Bayesian Inference for Dynamic Cointegration Models with application to Soybean Crush Spread

Maciej Marówka*, Gareth W. Peters†, Nikolas Kantas*, Guillaume Bagnarosa‡

Abstract

In crush spread commodity trading strategies it is a common practice to select portfolio positions not based on statistical properties, but instead based on physical refinery conditions and efficiency in extracting byproducts from crushing raw soybeans to get soymeal and soyoil. The selected portfolio positions based on knowledge of refinery efficiency are then used to provide a basis for constructing the so called spread series, which is investigated separately using a model with a linear Gaussian structure. In this paper we take a statistical approach instead based on forming portfolio positions following from the cointegration vector relationships in the price series, which we argue endogenously take into consideration the respective demand and supply equilibrium dynamic associated to each component of the soybean complex spread. We propose an extension of the standard Cointegrated Vector Autoregressive Model that allows for a hidden linear trend under an error correction representation. The aim of this paper is to perform Bayesian estimation of the optimal cointegration vectors jointly with latent trends and to this end we develop an efficient Markov Chain Monte Carlo (MCMC) algorithm. The performance of this method is illustrated using numerical examples with simulated observations. Finally, we use the proposed model and MCMC sampler to perform analysis for soybean crush data. We will find the evidence in favour of the model structure proposed and present empirical justification that cointegration portfolio selection based on physical features of soybean market is sensitive to different roll adjustment methods used in the industry.

1 Introduction

Cointegration is a property of multivariate time series whereby a linear transformation of a non-stationary time series leads to a stationary and lower dimensional projected time series, which we will refer to in the context of this manuscript as the spread series. The projection undertaken to obtain the spread series is formed from an inner product operation of each vector time series point with the appropriate cointegration basis. In the context of multi-commodity futures trading strategies, one can also refer to this as the *cointegration portfolio*. The topic is very popular in econometrics and finance as it is expected that asset prices with a mean-reverting spread will have some degree of cointegration. Similarly equilibrium relationships are expected to exist between different categories of financial instruments such as spot and futures prices for commodities, equity prices and dividends as well as the components of the crush spread. It is often sensible to assume that these time series are tied together in the long-term by a common stochastic trend, because any new piece of information will affect the prices in similar ways, Alexander et al. [2002].

In this paper, we will focus on Cointegrated Vector Autoregression (CVAR) models. Since their introduction in the seminal paper of Engle and Granger [1987], CVAR models have been widely studied in the econometrics and time series literature; see Koop [2005] and Johansen [1995] for reviews on Bayesian and likelihood based approaches respectively. Here, we will propose a CVAR model with a latent stochastic trend. This will be then used to study the cointegration relationship of soybean crush products. Adding latent or hidden components to a CVAR model is commonly used to allow for time varying model coefficients (Koop and Korobilis [2013], Primiceri [2005] and Koop et al. [2008]) or time varying cointegration ranks Jochmann and Koop [2015]. Here we use dynamic linear factors motivated by earlier modeling of commodity futures contracts. Following the discussion in Peters et al. [2013], we modify the stochastic model of Schwartz and Smith [2000], to allow both the long and short factors to be mean reverting. In addition, using ideas from earlier works on seasonal structure of agricultural commodities in Simon [1999], Mitchell [2010], we incorporate an additive structural model to account for monthly varying means. Both cointegration and dynamic factors based representations are common in statistical analysis of financial spread

*Department of Mathematics, Imperial College London, UK.
†Department of Statistical Science, University College London, UK.
‡ESC Rennes School of Business, France.

Electronic copy available at: https://ssrn.com/abstract=2960638

portfolios, e.g. Alexander [1999]. However, to our best knowledge, they have not been investigated in a single framework so far. Therefore, in this paper, we propose a model where both approaches can be treated jointly.

We will perform Bayesian inference for the parameters of the proposed model. Typically in Bayesian inference for CVAR models the focus lies on specifying appropriate priors for the cointegration matrix and the adjustment coefficients (each to be defined precisely later in Section 2) as well as rank selection, e.g. Sugita [2002]. Posterior distributions can be very sensitive to even seemingly un-informative priors such as uniform distributions on bounded domains and in addition there could be problems related to model identifiability. To deal with this, early approaches include Jeffrey's priors Kleibergen and Van Dijk [1994] or nested prior embeddings for different ranks as in Kleibergen and Van Dijk [1998]; see Koop [2005] for a review. More recently, the focus has shifted from cointegration vectors to treating the cointegration space as the subject of the inferential procedure: Villani [2000], Villani et al. [2005], Strachan [2012] and Koop et al. [2009].

We will adopt this approach and combine appropriate priors for the cointegration space with the likelihood coming from our proposed CVAR model. For the posterior distribution of interest, we will develop an efficient Markov Chain Monte Carlo algorithm. The algorithm can be viewed as an extension to the Gibbs sampler of Koop et al. [2009] by adding partial collapsing and Rao-Blackwellization steps, see Casella and Robert [1996], Doucet et al. [2000], Van Dyk and Park [2008] for the similar constructs. Both the proposed model and sampler should be of interest to applications with cointegration. Our approach enables joint estimation of the cointegration parameters together with latent trends. This is not standard in the literature (Vidyamurthy [2004] and Krauss [2016]). In particular, for spread trading applications, a common practice is to first estimate the cointegration parameters and then the remaining parameters separately, see Barrett and Kolb [1995], Simon [1999], Mitchell [2010] for an empirical approach using the physical properties of the assets under consideration or Miao [2014] and Triantafyllopoulos and Montana [2011] for a regression based approach.

In the real data application of this paper, we particularly focus on the soybean crush cointegration. The crush spread is the difference between the combined value of soymeal and soyoil and the value of the original soybeans. The soybean futures crush spread stems from the process by which soybeans are crushed into soymeal and soyoil. It is also commonly used by the processors in order to hedge their exposure to the purchase price of soybeans and the value of by-products, namely the gross processing margin. Furthermore, speculators exploit unusual levels and the mean reversion nature of the futures crush spread by buying meal and oil futures and selling soybean futures when the crush spread is low, or selling meal and oil futures and buying soybean futures when the crush spread is high. This strategy is evaluated on the basis of physical estimation that one unit of soybeans produces certain amount of soybean meal and soyoil, and traders often use the appropriate weights to construct the cointegration portfolio, see Johnson et al. [1991].

In the real data experiments, we found the strong evidence in favour of the dynamic trend structure proposed in this paper. Another practical contribution of this paper in the context of commodity trading is to use the proposed CVAR model to estimate the cointegration vectors instead of using physical considerations. In fact, we will present evidence that a portfolio selection based on physical considerations is usually sensitive to different price adjustment methods and unlike our method does not typically result in a mean reverting spread; hence in a practical setup would require more frequent portfolio weights rebalancing. Furthermore, we will show that different roll-adjustment techniques exhibit different robustness to the cointegration space estimation methods. This would hence invalidate most of the previous studies which propose to apply the two stage procedure by first estimating cointegration vectors and then focusing on statistical properties of the resulting spread process dynamics. This two stage procedure used in practice is inefficient from a statistical estimation perspective as it does not remove accurately the non-stationarity in the presence of stochastic trend and furthermore does not perform join estimation of important aspects of the model. Such inefficiency in estimation and lack of statistical rigor may lead to sub-optimal trading strategy performance.

Given the generality of the CVAR modeling our approach can be generalized to other commodity spreads, such as crack or spark spreads. Furthermore, Peters et al. [2011, 2010] analyzed other applications of CVAR models to algorithmic trading.

The remainder of the paper is organized as follows. In Section 2 we propose our CVAR model with time varying linear trend and additive structural components to account for unrestricted and contemporaneous shifts in the long-run equilibrium of underlying assets. In Section 3, we specify appropriate priors to be used for Bayesian inference. Section 4 develops the MCMC algorithm for the resulting posterior distributions. In Section 5 we illustrate our method in simulated data examples and in Section 6 we perform data analysis for the soybean crush spread. In Section 7 we conclude with a discussion and ideas for future investigations.

Notation We denote a Gaussian random $(n \times T)$ matrix by $Y \sim N_{n,T}(\mu, \Sigma, \Psi)$ with row dependence in $(n \times n)$ covariance matrix Σ and column dependence in $(T \times T)$ matrix Ψ. Matrix variate Inverse Wishart distribution with

Electronic copy available at: https://ssrn.com/abstract=2960638

the support on symmetric positive definite matrices $(n \times n)$ with $\nu > n-1$ degrees of freedom and scale matrix Ψ is $\mathcal{W}^{-1}(\nu, \Psi)$. Gamma and Inverse Gamma distributions will be denoted as $Gamma(\alpha_\gamma, \beta_\gamma)$ and $IG(\alpha_\gamma, \beta_\gamma)$ respectively with scale and shape parameters α_γ and β_γ.$N(\mu, \Sigma)$ will be multivariate normal with mean vector $\mu \in \mathbb{R}^{n\times 1}$ and covariance matrix $\Sigma \in \mathbb{R}^{n\times n}$. By $I_n \in \mathbb{R}^{n\times n}$ we will refer to the square identity matrix; $Vec(M)$ denotes the matrix vectorization operator which transform a matrix M into column vector in which columns of M are successively stacked. Furthermore we denote the kronecker product or tensor product between two matrices by $\otimes$ and Kronecker sum as $\oplus$. The space spanned by the columns of any $m \times n$ matrix A is denoted as $sp(A)$; if A is of full column rank $n < m$, then $A_\perp$ denotes $m \times (m-n)$ matrix of full column rank satisfying $A_\perp^T A = 0$. For any square matrix, A, $||A||_F$ is a Frobenius norm, $||A||^2 = tr\{A^T A\}$, and $\rho(A)$ its spectral radius (that is, the maximal absolute value of the eigenvalues of A). Cardinality of a set B is denoted by $|B|$. Unless otherwise specified all densities are w.r.t the Lebesgue measure on the appropriate space and $d\mathcal{H}$ denotes Hausdorff measure on the appropriate manifold. In particular, the Haussdorf measure is related to the Lebesque measure according to the expression:

$$\mathcal{H}(dM) = \sqrt{G(M)}\mathcal{L}(M),$$

which is a Lebesque measure scaled by the volume element $\sqrt{G(M)}$, where G is a Riemanian metric on the manifold. $\mathbb{I}_A$ is a usual indicator function with $\mathbb{I}_A(x) = 1$ if $x \in A$ and 0 otherwise. Also, by $x(j)$ and $X(i,j)$ we mean $j-th$ and (i,j) entry of multivariate vector $x \in \mathbb{R}^{n\times n}$ and matrix $X \in \mathbb{R}^{n\times n}$ respectively. We say furthermore that time series y_t is integrated of order 1 and denote it as $I(1)$ if $\Delta y_t = y_t - y_{t-1}$ is a weakly stationary process, i.e. integrated of order 0, $I(0)$. For a time series sequence $y_1, y_2, ..., y_T$, we will also use the concise notation $y_{1:T}$.

2 A CVAR model with a stochastic trend

We will present a CVAR model with stochastic linear trend in which an unrestricted static intercept term varies accross different months in order to reflect inter-temporal seasonal patterns often reflected in commodity futures on agricultural products. For the purpose of this presentation, we do not include autoregressive lags, however we note that analogously to the approach of Villani et al. [2005], they can be trivially incorporated in the simulation methodology. We assume that the observable vector $y_t \in \mathbb{R}^{n\times 1}$ is integrated of order 1, $I(1)$, with r linear cointegration relationships; where r is assumed to be known. In principle, it can also be estimated via model selection approaches, see discussion in Peters et al. [2010]. The observation equation of dynamic Error Correction Model (ECM) representation is given by,

$$\Delta y_t = \alpha\beta^T y_{t-1} + \mu_t + g(t) + \epsilon_t \tag{1}$$

where $t = 1, ..., T$. The object of interest for cointegration analysis is the full rank cointegration matrix $\beta \in \mathbb{R}^{n\times r}$. The matrix $\alpha \in \mathbb{R}^{n\times r}$ is referred to as cointegration adjustment matrix that determines the speed of the mean reversion of the spread series. We treat the multivariate stochastic component $\beta^T\mu_t \in \mathbb{R}^{n\times 1}$ to be the mean level of the implied spread series and model it as a latent stochastic process according to a $K-$dimensional dynamic factor specification:

$$\mu_t = Hx_t \tag{2}$$

$$x_t = Bx_{t-1} + \delta_t, \quad x_0 = 0 \tag{3}$$

with $B \in \mathbb{R}^{K\times K}$ and $H \in \mathbb{R}^{n\times K}$. The error vectors at all times t, $\epsilon_t \in \mathbb{R}^{n\times 1}$ and $\delta_t \in \mathbb{R}^{K\times 1}$, are assumed independent and identically distributed (i.i.d.) as $\epsilon_t \sim N(0, R)$ and $\delta_t \sim N(0, Q)$ with symmetric positive definite covariance matrices $R \in \mathbb{R}^{n\times n}$ and $Q \in \mathbb{R}^{K\times K}$. K is a number of dynamic factors and in order to be consistent with long-short factor representation in commodities models, Peters et al. [2013], we will take $K = 2$ in the real data analysis. To ensure mean reversion in hidden factors, we must assume $|\rho(B)| < 1$, Vermaak et al. [2004]. The parametric seasonal model we use in (1) is given by:

$$g(t) = \xi_1 + \sum_{i=2}^{m} \xi_i \mathbb{I}_{\{t\in d_i\}} \tag{4}$$

where $\xi_i \in \mathbb{R}^{n\times 1}$ and d_i denotes the set of time indices corresponding to the i-th month.

Remark 1. Note that this formulation ensures that the nonstationarity in the level series can be explained solely by the random walk terms such as the common factor representation of a cointegrated system in Stock and Watson [1988].

We may define the so called *spread series* as $z_t = \beta^T y_t$. Ignoring the latent path and seasonal components, the dynamics of this process can be written as:

$$z_t = Az_{t-1} + \beta^T \epsilon_t \tag{5}$$

where $A = I + \beta^T \alpha$ and we will refer to it as the *autoregression matrix* of the implied spread process. Therefore, the spread process under the ECM representation (1) is a $r-$dimensional first order VAR process. Note that the necessary condition for the stability of the ECM is $|\rho(A)| < 1$.

The model in (1) can be re-expressed in a matrix regression format:

$$Y = \alpha\beta^T Y_0 + \mu + \xi\Upsilon + E, \tag{6}$$

where we have used the following notation $Y = (\Delta y_1, \Delta y_2, ..., \Delta y_T)$, $Y_0 = (y_0, y_1, ..., y_{T-1})$, $E = (\epsilon_1, ..., \epsilon_T)$, $\mu = (\mu_1, ..., \mu_T)$, $\xi = (\xi_1, \xi_2, ..., \xi_m)$, and

$$\Upsilon = \begin{bmatrix} 1 & & 1 \cdots & & & & & & \cdots 1 \\ 0 & \underbrace{1 \quad 1 \quad \cdots \quad 1}_{|d_2|} & & 0 \cdots 0 & & & & & \cdots 0 \\ 0 & & 0 \cdots & & \underbrace{1 \quad 1 \quad \cdots \quad 1}_{|d_3|} & \cdots & & & \cdots 0 \\ \vdots & & \vdots & & \vdots & \ddots & & \vdots & \\ 0 & & 0 \cdots & & \cdots & \cdots & \underbrace{1 \quad 1 \quad \cdots \quad 1}_{|d_m|} & & \end{bmatrix}.$$

We will use also $X = (x_1, \ldots, x_T)$ to denote the joint unobserved stochastic factor. Expressing the model in this form will often provide a convenient and concise representation that facilitates convenient manipulations of the model likelihood in Bayesian Markov chain Monte Carlo context.

Remark 2. In the application later in Section 6, we will be using daily market closing data and re-calibrate the model annually on the 1st of January each year with one year period of learning data. Then the design matrix Υ is given as above, with $m = 12$ and each row having only one consecutive block of 1's. This is consistent with a common practice in finance to re-fit the models on a rolling basis. Extensions to longer time periods or other than the 1st of January - 31 December can be easily accommodated in our model by trivial changes in Υ. Note also that, the subsequent methodology does not depend on the particular structure of Υ.

