Factor-augmented Bayesian cointegration models

Factor Augmented Bayesian Cointegration Model: a case study on the Soybean Crush Spread

Maciej Marowka

Department of Mathematics, Imperial College London, UK.

E-mail: maciej.marowka14@imperial.ac.uk

Gareth W. Peters

Department of Actuarial Mathematics and Statistics, Heriot-Watt University, Edinburgh, UK.

Nikolas Kantas

Department of Mathematics, Imperial College London, UK.

Guillaume Bagnarosa

ESC Rennes School of Business, France.

Summary. In this paper we investigate how vector autoregressive (VAR) models can be used to study the soybean crush spread. By crush spread we mean a time series marking the difference between a weighted combination of the value of soymeal and soyoil to the value of the original soybeans. Commodity industry practitioners often use fixed prescribed values for these weights, which do not take into account any time varying effects or any financial market based dynamic features that can be discerned from futures price data. In this work we address this issue by proposing an appropriate time series model with cointegration. Our model consists of an extension of a particular VAR model used widely in econometrics. Our extensions are inspired by the problem at hand and allow for a time varying covariance structure and a time varying intercept to account for seasonality. To perform Bayesian inference we design an efficient Markov Chain Monte Carlo algorithm, which is based on the approach of Koop et al. [2009]. Our investigations on prices obtained from futures contracts data confirmed that the added features in our model are useful in reliable statistical determination of the crush spread. Although the interest here is on the soybean crush spread, our approach is applicable also to other tradable spreads such as oil and energy based crack or spark.

Keywords: Soybean Crush Spread, State Space models with Cointegration, Bayesian Inference, Markov Chain Monte Carlo

1. Introduction

Soybeans are the fourth most produced crop in the world and if one includes the trading of by-products, they are the most traded agricultural commodity representing in value 10% of the worldwide agricultural trade†. About 85% of the soybeans

†https://www.ers.usda.gov/amber-waves/2016/may/major-factors-affecting-global-soybean-and-products-trade-projections/

are either consumed once processed or transformed into soymeal, soyoil and hull. The soymeal being very rich in protein is mainly used for animal feed (hog, livestock, fish, etc.) 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. A particularly interesting feature of the soybean market is the fact that unlike other commodities, derivative components of soybeans are also traded on commodity futures markets by producers and speculators. While the prices of soybean, soymeal and soyoil are generally strongly correlated with each other, a move in 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 soy byproducts will necessarily correlate with an increase in the demand for the underlying beans; see Barrett and Kolb [1995], Simon [1999], Mitchell [2010].

Loosely speaking, the crush spread is a weighted combination of the value of soymeal and soyoil and the original soybeans, which models the loss or gain in value after soybeans are crushed into soymeal and soyoil. This crush spread relationship is useful for soymeal and soyoil manufacturers, who use it to minimise risk from fluctuations of the purchase price of soybeans. Similarly, it is used from commodities speculators, who often devise strategies that exploit deviations from stationarity. Of course, determining weights used to construct the crush spread is not straightforward. Industry practitioners typically use values based on physical considerations that one unit of soybeans produces certain amount of soybean meal and soyoil. For instance, the Chicago Board of Trade (CBOT) quotes the crush spread as $y_t(1) - 11y_t(2) - 2.2y_t(3)$ (Nelson Low [2015]). Here $y_t(1)$, $y_t(2)$ and $y_t(3)$ correspond to the quoted prices of futures contract associated respectively to soybean, soyoil and soymeal at a given time t. The intuition here is that typically when a bushel of soybeans weighing 60 pounds is crushed, it results to 11 pounds of soybean oil, 44 pounds of 48% protein soybean meal, 4 pounds of hulls, and 1 pound of waste, see Johnson et al. [1991] for more details.

Of course, using fixed weights to determine the crush spread does not take into account any time variation due to the aforementioned market dynamics or seasonality from the production process. In this paper, we aim to address this issue by proposing a carefully designed time series model that exhibits cointegration. Cointegration is a property of multivariate time series whereby a linear transformation of a non-stationary time series leads to a stationary lower dimensional time series, which we will refer to as the cointegrated *spread series*. Following the discussion above, it makes sense to model the crush spread time series as a cointegrated spread series. Whilst changes in supply and demand might cause the prices of the soy products to be non-stationary, the crush spread has been observed to exhibit stationarity due to the strong correlation between the three soy products (Barrett and Kolb [1995], Simon [1999], Mitchell [2010]).

We will use 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 et al. [2006] and Johansen [1995] for reviews. Cointegration models like CVAR have been popular with commodities modelling and modelling spreads, e.g. Dolatabadi et al. [2018]. Here, we will extend the standard CVAR model by adding a dynamic linear factor and a time varying intercept to account for seasonality in agriculture (Simon [1999], Mitchell [2010]). Adding a linear dynamic factor results to a flexible time varying (heteroscedastic) covariance model that is common in econometrics, e.g. Aguilar and West [1998], Geweke and Zhou [1996]. In addition it is a sensible modelling choice for commodity futures contracts, where often market dynamics seem to be adequately captured by factor models as shown in Peters et al. [2013]. Both factor and cointegration models have been used in commodities applications, but to our best knowledge they have not been investigated in a single framework so far. We propose to treat them jointly. The intuition is to specify the cointegration parameters, while at the same time we account for a temporal variations in the model not directly related with the cointegration relationship itself. This is not standard in related applications, where a common practice is to first estimate the cointegration parameters using likelihood based methods found in Johansen [1995] and then the remaining parameters separately; see Miao [2014] and Triantafyllopoulos and Montana [2011] for related studies.

We will perform Bayesian inference for the parameters of the proposed model. In the context of Bayesian cointegation modelling, the posterior distributions can be very sensitive to even seemingly un-informative priors and in addition problems related to model identifiability arise; see Koop et al. [2006] for a review. More recently, the focus has shifted towards inferring the cointegration as a space rather than a vector of parameters, Koop et al. [2006], Villani [2005], Strachan [2003]. 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 (Van Dyk and Park [2008]) and data augmentation steps (Frühwirth-Schnatter [1994], Carter and Kohn [1994]).

The remainder of the paper is organised 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 perform data analysis for the soybean crush spread. In Section 6 we conclude with a discussion and ideas for future investigations.

2. A CVAR model with dynamic factors

Let $y_t \in \mathbb{R}^n$ denote a vector of cointegrated time series. In the soybean case study, each component of y_t will consist of daily prices of different soy products (extracted from futures contracts) and $n = 3$. In the most simple setting a CVAR time series

can be written as:

$$y_t = y_{t-1} + \alpha\beta^T y_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, R), \quad t = 1, ..., T. \tag{1}$$

The matrix $\alpha\beta^T$ is a reduced rank matrix with rank r, which is decomposed as a product of a full rank cointegration matrix $\beta \in \mathbb{R}^{n\times r}$ and a matrix $\alpha \in \mathbb{R}^{n\times r}$ that is commonly referred to as cointegration adjustment matrix. Under this model, we assume that there are r linear cointegration relationships. The primary object of interest for cointegration analysis is to identify β, which is used to construct the stationary spread series $z_t = \beta^T y_t$. Then elements of α can be interpreted as the speed of the mean reversion of z_t. As we are interested in a particular application where $r = 1$, our methodology assumes r to be known. In principle, it can also be estimated via model selection approaches, e.g. Peters et al. [2010] and Sugita [2002]. Also, often CVAR models include higher order autoregressive lags, however for simplicity this is omitted when presenting (1).

