatrix} \beta_\perp + \begin{pmatrix} 0 & I \end{pmatrix} d\beta_\perp - \beta'_\perp \begin{pmatrix} 0 \\ I \end{pmatrix} - d\beta'_\perp \begin{pmatrix} 0 \\ I \end{pmatrix} = I_{n-r}$$

and by exploiting (38) and (43) this reduces to

$$\beta'_\perp d\beta_\perp = -d\beta'_\perp \beta_\perp - \begin{pmatrix} 0 & I \end{pmatrix} d\beta_\perp + d\beta'_\perp \begin{pmatrix} 0 \\ I \end{pmatrix}. \tag{45}$$

From orthogonality, it holds that $d\beta_{\perp} = -\beta^{+'}d\beta'_{\perp} - \beta_{\perp}d\beta'_{\perp}\beta_{\perp}$ and hence also $\beta'_{\perp}d\beta_{\perp} = -\beta'_{\perp}\beta^{+'}d\beta'_{\perp} - \beta'_{\perp}\beta_{\perp}d\beta'_{\perp}\beta_{\perp} = -d\beta'_{\perp}\beta_{\perp}$. Let us define

$$Q = \beta'_{\perp}d\beta_{\perp} = -d\beta'_{\perp}\beta_{\perp}. \quad (46)$$

Then, based on 45 and 46, we obtain

$$\begin{aligned} d\beta_{\perp} &= -\beta^{+'}d\beta'_{\perp}\beta_{\perp} + \beta_{\perp}Q \\ &= -\beta^{+'}d\beta'_{\perp}\beta_{\perp} + \beta_{\perp} \left[-d\beta'_{\perp}\beta_{\perp} - \begin{pmatrix} 0 & I \end{pmatrix} d\beta_{\perp} + d\beta'_{\perp} \begin{pmatrix} 0 \\ I \end{pmatrix} \right]. \end{aligned}$$

Collecting terms with $d\beta_{\perp}$ at the left-hand side and terms involving $d\beta$ at the right-hand side:

$$d\beta_{\perp} + \beta_{\perp}d\beta'_{\perp}\beta_{\perp} + \beta_{\perp} \begin{pmatrix} 0 & I \end{pmatrix} d\beta_{\perp} - \beta_{\perp}d\beta'_{\perp} \begin{pmatrix} 0 \\ I \end{pmatrix} = -\beta^{+'}d\beta'_{\perp}\beta_{\perp}. \quad (47)$$

In vec notation we obtain:

$$\begin{aligned} \left[I_{n-r} \otimes I_n + (\beta'_{\perp} \otimes \beta_{\perp})K + I_{n-r} \otimes \beta_{\perp} \begin{pmatrix} 0 & I \end{pmatrix} - \left(\begin{pmatrix} 0 & I \end{pmatrix} \otimes \beta_{\perp} \right) K \right] \text{vec}d\beta_{\perp} = \\ -(\beta'_{\perp} \otimes \beta^{+'})K \text{vec}d\beta, \end{aligned}$$

and thus

$$\begin{aligned} \frac{\partial \text{vec}(\beta_{\perp})}{\partial \text{vec}(\beta)'} &= - \left[I_{n-r} \otimes I_n + (\beta'_{\perp} \otimes \beta_{\perp})K + I_{n-r} \otimes \beta_{\perp} \begin{pmatrix} 0 & I \end{pmatrix} - \left(\begin{pmatrix} 0 & I \end{pmatrix} \otimes \beta_{\perp} \right) K \right]^{-1} \times \\ &\quad (\beta'_{\perp} \otimes \beta^{+'})K \\ &= - \left[I_{n(n-r)} + (\beta'_{\perp} \otimes \beta_{\perp})K + \begin{pmatrix} 0 & (I_{n-r} \otimes \beta_{\perp})(I - K) \end{pmatrix} \right]^{-1} (\beta'_{\perp} \otimes \beta^{+'})K. \end{aligned}$$

Equation (47) provides the workhorse for finding $\frac{\partial \text{vec}(\beta_{\perp})}{\partial \text{vec}(\beta_2)'}$ and $\frac{\partial \text{vec}(\alpha_{\perp})}{\partial \text{vec}(\alpha)'}$.

C Jacobian of the transformation from Π to (α, β, Σ) under orthogonal normalization

This discussion is based on Gatarek et al. (2010). 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

$$\Pi = USV' = \begin{pmatrix} U_1 & U_2 \end{pmatrix} \begin{pmatrix} S_1 & 0 \\ 0 & S_2 \end{pmatrix} \begin{pmatrix} V_1' \\ V_2' \end{pmatrix}, \quad (48)$$

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 $J(\Pi, (\alpha, \beta_2, \lambda))$

$$J(\Pi, (\beta_2, \alpha, \lambda)) = \begin{pmatrix} \frac{\partial \text{vec} \Pi}{\partial (\text{vec} \beta_2)'} & \frac{\partial \text{vec} \Pi}{\partial (\text{vec} \alpha)'} & \frac{\partial \text{vec} \Pi}{\partial (\text{vec} \lambda)'} \end{pmatrix}. \quad (49)$$