3 Bayesian Inference

In this section, we detail how the Bayesian inference can be used for estimation of parameters and latent factors. The model likelihood is expressed equivalently through the representation in (1) or alternatively the representation obtained in matrix variate form in (6). (3) provides a prior distribution for X given B, Q. To specify the posterior distribution of interest we need to choose suitable priors for the remaining variables, $\alpha, \beta, H, R, B, Q, \xi$. An interesting observation of the model given by (1)-(4) is that it results in a marginal likelihood that is invariant to certain linear transformations. For any orthogonal matrix $U \in \mathbb{R}^{K\times K}$ consider the transformation $T(H, B, Q) = (HU^{-1}, UBU^{-1}, UQU^T)$. It is straightforward to show that the marginal likelihood satisfies

$$p(Y|\alpha, \beta, H, B, Q, R, \xi) = p(Y|\alpha, \beta, T(H, B, Q), R, \xi).$$

One way to see this is use the transformed observations

$$\tilde{Y} = Y - \alpha\beta^T Y_0 - \xi\Upsilon \tag{7}$$

and perform the marginalization analytically using

$$p(Y|\alpha, \beta, H, R, B, Q, \xi) = \int p(\tilde{Y}|\alpha, \beta, H, R, \xi, X)p(X|B, Q)dX,$$

and noting that $d(G^{-1}x_t) = |G|dx_t = dx_t$. The invariance of this likelihood is a well known observation in the literature related to factor models, Aßmann et al. [2016], Bai and Wang [2015], Jackson et al. [2016].

Since the likelihood function is invariant under certain transformations, we will impose the same to hold for the prior on H, B, Q. This will mean that posterior inference will also possess this invariance property and will not depend on the particular parameterisation chosen. Selecting appropriate priors has been discussed in Blais [2009], where the author presents several ways of designing suitable priors for B, H and Q for Gaussian ECMs. Based on the findings of this author, in Table 1, we list the priors that we will use for B, Q, H, R, ξ. These are conjugate priors when combined with the likelihood in (6) which also preserve invariance with respect to the transformation T above. In addition, we are including hierarchy in the priors to avoid having to use a fixed and over sensitive parameter specification. We use hyper-parameters σ_H^2, σ_R^2 and σ_B^2, whose distributions are also summarized in Table 1.

Parameters	Prior
B	$N_{K,K}(0, \sigma_B^2 I_K, \sigma_B^2 I_K)$
Q	$\mathcal{W}^{-1}(\nu_Q, \sigma_Q^2 I_K)$
H	$N_{n,K}(0, \sigma_H^2 I_n, \sigma_H^2 I_K)$
R	$\mathcal{W}^{-1}(\nu_R, \sigma_R^2 I_n)$
ξ	$N_{n,m}(0, \sigma_\xi^2 I_n, \sigma_\xi^2 I_m)$

Hyper parameters	Prior
σ_H^2	$G(\alpha_H, \beta_H)$
σ_B^2	$IG(\alpha_B, \beta_B)$
σ_R^2	$IG(\alpha_R, \beta_R)$

Table 1: Prior distributions for Q, B, H, R, ξ (left) and distributions of hyper-parameters (right)

We proceed by specifying appropriate priors for α and β. Recent advances in the literature began to develop ways of eliciting priors over cointegration spaces (as opposed to directly defining a prior over the Euclidean parameter space defined by elements of matrix β) and deriving simulation methods for the corresponding posteriors, see discussion in Villani et al. [2005], Larsson and Villani [2001], Koop et al. [2008]. We will follow this approach and focus directly on $sp(\beta)$. An identification restriction which does not restrict the possible cointegration space is

$$\beta^T \beta = I_r,$$

where r is the cointegration rank. Formally, this restricts the matrix of cointegrating vectors to the Stiefel manifold $V_{n,r} := \{V \in \mathbb{R}^{n \times r} : \ V^T V = I_r\}$. These spaces are compact, and as such one can be certain that defining a uniform prior on them is well specified in the sense that it will form a class of proper priors. Similar to Koop et al. [2009], we will choose the prior distribution for $\beta \in V_{n,r}$ to be the matrix angular central Gaussian distribution with parameter P_τ [Gupta and Nagar, 1999], i.e.

$$dp(\beta) \propto |P_\tau|^{-r/2} |\beta^T (P_\tau)^{-1} \beta|^{-n/2} d\mathcal{H}(\beta) \tag{8}$$

where $P_\tau = HH^T + \tau H_\perp H_\perp^T$; $\tau \in [0, 1]$. P_τ determines the central location of $sp(\beta)$ and τ the amount of the dispersion around the central location. If $\tau = 1$, then $P_\tau = I_r$ and (8) defines a uniform prior on the manifold. In line with previous literature on CVAR and Bayesian regression models (e.g. Canova [2007], Villani et al. [2005]) we choose a shrinkage prior with zero prior mean for parameter α:

$$\alpha | \beta \sim N_{n \times r}(0, (v \beta^T P_{1/\tau} \beta)^{-1}, G), \tag{9}$$

where v controls the degree of shrinkage and can be made stochastic by defining hierarchical priors; the choice of G is flexible and gives $Vec(\alpha)|\beta \sim N(0, \Sigma_\alpha)$ with $\Sigma_\alpha = (v \beta^T P_{1/\tau} \beta)^{-1} \otimes G$.

Remark 3. The Gaussian Angular prior distribution enables us to choose both informative and non-informative priors. In practice, this is useful in crush spread strategy and more generally in the context of trading in commodities futures, as the cointegration portfolio allocation is usually driven by some physical considerations known a priori by practitioners; in context of crush spread it would be $H = (1, -11, -2.2)^T$, see discussion later in Section 6, and hence the practitioner can specify the prior centered around the line spanned by this vector and model directly the amount of dispersion around it. Note however, that in this paper, in order to compare the results obtained through the frequentist based approach, we will resort to the uniform prior on the relevant Stiefel manifold.

Note that in our case the semi-orthogonality restriction implies that the conditional posterior of β is non-standard. To mitigate this problem, the following transformation has been considered in Koop et al. [2009].

$$\beta \alpha^T = (\beta \kappa)(\alpha \kappa^{-1})^T = \mathcal{B} \mathcal{A}^T, \tag{10}$$

where

$$\kappa = (\alpha^T \alpha)^{1/2} = (\mathcal{B}^T \mathcal{B})^{1/2}, \tag{11}$$

$$\beta = \mathcal{B}(\mathcal{B}^T \mathcal{B})^{-1/2}. \tag{12}$$

and $\mathcal{A}$ is semi-orthogonal now, while $\mathcal{B}$ is unrestricted. Lemma 1 presented below provides the analytic expression for the priors on the transformed cointegration parameters $\mathcal{A}$ and $\mathcal{B}$ derived from the priors (8)-(9).

Lemma 1. *Given the hierarchical prior on (α, β) as in Eq. (8)-(9), the prior for $\mathcal{A}$ and $\mathcal{B}$ is given by:*

$$dp(\mathcal{A}) \propto |G|^{-r/2} |\mathcal{A}^T G^{-1} \mathcal{A}|^{-n/2} d\mathcal{H}(\mathcal{A}), \tag{13}$$

$$p(Vec(\mathcal{B})|\mathcal{A}) = N(0, \Sigma_{\mathcal{B}}) \tag{14}$$

where $\Sigma_{\mathcal{B}} = (\mathcal{A}^T G^{-1} \mathcal{A})^{-1} \otimes \nu P_\tau$.

The proof can be found in the technical appendix of Koop et al. [2009].

Remark 4. In (14) we present the prior for $Vec(\mathcal{B})$ instead of $\mathcal{B}$. To perform computations more conveniently when using Gaussian matrices we will often switch between a matrix parameter and its vectorized form and exploit the key property that for an arbitrary matrix Y if $Y \sim N_{n,T}(\mu, \Sigma, \Psi)$, then $Vec(Y) \sim N(Vec(M), \Sigma \otimes \Psi)$; Gupta and Nagar [1999, Theorem 2.2.1].

4 MCMC Sampler for Posterior Inference

Algorithm 1 A Partially Collapsed Gibbs Sampler to simulate from (15)

Initialization: Draw the parameters from their respective prior distributions
For $i = 1, 2, ..., N$: Sample

1. $Vec(B^{(i)}) \sim p_B(\cdot|Q^{(i-1)}, X^{(i-1)}, {\sigma_B^2}^{(i-1)})$,
2. ${\sigma_B^2}^{(i)} \sim p_{\sigma_B^2}(\cdot|B^{(i)})$,
3. $Q^{(i)} \sim p_Q(\cdot|B^{(i)}, X^{(i-1)})$
4. $Vec(H^{(i)}) \sim p_H(\cdot|Y, \alpha^{(i-1)}, \beta^{(i-1)}, R^{(i-1)}, X^{(i-1)}, \xi^{(i-1)}, {\sigma_H^2}^{(i-1)})$,
5. ${\sigma_H^2}^{(i)} \sim p_{\sigma_H^2}(\cdot|H^{(i)})$,
6. $R^{(i)} \sim p_R(\cdot|Y, \alpha^{(i-1)}, \beta^{(i-1)}, H^{(i)}, X^{(i-1)}, \xi^{(i-1)}, {\sigma_R^2}^{(i-1)})$,
7. ${\sigma_R^2}^{(i)} \sim p_{\sigma_R^2}(\cdot|R^{(i)})$
8. $Vec(\alpha^*) \sim p_\alpha(\cdot|Y, \beta^{(i-1)}, H^{(i)}, R^{(i)}, B^{(i)}, Q^{(i)}, \xi^{(i-1)})$ and compute $\mathcal{A}^* = \alpha^*(\alpha^{*T}\alpha^*)^{-1/2}$,
9. $Vec((\mathcal{B}^*)^T) \sim p_{\mathcal{B}}(\cdot|Y, \mathcal{A}^*, H^{(i)}, R^{(i)}, B^{(i)}, Q^{(i)}, \xi^{(i-1)})$. Compute $\beta^{(i)} = \mathcal{B}^*(\mathcal{B}^{*T}\mathcal{B}^*)^{-1/2}$ and $\alpha^{(i)} = \mathcal{A}^*(\mathcal{B}^{*T}\mathcal{B}^*)^{1/2}$
10. $Vec(\xi^{(i)}) \sim p_\xi(\cdot|Y, \alpha^{(i)}, \beta^{(i)}, H^{(i)}, R^{(i)}, B^{(i)}, Q^{(i)})$,
11. Sample $X^{(i)} \sim p(\cdot|Y, \alpha^{(i)}, \beta^{(i)}, \xi^{(i)}, H^{(i)}, R^{(i)}, B^{(i)}, Q^{(i)})$

We will use a MCMC sampling scheme for the posterior of model parameters. The posterior can be written as

$$\begin{aligned} p(\alpha, \beta, H, R, B, Q, \xi, X|Y) \propto\ & p(Y|\alpha, \beta, H, R, \xi, X) p(X|B, Q) \\ & \times p(\alpha|\beta) p(\beta) p(\xi) p(B) p(Q) p(H) p(R) \end{aligned} \tag{15}$$

For this posterior distribution, it is possible to derive a Gibbs Sampler. In numerical examples presented later in Section 5, we have noticed poor mixing and efficiency for a standard Gibbs Sampler. We will tackle this issue, by using a popular approach in Statistics (Van Dyk and Park [2008]) that alters certain steps in the Gibbs sampling

methodology. This is the so called partially collapsed Gibbs Sampler, which for this posterior could be performed by splitting the parameters in three blocks: (B, Q, H, R), (α, β, ξ), X and sampling iteratively as presented below. For the more detailed discussion on Gibbs sampling strategies in state space models see Carter and Kohn [1994].

$$\begin{aligned}(B, Q, H, R) &\sim p(\cdot|Y, (\alpha, \beta, \xi), X), &(16)\\ (\alpha, \beta, \xi) &\sim p(\cdot|Y, (\alpha, \beta, \xi)), &(17)\\ X &\sim p(\cdot|Y, (B, Q, H, R), (\alpha, \beta, \xi)). &(18)\end{aligned}$$

In general, care needs to be taken when constructing a partially collapsed Gibbs sampler, since the order of the Gibbs steps can significantly affect the stationary distribution, see Van Dyk and Park [2008] for details related to the validity of this scheme. For example in our case changing the order of (17) with (18) would result in a sampler that is not invariant to (15). The sampling step in (17) is invariant to $p(\alpha, \beta, H, R, B, Q, \xi|Y)$ with X marginalised out, so can be viewed as a Rao-Blackwellization step; see Casella and Robert [1996], Doucet et al. [2000], Peters et al. [2013] for a general background.

From the conditional distributions in (16), only (18) can be sampled explicitly using standard Kalman smoothing techniques, which for the convenience of the reader are included in the supplementary material in Appendix (B). Here, we will construct nested Gibbs updates that update each variables within the blocks (B, Q, H, R) and (α, β, ξ) so that these are left invariant to (16) and (17) respectively. In Algorithm 1 we present a partially collapsed Gibbs sampler that enables to simulate from (15). Note that conditional on X, the updates of (B, Q) are independent of conditional updates of (H, R). Steps 1-7 in Algorithm 1 simulate according to (16). In steps 8-10, we are sampling from $p(\alpha|Y, \beta, \xi)$, $p(\beta|Y, \alpha, \xi)$, $p(\xi|Y, (\alpha, \beta))$ respectively. This sequence of steps is invariant to (17), so can be viewed as Gibbs style implementation of (17) that is invariant to $p(\alpha, \beta, H, R, B, Q, \xi|Y)$ and hence valid to use in the sampler. In particular, steps 8 and 9 implement the $\kappa-$Gibbs Cointegration sampler of Koop et al. [2009]. The expressions of all distributions used in this algorithm are stated in the Proposition 1 and are derived in the Appendix A.1.

Proposition 1. *Under the model in* (1)-(4) *and for the prior choices of* (8),(9) *and Table 1, the conditional distributions used in Algorithm 1 are as follows:*

$$\begin{aligned}p_B(Vec(B)|Q, X) =& N(\mu^B_{post}, \Sigma^B_{post}),\\ p_Q(Q|B, X) =& \mathcal{W}^{-1}(\nu_Q + T, \tilde{X}\tilde{X}^T + \sigma^2_Q I_K),\\ p_H(Vec(H)|\alpha, \beta, Y, \xi) =& N(\mu^H_{post}, \Sigma^H_{post}),\\ p_R(R|Y, \alpha, \beta, \xi) =& \mathcal{W}^{-1}(\nu + T, \tilde{Y}\tilde{Y}^T + \sigma^R I_n),\\ p_\alpha(Vec(\alpha)|Y, \beta, H, R, B, Q, \xi) =& N(\mu^\alpha_{post}, \Sigma^\alpha_{post}), &(19)\\ p_{\mathcal{B}}(Vec(\mathcal{B}^T)|Y, \mathcal{A}, H, R, B, Q, \xi) =& N(\mu^{\mathcal{B}}_{post}, \Sigma^{\mathcal{B}}_{post}), &(20)\\ p_\xi(Vec(\xi)|Y, \alpha, \beta, H, R, B, Q) =& N(\mu^\xi_{post}, \Sigma^\xi_{post}),\\ p_{\sigma^2_R}(\sigma^2_R|R) =& G(\frac{n\nu_R}{2} + \alpha_{\sigma^2_R}, \beta_{\sigma^2_R} + \frac{1}{2}tr(R^{-1})),\\ p_{\sigma^2_H}(\sigma^2_H|H) =& IG(\alpha_{\sigma^2_H} + \frac{nK}{2}, \beta_{\sigma^2_H} + \frac{1}{2}Vec(H)^T Vec(H)),\\ p_{\sigma^2_B}(\sigma^2_B|B) =& IG(\alpha_{\sigma^2_B} + \frac{K^2}{2}, \beta_{\sigma^2_B} + \frac{1}{2}Vec(B)^T Vec(B)).\end{aligned}$$

The parameters of each distribution are given b

$$\begin{aligned}\left(\Sigma^B_{post}\right)^{-1} &= (X_0^T \otimes I_K)^T(\oplus_{t=1}^T Q^{-1})(X_0^T \otimes I_K)^T + (\sigma_B)^{-2} I_{K^2},\\ \mu^B_{post} &= \Sigma^B_{post}((X_0^T \otimes I_K)^T(\oplus_{t=1}^T Q^{-1})vec(X)),\\ \left(\Sigma^H_{post}\right)^{-1} &= (X^T \otimes I_K)^T(\oplus_{t=1}^T R^{-1})(X^T \otimes I_K)^T + (\sigma_H)^{-2} I_{K^2},\\ \mu^H_{post} &= \Sigma^H_{post}((X^T \otimes I_K)^T(\oplus_{t=1}^T R^{-1})Vec(\tilde{Y})),\\ \left(\Sigma^{\cdot}_{post}\right)^{-1} &= M_{\cdot}^T \widetilde{V}^{-1} M_{\cdot} + (\Sigma_{\cdot})^{-1},\\ \mu^{\cdot}_{post} &= \Sigma^{\cdot}_{post}(M_{\cdot}^T \widetilde{V}^{-1} \widetilde{y}),\end{aligned}$$

where we use $\cdot$ *for either* $\alpha, \mathcal{B}, \xi$ *and define* $\widetilde{V} = \oplus_{t=1}^T R + \widetilde{H} V \widetilde{H}^T$, $\widetilde{y} = Vec(Y) - (\Upsilon^T \otimes I_n)\xi$, $M_\alpha = Y_0^T \beta \otimes I_n$,

$M_{\mathcal{B}} = Y_0^T \otimes \alpha$, $\tilde{X} = (x_1 - Bx_0, \ldots, x_T - Bx_{T-1})$, $X_0 = (x_0, \ldots, x_{T-1})$, $\widetilde{H} = (I_T \otimes H)$, $V = (\widetilde{A}^T \oplus_{t=1}^{T} Q^{-1} \widetilde{A})^{-1}$ *and* $\widetilde{A} = I_{TK} - (I_T \otimes B)PM$, *with* $P = \begin{bmatrix} 0_{K\times K(T-1)} & I_{K\times K} \\ I_{K(T-1)\times K(T-1)} & 0_{K(T-1)\times K} \end{bmatrix}$ *and* $M = \begin{bmatrix} I_{K(T-1)\times K(T-1)} & 0_{K\times K(T-1)} \\ 0_{K(T-1)\times K} & 0_{K\times K} \end{bmatrix}$.