Following the comments in the Introduction, we modify (1) to be appropriate to agricultural futures contracts by proposing:

$$y_t = y_{t-1} + \alpha\beta^T y_{t-1} + \mu_t + g_t + \epsilon_t, \quad \epsilon_t \sim N(0, R), \tag{2}$$

where we have added a time varying intercept g_t (that could be held fixed for a given interval of time, e.g. weekly or monthly) and a k-dimensional dynamic linear factor μ_t that obeys:

$$\mu_t = Hx_t \tag{3}$$

$$x_t = Bx_{t-1} + \delta_t, \quad \delta_t \sim N(0, Q), \quad x_0 = 0 \tag{4}$$

where $B \in \mathbb{R}^{k\times k}$ and $H \in \mathbb{R}^{n\times k}$ and the error vectors ϵ_t and δ_t are i.i.d. The factor μ_t is common to all n products under consideration as any new piece of information will affect their prices in similar ways; see Alexander et al. [2002] for a discussion related to commodities and financial applications. For g_t we will assume:

$$g_t = \xi_1 + \sum_{i=2}^{m} \xi_i \mathbb{I}_{\{t\in\mathcal{I}_i\}}, \tag{5}$$

where $\mathcal{I}_i$ denotes a specific interval in time such as the i month or week. For the soybean application we will use monthly varying intercept. This choice is motivated from seasonality intuition in agriculture and empirical studies (Simon [1999], Mitchell [2010]) and we noticed better performance than other alternatives, e.g. weekly, quarterly or yearly. This monthly step function also agrees with the fact that grains futures markets are highly influenced by various economic reports providing assessment of the expected soybean production around the world. These are published early in each month by the United States Department of Agriculture.

In terms of the model structure, since μ_t and x_t are zero mean, they can be thought of as a means of setting a time varying covariance of y_t in a recursive manner. We will assume $k \leq n$ so the covariance of y_t is driven by a smaller number of dynamic factors. For the soybean crush, we will use $k = 2$.

This choice follows an intuition from an application's perspective that commodity assets tend to be modelled using a two factor model, see Schwartz and Smith [2000] or Peters et al. [2013] for more details. Adding these two factors can capture features due to complex market dynamics, e.g. to represent the spot price and the convenience yield or equivalently the short term and long term components of each commodities taken individually. The short term factor could be interpreted as a financially driven factor where for instance the speculation or the hedging pressure are playing an important role, while the long term component is instead related to the physical cycle of the commodities with the long term impact of investments in production capacity or the level of inventories.

The model in (2)-(4) is a linear Gaussian state space model and defines a Gaussian likelihood function $p(y|\alpha, \beta, H, B, Q, R, \xi, x)$, where we denote $y = (y_0, y_1, ..., y_T)$ and similarly for ϵ, μ, ξ and x. This likelihood will be used for inferring parameters $\alpha, \beta, H, B, Q, R, \xi$. An important observation here is that in the likelihood some parameters appear as products, e.g. $\alpha\beta^T$, Hx_t, Bx_{t-1}, etc. This results to a parameterisation with identifiability issues. One would obtain the same likelihood by replacing the aforementioned products with $\alpha V V^{-1}\beta^T$, $HUU^{-1}x_t$ for any non-singular matrices V, U of appropriate dimensions. The same holds for the marginal likelihood when V, U are orthogonal, and x in (3)-(4) is integrated out, i.e.

$$p(y|\alpha, \beta, H, B, Q, R, \xi) = p(y|\alpha V^{-1}, V\beta, HU^{-1}, UBU^{-1}, UQU^T, R, \xi). \quad (6)$$

This is a well known observation in the literature related to both cointegration (Koop et al. [2006], Villani [2005], Larsson and Villani [2001], Koop et al. [2011]) and factor models (Aßmann et al. [2016], Bai and Wang [2015], Jackson et al. [2016], Chan et al. [2018]). The issue is akin to the notorious label switching problem in mixture models, e.g. Celeux et al. [2000], Jasra et al. [2005], Geweke [2007]. Consequently, care should be taken when choosing specific parameterisations (such as restrictions or normalisations for matrices) and priors or loss functions for Bayesian inference. A related implication of this issue, is that the parameterisations and priors chosen can affect the mixing of MCMC samplers or increase the chance that they can get stuck in local modes; see Pitt and Shephard [1999], Papaspiliopoulos et al. [2007]. As a result, when choosing priors below or designing MCMC for the resulting posterior distributions, we put emphasis in extensions that drastically improve sampling efficiency and hence lead to more reliable inference.

3. Bayesian Inference

To perform Bayesian inference for the parameters in (2), we will use the following posterior distribution:

$$\begin{aligned} p(\alpha, \beta, H, R, B, Q, \xi, x|y) \propto\ & p(y|\alpha, \beta, H, R, \xi, x)p(x|B, Q) \\ & \times p(\alpha|\beta)p(\beta)p(\xi)p(B)p(Q)p(H)p(R), \end{aligned} \quad (7)$$

where equation (4) provides an expression for $p(x|B, Q)$. For the remaining variables, we will take into account the identifiability issues mentioned above and allow

for tractable full or partial conditional distributions that allow implementation of Gibbs sampling. Below, we discuss separately the priors for α, β with those for H, B, Q, R, ξ.

3.1. Priors for α, β

Choosing parameterisations and priors for the cointegration matrices has been well studied; see Koop et al. [2006] for a review. In addition to issues related to model identifiability, posterior distributions for $\alpha\beta^T$ can be very sensitive to even seemingly un-informative priors such as uniform distributions on bounded domains. To deal with this in Bayesian cointegation modelling, early approaches include Jeffrey's priors as in Kleibergen and Van Dijk [1994] or nested prior embeddings for different ranks as in Kleibergen and Van Dijk [1998]. Recent advances in the literature, such as Villani [2005], Larsson and Villani [2001], Koop et al. [2011], developed 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.

We will follow this approach and focus directly on the cointegration space, that is equivalent to the column span of β, $sp(\beta)$. 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 $V_{n,r} := \{V \in \mathbb{R}^{n\times r} : \ V^TV = I_r\}$. These spaces are compact, and as such one can be certain that defining a uniform prior on them 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.

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

where $d\mathcal{H}(\cdot)$ denotes the Haussdorf measure (Chikuse [2012]), $P_\tau = \tilde{\beta}\tilde{\beta}^T + \tau\tilde{\beta}_\perp\tilde{\beta}_\perp^T$ with $\tau \in [0, 1]$ and $\tilde{\beta}_\perp$ denotes the orthogonal complement of $\tilde{\beta}$. 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. The Gaussian Angular prior distribution enables us to choose both informative and non-informative priors. Hence the practitioner can specify the prior centred around the line spanned by H and model directly the amount of dispersion around it. For the soybean crush spread, using the physical considerations mentioned earlier we could set $\tilde{\beta}$ to $\beta^C = (1, -11, -2.2)^T$.