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

$$\frac{\partial \text{vec} \Pi}{\partial (\text{vec} \alpha)'} = (I_p \otimes \beta) + \frac{\partial \text{vec} \Pi}{\partial (\text{vec} \alpha_\perp)'} \frac{\partial \text{vec} \alpha_\perp}{\partial (\text{vec} \alpha)'}, \quad (50)$$

and

$$\frac{\partial \text{vec} \Pi}{\partial (\text{vec} \alpha_\perp)'} = (I \otimes \beta_\perp \lambda). \quad (51)$$

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

$$\frac{\partial \text{vec} \alpha_\perp}{\partial (\text{vec} \alpha)'} = \left((c(\alpha c)^{-1} \otimes (c(\alpha c)^{-1} \alpha c_\perp)') - (c(\alpha c)^{-1} \otimes c_\perp') \right) K_{r,p}, \quad (52)$$

so that

$$\frac{\partial \text{vec} \Pi}{\partial (\text{vec} \alpha)'} = (I_p \otimes \beta) + \left(c(\alpha c)^{-1} \otimes \beta_\perp \lambda c_\perp' ((c(\alpha c)^{-1} \alpha)' - I_p) \right) K_{r,p}. \quad (53)$$

Then for β_2 we obtain

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

²In 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 .

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

$$\begin{aligned} 0 &= K_r(I_r \otimes \beta'_1)d\text{vec}\beta_1 + (I_r \otimes \beta'_1)d\text{vec}\beta_1 + K_r(I_r \otimes \beta'_2)d\text{vec}\beta_2 + (I_r \otimes \beta'_2)d\text{vec}\beta_2 \\ &= (K_r + I_r)((I_r \otimes \beta'_1)d\text{vec}\beta_1) + (K_r + I_r)((I_r \otimes \beta'_2)d\text{vec}\beta_2) \\ &= 2N_r((I_r \otimes \beta'_1)d\text{vec}\beta_1) + 2N_r((I_r \otimes \beta'_2)d\text{vec}\beta_2), \end{aligned} \tag{54}$$

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

$$0 = 2N_r(I_r \otimes \beta'_1)d\text{vec}\beta_1 + 2N_r(I_r \otimes \beta'_2)d\text{vec}\beta_2 \tag{55}$$

and

$$\frac{\partial \text{vec}\beta_1}{\partial (\text{vec}\beta_2)'} = -(N_r(I_r \otimes \beta'_1))^{-1}(N_r(I_r \otimes \beta'_2)). \tag{56}$$

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

$$\begin{aligned} \frac{\text{vec}\beta_\perp}{\partial (\text{vec}\beta)'} &= \frac{\partial \text{vec}\beta_\perp}{\partial (\text{vec}\tilde{X})'} \frac{\partial \text{vec}\tilde{X}}{\partial (\text{vec}\beta)'} \\ &= - \left(I \otimes \begin{pmatrix} 0 \\ I_{p-r} \end{pmatrix} \right) ((I_p + U)' \otimes (I_p + \tilde{X})^{-1}) \times \left[- \left(I \otimes \begin{pmatrix} I_r \\ 0 \end{pmatrix} \right) ((I_p + \tilde{X})' \otimes (I_p + U)^{-1}) \right], \end{aligned} \tag{57}$$

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

$$\frac{\partial \text{vec}\Pi}{\partial (\text{vec}\lambda)'} = \alpha'_\perp \otimes \beta_\perp. \tag{58}$$

D 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.

	$k = 0$					$k = 1$							
	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.

	$k = 0$						$k = 1$						
	CSA						ISA						
	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.

	$k = 0$					$k = 1$						
	CSA					ISA						
	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	0.48	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	0.52	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	0.36
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
NI-OSG	0.46	0.43	0.43	0.4	0.4	0.36	0.42	0.48	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
MO-UPS	0.52	0.52	0.47	0.42	0.39	0.42	0.5	0.58	0.58	0.53	0.46	0.46
NLR	0.38	0.39	0.42	0.45	0.45	0.45	0.49	0.42	0.44	0.47	0.49	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
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
	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	0.4
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
DUK-OSG	0.4	0.44	0.44	0.38	0.42	0.36	0.4	0.49	0.52	0.51	0.45	0.36
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
NI-OSG	0.39	0.32	0.36	0.36	0.31	0.27	0.37	0.44	0.39	0.41	0.39	0.23
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
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
NLR	0.45	0.46	0.5	0.54	0.57	0.51	0.5	0.49	0.52	0.56	0.61	0.63
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
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

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.

	$k = 0$						$k = 1$					
	CSA						ISA					
	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	0.49	0.46	0.42	0.37
DUK-IBM	0.45	0.48	0.49	0.46	0.43	0	0.5	0.54	0.53	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	0.36
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
NI-OSG	0.41	0.39	0.39	0.38	0.32	0.33	0.44	0.45	0.4	0.4	0.39	0.36
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
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
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
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
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
	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
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
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
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
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
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
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
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
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
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