4.1 Point estimators for Cointegration Vectors

In this section we discuss the use of MCMC simulation in order to obtain point estimates for α, β. An MCMC generates samples for $\beta^{(i)}$ that lie on the Stiefel manifold parameterisation due to (8). We are essentially interested to provide a Bayesian point estimate of the posterior distribution for $sp(\beta)$. We will use the Posterior Mean Cointegration Estimator (PMCS) proposed in Villani [2006] (see Bernardo and Smith [2001] for background on Bayesian point estimation). The PMCS estimator is defined as

$$\hat{\beta} = \underset{\tilde{\beta} \in \mathbb{V}_{n,r}}{\arg\min}\ \mathbb{E}[d(\beta, \tilde{\beta})|Y].$$

It was proven that the PMCS estimator is $\hat{\beta} = (v_1, ..., v_r)$, where v_i are the eigenvectors of $\mathbb{E}(\beta\beta^T|Y)$ corresponding to the $i-th$ largest eigenvalue. Using the fact that the Stiefel manifold is a compact space, it can be shown that $\mathbb{E}(\beta\beta^T|Y)$ exists but does not admit a closed form expression. We can use MCMC and approximate it as $\frac{1}{N}\sum_{i=1}^{N} \beta^{(i)}\beta^{(i)T}$ using the output of Algorithm 1.

Furthermore, to quantify the accuracy of a point estimate we will use a loss function being the Frobenius projective distance measure between cointegrated spaces proposed in Villani [2006] and defined as:

$$d(\beta, \beta^*) = ||\beta\beta^T - \beta^*(\beta^*)^T||_F, \tag{21}$$

where we assume that both β and β^* are semi-orthogonal and $||\cdot||_F$ denotes standard matrix Frobenius norm. Note that, by orthogonality of these matrices, the metric defined in such a way is a Frobenius distance between projection matrices on their corresponding, uniquely determined column spaces $sp(\beta)$ and $sp(\beta^*)$.

In the subsequent numerical examples, we will investigate the effect of rank misspecification on the PMCS estimator. To do this we will employ a slightly modified version of the alternative distance measure introduced in Larsson and Villani [2001]. Let $\beta_1 \in \mathbb{R}^{p\times r_1}$ and $\beta_2 \in \mathbb{R}^{p\times r_2}$ be two arbitrary orthonormal matrices of full rank and $r_1 \geq r_2$. The distance between β_1 and β_2 which we utilize in one of the experiments is based on the following decomposition of β_2 is given by

$$\beta_2 = \beta_1\gamma_1 + \beta_{1\perp}\gamma_2,$$

where $\gamma_1 \in \mathbb{R}^{r_1\times r_2}$ and $\gamma_2 \in \mathbb{R}^{(p-r_1)\times r_2}$. These matrices can then be explicitly expressed as $\gamma_1 = \beta_1^T\beta_2$ and $\gamma_2 = \beta_{1\perp}^T\beta_2$. The alternative distance between $sp(\beta_1)$ and $sp(\beta_2)$ is then given by

$$l(\beta_1, \beta_2) = tr(\beta_2\beta_{1\perp}\beta_{1\perp}^T\beta_2)^{1/2}. \tag{22}$$

Unlike in Larsson and Villani [2001], we allow β_1 and β_2 to be of different dimensions. Hence, the expression in (22) does not define a metric for $r_1 \neq r_2$. It can be easily verified that if $\beta_2 \in sp(\beta_1)$, then $l(\beta_1, \beta_2) = 0$, whereas $l(\beta_2, \beta_1) = 1$. In fact, $l(\beta_1, \beta_2) = 0$ if and only if $\beta_2 \in sp(\beta_1)$. This means (22) can be conveniently used to check how far apart $\widehat{\beta}$ lies from $sp(\beta)$ (or similarly β from $\widehat{sp(\beta)}$) if a misspecified rank is smaller or larger than the true rank. If however dimensions of the respective matrices match, i.e. $r_1 = r_2$, then $l(\cdot, \cdot)$ defines a metric and possesses convenient property that $l(\beta_1, \beta_2) \leq min(r, (p-r))$.

5 Numerical examples using simulated data

In this section we present a detailed case study of the performance of the proposed Bayesian estimation procedure using simulated data sets from model equations (1)-(4). The numerical examples presented in this section aim to assess the following aspects: (a) the accuracy of estimation, (b) the improvement of Algorithm 17 over a standard Gibbs sampler and (c) the impact of the rank misspecification. In most experiments we ran Algorithm 1 for $N = 10^5$ and discarded 10^4 samples for burn-in. Furthermore, throughout this section we will use estimated posterior means as point estimates.

In the first experiment, we demonstrate the performance of Algorithm 1 for a simulated data-set using $n = 4$, $R = diag(1, 1, 1, 1)$ and $Q = diag(1, 1)$. Posterior samples and true values for μ are presented in Figure 1, posterior densities for $\Pi = \alpha\beta^T$ in Figures 2 and box plots for the seasonal components ξ in Figure 3. The estimation seems to be quite accurate. Furthermore, for a sample output $\Pi^{(i)} = \alpha^{(i)}\beta^{(i)T}$ we compare Algorithm 1 with a standard

0.40	0.63	0.44	0.31
0.74	0.70	0.48	0.53
0.62	0.51	0.49	0.39
0.40	0.46	0.37	0.35

0.34	0.31	0.24	0.17
0.33	0.14	0.22	0.15
0.20	0.09	0.10	0.10
0.19	0.08	0.09	0.10

Table 2: ESS for each entry of $\alpha\beta^T$ with Algorithm 1 (left table) and standard Gibbs sampler (right table).

$(r, \tilde{r})$	ESS	AD	PSV	MD	ED
$(1,1)$	0.96	0.072	0.003	0.06	0.18
$(1,2)$	0.88	0.82	0.04	0.03	0.33
$(1,3)$	0.95	0.71	0.10	0.006	0.34
$(2,1)$	0.97	0.094	0.02	0.03	0.44
$(2,2)$	0.83	0.093	0.001	0.03	0.49
$(2,3)$	0.96	0.57	0.08	0.007	0.48
$(3,1)$	0.98	0.098	0.19	0.008	1.52
$(3,2)$	0.99	0.20	0.02	0.008	0.49
$(3,3)$	0.98	0.042	$9.73e-5$	0.008	0.48

Table 3: Posterior cointegration space sampler diagnostics.

standard Gibbs sampler (without collapsing or Rao-Blackwellization) in terms of autocorrelation functions in Figure 4; Effective Sample Size (ESS) is presented in Table 2. Algorithm 1 shows a clearly superior performance.

We proceed to investigate the impact of rank misspecification on the cointegration space point estimation. Let r denote the true rank that generated the data and $\tilde{r}$ the one used to run the sampler. In Table 3, we present the results obtained from Algorithm 1 for all possible combinations for $(r, \tilde{r})$. For each pair $(r, \tilde{r})$, we report the following diagnostics:

- ESS for $\left(d(\beta^{(i)}, \beta)\right)_{i=1}^{N}$, where β is the true cointegration space.
- Estimated expected distance between sampled subspaces $AD = \frac{1}{N}\sum_{i=1}^{N} d(\beta^{(i)}, \beta^{(i-1)})$.
- The projective Frobenius span variation (FSV) is defined as $\tau_{sp}^2 = \frac{\mathbb{E}[d(\beta,\hat{\beta})|Y]}{r(p-r)/p}$ where $\hat{\beta}$ is the PMCS estimate of β. We evaluate this diagnostics using $\widehat{\tau_{sp}^2} = \frac{r-\sum_{i=1}^{r}\lambda_i}{r(p-r)/p}$ where λ_i is the estimate of the $i-th$ largest eigenvalue of $\mathbb{E}(\beta\beta^T)$.
- Estimates for $\mathbb{E}[l(\beta, sp(\hat{\beta}))\Big| Y]$ if $r \geq \tilde{r}$ and $\mathbb{E}[l(sp(\hat{\beta}), \beta)\Big| Y]$ if $r < \tilde{r}$, denoted as MD.
- The estimated Euclidean distance between the true known value of Π and its posterior mean estimate $\hat{\Pi}$, $ED = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{m}\left(\Pi_{i,j} - \hat{\Pi}_{i,j}\right)^2}$.

Table 3 clearly demonstrates that although the sampler efficiency and performance is good in all possible ranks configurations, AD and FSV are smallest when $r = \tilde{r}$. FSV seems to be a clear indicator whether the correct rank is used. The numerical results seem to indicate that MCMC mixing is higher when rank is over-specified. When the estimation of the space is concerned then the small values in MD column show that whenever we misspecify the true rank in the simulation procedure (to the lower/higher than its true value used for synthetic data generation), the estimated/true space is in all cases lying very close to its projection onto the true/estimated space - what is manifested by the fact that is close to be the subspace of the true/estimated higher dimensional space. Finally, the values of ED indicate that the different posterior means for the cointegration space are not very sensitive to the rank misspecification.

6 The Soy Bean Crush Spread

In this section we study an important real data example of cointegration in agricultural commodities trading pertaining to what is known as the *Soybean Crush*. We will present a detailed analysis on a data set obtained

from the constituents of the soy complex known as the soybean crush, comprising futures contracts on the soybean bushels, soymeal and soyoil. We investigate annual futures price series data of crush spread components, that is soybean, soyoil and soymeal over the period of fifteen years, 2001-2015. For each of the futures, we use suitable price process adjustments to tradeable assets based both on the nearest and most active contracts across the futures curve. For the comprehensive review of futures prices rolling techniques applied in the industry, see Carchano and Pardo [2009]

To understand the importance of the soybean market and the traded relationship known as the soybean crush, we first comment on the size and scope of the soybean commodity futures market. While soybeans are representing in volume the fourth more important crop production in the world. If you also include the trading of by-products, soybean turns out to be the most traded agricultural commodity and represents in value 10% of the worldwide agricultural trade [1]. From a supply perspective, about 80% of the soybean exportations are insured by the United States and Brazil. The demand dynamic on the other hand is mainly driven by the demand of by-products. About 85% of the soybeans are indeed consumed once processed and transformed into soymeal, soyoil and hull. The soymeal being very rich in protein is mainly used for animal feed (hog, livestock, fish...) and is thus directly related to the worldwide meat and fish consumption. Soyoil is used as edible oil for cooking and more recently for biodiesel. Another interesting feature of the soybean market is the fact that unlike other commodities, derivative components of soybeans are also bought and sold by hedgers and speculators on futures markets. While the prices of soybean, soymeal and soyoil are generally strongly correlated with each other, a move of the price of meal and oil against the other can be a result of a shock on the demand or supply of one of the aforementioned markets. In addition, since both soybean meal and soybean oil are a product of the same processing procedure, the total production of each will always remain constant relative to the other and therefore, ceteris paribus, a net increase in demand for one of the derivatives will necessarily correlate with an increase in the demand for the underlying beans.

The Chicago Board of Trade (CBOT) [2] recognizes the crush spread as an important aspect of the soybean market, Nelson Low [2015]. Loosely speaking, the crush spread is the difference between the combined value of the products corresponding to the soymeal and soyoil and the value of the unrefined raw soybeans. It is a measurement of the profit margin for the soybean processor. A soybean processor will be interested in the crush spread as part of a hedging strategy, and a speculator will look at the crush spread for trading opportunities. The crush spread is quoted as the difference between the combined sales value of soybean meal and oil and the price of soybeans. Based on the notation in (1) let at time t, $y_t(1)$, $y_t(2)$ and $y_t(3)$ correspond to the quoted prices of futures contract associated respectively to soybean, soyoil and soymeal. To compute the value of the crush spread, practitioners often use $y_t(1) - 11y_t(2) - 2.2y_t(3)$ (typically when a bushel of soybeans weighing 60 pounds is crushed, the typical result is 11 pounds of soybean oil, 44 pounds of 48% protein soybean meal, 4 pounds of hulls, and 1 pound of waste). Thus we will refer to the vector $\beta^C = (1, -11, -2.2)^T$ as the CBOT physical cointegration relation which we will benchmark against the estimates from the model in (1)-(4) and the data.

Most of the literature on soybean complex spread have the common feature that the trade timing and sizing of each position are heuristically specified or based on physical processes. However if this consideration is relaxed, one can consider what the optimal relationships to trade should be based on a statistical model. In general, it is often assumed that the portfolio weights in cointegation based trading is known apriori or estimated based on simple non-parametric techniques, see Krauss [2016], Miao [2014], Vidyamurthy [2004]. To the best of our knowledge, cointegration space estimation under a state space extension of cointegration financial models like (1) has not been attempted earlier, with notable exemptions Peters et al. [2011] and Marówka [2015]. In the next section, present an in-sample statistical analysis concentrated on the estimation of the cointegration relationship by fitting the model (1)-(4) for daily data for each year in the period 2001-2015.

6.1 Data Analysis of Soybean Crush

The data set is a panel of futures curves for soybean, soyoil and soymeal reported daily for fifteen years from 2001 until 2015. We perform estimation separately for each year in order to take into consideration the four seasons and the associated soybean cycle of production and processing. Here y_t in (1) is the adjusted price series obtained by applying two most common rollover strategies used in the industry: switching to the next month (front) contract on the day before maturity (which we will further refer to as *maturity* technique) and switching to the most active contract across the futures curve (*most active* technique). Details both strategies are provided in Lucia and Pardo [2010], Carchano and Pardo [2009]. For the priors we will use $v = 1, \tau = 1$ and $G = I$ implying uninformative

[1] https://www.ers.usda.gov/amber-waves/2016/may/major-factors-affecting-global-soybean-and-products-trade-projections/
[2] https://www.cmegroup.com/trading/agricultural/files/pm374-cbot-soybeans-vs-dce-soybean-meal-and-soybean-oil.pdf

uniform prior on the Stiefel manifold (and hence cointegration space) and regarding the remaining model parameters we select

$$\sigma_\xi^2 = 10^4; \nu_Q = 4; \nu_R = 5; \sigma_Q = 1; \alpha_H = \alpha_B = \alpha_R = 10^{-3}; \beta_H = \beta_B = \beta_R = 10^{-3}$$

We consider estimation of β and B and investigate how their estimates vary across the years. Note that B can have a significant effect on the estimation of β; see Appendix D for the relevant motivations. The results are presented in Tables 4 and 5. Therein, we also include comparisons with β^C as well as β^J and β^V which are respectively the estimators from the popular method by Johansen, see Johansen [1995] and Bayesian point (PMCS) estimator under vanilla model without the latent trend; we normalize β so that its first entry corresponding to the soybean price series is 1.