Under this parameterisation α will be modelled as an unrestricted matrix. Similar to Villani [2005] we choose a zero mean prior:

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

where v controls the degree of shrinkage and can be set using hierarchical priors; the choice of G is flexible and gives $Vec(\alpha)|\beta \sim N(0, \Sigma_\alpha)$ with $\Sigma_\alpha = (v\beta^T P_{1/\tau}\beta)^{-1} \otimes G$. Here $\otimes$ denotes the Kronecker matrix product. Note that we will often switch between a matrix parameter and its vectorised form. This assists in performing computations more conveniently when using Gaussian matrices and exploits

Table 1. Prior distributions for Q, B, H, R, ξ and distributions of hyper-parameters

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

a standard property that for an arbitrary matrix Y if $Y \sim N_{n,T}(\mu, \Sigma, \Psi)$, then $Vec(Y) \sim N(Vec(M), \Sigma \otimes \Psi)$; see Gupta and Nagar [1999, Theorem 2.2.1].

3.2. Priors for H, B, Q, R, ξ

Based on our discussion so far it is clear that unless particular restrictions in the matrices are specified, then H, B, Q are not uniquely identifiable. Similar to Chan et al. [2018], one could use similar restrictions like for α, β to address this. In this paper, we will take an alternative and simpler approach. Our motivation for including H, B, Q in (2) is based on our desire to use a flexible and simple heteroscedastic model. At the same time μ can be identified and can still be used for inferring the covariance structure of y. As μ is a deterministic function of B, H, Q and x, we will choose priors for B, H, Q that as in $p(y_t|y_{t-1}, \xi_t, x_t, \alpha, \beta, H, R)$ preserve invariance with respect to transformation with an orthonormal matrix and its inverse. The approach is as in Geweke [2007] or Blais [2017] and does not require choosing particular restrictions for H, B, Q. In Table 1, we list the priors that we will use for B, Q, H, R, ξ. Note that these are conjugate priors.

Furthermore, in the experiments not presented in this paper, we realised that the mixing efficiency of the sampler corresponding μ are sensitive to the choice of hyper-parameters. This is not uncommon for linear state space models, see Frühwirth-Schnatter and Wagner [2010] for similar observations. To alleviate this problem, we will use hyper-priors for σ_H^2, σ_R^2 and σ_B^2. These are also presented in Table 1.

4. MCMC Sampling for Posterior Inference

We will use a MCMC sampling scheme to simulate from the posterior of model parameters. For the posterior in (7), it is possible to derive a standard Gibbs Sampler based on data augmentation; see Carter and Kohn [1994], Frühwirth-Schnatter [1994], De Jong and Shephard [1995]. In numerical examples on simulated data presented in the supplementary material, we have noticed poor mixing for such a Gibbs Sampler. We will tackle this issue, by using partial collapsing explained in detail in Park and van Dyk [2009], Van Dyk and Park [2008]. For the posterior in (7), we will perform this by splitting the parameters in three blocks (B, Q, H, R),

(α, β, ξ), x and sampling iteratively as presented below.

$$(H, R, B, Q)^* \sim p(\cdot|y, (\alpha, \beta, \xi)^{i-1}, (H, R, B, Q)^{i-1}, x^{i-1}), \tag{10}$$

$$(\alpha, \beta, \xi)^* \sim p(\cdot|y, (\alpha, \beta, \xi)^{i-1}, (H, R, B, Q)^{i}), \tag{11}$$

$$x^* \sim p(\cdot|y, (\alpha, \beta, \xi)^{i}, (H, R, B, Q)^{i}), \tag{12}$$

where superscript i is the MCMC iteration index. Notice that the sampling requires specific partial conditional distributions in contrast to the standard Gibbs sampler that would use instead full conditionals in (10)-(12). Using this partial collapsing results in a sampler that mixes significantly faster.

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. In particular, in our case, changing the order of (11) with (12) would result in a sampler that is not invariant anymore to the required distribution in (7). The sampling step in (10) simulates from the kernel leaving $p(H, R, B, Q|y, (\alpha, \beta, \xi)^{i-1}, x^{i-1})$ invariant. Whereas, the sampling step in (11) is invariant to $p(\alpha, \beta, \xi|y, (H, R, B, Q)^i)$ with x marginalised out, so can be viewed as a conditional update with Rao-Blackwellisation.

In Algorithm 1 we present our MCMC approach. We found the resulting sampler to be very efficient and much better than a simpler version based on data augmentation and similar to Carter and Kohn [1994], Frühwirth-Schnatter [1994], De Jong and Shephard [1995]. Results on simulated y together with the expressions and derivations of all posterior conditional distributions are presented in the web-based supplementary material document.

As regards to the particular steps in Algorithm 1, (12) can be sampled explicitly using standard Kalman smoothing techniques. In addition, 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 (10), with 2, 5 and 7 corresponding to updating the hyper-parameters. 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 (11), so can be viewed as a Gibbs steps implementation of (11) that is invariant to $p(\alpha, \beta, H, R, B, Q, \xi|y)$ and hence a valid sampler.

In particular, steps 8 and 9 implement the $\kappa-$Gibbs Cointegration sampler of Koop et al. [2009], which in our setup corresponds to a Gibbs sampler invariant to $p(\alpha, \beta|y, H, R, B, Q, \xi, x)$. In Koop et al. [2009] the authors alternate between two parameterisations (α, β) and $(\mathcal{B} := \beta\kappa, \mathcal{A} := \alpha\kappa^{-1}, \kappa = (\alpha^T\alpha)^{1/2})$ in order to implement the Gibbs sampler of Liu and Wu [1999] by alternating simulations of $p(\alpha|y, \beta)$ and $p(\mathcal{B}|y, \mathcal{A}, \kappa)$; (we are omitting here the conditioning on H, R, B, Q, ξ, x that is specific to our problem). The resulting sampler is very efficient and the posterior conditionals admit simple Gaussian expressions. These are based on the observations that $\beta\alpha^T = \mathcal{B}\mathcal{A}^T$, $\kappa = (\alpha^T\alpha)^{1/2} = (\mathcal{B}^T\mathcal{B})^{1/2}$, and $\beta = \mathcal{B}(\mathcal{B}^T\mathcal{B})^{-1/2}$. As a result, $\mathcal{A}$ is semi-orthogonal, while $\mathcal{B}$ is unrestricted. This allows the prior on

(α, β) shown in (8)-(9) to be re-expressed equivalently for $\mathcal{A}$ and $\mathcal{B}$ as:

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

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

where $\Sigma_{\mathcal{B}} = (\mathcal{A}^T G^{-1}\mathcal{A})^{-1} \otimes \nu P_\tau$. Detailed derivations can be found in the technical appendix of Koop et al. [2009].

