Bayesian Analysis (2011) **6**, Number 4, pp. 755–792

Bayesian Cointegrated Vector Autoregression Models Incorporating α-stable Noise for Inter-day Price Movements Via Approximate Bayesian Computation.

Gareth W. Peters*, Balakrishnan Kannan†, Ben Lasscock†, Chris Mellen† and Simon Godsill‡

Abstract. We consider a statistical model for pairs of traded assets, based on a Cointegrated Vector Auto Regression (CVAR) Model. We extend standard CVAR models to incorporate estimation of model parameters in the presence of price series level shifts which are not accurately modeled in the standard Gaussian error correction model (ECM) framework. This involves developing a novel matrix-variate Bayesian CVAR mixture model, comprised of Gaussian errors intra-day and α-stable errors inter-day in the ECM framework. To achieve this we derive conjugate posterior models for the Scale Mixtures of Normals (SMiN CVAR) representation of α-stable inter-day innovations. These results are generalized to asymmetric intractable models for the innovation noise at inter-day boundaries allowing for skewed α-stable models via Approximate Bayesian computation.

Our proposed model and sampling methodology is general, incorporating the current CVAR literature on Gaussian models, whilst allowing for price series level shifts to occur either at random estimated time points or known *a priori* time points. We focus analysis on regularly observed non-Gaussian level shifts that can have significant effect on estimation performance in statistical models failing to account for such level shifts, such as at the close and open times of markets. We illustrate our model and the corresponding estimation procedures we develop on both synthetic and real data. The real data analysis investigates Australian dollar, Canadian dollar, five and ten year notes (bonds) and NASDAQ price series. In two studies we demonstrate the suitability of statistically modeling the heavy tailed noise processes for inter-day price shifts via an α-stable model. Then we fit the novel Bayesian matrix variate CVAR model developed, which incorporates

*Department of Mathematics and Statistics, University of New South Wales, Sydney, 2052, Australia, garethpeters@unsw.edu.au

†Baronia Capital Pty. Ltd., 12 Holtermann St., Crows Nest, NSW 2065, Australia

‡Department of Electrical Engineering, University of Cambridge, UK, sjg@eng.cam.ac.uk

 DOI:10.1214/11-BA628

a composite noise model for α-stable and matrix variate Gaussian errors, under both symmetric and non-symmetric α-stable assumptions.

Keywords: Cointegrated Vector Autoregression, α-stable, Approximate Bayesian Computation

1 Introduction

In this paper we consider estimation of Bayesian models for pairs trading strategies. Recent empirical studies by Bock and Mestel (2009) and Gatev et al. (2006) have shown that, in spite of the increasing volume of statistical arbitrage quantitative funds performing algorithmic trading, statistical pairs trading still seems to be consistently assessed as a profitable trading strategy. This provides motivation to further develop Bayesian cointegrated vector autoregression (CVAR) pairs trading models for practical financial applications.

CVAR models have been studied widely in the econometric literature, see Engle and Granger (1987) and Sugita (2009). For the error correction representation of a co-integrated series, see Granger and Weiss (2001), Strachan and Inder (2004) and the overview of Koop et al. (2006). Bayesian analysis of CVAR models has been addressed in several papers, see Bauwens and Lubrano (1994), Geweke (1996), Kleibergen and Van Dijk (2009), Ackert and Racine (1999), Strachan (2003), Sugita (2002) and Peters et al. (2010a). Typically CVAR models are fitted to low frequency data on time intervals of daily, monthly or yearly data. In this paper we explore their utility in higher frequency data modeling for pairs trading. In the process we develop a novel Bayesian model to overcome associated complications that arise when modeling on an intra-day sampling period. In particular we demonstrate that, when estimating matrix-variate parameters for CVAR models using data which is sampled at a frequency less than one day, the accuracy and robustness of the statistical model fit and portfolio weights estimation is strongly affected by level shifts or jumps in price series due to inter-day price movements.

In practice level shifts in price series occur as a result of the time delays between the open and close of markets for each asset in a traded pair. These can not solely be accounted for by the evolution of the statistical model during the time period in which either market is closed. Instead, these level shifts in each price series are a result of complicated economic and social market factors, we do not attempt to explain these with an economic rationale, see discussions in Granger and Hyung (2004), Wang and Zivot (2000) and Mills and Markellos (2008). Instead we analyze the consequences of failing to account for these level shifts in CVAR parameter estimation. To achieve this we develop a composite matrix-variate α-stable and Gaussian statistical model that allows us to account for such features in our high frequency data.