To assess whether two values of β are different we use a threshold for distance measures exceeding 14%; see Appendix C for details. Under the maturity adjustment technique, both the estimates of Johansen and of our method are relatively close to each other with regards to the metric l described in expression (22). The significant difference in estimated cointegration vectors occur only in 2004 and 2006. However in 2004, the estimated spectral radius of A in (5) is 1.02 what indicates non-stationarity and possibly an over-specification of the cointegration rank. This discrepancy is even more frequent when we compare the results obtained under our extended specification with the pure 'vanilla' Bayesian cointegration model without the stochastic trend. In this case, we observe significant distortion in estimation results in four years. The pointwise estimation of the cointegration space is thus relatively influenced here by the choice of the inference method. For the most active rolling strategy, the differences between Johansen and Bayesian estimates are even more considerable. In six and eight out of fifteen years of the analysis, the space discrepancy is above our chosen threshold of 14% for the comparison between our model and Johansen and 'vanilla' Bayesian estimators respectively. This suggest, that as far as the 'most active' roll adjustment technique is concerned, the point estimation of the cointegration space itself is not robust to the estimation technique applied. Consequently, we can observe that the modeling technique proposed in this paper is more useful for data obtained under 'most active' transformation. We find furthermore that the Bayesian PMCS estimator provides more stable estimates comparing to the Johansen method. This can be monitored using distance l between current and previous year cointegration vector estimates. The values of consecutive years l are reported in bold if they exceed the level of 14%. Our approach provides relatively stable results from 2001 until 2008 for front contract based adjustment (maturity) and after that time, the cointegration spaces seem to vary more. For the most active adjustment this stability in cointegration estimation is retained until 2010. Note also, that the estimated $\rho(A)$ being close to 1 introduces variation in points point estimates for β between years.

Columns 3 and 4 in Tables 4 and 5 show that the estimated posterior variance of $\rho(B)$ is much larger compared to that of $\rho(A)$, which is directly linked to the strength of cointegration relation. The resulting Bayesian posterior mean and variance of $\rho(B)$ suggest stationary latent paths μ_t. In our methodology we do not restrict B in any way to induce this. In addition, we observe that the cointegration space parameters can solely account for any non-stationarity present in the time series data under consideration.

In column 7 of Tables 4 and 5, we report the discrepancy between our estimated β and the CBOT quoted cointegration relation β^C. We found that this discrepancy is large in eight and five out of fifteen years under investigation for maturity and most active adjustments respectively. This motivates the use of model based statistical analysis as the one presented here. Moreover, in combination with Figure 5 where we provide the illustration of the dynamics of the spread series obtaied under the PMCS estimates $\hat{\beta}$ and CBOT relation β^C , we can infer that the portfolio selection based on naive physical considerations is usually sensitive to different price adjustment methods and unlike our method does not result in a typical for cointegration based trading mean reverting spread. In practical terms this would require more frequent position changing and consequently incuring higher trading costs. The results presented in Figure 5 pertains to one year, 2009, only. However, the similar pattern and conclusions can be drawn from observing the results from other years under investigation.

For the case of the most active adjustment price, in every year except 2011 our estimator for β has the same signs as β^C. For the maturity adjustment, this occurs in 2009 and 2011. This difference in the sign is also confirmed by the Johansen estimator. Paying attention to the signs of β is important from a financial perspective. Based on β^C only, a trader typically will take a long position in soybean and a short one for soyoil and soymeal, but this strategy could be refined using our results.

6.2 Dynamic Conditional Correlation Model

The model in (1)-(4) accounts for a mean model, but there could be some dynamic correlations present in the soybean crush. We proceed by investigating any implied volatility dynamics using the Dynamic Conditional Correlation

(DCC) methodology of Engle [2002]. Consider the estimated residual series from (1):

$$\hat{\epsilon}_t = \Delta y_t - \widehat{\alpha\beta^T} y_{t-1} - \widehat{\mu}_t - \widehat{\xi}_1 - \sum_{i=2}^{m} \widehat{\xi}_i \mathbb{I}_{\{t \in d_i\}},$$

where the estimators ($\widehat{\alpha\beta^T}$, $\widehat{\mu}_t$, $\widehat{\xi}$) are estimated posterior means using Algorithm (1).

We will specify a DCC model for fitting $\hat{\epsilon}_t$. Let $\epsilon_t \sim N(0, R_t)$, where R_t now is different than (1) and varies according to $R_t = \Psi_t \Omega_t \Psi_t$ with $\Psi_t = diag(\psi_t(1,1)^{1/2}, \psi_t(2,2)^{1/2}, ..., \psi_t(n,n)^{1/2})$ and each $\psi_t(i,i)$ evolves as a univariate variance from Gaussian GARCH(1,1) model of $\hat{\epsilon}_t(i)$. We will use

$$\Omega_t = diag(\omega_t(1,1)^{1/2}, \omega_t(2,2)^{1/2}, ..., \omega_t(n,n)^{1/2}) \Xi_t diag(\omega_t(1,1)^{1/2}, \omega_t(2,2)^{1/2}, ..., \omega_t(n,n)^{1/2})$$

where $\Xi_t = (\omega_t(i,j))_{\,i,j \in \{1,...n\}}$ and we set

$$\Xi_t = (1 - k_1 - k_2)\overline{\boldsymbol{\Xi}} + k_1 u_{t-1} u_{t-1}^T + k_2 \Xi_{t-1}$$

where $u_t(i) = \hat{\epsilon}_t(i)/\sqrt{\psi_t(i,i)}$, $\overline{\boldsymbol{\Xi}}$ is the empirical covariance matrix of u_t and k_1,k_2 and non-negative parameters satisfying $k_1 + k_2 < 1$. To fit this model for $\hat{\epsilon}_t$ we apply the pseudo-Maximum-Likelihood method of Orskaug [2009] using the R package of Nakatani [2010].

The results for the running volatilities and pairwise correlations of the marginal components of the soybean crush data for 2009 under most active roll-over strategy are presented in Figures 6 and 7 respectively. For brevity, we present the results only from this data-set, but the results are similar for other years and the maturity methods of price adjustments. In Figure 6, we observe that the volatility patterns of all three components are very similar and that the spikes in the soyoil and soymeal components occur at the same times regardless of the rank used to fit the model. Also, the pattern of volatility time series for the soybean component is almost identical in all possible ranks. All volatilities are inflated for a rank different than 1; which also corresponds to the lowest BIC after the DCC model fit. Finally, based on the results in Figure 7, the conditional correlations are almost constant for the misspecified rank cases and exhibit some seasonal pattern for rank equal 1.

7 Conclusions

We proposed an extension to a popular cointegration model to allow for joint Bayesian estimation of the cointegration parameters and stochastic trends. To perform Bayesian inference, we developed a novel MCMC algorithm that extends the Gibbs Sampler of Koop et al. [2009] and can be implemented in a computationally efficient manner. We used this sampler to perform data analysis on the soybean crush spread. We showed that our methodology is both accurate and robust to different roll adjustment methods used in the industry, which is not true for portfolio selection based on physical features of soybean market. In the final steps of our investigations we used a DCC approach and illustrated seasonality both in the running correlations and volatilities. This motivates further work towards stochastic volatility modeling and further extensions of the sampling methodology using particle MCMC methods Andrieu et al. [2010], Pitt et al. [2012].

References

Carol Alexander. Optimal hedging using cointegration. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 357(1758):2039–2058, 1999.

Carol Alexander, Ian Giblin, and Wayne Weddington. Cointegration and asset allocation: A new active hedge fund strategy. *ISMA Centre Discussion Papers in Finance Series*, page 40, 2002.

Christophe Andrieu, Arnaud Doucet, and Roman Holenstein. Particle markov chain monte carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(3):269–342, 2010.

Year	$\hat{\beta_\tau}$(Soybean, Soyoil, Soymeal)	$\tau_{\hat{s}p}$	$\hat{\rho}(A)$ $(Var(\rho(A)\|y))$	$\hat{\rho}(B)$ $(Var(\rho(B)\|y))$	$l(\hat{\beta_\tau}, \hat{\beta}_{\tau-1})$	$l(\beta^J_\tau, \beta^J_{\tau-1})$	$l(\hat{\beta}_\tau, \beta^C)$	$l(\hat{\beta}_\tau, \beta^J_\tau)$	$l(\hat{\beta_\tau}, \beta^V)$
2001	(1 −35.40 −0.96)	0.0002	0.91(0.01)	0.51(0.19)	-	-	**0.17**	0.001	0.018
2002	(1 −18.33 −2.20)	0.0017	0.92(0.02)	0.45(0.16)	0.09	0.07	0.09	0.02	0.062
2003	(1 −21.45 −1.31)	0.0001	0.86(0.02)	0.21(0.11)	0.06	0.03	**0.14**	0.006	0.026
2004	(1 35.40 −0.96)	0.24	1.02(0.01)	0.33(0.11)	**0.34**	**0.19**	**0.22**	**0.14**	**0.381**
2005	(1 −27.54 −1.98)	0.001	0.94(0.009)	0.26(0.10)	**0.33**	**0.19**	**0.14**	0.02	0.025
2006	(1 −17.28 −2.15)	0.004	0.94(0.012)	0.34(0.09)	0.06	**0.29**	0.07	**0.25**	**0.206**
2007	(1 −13.11 −1.95)	0.0002	0.87(0.02)	0.72(0.12)	0.02	**0.19**	0.05	0.027	0.028
2008	(1 −13.34 −1.90)	3.45e-05	0.74(0.04)	0.32(0.18)	0.02	0.006	0.06	0.006	0.012
2009	(1 279.56 −2.58)	0.06	0.98(0.01)	0.27(0.11)	**0.16**	**0.19**	**0.22**	0.02	0.052
2010	(1 −14.86 −1.86)	0.0006	0.92(0.01)	0.33(0.11)	**0.15**	**0.15**	0.07	0.01	0.071
2011	(1 323.12 −38.99)	0.017	0.96(0.01)	0.41(0.16)	**0.25**	**0.21**	**0.32**	0.02	**0.216**
2012	(1 −15.54 −2.70)	0.025	0.97(0.01)	0.34(0.14)	**0.29**	**0.36**	0.03	0.09	0.138
2013	(1 −68.13 −4.72)	0.18	0.99(0.01)	0.44(0.11)	**0.26**	0.11	**0.14**	0.05	0.086
2014	(1 −31.23 −1.6217)	0.08	0.99(0.02)	0.39(0.17)	0.02	0.06	**0.16**	0.02	0.019
2015	(1 −13.11 −1.95)	0.0002	0.87(0.02)	0.72(0.12)	0.1	0.1	0.05	0.06	**0.453**

Table 4: Cointegration Estimation results for the maturity adjusted price series. Here$\hat{\beta_\tau}$ is the PMCS estimator and τ is used to denote the year that provided the data for estimation, β^J_τ is the Johansen's estimator and β^V_τ is the PMCS estimator obtained for the vanilla Bayesian cointegration formulation without the stochastic trend. Values of l exceeding 14% are reported in the bold font.

Year	$\hat{\beta_\tau}$(Soybean, Soyoil, Soymeal)	$\tau_{\hat{s}p}$	$\hat{\rho}(A)$; $(Var(\rho(A)\|y))$	$\hat{\rho}(B)$ $(Var(\rho(B)\|y))$	$l(\hat{\beta_\tau}, \hat{\beta}_{\tau-1})$	$l(\beta^J_\tau, \beta^J_{\tau-1})$	$l(\hat{\beta}_\tau, \beta^C)$	$l(\hat{\beta}_\tau, \beta^J_\tau)$	$l(\hat{\beta_\tau}, \beta^V)$
2001	(1 −21.20 −1.41)	0.0001	0.86(0.01)	0.60(0.21)	-	-	0.13	0.01	0.026
2002	(1 −20.58 −3.23)	0.02	0.95(0.01)	0.41(0.19)	0.08	0.017	0.05	0.12	**0.154**
2003	(1 −54.53 −0.18)	0.005	0.93(0.03)	0.58(0.21)	**0.15**	0.001	**0.20**	0.04	0.048
2004	(1 −15.91 −2.21)	0.001	0.92(0.01)	0.35(0.15)	**0.14**	**0.926**	0.06	**0.96**	**0.299**
2005	(1 −17.52 −2.14)	0.001	0.93(0.01)	0.41(0.16)	0.01	**0.921**	0.08	0.10	0.0199
2006	(1 −28.58 −1.53)	0.008	0.95(0.009)	0.39(0.16)	0.07	**0.76**	**0.15**	**0.73**	**0.794**
2007	(1 −12.15 −2.00)	0.0002	0.82(0.03)	0.81(0.09)	0.11	**0.64**	0.03	0.02	0.033
2008	(1 −11.10 −2.14)	7.163362e-05	0.78(0.03)	0.42(0.14)	0.03	0.006	0.007	0.02	0.018
2009	(1 −12.92 −2.03)	0.0001	0.87(0.02)	0.43(0.14)	0.03	0.06	0.04	0.10	0.028
2010	(1 −11.92 −2.15)	0.0005	0.90(0.01)	0.55(0.20)	0.02	0.09	0.02	0.01	**0.156**
2011	(1 0.59 −3.29)	0.58	1.00(0.01)	0.57(0.16)	**0.99**	**0.80**	**0.99**	**0.62**	**0.991**
2012	(1 −10.04 −2.96)	0.005	0.97(0.007)	0.53(0.19)	**0.99**	**0.74**	0.08	0.01	**0.311**
2013	(1 −4.39 −1.80)	0.11	1.01(0.007)	0.51(0.17)	**0.15**	**0.29**	**0.22**	**0.44**	**0.215**
2014	(1 −15.88 −1.41)	0.29	0.99(0.03)	0.84(0.08)	**0.36**	**0.15**	**0.15**	**0.17**	0.033
2015	(1 −21.20 −2.69)	0.02	0.95(0.02)	0.52(0.22)	0.11	0.02	0.02	**0.14**	**0.155**

Table 5: Cointegration Estimation results for the most active adjusted price series. See caption of Table 4 for details.

Christian Aßmann, Jens Boysen-Hogrefe, and Markus Pape. Bayesian analysis of static and dynamic factor models: An ex-post approach towards the rotation problem. *Journal of Econometrics*, 2016.

Jushan Bai and Peng Wang. Identification and bayesian estimation of dynamic factor models. *Journal of Business & Economic Statistics*, 33(2):221–240, 2015.

W Brian Barrett and Robert W Kolb. Analysis of spreads in agricultural futures. *Journal of Futures Markets*, 15 (1):69–86, 1995.

José M Bernardo and Adrian FM Smith. *Bayesian theory*. Wiley Series in Probability and Statistics, 2001.

Sébastien Blais. Forecasting with weakly identified linear state-space models. 2009.

Fabio Canova. *Methods for applied macroeconomic research*, volume 13. Princeton University Press, 2007.

Óscar Carchano and Ángel Pardo. Rolling over stock index futures contracts. *Journal of Futures Markets*, 29(7): 684–694, 2009.

Chris K Carter and Robert Kohn. On gibbs sampling for state space models. *Biometrika*, 81(3):541–553, 1994.

George Casella and Christian P Robert. Rao-blackwellisation of sampling schemes. *Biometrika*, 83(1):81–94, 1996.

Arnaud Doucet, Nando De Freitas, Kevin Murphy, and Stuart Russell. Rao-blackwellised particle filtering for dynamic bayesian networks. In *Proceedings of the Sixteenth conference on Uncertainty in artificial intelligence*, pages 176–183. Morgan Kaufmann Publishers Inc., 2000.

Robert Engle. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 20(3):339–350, 2002.

Robert F Engle and Clive WJ Granger. Co-integration and error correction: representation, estimation, and testing. *Econometrica: journal of the Econometric Society*, pages 251–276, 1987.

Arjun K Gupta and Daya K Nagar. *Matrix variate distributions*, volume 104. CRC Press, 1999.

Laura E Jackson, M Ayhan Kose, Christopher Otrok, Michael T Owyang, Lucio Sarno, and Daniel L Thornton. Specification and estimation of bayesian dynamic factor models: A monte carlo analysis with an application to global house price comovement. *Dynamic Factor Models (Advances in Econometrics, Volume 35) Emerald Group Publishing Limited*, 35:361–400, 2016.

Jitesh Jain, Hong Li, Stephen Cauley, Cheng-Kok Koh, and Venkataramanan Balakrishnan. Numerically stable algorithms for inversion of block tridiagonal and banded matrices. 2007.

Markus Jochmann and Gary Koop. Regime-switching cointegration. *Studies in Nonlinear Dynamics & Econometrics*, 19(1):35–48, 2015.

Soren Johansen. *Likelihood-based inference in cointegrated vector autoregressive models*. Oxford university press, 1995.

Robert L. Johnson, Carl R. Zulauf, Scott H. Irwin, and Mary E. Gerlow. The soybean complex spread: An examination of market efficiency from the viewpoint of a production process. *Journal of Futures Markets*, 11(1):25–37, 1991. ISSN 1096-9934. doi: 10.1002/fut.3990110104. URL `http://dx.doi.org/10.1002/fut.3990110104`.

Frank Kleibergen and Herman K Van Dijk. On the shape of the likelihood/posterior in cointegration models. *Econometric Theory*, 10(3-4):514–551, 1994.