Algorithm 1 A Partially Collapsed Gibbs Sampler to simulate from (7)

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

- 1. $Vec(B^i) \sim p(\cdot|Q^{i-1}, x^{i-1}, \sigma_B^{2\,i-1})$,
- 2. $\sigma_B^{2\,i} \sim p(\cdot|B^i)$,
- 3. $Q^i \sim p(\cdot|B^i, x^{i-1})$,
- 4. $Vec(H^i) \sim p(\cdot|y, \alpha^{i-1}, \beta^{i-1}, R^{i-1}, x^{i-1}, \xi^{i-1}, \sigma_H^{2\,i-1})$,
- 5. $\sigma_H^{2\,i} \sim p(\cdot|H^i)$,
- 6. $R^i \sim p(\cdot|y, \alpha^{i-1}, \beta^{i-1}, H^i, x^{i-1}, \xi^{i-1}, \sigma_R^{2\,i-1})$,
- 7. $\sigma_R^{2\,i} \sim p(\cdot|R^i)$,
- 8. $Vec(\alpha^*) \sim p(\cdot|y, \beta^{i-1}, H^i, R^i, B^i, Q^i, \xi^{i-1})$. Compute $\mathcal{A}^* = \alpha^*(\alpha^{*T}\alpha^*)^{-1/2}$,
- 9. $Vec((\mathcal{B}^*)^T) \sim p(\cdot|y, \mathcal{A}^*, H^i, R^i, B^i, Q^i, \xi^{i-1})$. 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. $Vec(\xi^i) \sim p(\cdot|y, \alpha^i, \beta^i, H^i, R^i, B^i, Q^i)$,
- 11. Sample $x^i \sim p(\cdot|y, \alpha^i, \beta^i, \xi^i, H^i, R^i, B^i, Q^i)$

4.1. Bayesian Point Estimation

Obtaining point estimates of the cointegration space needs attention as the posterior distribution is defined on a Stiefel manifold. Following Villani [2006], we use the following loss function based on the Frobenius norm,

$$d(\beta, \beta^*) = ||\beta\beta^T - \beta^*(\beta^*)^T||_F;$$

where β and β^* are semi-orthogonal. To provide a Bayesian point estimate for $sp(\beta)$, we use the Posterior Mean Cointegration Space estimator (PMCS) proposed in Villani [2006]. The PMCS estimator is defined as

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

In Villani [2006], it was showed that the PMCS estimator can be computed as

$$\hat{\beta} = (\mathtt{v}_1, ..., \mathtt{v}_r), \tag{15}$$

where $\mathtt{v}_i$ is the eigenvector of $\mathbb{E}[\beta\beta^T \big| y]$ corresponding to the i-th largest eigenvalue. Given the Stiefel manifold is a compact space, all finite moments of the elements of β exist in the orthonormal normalization, which implies existence of $\mathbb{E}[\beta\beta^T \big| y]$. As $\mathbb{E}[\beta\beta^T \big| y]$ is not tractable, we resort to MCMC: after N iterations one can use $\frac{1}{N}\sum_{i=1}^{N} \beta^i \beta^{iT}$ and estimate $\hat{\beta}$ using its eigenvectors.

5. The Soy Bean Crush Spread

We use daily observations of adjusted prices over a period between 2001-2015 and perform estimation separately for each year. This takes into consideration the four seasons and the associated soybean cycle of production and processing and allows for a comparison of our model and estimation results under different years each of which could be influenced differently by events such as weather or market dynamics. In addition, we are aligned with a campaign calendar which is split in two important seasons: the planting season and the harvest season. The data for this case studies is obtained from the futures contracts (see below for details) related to the American soybean market. As a result, we matched our data truncation periods with the agronomical calendar of the US soybean. For the cointegration prior we will mostly use $v = 1, \tau = 1$ and $G = I$ that is a uninformative uniform prior. This is convenient when comparing our results with likelihood based estimates (Johansen [1995]), but note that our model is still diffferent due to other priors and μ_t, g_t. Clearly one could use our methodology with informative priors either by setting $\bar{\beta}$ to the physical weightings specified by CBOT, $\beta^C = (1, -11, -2.2)$ or an estimate from a previous year. This will be used when performing some model comparisons below.

Regarding the remaining model parameters we select

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

ensuring vague inverse-Gamma priors and a relatively weak prior belief induced by hyper-parameters. For the time varying intercept we use ξ_i to correspond to monthly intercept. We also assume that there is one cointegration relationship, so $r = 1$. This is natural here as the only source of the cointegration relation is driven by the physical refinery conditions. Additionally, we will use $k = 2$, which as mentioned earlier in the Introduction is a common assumption used in commodities and is motivated by the long/short term model introduced in Schwartz and Smith [2000] and discussed recently in Peters et al. [2013].

5.1. The data

The raw data set is a panel of the CBOT futures price curves from Bloomberg for soybean (with corresponding Bloomberg ticker S), soyoil (BO) and soymeal (SM) reported daily together with the corresponding volumes and is available by request from any of the authors. Before doing any statistical analysis we need to determine

y_t used in (2). We use the adjusted price series obtained by applying a commonly used strategy in the industry called the rollover strategy. For each day t it assigns y_t to the closest quoted contract with the highest volume, i.e. the most active contract across the futures curve is being traded. Details of this and other methods are provided in Lucia and Pardo [2010], Carchano and Pardo [2009].

5.2. Estimation results

We begin with a discussion on estimation of β that determines the cointegration spread. In Table 2, we present the PMCS estimates, $\hat{\beta}$, for each year. For convenience we set the first entry being normalised to 1. An interesting observation here is that in 2011 there is a sign switch for the estimated $\beta(2)$ compared to all other years and CBOT quoted values. Perhaps this is due to different trading positions of market participants. This could be also explained by the fact that Argentina, the third biggest producer of soybean, suffered the most important drought of the last two decades during the 2011/2012 campaign, which reduced the yield of about 20%. Note that the planting and harvest seasons of the southern hemisphere are six month later than the US one, so this can impact late in the year the soybean cointegration relationship, which is mainly driven by the summer US production season. We also present results for the posterior mean of the spectral radius of B and $A = I_r + \beta^T\alpha$ each denoted as $\rho(B)$ and $\rho(A)$ respectively. Note that both B and A are autoregression matrices for x_t and z_t respectively, so we expect the spectral radii to be less than 1 so that a posteriori suggests μ_t and z_t are stationary. This seems to be the case with one exception for year 2004. Note that in our modelling we do not use any prior restrictions to impose this. In Figure 1, we plot posterior densities $\rho(B)$ for selected years, whose support is in $(0, 1)$, which is evidence for using μ_t in our model. In Figure 1, we only present years 2010-2013, however the results for other years are very similar. Finally, in the last column of Table 2 we record the the posterior cointegration space variation, $\hat{\tau}^2_{sp\beta}$, proposed in Villani [2006]. This can be interpreted as a measure of variability in posterior distribution of the cointegation space that takes values between 0 and 1.