We model the level shifts in each price series via the class of α-stable models, see Zolotarev (1986), Samorodnitsky and Taqqu (1994), Qiou and Ravishanker (1998) and Nolan (1997). This class of models is of particular interest as they are flexible in terms of skew and kurtosis, whilst also admitting Gaussian distributions as a family member. The statistical challenge then lies in the development of an integrated matrix-variate CVAR model with standard Gaussian errors for intra-day observations and α-stable errors for inter-day observations capturing the price series level shifts. In this regard the matrix-variate model we develop, which incorporates under a unified likelihood model these two error structures, represents a novel contribution to CVAR model development. We then extend this result into a Bayesian CVAR framework. In doing so we consider two cases, the first involves the class of symmetric α-stable models which after a novel transformation of the data and representation of the matrix-variate α-stable errors that we develop, produce closed form matrix-variate conjugate posterior models. The second case involves a non-symmetric α-stable model which results in an intractable likelihood model which we tackle utilizing Approximate Bayesian Computation (ABC) methodology, see Peters et al. (2008), Tavaré et al. (1997), Fearnhead and Prangle (2010), Beaumont et al. (2009), Del Moral et al. (2011) and the review of Sisson and Fan (2010).

1.1 Contribution and Structure

The novelty of this paper springs from two aspects related to matrix-variate Bayesian model construction and the resulting Markov chain sampling and estimation frameworks. First we develop a novel model for Bayesian co-integration, incorporating a mixture of matrix-variate and matrix α-stable observation errors under an error correction model (ECM) framework. We consider two distinct cases, the symmetric and asymmetric α-stable models. In the symmetric model a non-standard version of the scale mixture of normals (SMiN) representation of the matrix variate α-stable model is developed.

We utilize this SMiN representation to derive a new result for representation of a matrix-variate likelihood of the CVAR model. This representation combines a Gaussian matrix variate CVAR likelihood with matrix-variate α-stable errors at inter-day time points, which are combined into a single matrix-variate Gaussian likelihood. To achieve this we developed a novel transformation that, when applied to the intra and inter-day price series, results in the transformed price series going from a composite model of matrix-variate α-stable and Gaussian errors, to being distributed according to a single matrix-variate Gaussian likelihood. We then derive all the results for the properties of the resulting combined matrix-variate Gaussian likelihood, including relevant moments and the uniqueness of the inverse transform in order to ensure estimation of the model parameters can be obtained in the original untransformed model. The advantage of this representation is that it allowed us to obtain a conjugate family of matrix-variate Bayesian models for the proposed CVAR model. We then sample this posterior via an adaptive matrix-variate Metropolis within Gibbs sampler.

In the non-symmetric α-stable setting the likelihood becomes analytically intractable to write down in closed form. Therefore we develop an ABC based estimation and sampling procedure for estimation of the resulting matrix-variate Bayesian model. In particular the Markov chain proposal which we develop for this matrix-variate model utilizes the Scale Mixture of Normals (SMiN) CVAR posterior model developed for the symmetric case. We demonstrate the utility of this proposal in several simulation studies and show that it works well in this ABC-MCMC sampling framework.

In the data analysis, we begin by studying the statistical properties of the inter-day level shifts in the differenced price series. These are obtained from the difference between the open and close prices of the times when both markets are trading for each pair of assets considered. The multivariate α-stable model is fitted to the intra-day price level shifts over a range of currency and index pairs, each for 30 contract segments dating back to 1999 on 10 minute sampled price data. This totals around 30,000 combined intra and inter-day observations for each pair. This provides us with a statistical model of the inter-day left shifts via generalized α-stable models for each asset pair. We demonstrate that in most cases the standard assumption of Gaussian residuals for these time periods is inadequate. In particular several assets demonstrate that significantly heavy tailed distributions are appropriate for capturing the inter-day price deviations resulting from these level shifts. This contradicts typical statistical assumptions of constant homoskedastic, multi-variate Gaussian innovation noise, made when fitting the basic CVAR models, widely utilized when trading pairs of assets.