Frank Kleibergen and Herman K Van Dijk. Bayesian simultaneous equations analysis using reduced rank structures. *Econometric Theory*, 14(06):701–743, 1998.

Gary Koop. Bayesian approaches to cointegration. *PhD Thesis. Stockholm University*, 2005.

Gary Koop and Dimitris Korobilis. Large time-varying parameter vars. *Journal of Econometrics*, 177(2):185–198, 2013.

Gary Koop, Roberto León-González, and Rodney W Strachan. Efficient posterior simulation for cointegrated models with priors on the cointegration space. *Econometric Reviews*, 29(2):224–242, 2009.

GM Koop, Roberto Leon-Gonzalez, and Rodney W Strachan. Bayesian inference in the time varying cointegration model. 2008.

Christopher Krauss. Statistical arbitrage pairs trading strategies: Review and outlook. *Journal of Economic Surveys*, 2016.

Rolf Larsson and Mattias Villani. A distance measure between cointegration spaces. *Economics Letters*, 70(1): 21–27, 2001.

Julio J Lucia and Angel Pardo. On measuring speculative and hedging activities in futures markets from volume and open interest data. *Applied Economics*, 42(12):1549–1557, 2010.

Maciej Marówka. Cointegrated Vector Autoregression models with Nonlinear Dynamic Dependence Structure. Master's thesis, University College London, 2015. URL `http://media.wix.com/ugd/34a804_74a697c96a2c4c36bb126061480c9d93.pdf`.

George J Miao. High frequency and dynamic pairs trading based on statistical arbitrage using a two-stage correlation and cointegration approach. *International Journal of Economics and Finance*, 6(3):96, 2014.

John B Mitchell. Soybean futures crush spread arbitrage: Trading strategies and market efficiency. *Journal of Risk and Financial Management*, 3(1):63–96, 2010.

Tomoaki Nakatani. ccgarch: An R package for modelling multivariate garch models with conditional correlations. *R package version 0.2. 0*, 2010.

Tina Liao Nelson Low. CBOT Soybeans vs. DCE Soybean Meal and Soybean Oil Crush Spread. Technical report, Chicago Board of Trade, 2015. URL `https://www.cmegroup.com/trading/agricultural/files/pm374-cbot-soybeans-vs-dce-soybean-meal-and-soybean-oil.pdf`.

Elisabeth Orskaug. Multivariate DCC-GARCH model:-with various error distributions. *Master's thesis, Institutt for matematiske fag*, 2009.

Gareth W. Peters, Balakrishnan Kannan, Ben Lasscock, and Chris Mellen. Model selection and adaptive markov chain monte carlo for bayesian cointegrated {VAR} models. *Bayesian Analysis*, 5(3):465–491, 2010.

Gareth W. Peters, Balakrishnan Kannan, Ben Lasscock, Chris Mellen, and Simon Godsill. Bayesian cointegrated vector autoregression models incorporating alpha-stable noise for inter-day price movements via approximate bayesian computation. *Bayesian Analysis*, 6(4):755–792, 2011.

Gareth W. Peters, Mark Briers, Pavel Shevchenko, and Arnaud Doucet. Calibration and filtering for multi factor commodity models with seasonality: incorporating panel data from futures contracts. *Methodology and Computing in Applied Probability*, 15(4):841–874, 2013.

Michael K Pitt, Ralph dos Santos Silva, Paolo Giordani, and Robert Kohn. On some properties of markov chain monte carlo simulation methods based on the particle filter. *Journal of Econometrics*, 171(2):134–151, 2012.

Giorgio E Primiceri. Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies*, 72(3):821–852, 2005.

Herbert E Rauch, CT Striebel, and F Tung. Maximum likelihood estimates of linear dynamic systems. *AIAA journal*, 3(8):1445–1450, 1965.

Eduardo Schwartz and James E Smith. Short-term variations and long-term dynamics in commodity prices. *Management Science*, 46(7):893–911, 2000.

David P Simon. The soybean crush spread: Empirical evidence and trading strategies. *Journal of Futures Markets*, 19(3):271–289, 1999.

James H Stock and Mark W Watson. Testing for common trends. *Journal of the American statistical Association*, 83(404):1097–1107, 1988.

Rodney W Strachan. Valid bayesian estimation of the cointegrating error correction model. *Journal of Business & Economic Statistics*, 2012.

Katsuhiro Sugita. Testing for cointegration rank using bayes factors. *Warwick Economic Research Papers*, 2002.

Kostas Triantafyllopoulos and Giovanni Montana. Dynamic modeling of mean-reverting spreads for statistical arbitrage. *Computational Management Science*, 8(1-2):23–49, 2011.

David A Van Dyk and Taeyoung Park. Partially collapsed gibbs samplers: Theory and methods. *Journal of the American Statistical Association*, 103(482):790–796, 2008.

Jaco Vermaak, Christophe Andrieu, Arnaud Doucet, and SJ Godsill. Reversible jump markov chain monte carlo strategies for bayesian model selection in autoregressive processes. *Journal of Time Series Analysis*, 25(6):785–809, 2004.

Ganapathy Vidyamurthy. *Pairs Trading: quantitative methods and analysis*, volume 217. John Wiley & Sons, 2004.

Mattias Villani. *Aspects of Bayesian Cointegration*. PhD thesis, 2000.

Mattias Villani. Bayesian point estimation of the cointegration space. *Journal of Econometrics*, 134(2):645–664, 2006.

Mattias Villani et al. Bayesian reference analysis of cointegration. *Econometric Theory*, 21(2):326–357, 2005.

A Appendix to Section 4

Proposition 2. *For $X = (x_1, \ldots, x_T)$ given by (3), we have that $Vec(X) \sim N(0, V)$, with $V = (\widetilde{A}^T \oplus_{t=1}^{T} Q^{-1} \widetilde{A})^{-1}$ and $\widetilde{A} = I_{TK} - (I_T \otimes B)PM$, with $P = \begin{bmatrix} 0_{K\times K(T-1)} & I_K \\ I_{K(T-1)} & 0_{K(T-1)\times K} \end{bmatrix}$ and $M = \begin{bmatrix} I_{K(T-1)} & 0_{K\times K(T-1)} \\ 0_{K(T-1)\times K} & 0_{K\times K} \end{bmatrix}$.*

Proof. We can rewrite multivariate time series model 3 in the matrix regression format by stacking vectors of the series realizations x_t in the columns of matrix X,

$$X = BX_0 + \delta_{1:T},$$

where $X_0 = (x_0, \ldots, x_{T-1})$. After applying the vectorization we get

$$Vec(X) = (I_T \otimes B)Vec(X_0) + W$$

where $W \sim N(0, \oplus_{i=1}^{T} Q)$. Recall the permutation matrix P is invertible and hence we can write

$$\begin{aligned} Vec(X) &= (I_T \otimes B)PP^{-1}Vec(X_0) + W \\ Vec(X) &= \widetilde{B}Vec(X_0') + W \\ Vec(X) &= \widetilde{C}Vec(X) + W \\ \widetilde{A}Vec(X) &= W \end{aligned} \tag{23}$$

where we let $\widetilde{B} = (I_T \otimes B)P$, $X_0' = (x_1, \ldots, x_{T-1}, x_0)$ $\widetilde{C} = \widetilde{B} \begin{bmatrix} I_{K(T-1)} & 0_{K(T-1)\times K} \\ 0_{K\times K(T-1)} & 0_{K\times K} \end{bmatrix}$ and $\widetilde{A} = I_{TK} - \widetilde{C}$. Recall in (1)-(4) X is zero so that we can multiply matrix $\widetilde{B}$ by identity matrix with last K diagonal entries equal to 0. Finally, due to (23) we get

$$\begin{aligned} p(Vec(X)) &\propto \exp(-\frac{1}{2}(\widetilde{A}Vec(X))^T \oplus_{t=2}^{T} Q^{-1}((\widetilde{A}Vec(X)) \\ &\propto \exp(-\frac{1}{2}Vec(X)^T(\widetilde{A}^T \oplus_{t=1}^{T} Q^{-1}\widetilde{A})Vec(X)). \end{aligned}$$

and so $Vec(X) \sim N(0, V)$, where $V = (\widetilde{A}^T \oplus_{t=1}^{T} Q^{-1}\widetilde{A})^{-1}$. □

Note that we can use Proposition 2 to sample cointegration parameters unconditionally on hidden states X, (19), (20). This significantly reduces dimensionality in Gibbs updates.

A.1 Proof of Proposition 1

We will make extensive use of the lemma below, which allows for analytic representation of the full and marginalized conditionals in Gaussian matrix variate likelihood model.

Lemma 2. *Consider a Gaussian vector regression model with output $\mathcal{Y} \in \mathbb{R}^{n\times 1}$ and input $\mathcal{X} \in \mathbb{R}^{n\times n}$ where*

$$\mathcal{Y} = \mathcal{X}\theta + \epsilon$$

and $\theta \in \mathbb{R}^{n\times 1}$ and $\epsilon \in \mathbb{R}^{n\times 1}$ is an innovation error vector such that $\epsilon \sim N(0, \Sigma_{\mathcal{Y}})$. Suppose that a conjugate prior of θ is a Gaussian vector $p(\theta) \sim N(m_\theta, \Sigma_\theta)$. Then $p(\theta|\mathcal{Y}) \sim N(\widetilde{m_\theta}, \widetilde{\Sigma_\theta})$, where

$$\widetilde{\Sigma_\theta} = (\mathcal{X}^T\Sigma_{\mathcal{Y}}^{-1}\mathcal{X} + \Sigma_\theta^{-1})^{-1},$$
$$\widetilde{m_\theta} = M_\theta((15)\mathcal{X}^T\Sigma_{\mathcal{Y}}^{-1}\mathcal{Y} + \Sigma_\theta^{-1}m_\theta).$$

The proof is omitted as it follows by straightforward linear algebra.

Lemma 3. *Suppose $\boldsymbol{\mathcal{Y}} := [\mathcal{Y}_1, \mathcal{Y}_2, ..., \mathcal{Y}_n]$ is a matrix whose columns are i.i.d samples from $N(0, \Sigma)$. If $\Sigma \sim \mathcal{W}^{-1}(\nu, \Psi)$, then $p(\Sigma|\boldsymbol{\mathcal{Y}}) \sim \mathcal{W}^{-1}(\nu + n, \Psi + \boldsymbol{\mathcal{Y}}\boldsymbol{\mathcal{Y}}^T)$.*

See Gupta and Nagar [1999] for a proof.
We proceed with the proof of Proposition 1.

Proof. (of Proposition 1).
$p_B(Vec(B)|Q, x_{1:T})$: As in the proof of Proposition 2 above,

$$x_t = Bx_{t-1} + \delta_t \ \forall_{t=1,...,T}$$

can be rewritten in matrix regression format as:

$$X = BX_0 + W$$

with $W = (\delta_1, \delta_2, \ldots, \delta_T)$. Applying matrix vectorization operator:

$$Vec(X) = (X_0^T \otimes I_K)Vec(B) + w, \tag{24}$$

where $w \sim N(0, \oplus_{t=1}^T Q)$. Hence by multivariate normal conjugacy implied by gaussian likelihood in (24) and the results of Lemma 2, the conditional posterior of $Vec(B)$ is also multivariate normal

$$Vec(B)|Q, X \sim N(\mu_{post}^B, \Sigma_{post}^B), \tag{25}$$

where

$$\Sigma_{post}^B = ((X_0^T \otimes I_K)^T(\oplus_{t=1}^T Q^{-1})(X_0^T \otimes I_K)^T + (\sigma_B)^{-2}I_{K^2})^{-1}$$
$$\mu_{post}^B = \Sigma_{post}^B((X_0^T \otimes I_K)^T(\oplus_{t=1}^T Q^{-1})vec(X)).$$

We proceed with $p_Q(Q|B, x_{1:T})$. We can simulate from full conditional update by exploiting conjugacy class relations of normal and inverse Wishart distributions as well as conditional independence of Q and observed time series realizations given latent path. The hidden model likelihood can be written as

$$\widetilde{x}_t = Q^{1/2}r_t, \quad \widetilde{x}_t = x_t - Bx_{t-1},\ x_0 = 0,$$

with r_t i.i.d. multivariate standard random normal. Since the prior for Q is $\mathcal{W}^{-1}(\nu_Q, \sigma_Q^2 I_K)$, then by the Lemma 3 we have

$$Q|B, X \sim \mathcal{W}^{-1}(\nu_Q + T, \widetilde{X}\widetilde{X}^T + \sigma_Q^2 I_K), \tag{26}$$

with $\widetilde{X} = X - BX_0$.

For computing $p_H(Vec(H)|\alpha, \beta, Y, x_{1:T}, \xi)$: Note that, conditioning on the realization of latent path, observation model parameters are independent of hidden static parameters. Recall, $\tilde{Y} = Y - \alpha\beta^T Y_0 - \xi\Upsilon$ and noting vectorized regression format as

$$Vec(\widetilde{Y}) = (X^T \otimes I_K)vec(H) + \widetilde{E},$$

where $\widetilde{E} \sim N(0, \oplus_{t=1}^{T} R)$, we can exploiting the same conjugacy properties and Lemma 2 and 3 to obtain

$$Vec(H)|Y, \alpha, \beta, X, \xi \sim N(\mu_{post}^{H}, \Sigma_{post}^{H}),$$

where

$$\Sigma_{post}^{H} = ((X^T \otimes I_K)^T (\oplus_{t=1}^{T} R^{-1})(X^T \otimes I_K)^T + (\sigma_H)^{-2} I_{K^2})^{-1},$$

$$\mu_{post}^{H} = \Sigma_{post}^{B}((X^T \otimes I_K)^T (\oplus_{t=1}^{T} R^{-1}) vec(\widetilde{Y})).$$

$p_R(R|Y, \alpha, \beta, x_{1:T}, \xi)$: Conditional on latent path and all remaining parameters, $\widetilde{Y} - HX$ can be regarded as T vector variate normal samples with zero mean and variance R. Hence the full conditional of R follows directly from Lemma 3:

$$R|\alpha, \beta, Y, X, \xi \sim \mathcal{W}^{-1}(\nu + T, (\widetilde{Y} - HX)(\widetilde{Y} - HX)^T + \sigma_R^2 I_n). \tag{27}$$

$p_\alpha(Vec(\alpha)|Y, \beta, H, R, B, Q, \xi)$, $p_{\mathcal{B}}(Vec(\mathcal{B}^T)|Y, \mathcal{A}, H, R, B, Q, \xi)$: To obtain coitegration parameters' marginalized conditionals of partially collapsed Gibbs algorithm, we utilize convolution property of normal distribution, Proposition 2 and Lemma 1 together with the following representation:

$$\begin{aligned}
Vec(Y) &= (Y_0^T \beta \otimes I_n) Vec(\alpha) + Vec(\mu) + (\Upsilon^T \otimes I_n) Vec(\xi) + \widetilde{E} \\
&= (Y_0^T \otimes \mathcal{A}) Vec(\mathcal{B}^T) + Vec(\mu) + (\Upsilon^T \otimes I_n) Vec(\xi) + \widetilde{E} \\
&= (Y_0^T \beta \otimes I_n) Vec(\alpha) + (\Upsilon^T \otimes I_n) Vec(\xi) + \widetilde{\boldsymbol{E}} \\
&= (Y_0^T \otimes \mathcal{A}) Vec(\mathcal{B}^T) + (\Upsilon^T \otimes I_n) Vec(\xi) + \widetilde{\boldsymbol{E}},
\end{aligned} \tag{28}$$

where $\widetilde{\boldsymbol{E}} \sim N(0, \widetilde{V})$; $\widetilde{V} = \oplus_{t=1}^{T} R + \widetilde{H} V \widetilde{H}^T$ with $\widetilde{H}$ and V defined in proof of Proposition 2 above. If we define further:

$$\begin{aligned}
\widetilde{y} &= Vec(Y) - (\Upsilon^T \otimes I_n)\xi; \\
M_\alpha &= (Y_0^T \beta \otimes I_n); \\
M_{\mathcal{B}} &= (Y_0^T \otimes \alpha).
\end{aligned}$$

Then, it follows from Lemma 2 that collapsed conditionals (with marginalized latent paths) for vectorized components of cointegration model, $Vec(\alpha)$ and $Vec(\mathcal{B}^T)$ are distributed as vector variate normals, p_α, p_β respectively, with variances and means as provided below:

$$p_\alpha \sim N(\mu_{post}^{\alpha}, \Sigma_{post}^{\alpha}), \tag{29}$$

$$p_{\mathcal{B}} \sim N(\mu_{post}^{\mathcal{B}}, \Sigma_{post}^{\mathcal{B}}). \tag{30}$$

$$\begin{aligned}
\Sigma_{post}^{\alpha} &= (M_\alpha^T \widetilde{V}^{-1} M_\alpha + (\Sigma_\alpha)^{-1})^{-1}, \\
\mu_{post}^{\alpha} &= \Sigma_{post}^{\alpha}(M_\alpha^T \widetilde{V}^{-1} \widetilde{y}).
\end{aligned} \tag{31}$$

$$\begin{aligned}
\Sigma_{post}^{\mathcal{B}} &= (M_{\mathcal{B}}^T \widetilde{V}^{-1} M_{\mathcal{B}} + (\Sigma_{\mathcal{B}})^{-1})^{-1}, \\
\mu_{post}^{\mathcal{B}} &= \Sigma_{post}^{\mathcal{B}}(M_{\mathcal{B}}^T \widetilde{V}^{-1} \widetilde{y}).
\end{aligned} \tag{32}$$

Using (28) again, we can define

$$\begin{aligned}
\mathcal{Y} &= Vec(Y) - (Y_0^T \otimes \mathcal{A}) Vec(\mathcal{B}^T), \\
M_\xi &= (\Upsilon^T \otimes I_n);
\end{aligned}$$

Then analogously as in (31), (32) and by Lemma 2,

$$p_\xi(Vec(\xi)|Y, \alpha, \beta, H, R, B, Q) = N(\mu_{post}^{\xi}, \Sigma_{post}^{\xi})$$

where

$$\begin{aligned}\Sigma^{\xi}_{post} &= (M_{\xi}^T \widetilde{V}^{-1} M_{\xi} + (\Sigma_{\xi})^{-1})^{-1} \\ \mu^{\xi}_{post} &= \Sigma^{\xi}_{post}(M_{\xi}^T \widetilde{V}^{-1} \tilde{y}),\end{aligned}$$

and $\Sigma_{\xi} = \sigma_{\xi}^2 I_{mn}$.