Notice that the results vary significantly from year to year. In the top panel of Figure 2 we show the boxplots from the MCMC output corresponding to the 95% credible intervals of $\beta(2)$ and $\beta(3)$ for each year; we have transformed the MCMC output so that $\beta(1) = 1$. For convienience we also plot, the CBOT value, β^C, and for the sake of illustration we also include in the bottom panel a restricted (to $\mathbb{V}_{n,r}$) maximum likelihood estimate, β^J, commonly referred to as the Johansen method (Johansen [1995]). In the bottom panel of Figure 2 we also include confidence intervals from a bootstrap procedure. Both the Johansen estimate and bootstrap intervals were implemented using the `tsDyn` CRAN package, see Stigler [2010]. Both β^C and β^J are common choices considered by the pracitioners, see Miao [2014] and Dunis et al. [2006] for some recent case studies. Note that β^J does not account for μ_t and uses a constant for g_t and hence is expected to produce different to our estimates. In addition, we note that there is considerable variability in the bootstrap estimates. We observe some difference in estimation results between $\hat{\beta}$ and β^J in some of the years under consideration. In addition, bootstrap bands

Table 2. MCMC estimates for $\hat{\beta}$, $\rho(B)$, $\rho(A)$, $\hat{\tau}^2_{sp\beta}$.

Year	$(\hat{\beta}(2), \hat{\beta}(3))$	$\hat{\rho}(A)$ $(Var[\rho(A)\|y])$	$\hat{\rho}(B)$ $(Var[\rho(B)\|y])$	$\hat{\tau}^2_{sp\beta}$
2001	$(-21.20, \quad -1.41)$	0.91(0.01)	0.51(0.19)	0.0001
2002	$(-20.58, \quad -3.23)$	0.92(0.02)	0.45(0.16)	0.0179
2003	$(-43.09, \quad -0.18)$	0.86(0.02)	0.21(0.11)	0.0041
2004	$(-15.91, \quad -2.21)$	1.02(0.01)	0.33(0.11)	0.0015
2005	$(-17.52, \quad -2.14)$	0.94(0.009)	0.26(0.10)	0.0015
2006	$(-28.58, \quad -1.53)$	0.94(0.012)	0.34(0.09)	0.0072
2007	$(-12.15, \quad -2.00)$	0.87(0.02)	0.72(0.12)	0.0001
2008	$(-11.10, \quad -2.14)$	0.74(0.04)	0.32(0.18)	0.0006
2009	$(-12.92, \quad -2.03)$	0.98(0.01)	0.27(0.11)	0.0001
2010	$(-11.92, \quad -2.15)$	0.92(0.01)	0.33(0.11)	0.0004
2011	$(0.59, \quad -3.29)$	0.96(0.01)	0.41(0.16)	0.5908
2012	$(-10.04, \quad -2.96)$	0.97(0.01)	0.34(0.14)	0.0057
2013	$(-4.39, \quad -1.80)$	0.99(0.01)	0.44(0.11)	0.1126
2014	$(-15.88, \quad -1.41)$	0.99(0.02)	0.39(0.17)	0.3066
2015	$(-11.67, \quad -2.73)$	0.87(0.02)	0.72(0.12)	0.0294

on the Johansen estimated frequentist point estimators tend to be wider compared to our Bayesian credible bands with the exception of years 2003, 2006 and 2014. Naturally, these can be attributed to the different models used, but both estimation procedures tend to agree in that β varies from year to year and in most cases end up with the same sign.

An interesting finding is that not only the cointegration relationship changes yearly, but so does the variance of the corresponding posterior distributions of β presented in Figure 2. In years 2007-2011, there is evidently low variance that is most likely related to the global financial crisis and attributed to the lower level of market speculation. We attribute this to the fact that during market meltdown investors would prefer to stick to the physical considerations that one unit of soybeans produces certain amount of soymeal and soyoil and not the price expectations for each byproduct relative to the bean price.

It is documented in the literature that the dependence among the commodities was even stronger during this period because of the financialisation of the commodities market, see Henderson et al. [2014]. It is also why all other estimators are more closely aligned to the CBOT quotes which corresponds to the physical, trading activity independent, refinery conditions of processing soybeans. For the years with higher level of variance (2003, 2006, 2014) it is harder to assess reasons behind this. It could be attributed to varying production levels (e.g. the high level of soybean annual production in 2014), higher soybean annual average consumption growth from Asia or agriculture boom starting around 2003, deregulation incentives in the U.S. during 2006 or even trends in international prices of metals and fuel that affect agriculture.

5.3. Model comparisons

We proceed by considering different model specifications for (2) and the prior for β:

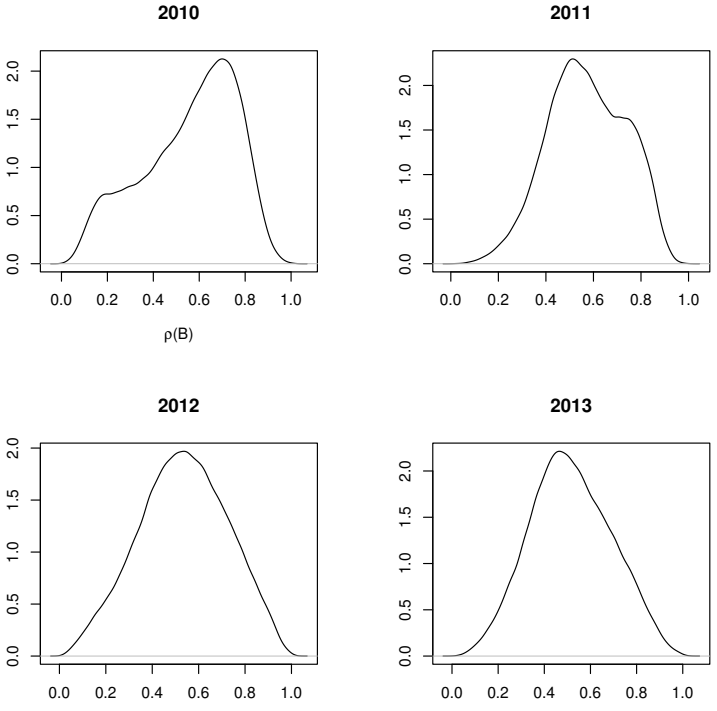

Fig. 1. Posterior density of the spectral radius of autoregression matrix B.