Next we study the impact of naively applying the standard Johansen procedure (Johansen (1988)) and the Bayesian model of Peters et al. (2010a) to price series which contain these intra-day level shifts. We focus on the impact on the CVAR parameter estimates if one fails to adequately account for the level shifts present in the data. We demonstrate that estimation of the matrix-variate parameters of a CVAR model is adversely affected by level shifts in the price series, on both synthetic and actual data. In such situations, trading systems, utilizing such parameter estimates, will in turn be sensitive to the changes in parameter estimates, arising from the level shifts at day break boundaries. If this issue is not addressed, this could result in regular changes to portfolio allocations, resulting in additional transaction costs and other complications related to trade volumes. Therefore, in this paper we demonstrate that the underlying CVAR model will be a suitable model for the intra-day price series in which the parameter estimation can be made less sensitive through appropriately modeling the price level shifts in the inter-day prices at open and close of markets.

Finally, we demonstrate the ability of the Bayesian models we developed, which incorporate a novel formulation of a composite Gaussian and SMiN matrix variate α-stable model, to account for level shifts when performing posterior parameter estimation on several synthetic and real data studies.

2 Standard Gaussian CVAR Error Correction Model

We first briefly review the standard matrix-variate Gaussian model before extending to the matrix-variate α-stable setting. For the error correction representation of a co-integrated series, see Granger and Weiss (2001), Strachan and Inder (2004) and the overview of Koop et al. (2006). We denote the vector observation at time t by $\boldsymbol{x}_t$ and we assume $\boldsymbol{x}_t$ is an integrated of order 1, $(n \times 1)$-dimensional vector with r linear cointegrating relationships. The error vector at time t, $\boldsymbol{\epsilon}_t$ is assumed time independent and zero mean multivariate Gaussian distributed, with covariance Σ. The Error Correction Model (ECM) representation is given by,

$$\triangle \boldsymbol{x}_t = \boldsymbol{\mu} + \boldsymbol{\alpha}\boldsymbol{\beta}' \boldsymbol{x}_{t-1} + \sum_{i=1}^{p-1} \Psi_i \triangle \boldsymbol{x}_{t-i} + \boldsymbol{\epsilon}_t, \tag{1}$$

where $t = p, p{+}1, \ldots, T$ and p is the number of lags. Furthermore, the matrix dimensions are given by: $\boldsymbol{\mu}$ and $\boldsymbol{\epsilon}_t$ each $(n \times 1)$, Ψ_i and Σ each $(n \times n)$, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are each $(n \times r)$. We can now re-express the model in Equation (1) in a multivariate regression format, as follows

$$Y = X\Gamma + Z\boldsymbol{\beta}\boldsymbol{\alpha}' + E = WB + E, \tag{2}$$

where,

$$Y = \begin{pmatrix} \triangle\boldsymbol{x}_p & \triangle\boldsymbol{x}_{p+1} & \dots & \triangle\boldsymbol{x}_T \end{pmatrix}', Z = \begin{pmatrix} \boldsymbol{x}_{p-1} & \boldsymbol{x}_p & \dots & \boldsymbol{x}_{T-1} \end{pmatrix}'$$
$$E = \begin{pmatrix} \boldsymbol{\epsilon}_p & \boldsymbol{\epsilon}_{p+1} & \dots & \boldsymbol{\epsilon}_T \end{pmatrix}', \Gamma = \begin{pmatrix} \boldsymbol{\mu} & \Psi_1 & \dots & \Psi_{p-1} \end{pmatrix}'$$
$$X = \begin{pmatrix} 1 & \triangle\boldsymbol{x}'_{p-1} & \dots & \triangle\boldsymbol{x}'_1 \\ 1 & \triangle\boldsymbol{x}'_p & \dots & \triangle\boldsymbol{x}'_2 \\ \vdots & \vdots & \dots & \vdots \\ 1 & \triangle\boldsymbol{x}'_{T-1} & \dots & \triangle\boldsymbol{x}'_{T-p+1} \end{pmatrix}, W = \begin{pmatrix} X & Z\boldsymbol{\beta} \end{pmatrix}, B = \begin{pmatrix} \Gamma' & \boldsymbol{\alpha} \end{pmatrix}'.$$

Here, we let t be the number of rows of Y, hence $t = T - p + 1$, producing X with dimension $t \times (1 + n(p-1))$, Γ with dimension $((1 + n(p-1)) \times n)$, W with dimension $t \times k$ and B with dimension $(k \times n)$, where $k = 1 + n(p-1) + r$. The parameters μ represent the trend coefficients, and Ψ_i is the i^{th} matrix of autoregressive coefficients and the long run multiplier matrix is given by $\Pi = \boldsymbol{\alpha}\boldsymbol{\beta}'$.