$p_{\sigma_R^2}(\sigma_R^2|R)$: Assuming Inverse Wishart prior of $R \sim \mathcal{W}^{-1}(\nu_R, \sigma_R^2 I_n)$ with hyperparameter σ_R^2 following gamma distribution $\sigma_R^2 \sim Gamma(\alpha_R, \beta_R)$, then by straightforward algebra:

$$\begin{aligned}p(\sigma_R^2|R, y_{1:T}) &= p(\sigma_R^2|R) \\ &\propto (\sigma_R^2)^{\frac{\nu_R n}{2}} \exp\left(-\frac{1}{2}\sigma_R^2 tr(R^{-1})\right) \cdot (\sigma_R^2)^{\alpha_R - 1} \exp(-\beta_R \sigma_R^2) \\ &= (\sigma_R^2)^{\frac{\nu_R n}{2} + \alpha_R - 1} \exp\left(-\sigma_R^2\left(\beta_R + \frac{1}{2}tr(R^{-1})\right)\right)\end{aligned}$$

so it follows clearly that $p_{\sigma_R^2}(\sigma_R^2|R)$ is a $Gamma\left(\frac{n\nu_R}{2} + \alpha_{\sigma_R^2}, \beta_{\sigma_R^2} + \frac{1}{2}tr(R^{-1})\right)$.

$p_{\sigma_H^2}(\sigma_H^2|H)$: Assuming multivariate normal prior of $Vec(H) \sim N(0, \sigma_H^2 I_{nK})$ with hyperparameter σ_H^2 following inverse gamma distribution $\sigma_H^2 \sim IG(\alpha_H, \beta_H)$, it follows that

$$\begin{aligned}p(\sigma_H^2|H, y_{1:T}) &= p(\sigma_H^2|H) \\ &\propto (\sigma_H^2)^{-nK/2} \exp\left(-\frac{1}{2\sigma_H^2} Vec(H)^T Vec(H)\right) \cdot \left(\sigma_H^2\right)^{-\alpha_H - 1} \exp(-\frac{\beta_H}{\sigma_H^2}),\end{aligned}$$

so $p_{\sigma_H^2}(\sigma_H^2|H)$ is $IG\left(\alpha_{\sigma_H^2} + \frac{nK}{2}, \beta_{\sigma_H^2} + \frac{1}{2}Vec(H)^T Vec(H)\right)$

$p_{\sigma_B^2}(\sigma_B^2|B)$: By analogous argument as above we have $p_{\sigma_B^2}(\sigma_B^2|B) = IG(\alpha_{\sigma_B^2} + \frac{K^2}{2}, \beta_{\sigma_B^2} + \frac{1}{2}Vec(B)^T Vec(B))$. □

B Supplementary material for implementing Algorithm 1Sampling from $p(X|Y, B, H, Q, R, \alpha, \beta)$

Sampling from $p(x_{1:T}|y_{1:T}, B, H, Q, R, \alpha, \beta)$ can be cast as a smoothing problem for a linear Gaussian state space model of the form:

$$\begin{aligned}x_t &= Bx_{t-1} + \delta_t \\ \tilde{y}_t &= Hx_t + \epsilon_t\end{aligned}$$

A standard approach is to use the Rauch–Tung–Striebel smoother Rauch et al. [1965]. This forward backward scheme for each t computes $p(x_t|y_{1:t}, B, H, Q, R, \alpha, \beta)$, which is a Gaussian with mean $x_{t|t}$ and covariance$P_{t|t}$. The steps are presented in Algorithm 2.

B.1 Computational considerations for fast matrix inversions

The most computationally intensive procedure of the algorithm proposed is the inversion of the matrix $\oplus_{t=1}^T R + \widetilde{H} V \widetilde{H}^T$ (used in (31) -(32)) where V is defined in Proposition 2 and $\widetilde{H} := (I \otimes H)$. A naive implementation would result in $O(T^3)$ cost, but we can perform the inversion with $O(T^2)$ cost by exploring the particular structure of the matrix. To perform this operation efficiently on can denote $\widetilde{R} := \oplus_t^T R$ and apply Woodbury matrix identity to obtain

$$(\widetilde{R} + \widetilde{H} V \widetilde{H}^T)^{-1} = \widetilde{R}^{-1} - \widetilde{R}^{-1}\widetilde{H}(V^{-1} + \widetilde{H}^T \widetilde{R}^{-1} \widetilde{H})^{-1} \widetilde{H}^T \widetilde{R}^{-1}.$$

Both $\widetilde{R}^{-1}$and $\widetilde{H}$ are by construction block diagonal and $(\widetilde{R}^{-1}\widetilde{H})(\widetilde{H}(V^{-1} + \widetilde{H}^T \widetilde{R}^{-1}\widetilde{H})^{-1})$ can be written as a product of a block diagonal general $Tn \times Tn$ matrix, which can be computed in $O(T^2)$ operations. Recall in Proposition 2, $V^{-1} = (\widetilde{A}^T \oplus_{t=1}^T Q^{-1} \widetilde{A})$. Straightforward manipulations show that

$$V^{-1} = \widetilde{Q}^{-1} - \widetilde{Q}^{-1}\widetilde{C} - \widetilde{C}^T \widetilde{Q}^{-1} + \widetilde{C}^T \widetilde{Q}^{-1} \widetilde{C}.$$

Algorithm 2 Forward Filter Backwards Smoother for obtaining a sample $x'_{1:T}$ from the joint smoothing density $p(x_{1:T}|B, H, Q, R, y_{1:T})$

Initialization $\boldsymbol{x}_{0|0} = 0$, $\Sigma^{\boldsymbol{x}}_{0|0} = \mathbf{0}$

1. ***(Forward Filter)*** **For** $t \in \{1, ..., T\}$ **compute**:
   (a) $\boldsymbol{x}_{t|t-1} = B\boldsymbol{x}_{t|(t-1)}$
   (b) $\Sigma^{\boldsymbol{x}}_{t|t-1} = B\Sigma^{\boldsymbol{x}}_{t-1|t-1}B^T + Q$
   (c) $y_{t|t-1} = H\boldsymbol{x}_{t|t-1}$
   (d) $\Sigma^{y}_{t|t-1} = H\Sigma^{\boldsymbol{x}}_{t|t-1}H^T + R$
   (e) $K_t = \Sigma^{\boldsymbol{x}}_{t|t-1}H^T(\Sigma^{y}_{t|t-1})^{-1}$
   (f) $\boldsymbol{x}_{t|t} = \boldsymbol{x}_{t|t-1} + K_t\left(\tilde{y}_t - y_{t|t-1}\right)$
   (g) $\Sigma^{\boldsymbol{x}}_{t|t} = \Sigma^{\boldsymbol{x}}_{t|t-1} - K_t\Sigma^{y}_{t|t-1}K_t^T$.
2. **For** $t = T$ **simulate** x_T from the filtering distribution $p(x_T|y_{1:T}) = N(x_{T|T}, \Sigma_{T|T})$.
3. **(Backward smoother) For** $t \in \{T-1, ..., 1\}$ **do:**
   (a) Step 1: Compute smoothing mean $m_{t|T}$ and covariance $P_{t|T}$

$$P_{t|T} = (B^TQ^{-1}B + {\Sigma^{\boldsymbol{x}}_{t|t}}^{-1})^{-1}$$

$$m_{t|T} = P_{t|T}(B^TQ^{-1}x_{t+1} + {\Sigma^{\boldsymbol{x}}_{t|t}}^{-1}\boldsymbol{x}_{t|t})$$

   (b) Step 2: Simulate $x'_t \sim N(m_{t|T}, P_{t|T})$.

with

$$\widetilde{Q}^{-1}\widetilde{C} = \widetilde{Q}(I_T \otimes B)P\begin{bmatrix} I_{2(T-2)} & 0_{2(T-2)\times 2} \\ 0_{2\times 2(T-2)} & 0_{2\times 2}\end{bmatrix}$$

being a product of block diagonal matrices. Similarly, $\widetilde{Q}^{-1}\widetilde{C}$ is a block upper diagonal and $\widetilde{C}^T\widetilde{Q}^{-1}\widetilde{C}$ is block diagonal. So V^{-1} is a sum of block diagonals and upper and lower block diagonal matrices, i.e. is block tridiagonal. Therefore, $(V^{-1} + \widetilde{H}^T\widetilde{R}^{-1}\widetilde{H})$ as a symmetric block tridiagonal matrix can be inverted in $O(T^2)$ operations. In our implementation we used Jain et al. [2007] who derive a compact representation for the inverses of block tridiagonal and banded matrices that can be implemented very efficiently. Consequently, as $\widetilde{R}^{-1}\widetilde{H}$ and its inverses are block diagonal, such multiplications are of computational complexity of order $O(T^2)$.

As $T \gg n$ we omit here the influence of $O(n^3)$ cost of single block inversions as they would be dominated by quadratic term in T. Finally, we note that a standard Gibbs sampler requires only the inversion of $(\oplus_{t=1}^{T}R)^{-1}$ which requires a cost of $O(T)$. Nevertheless, in simulations not presented here and using the same computational cost for each method, we observed a very large gain in terms of efficiency for Algorithm 1 compared to standard Gibbs.

C Supplementary material for Section 6.1: thresholds for comparing β estimates

In this simulation , we aim to provide a value of space distances $l(\beta_1, \beta_2)$ which can be used to determine when $sp(\beta_1)$ and $sp(\beta_2)$ could be considered statistically different at a 1% significance level. For this purpose, we set $\beta_1 = \frac{1}{\sqrt{3}}(1, 1, 1)^T$ and generate 10^5 uniform realizations of β_2 from the prior distribution in (8). We assume $\beta_1, \beta_2 \in \mathbb{R}^{3\times 1}$ in order to be ensure the cointegration rank is 1. For each simulated value of β_2, we compute $d(\beta_1, \beta_2)$. In Figure 8, we present the histogram of these simulated distances $d(\beta_1, \beta_2)$. It can be found that the 1% quantile of the simulated distribution of $d(\beta_1, \beta_2)$ under the prior is 0.14. The prior is a uniform distribution on a Stiefel manifold, so we could use this value to be the threshold for determining whether $sp(\beta_1) \neq sp(\beta_2)$.

D On Estimation using Johansen method for (1)-(4)

Often in the literature, the Johansen's method is used first to compute β and then more advanced model fitting is performed in a second stage. We will show that for our model in (1)-(4) this can lead to higher uncertainty and bias in the estimation. Here we will apply Johansen's method by assuming constant μ. In particular, we will set $\xi = 0$, $\Pi = \alpha\beta^T = \beta diag((\varsigma_A - 1))\beta^T$. Values of ς_A close to 0 correspond to fast mean reversion whereas values close to 1 to the spread behaving like a random walk. Also we will set $B = diag(\varsigma_B)$ and consider a range of values for $(\varsigma_A, \varsigma_B)$ which appear in Section 6.1 to be the values that the real soybean crush data exhibits. Using B means there is some mispecification when applying Johansen's method with μ being constant.

We will observe how the distance l in (22) between a true value of cointegration space and the Johansen's estimator varies across the different $(\varsigma_A, \varsigma_B)$. Here we assume $n = 4$ and $r = 3$ hence $l \leq 1$. For each different values of $(\varsigma_A, \varsigma_B)$ we will use 10^4 independently generated data-sets, which are simulated from the observation model (1), whilst ensuring that the autocorrelation matrix A of $z_t = \beta^T y_t$ in (5) has spectral radius smaller than 1. We use 100 uniform random samples of β, for each of which we generate 100 random realizations of noise process ϵ_t. From Figure 9, we see that varying ς_B for a given ς_A has a significant impact on the Johansen's estimator. As expected, increasing ς_B brings about increased level of estimated cointegration space distortion and wider confidence intervals for $l(\hat{\beta}^J, \beta_{true})$. The same behavior holds when ς_A is fixed and ς_B is increasing.

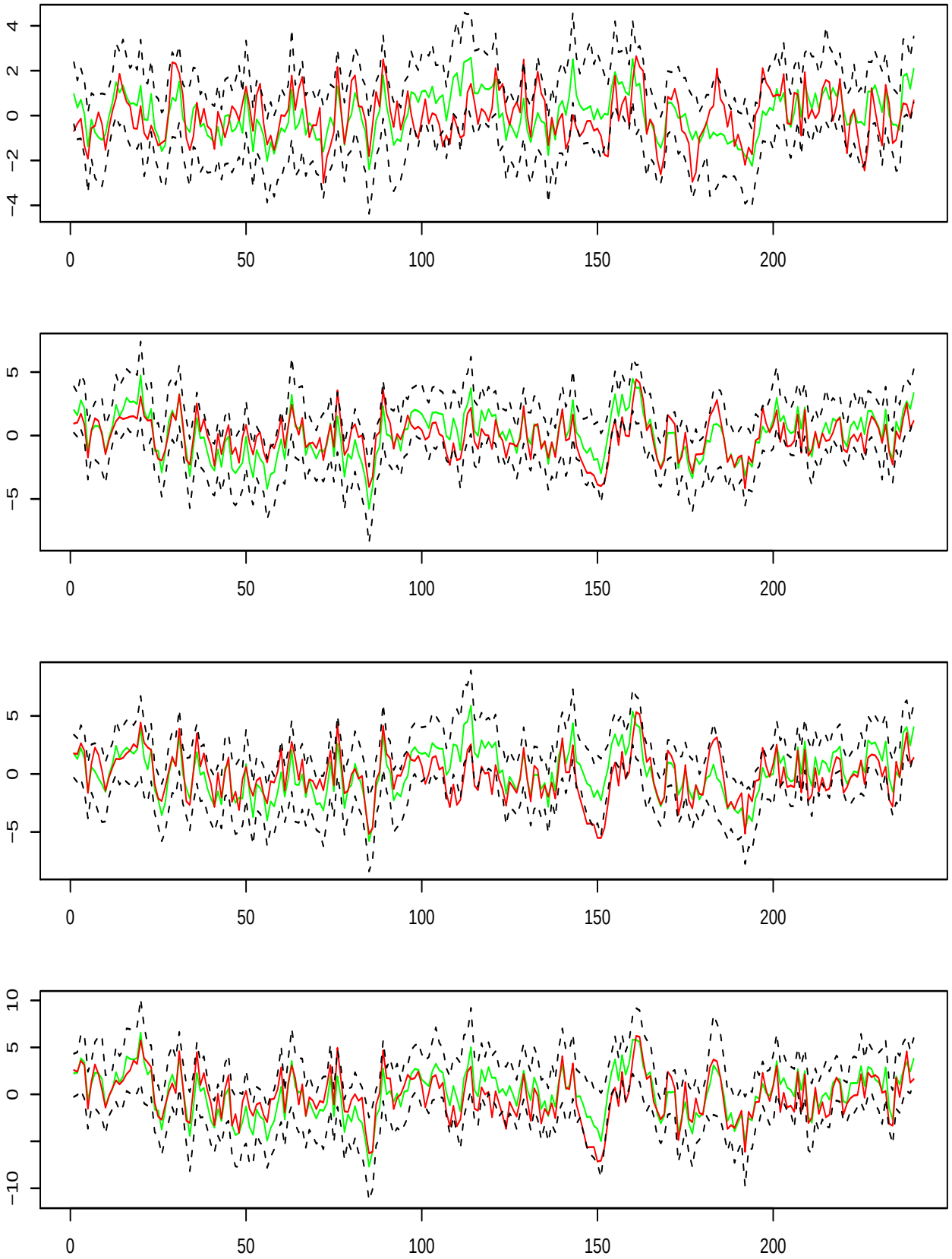


Figure 1: Estimation of $\mu_t(l)$ for $l = 1, \ldots, 4$: in each panel we show the true $\mu_t(l)$ used to simulate the data (red line), the posterior mean estimate from the Algorithm 1 (green line) and 95% posterior confidence intervals (dashed lines).