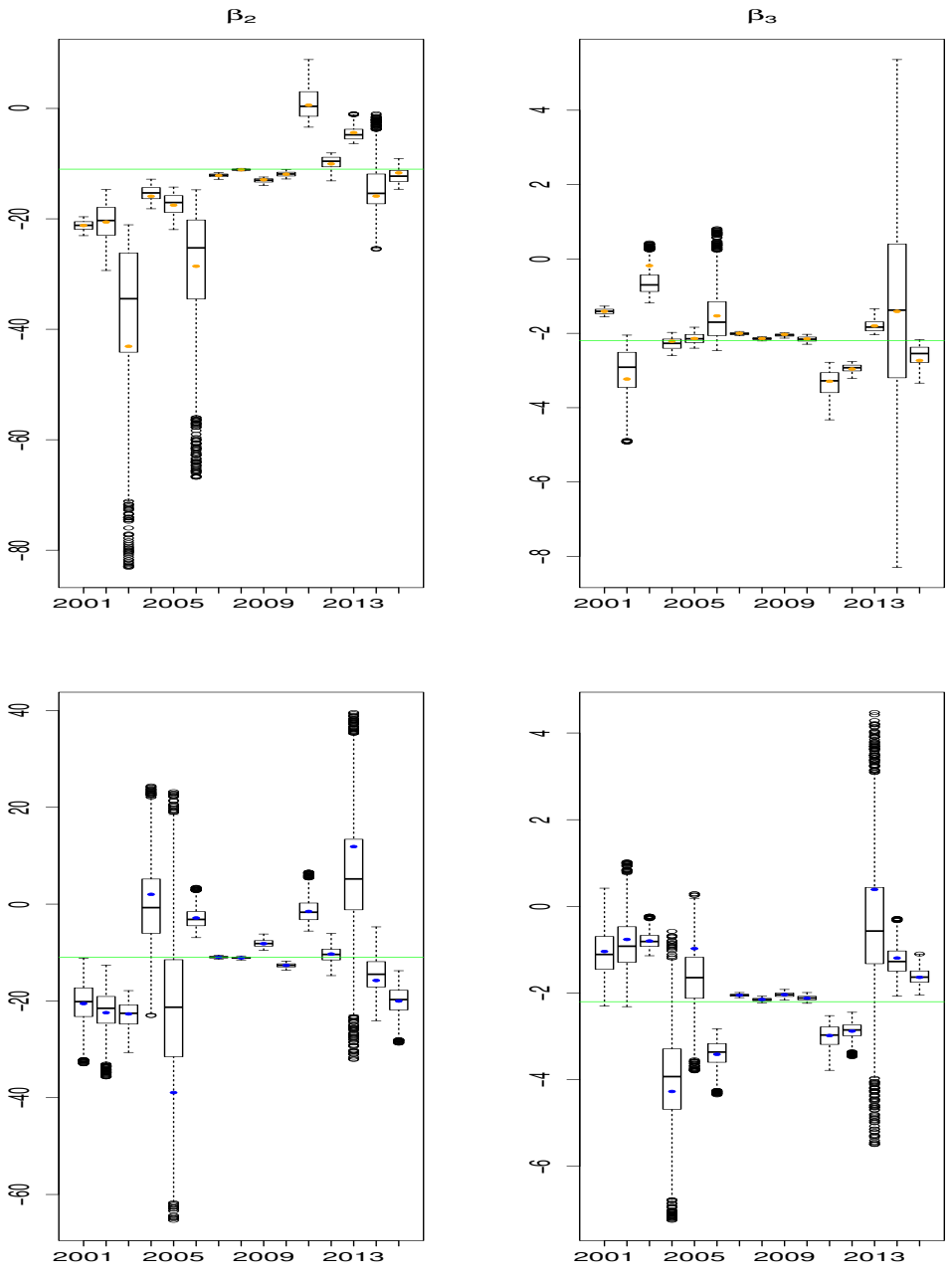

Fig. 2. Top panels: box-plots represent posterior credible intervals for each year. Orange dots correspond to the PMCS estimators. Bottom panels: dots correspond to Johansen restricted maximum likelihood estimates and box-plots are bootstrap confidence intervals. In both plots green lines denote β^C.

- $\mathcal{M}_1^I$ uses full model for (2) with an informative prior on β using the PMCS estimator of the previous year,
- $\mathcal{M}_1^U$ as above but with a uniform prior on β,
- $\mathcal{M}_1^{L1}$ and $\mathcal{M}_1^{L2}$ is as $\mathcal{M}_1^U$ but with one and two additional autoregressive terms resp., i.e. $\Gamma_1 \cdot (y_{t-1} - y_{t-2})$, $\Gamma_2 \cdot (y_{t-2} - y_{t-3})$.
- $\mathcal{M}_2$ is as $\mathcal{M}_1^U$ but without factors, i.e. fixed $\mu_t = 0$.
- $\mathcal{M}_3$ uses (2) with $\mu_t = 0$ and $\alpha, \beta = 0$, i.e. using only a time varying intercept ξ with no cointegration,
- $\mathcal{M}^C$ uses (2) with β known and fixed to the CBOT quoted values,
- $\mathcal{M}^F$ is a dynamic factor model containing μ and ξ only, i.e. $\alpha, \beta = 0$.

We wish to compare the models above in order to assess the various features in (2), i.e. specification of priors, the inclusion of factors μ, whether additional lags are appropriate, etc. We will assume equal probability a priori for each model, so in Table 3 we present estimates of the log marginal likelihoods $p(y|\mathcal{M})$ as computed from separate MCMC runs. For the estimators we use the stabilized harmonic mean estimator presented in Raftery et al. [2007]. In the supplementary material we also present diagnostics and a comparison w.r.t the naive harmonic mean estimator.

The first conclusion from Table 3 is that one of $\mathcal{M}_1^U$ and $\mathcal{M}_1^I$ are the best performers for every year except 2005, 2009, 2010 and 2014. Cointegation is definitely useful and only in 2014 a model without cointegration ($\mathcal{M}^F$) is marginally better. Note also, that in Table 2, 2014 has a quite high value for $\hat{\tau}^2_{sp\beta}$, so this variability in the posterior for β could be the cause of this. In fact, 2014 was a year of higher than usual production, so our approach seems to adapt to changes in market supply and demand. In the other years (2005, 2009, 2010) cointegration models with higher order lags seem to improve the marginal likelihood, but of course this comes at some loss in parsimony. In addition, for most years using an informative prior $\mathcal{M}_1^I$ gives the best model. When comparing $\mathcal{M}_1^\star$ (models with a cointegation relationship and factors) with $\mathcal{M}_2$, we notice that removing factors results in much lower marginal likelihood. It is interesting to note that the plain cointegration model of $\mathcal{M}_2$ is more close (and often worse) than a baseline model with a time varying intercept, i.e. $\mathcal{M}_3$. Overall, this suggests using factors here is very beneficial, which do produce a worse but decent goodness of fit even without cointegration included.

The next question to address is how many lags shall we use. We have already seen that in Table 3 using no lags performs better in most years, i.e. $\mathcal{M}_1^{I/U}$ has higher evidence than $\mathcal{M}_1^{L\cdot}$. One could then compare $\mathcal{M}_1^{L1}$ with $\mathcal{M}_1^{L2}$ and notice that the former outperforms the latter in nine out of fifteen years. It seems overall that increasing the number of lags does not come with an increased goodness of fit for most years, and in the interest of parsimony we did not consider investigating higher order lags. Furthermore, comparing $\mathcal{M}_1^{L\cdot}$ with the simpler models $\mathcal{M}_1^C$, $\mathcal{M}_1^2$ and $\mathcal{M}_3$, we still obtain better performance when lags are used as we did when no lags were used.

Table 3. Log Marginal Likelihoods for different models.

Year	$\mathcal{M}_1^U$	$\mathcal{M}_1^I$	$\mathcal{M}^F$	$\mathcal{M}_1^{L1}$	$\mathcal{M}_1^{L2}$	$\mathcal{M}^C$	$\mathcal{M}_3$	$\mathcal{M}_2$
2001	−1041.11	–	−1048.94	−1053.06	−1044.00	−1096.56	−1162.37	−1158.57
2002	−1075.51	−1063.10	−1077.48	−1075.72	−1081.74	−1090.35	−1187.56	−1182.18
2003	−1279.71	−1298.10	−1279.79	−1319.02	−1301.06	−1343.89	−1361.9	−1363.39
2004	−1596.86	−1582.86	−1602.62	−1596.91	−1593.68	−1681.42	−1637.9	−1632.33
2005	−1281.74	−1280.72	−1284.90	−1272.14	−1275.55	−1337.16	−1365.79	−1368.66
2006	−1123.11	−1106.25	−1120.47	−1186.90	−1119.48	−1154.17	−1218.79	−1214.94
2007	−1363.54	−1316.18	−1382.82	−1350.60	−1353.94	−1432.41	−1475.61	−1476.52
2008	−2026.45	−1987.08	−2042.97	−2026.29	−2024.49	−2083.34	−2073.28	−2048.79
2009	−1733.89	−1728.94	−1747.95	−1730.49	−1725.89	−1777.41	−1806.99	−1804.51
2010	−1567.58	−1567.38	−1578.49	−1563.32	−1566.98	−1600.08	−1631.47	−1634.52
2011	−1600.11	−1645.78	−1642.09	−1636.58	−1642.06	−1681.60	−1688.46	−1691.76
2012	−1765.74	−1748.68	−1765.91	−1765.93	−1768.14	−1799.89	−1793.40	−1795.73
2013	−1624.68	−1592.69	−1622.22	−1629.18	−1635.12	−1658.45	−1668.47	−1673.75
2014	−1732.26	−1731.64	−1728.91	−1743.33	−1747.74	−1734.68	−1762.15	−1766.62
2015	−1522.72	−1517.11	−1528.55	−1523.56	−1540.27	−1544.85	−1569.84	−1575.18

Finally, we want to compare our proposed model with using CBOT values. Clearly, $\mathcal{M}_1^{I/U}$ (and $\mathcal{M}_1^{L\cdot}$) have significantly higher evidence that $\mathcal{M}^C$. We note that this is a favourable comparison for using CBOT quoted values. In results not shown here, we noticed using CBOT values for β without factors had much worse performance than what is presented in Table 3.

Furthermore, in Figure 3, in both cases when μ_t is used or not we present the resulting posterior credible intervals of $\beta(2)$ for every year. Using μ_t results to more confident estimation and tighter posteria. The results for $\beta(3)$ are similar so omitted. Finally, in Figure 4, for the same two cases we attempt to visualise the predictions of the spread z_t on a daily basis using current years price, but the model parameters sampled from the previous year's posterior. We plot the 95% quantiles for $p(z_t|y_{1:t-1}, \tilde{y})$ with $\tilde{y}$ being the previous year's data and y_t representing the current years prices for day t. Again including μ_t results in tighter prediction intervals. For brevity, we present only representative results from year 2005 and 2006 and in the supplementary material one can find the same plot with the results from $\mathcal{M}^C$ added.

6. Conclusions

There are several advantages in using a CVAR model to estimate the cointegration vectors instead of using physical considerations related to the crush spread. For instance, in 2014 our model is able to adapt to changes in market supply and demand, showing considerably higher variability in the corresponding posterior compared to previous and following years. Our proposed model is more flexible than common CVAR models by including effects related to seasonality and being able to capture time varying behaviour related to aspects that are difficult to model such as market dynamics, climate variation and level of speculation. Our data analysis for the soy

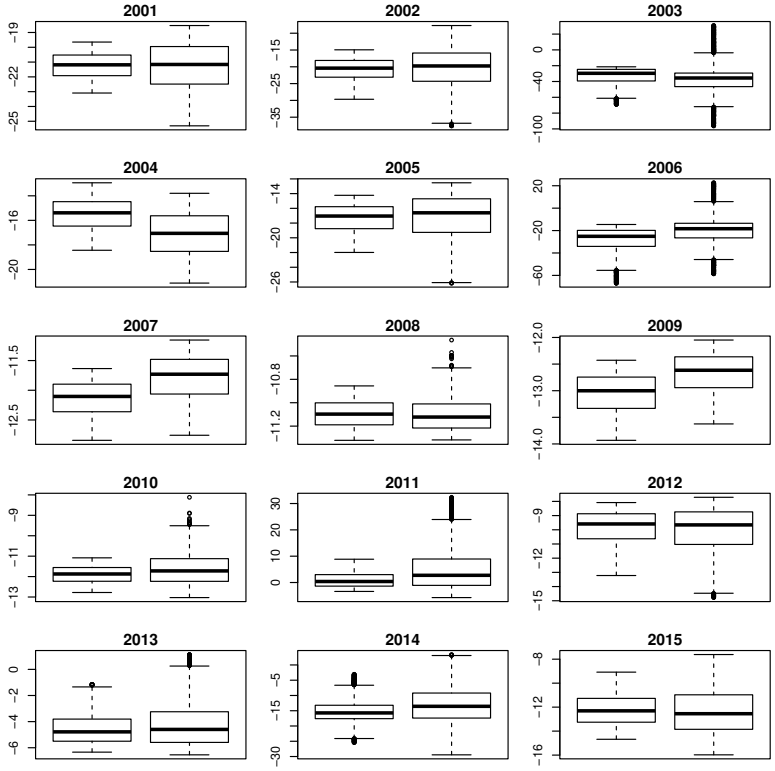

Fig. 3. Marginal posterior distributions of $\beta(2)$ obtained under the models $\mathcal{M}_1^U$ (left in each panel) and $\mathcal{M}_2$ (right in each panel).

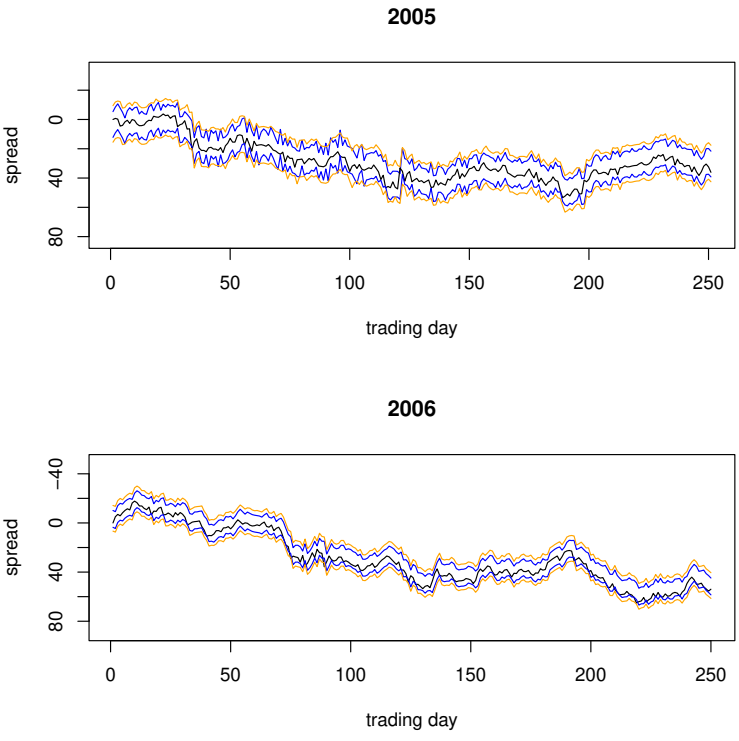

Fig. 4. Out of sample predictive posterior bands for or $p(z_t|y_{1:t-1}, \tilde{y})$ under the models $\mathcal{M}_1^U$ (orange lines) and $\mathcal{M}_2$ (blue lines).

beans crush spread indicates that our proposed model leads to accurate estimation and more confident predictions for the spread series. We emphasise that a reliable statistical analysis of this type is possible only after the development of a dedicated efficient MCMC sampler presented here.