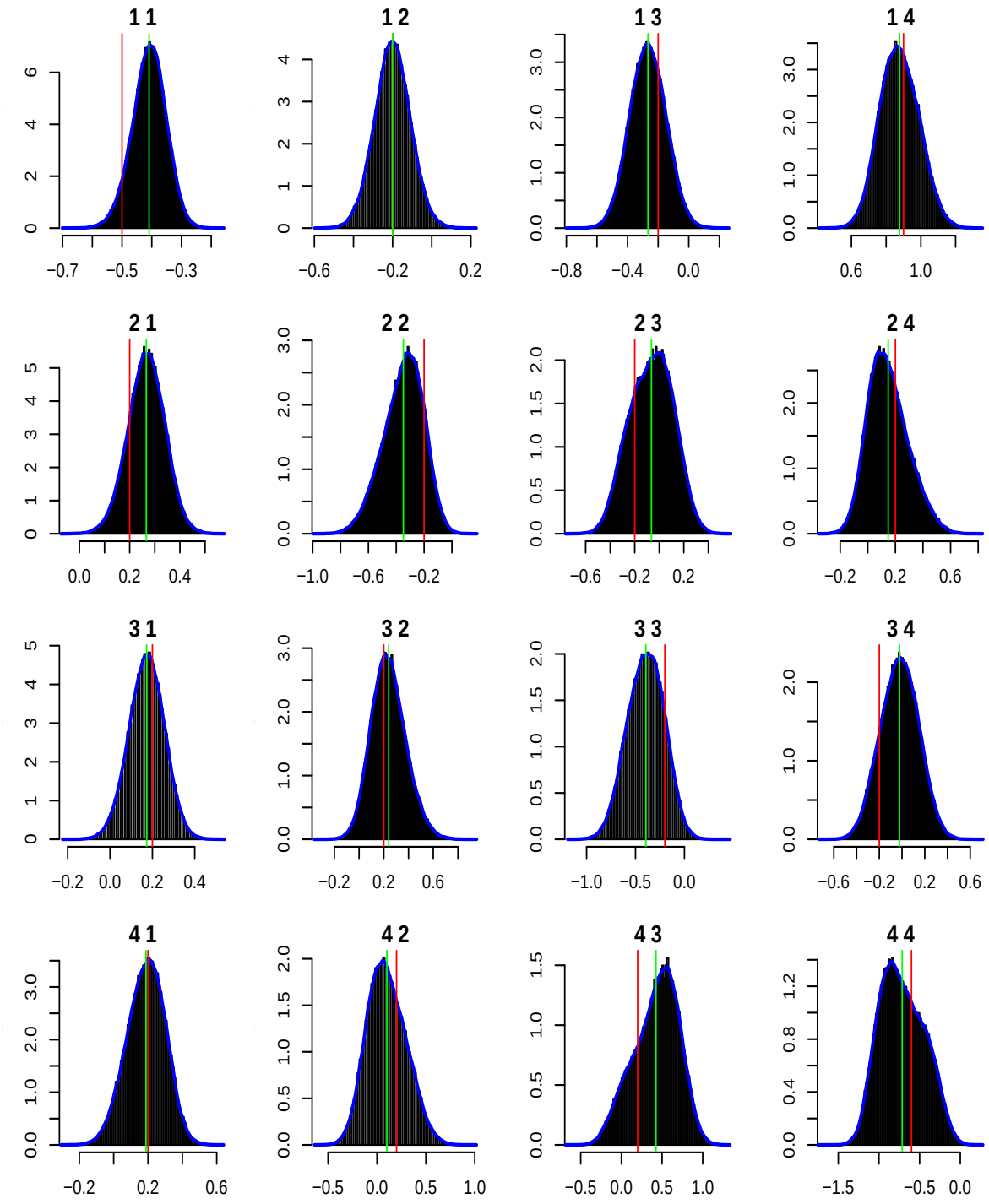


Figure 2: Histograms with estimated density curves of all entries of the long-run multiplier matrix $\alpha\beta^T$. The red horizontal lines mark the true value and the green ones the estimated posterior mean. Blue line is the fitted posterior density curve.

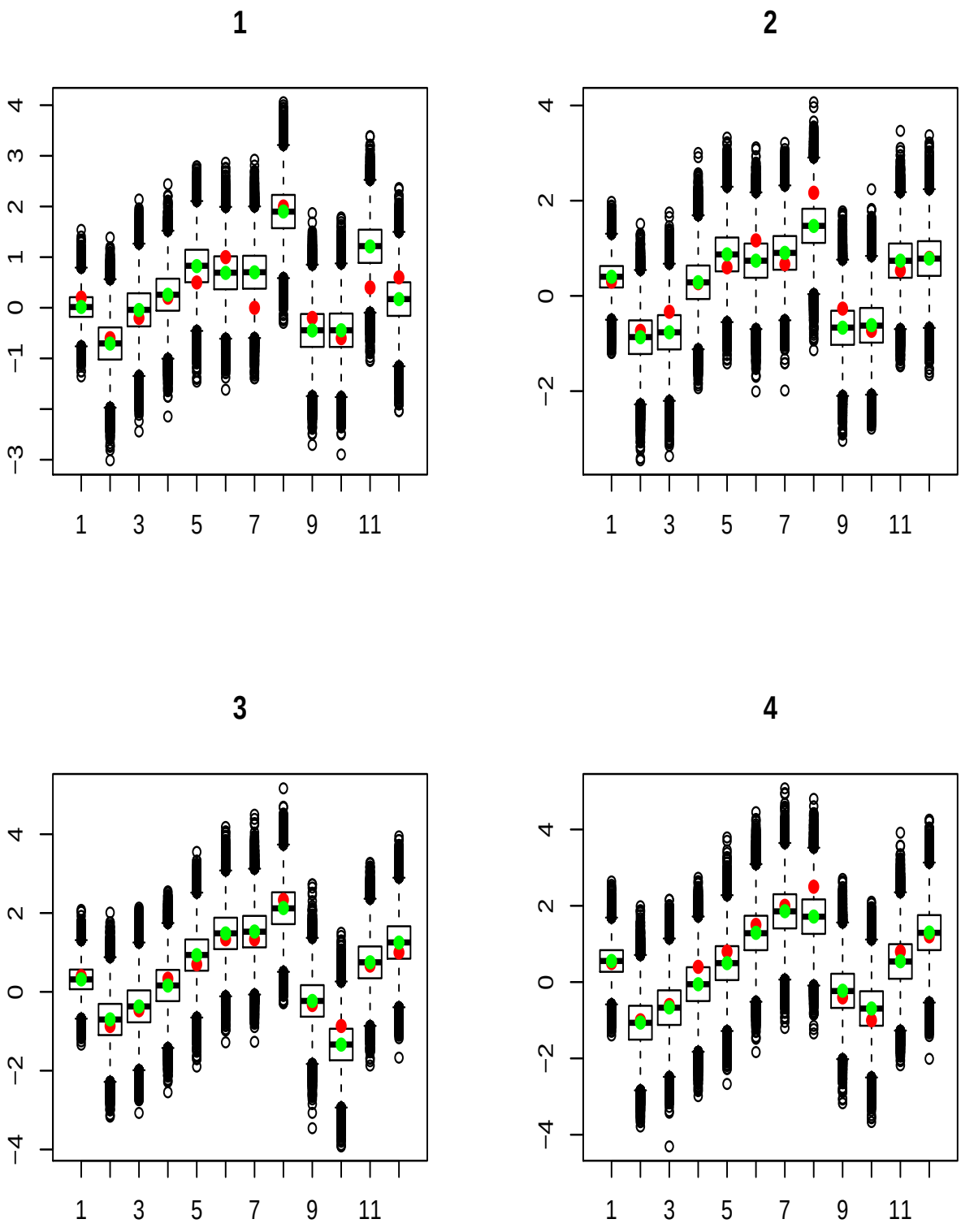


Figure 3: Boxplots for posterior of ξ: the green dot indicates the true value and the red one the posterior mean.

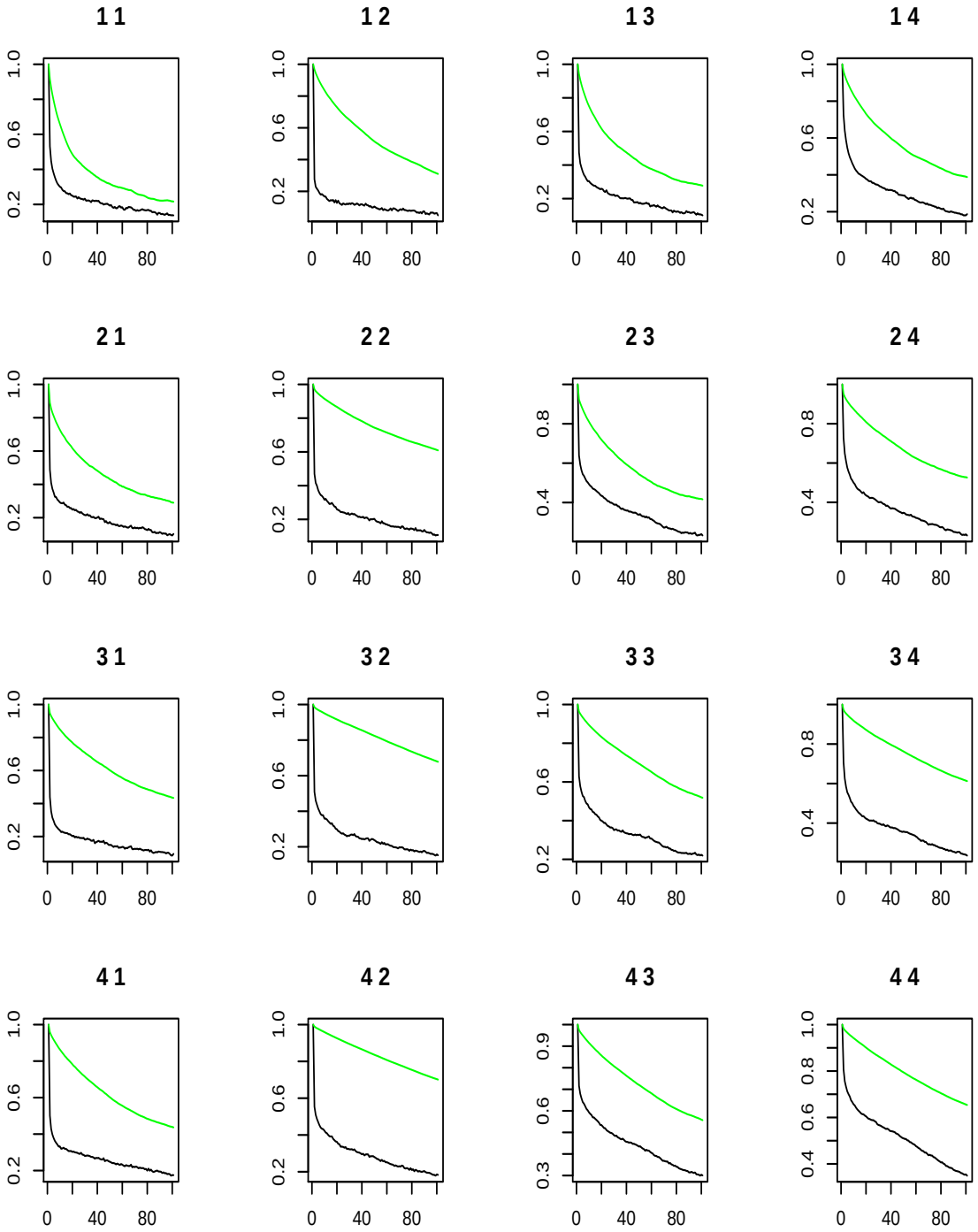


Figure 4: Autocorrelation of $\alpha^{(i)}\beta^{(i)T}$ against lag: black is for Algorithm 1 and green for standard standard Gibbs sampler.

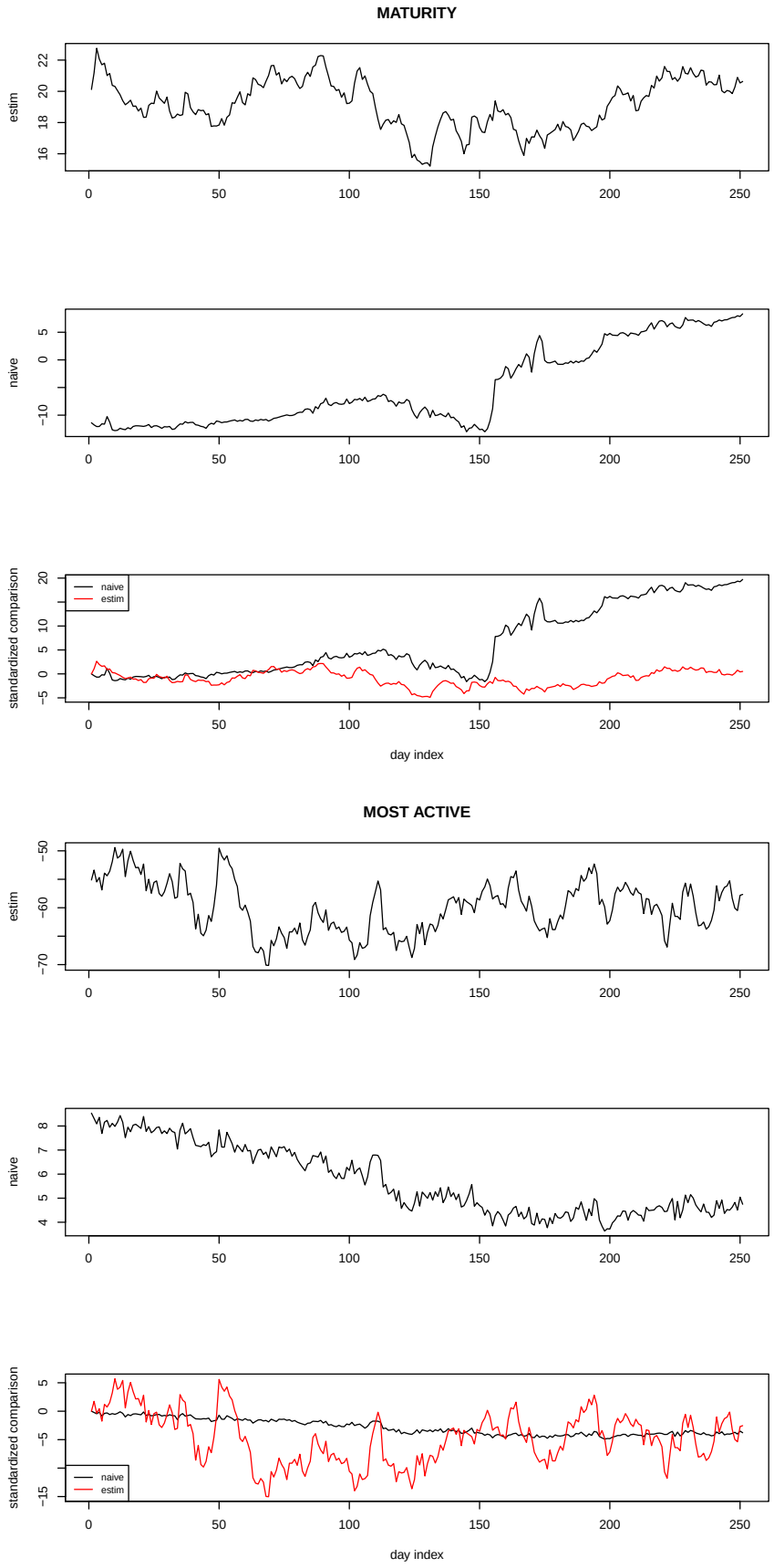
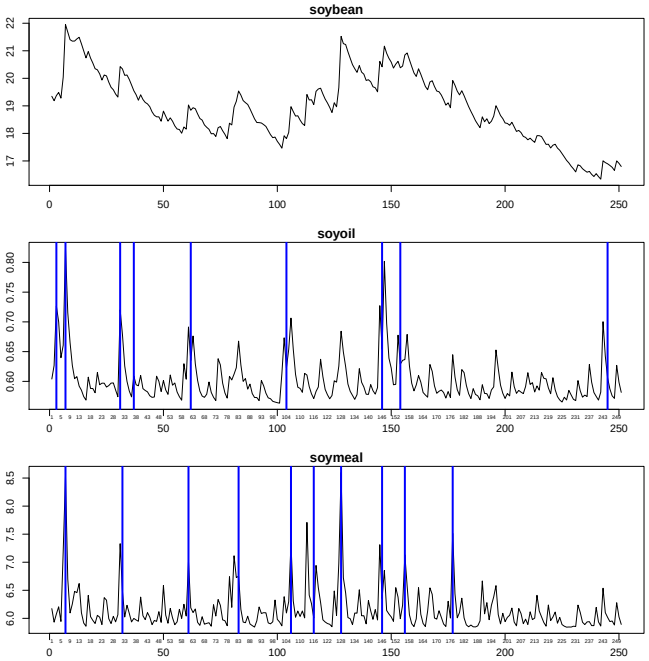
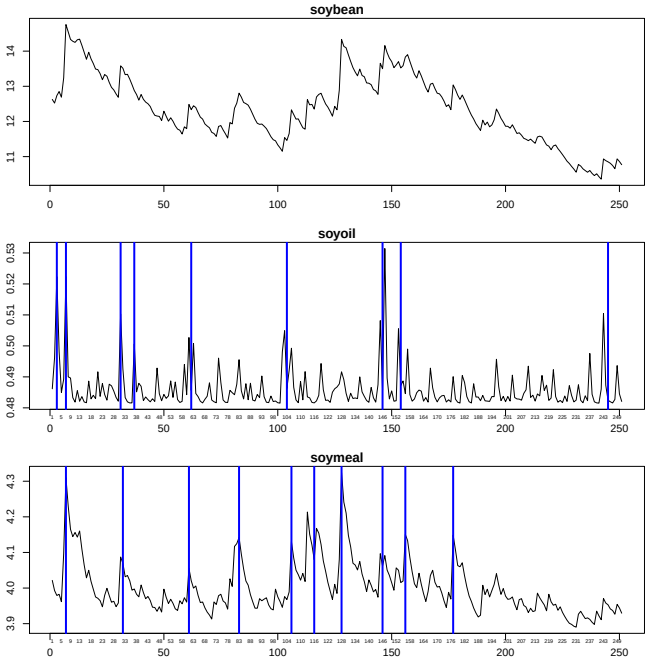


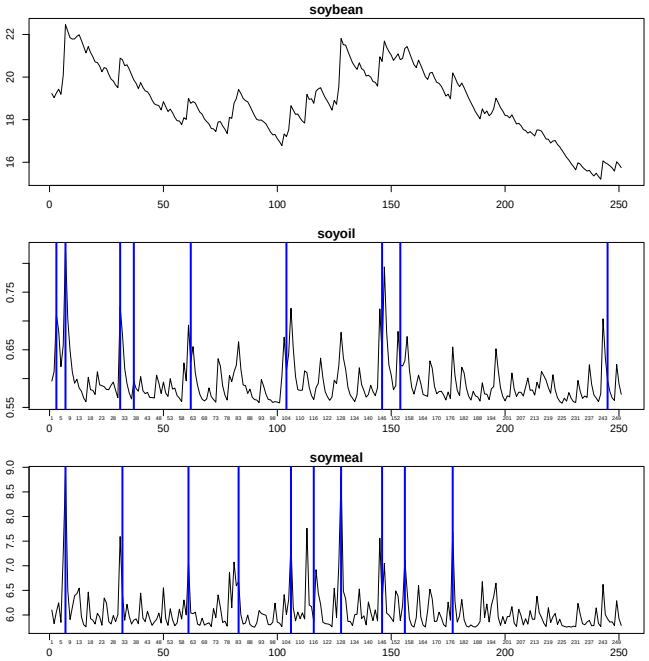
Figure 5: Results obtained for 2009; comparison of estimated spread series (estim) $z_t = \hat{\beta}^T y_t$ with $\hat{\beta}$ vectors computed as in Tables **??** and 5 with the physical relation based spread series (naive) $y_{t,1} - 11y_{t,2} - 2.2y_{t,3}$. As the returns from futures trading are computed as a difference of the closing and opening position prices, we also illustrate the standardized series obtained by translating the series so that the first value is set to 0.



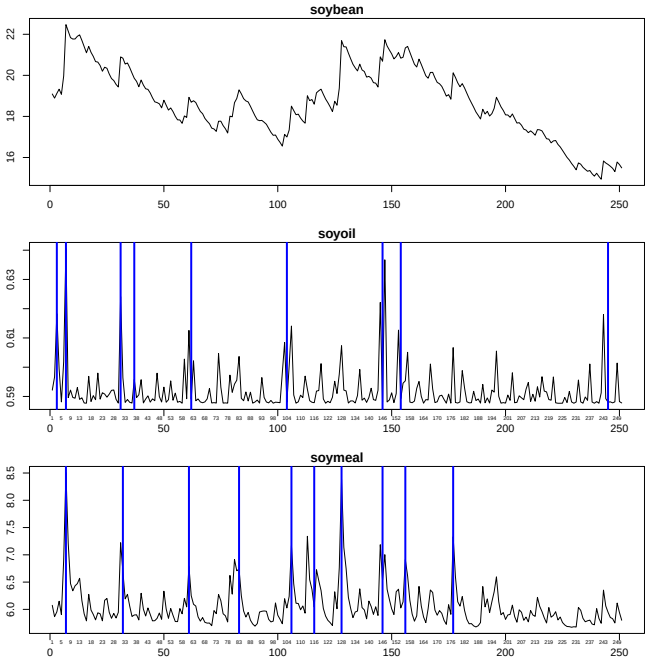
(a) rank 0; BIC = 3034.05



(b) rank 1; BIC = 2867.40

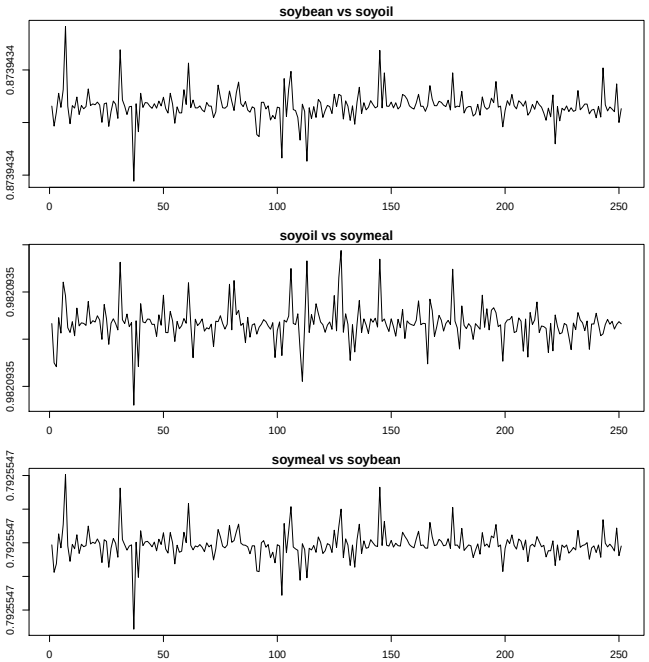


(c) rank 2; BIC = 2999.71

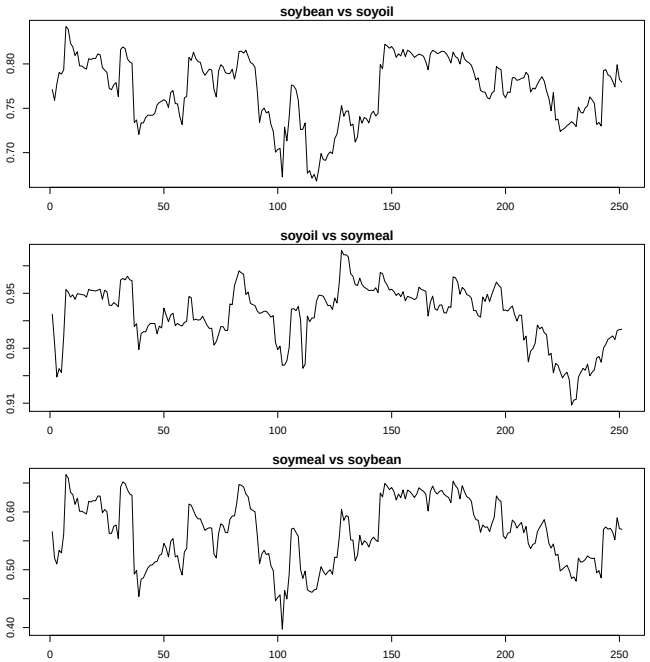


(d) rank 3; BIC = 3039.23

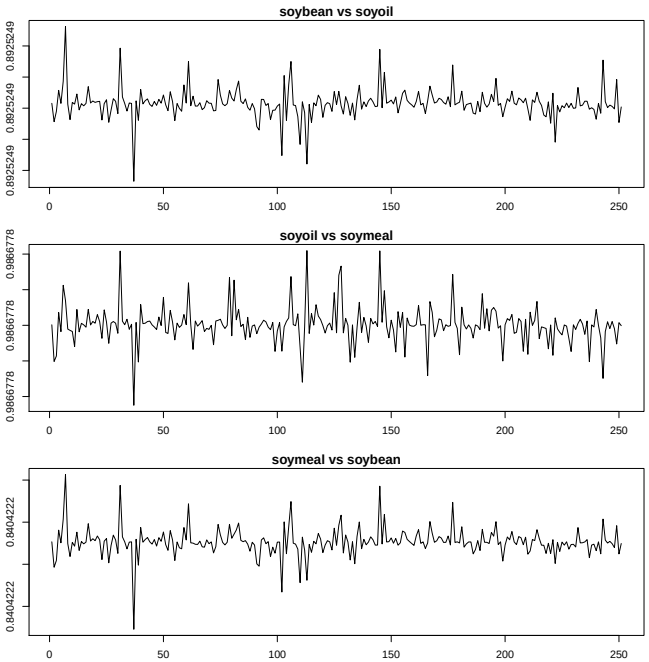
Figure 6: Running volatility estimation from the fitted DCC model. Thick blue vertical line denotes the times at which the volatility spikes occur for the rank 1 fit.



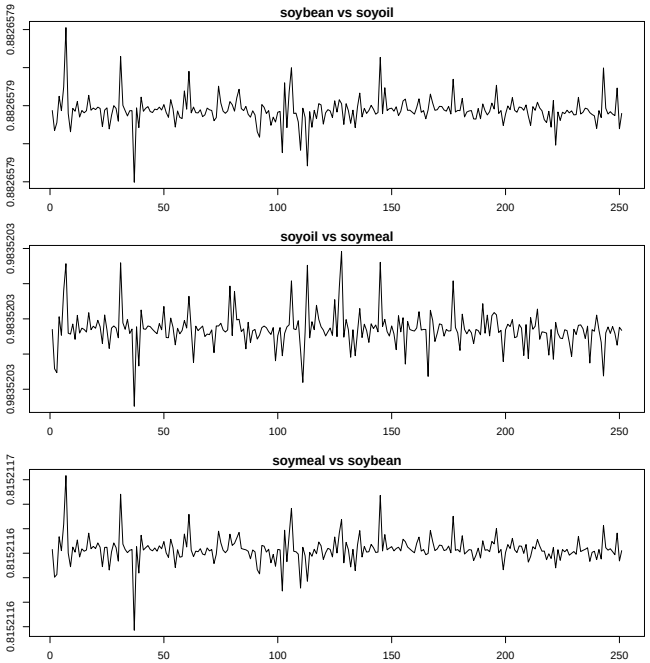
(a) rank 0



(b) rank 1



(c) rank 2



(d) rank 3

Figure 7: Pairwise conditional correlations between marginals of the crush spread obtained from the DCC model under different cointegration ranks.

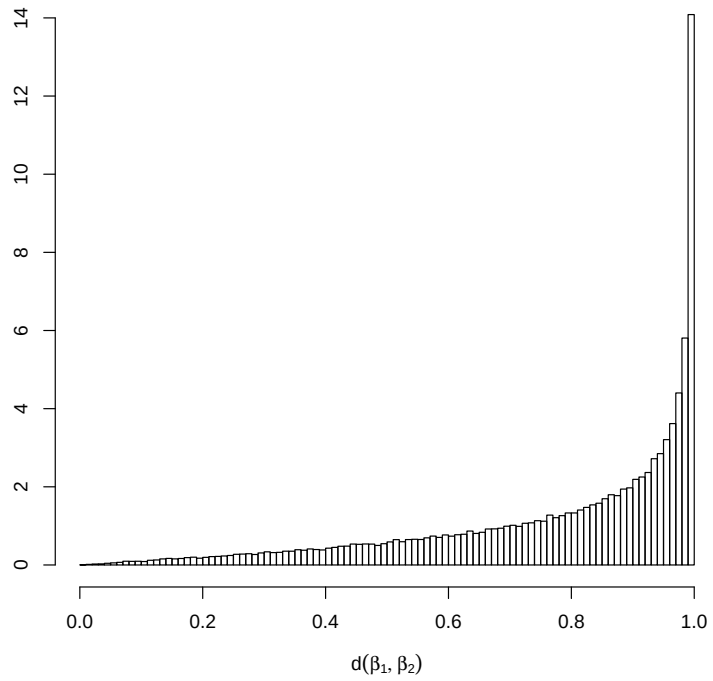


Figure 8: Empirical Distribution of $l((1,1,1)^T, \beta)$ under β being drawn uniformly from the prior.

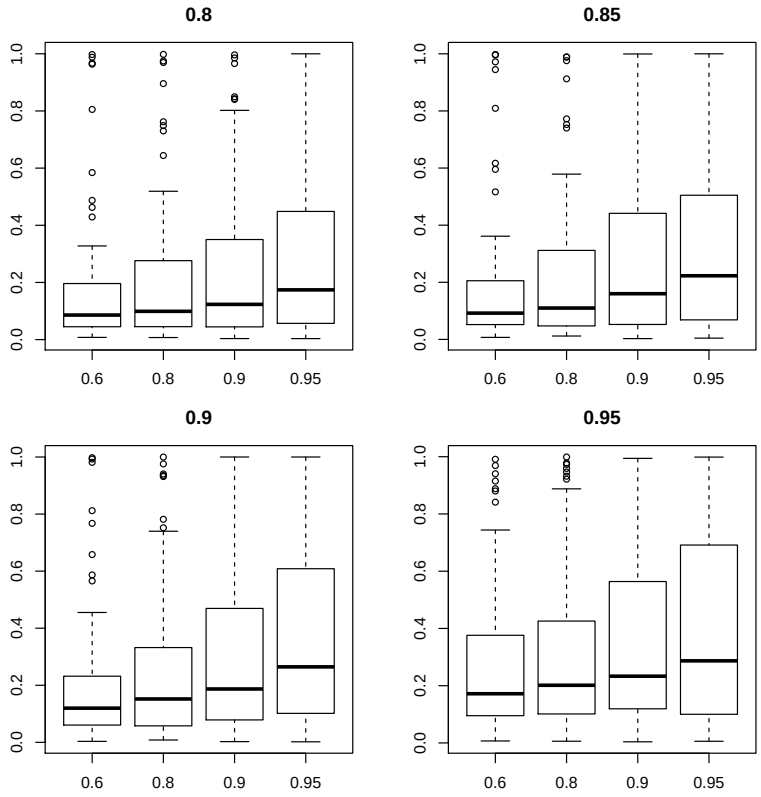


Figure 9: Monte Carlo experiment for $l(\hat{\beta}^J, \beta_{true})$. Each panel corresponds to a different value of $\varsigma_A \in \{0.8, 0.85, 0.9, 0.95\}$. and we show box-plots for each $\varsigma_B \in \{0.6, 0.8, 0.9, 0.95\}$.