Given the generality of the CVAR modelling our approach can be generalised to other commodity spreads, such as crack or spark spreads. Future work could also look at non-linear stochastic volatility models and extensions of the sampling methodology using particle MCMC methods as in Andrieu et al. [2010], Pitt et al. [2015]. In the soybean case study it was reasonable to assume for the cointegration rank $r = 1$. However, in other application contexts this needs to be estimated, so extensions of this work can include joint estimation of the cointegration rank, the cointegration parameters and the number of factors k or the number of lags. A starting point in this direction could be to consider extensions based on the fractional Bayesian method of Corander and Villani [2004].

Acknowledgements

M. M. was funded under a EPSRC studentship. N.K. was funded in part by JP-Morgan Chase & Co under a J.P. Morgan A.I. Research Awards 2019. Any views or opinions expressed herein are solely those of the authors listed, and may differ from the views and opinions expressed by JPMorgan Chase & Co. or its affiliates. This material is not a product of the Research Department of J.P. Morgan Securities LLC. This material does not constitute a solicitation or offer in any jurisdiction.

References

Omar Aguilar and Mike West. Bayesian dynamic factor models and portfolio allocation. *Journal of Business & Economic Statistics*, 18(3):338–357, 1998.

Carol Alexander, Ian Giblin, and Wayne Weddington. Cointegration and asset allocation: A new active hedge fund strategy. *ISMA Centre Discussion Papers in Finance Series*, 40(2), 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.

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*, 192(1):190 – 206, 2016.

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

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

Sébastien Blais. Normalizing empirically underidentified linear state-space models. Technical report, 2017.

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.

Gilles Celeux, Merrilee Hurn, and Christian P Robert. Computational and inferential difficulties with mixture posterior distributions. *Journal of the American Statistical Association*, 95(451):957–970, 2000.

Joshua Chan, Roberto Leon-Gonzalez, and Rodney W Strachan. Invariant inference and efficient computation in the static factor model. *Journal of the American Statistical Association*, 2018.

Yasuko Chikuse. *Statistics on special manifolds*, volume 174. Springer Science & Business Media, 2012.

Jukka Corander and Mattias Villani. Bayesian assessment of dimensionality in reduced rank regression. *Statistica Neerlandica*, 58(3):255–270, 2004.

Piet De Jong and Neil Shephard. The simulation smoother for time series models. *Biometrika*, 82(2):339–350, 1995.

Sepideh Dolatabadi, Paresh Kumar Narayan, and Nielsen. Economic significance of commodity return forecasts from the fractionally cointegrated VAR model. *Journal of Futures Markets*, 38(2):219–242, 2018.

Christian L Dunis, Jason Laws, and Ben Evans. Modelling and trading the soybean-oil crush spread with recurrent and higher order networks: A comparative analysis. *Neural Network World*, 16(3):193, 2006.

Robert F Engle and Clive WJ Granger. Co-integration and error correction: representation, estimation, and testing. *Econometrica: journal of the Econometric Society*, (2):251–276, 1987.

Sylvia Frühwirth-Schnatter. Data augmentation and dynamic linear models. *Journal of time series analysis*, 15(2):183–202, 1994.

Sylvia Frühwirth-Schnatter and Helga Wagner. Stochastic model specification search for gaussian and partial non-gaussian state space models. *Journal of Econometrics*, 154(1):85–100, 2010.

John Geweke. Interpretation and inference in mixture models: Simple mcmc works. *Computational Statistics & Data Analysis*, 51(7):3529–3550, 2007.

John Geweke and Guofu Zhou. Measuring the pricing error of the arbitrage pricing theory. *The review of financial studies*, 9(2):557–587, 1996.

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

Brian J Henderson, Neil D Pearson, and Li Wang. New evidence on the financialization of commodity markets. *The Review of Financial Studies*, 28(5):1285–1311, 2014.

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.

Ajay Jasra, Chris C Holmes, and David A Stephens. Markov chain monte carlo methods and the label switching problem in bayesian mixture modeling. *Statistical Science*, 20(1):50–67, 2005.

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.

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, Rodney Strachan, Herman van Dijk, and Mattias Villani. Bayesian approaches to cointegration. In *Palgrave Handbook of Econometrics: Volume 1: Econometric Theory*. Palgrave Macmillan Ltd, 2006.

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.

Gary Koop, Roberto Leon-Gonzalez, and Rodney W Strachan. Bayesian inference in a time varying cointegration model. *Journal of Econometrics*, 165(2):210–220, 2011.

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

Jun S Liu and Ying Nian Wu. Parameter expansion for data augmentation. *Journal of the American Statistical Association*, 94(448):1264–1274, 1999.

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.

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.

Tina Liao Nelson Low. CBOT Soybeans vs. DCE Soybean Meal and Soybean Oil Crush Spread. Technical report, Chicago Board of Trade, 2015.

Omiros Papaspiliopoulos, Gareth O Roberts, and Martin Sköld. A general framework for the parametrization of hierarchical models. *Statistical Science*, 22(1): 59–73, 2007.

Taeyoung Park and David A van Dyk. Partially collapsed gibbs samplers: Illustrations and applications. *Journal of Computational and Graphical Statistics*, 18(2): 283–305, 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, 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 and Neil Shephard. Analytic convergence rates and parameterization issues for the gibbs sampler applied to state space models. *Journal of Time Series Analysis*, 20(1):63–85, 1999.

Michael K Pitt, Jamie Hall, and Robert Kohn. Bayesian inference for latent factor GARCH models. *arXiv preprint arXiv:1507.01179*, 2015.

Adrian E Raftery, Michael A Newton, Jaya M Satagopan, and Pavel N Krivitsky. Estimating the integrated likelihood via posterior simulation using the harmonic mean identity. *Bayesian Statistics*, 8, 2007.

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.

Matthieu Stigler. *Threshold cointegration: overview and implementation in R*, 2010. R package version 0.7-2.

Rodney W Strachan. Valid bayesian estimation of the cointegrating error correction model. *Journal of Business & Economic Statistics*, 21(1):185–195, 2003.

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.

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

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