The latter long run multiplier matrix is an important quantity of this model, its properties include: if Π is a zero matrix, $\boldsymbol{x}_t$ contains n unit roots; if Π has full rank, univariate series in $\boldsymbol{x}_t$ are trend-stationary; and co-integration occurs when Π is of rank $r < n$. The matrix $\boldsymbol{\beta}$ contains the co-integration vectors, reflecting the stationary long run relationships between the univariate series within $\boldsymbol{x}_t$ and the $\boldsymbol{\alpha}$ matrix contains the adjustment parameters, specifying the speed of adjustment to equilibria $\boldsymbol{\beta}'\boldsymbol{x}_t$.

According to Gupta and Nagar (1999) [Theorem 2.2.1] we see that if we have an $(n \times T)$ random matrix-variate Gaussian $Y' \sim N_{n,T}(M, \Sigma, \Psi)$ with row dependence captured in an $(n \times n)$ covariance matrix Σ and column dependence captured in a $(T \times T)$ matrix Ψ, then the vectorized form, in which the columns are stacked on top of each other to make an $nT \times 1$ random vector, is multivariate Gaussian and denoted by $Vec(Y) \sim N_{nT}(Vec(M), \Sigma \otimes \Psi)$. Here we denote the Kronecker or tensor product between two matrices by the $\otimes$ operator. This allows us to represent the matrix-variate likelihood for this regression, for the model parameters of interest B, Σ and $\boldsymbol{\beta}$, by

$$\begin{aligned} L(B, \Sigma, \boldsymbol{\beta}; Y) &\propto |\Sigma \otimes I_t|^{-0.5} \exp\left(-0.5Vec(Y - WB)'(\Sigma^{-1} \otimes I_t^{-1})Vec(Y - WB)\right) \\ &\propto |\Sigma|^{-0.5t} \exp\left(-0.5tr[\Sigma^{-1}(\hat{S} + R)]\right), \end{aligned} \tag{3}$$

where $\Sigma = Cov(\boldsymbol{\epsilon})$ and $R = (B - \hat{B})'W'W(B - \hat{B})$, $\hat{S} = (Y - W\hat{B})'(Y - W\hat{B})$, $\hat{B} = (W'W)^{-1}W'Y$.

3 Mixture matrix-variate α-stable and Gaussian CVAR

Noise modeling via α-stable distributions has been suggested in several areas, such as wireless communications and in financial data analysis, see Fama and Roll (1968), Godsill (2000), Neslehova et al. (2006) and Peters et al. (2010a). α-stable distributions possess several useful properties, including infinite mean and infinite variance, skewness and heavy tails, see Zolotarev (1986) and Samorodnitsky and Taqqu (1994). One can think of α-stable distributions as generalizations of the Gaussian distribution, which are defined as the class of location-scale distributions which are closed under convolutions. In this paper we focus on the S0 parameterization, see Peters et al. (2010a) for details.

The univariate α-stable distribution is typically specified by four parameters: $\alpha \in (0, 2]$ determining the rate of tail decay; $\beta \in [-1, 1]$ determining the degree and sign of asymmetry (skewness); $\gamma > 0$ the scale (under some parameterizations); and $\delta \in \mathbb{R}$ the location. The parameter α is termed the characteristic exponent, with small and large α implying heavy and light tails respectively. Gaussian ($\alpha = 2, \beta = 0$) and Cauchy ($\alpha = 1, \beta = 0$) distributions provide the only analytically tractable sub-members of this family. In general, as α-stable models admit no closed form expression for the density which can be evaluated point-wise (excepting Gaussian, Cauchy and Levy members), inference typically proceeds via the characteristic function, see discussions in Peters et al. (2010a). Though intractable to evaluate point-wise, simulation of random variates is very efficient, see Chambers et al. (1976). This observation is crucial to the ABC based approach we develop in Section 4.

The advantage of modeling the inter-day level shifts between the open and close of a market via an α-stable statistical model is that the CVAR model matrix-variate parameter estimation is improved substantially. This is demonstrated on real and synthetic data sets in Section 6. In addition, by considering an α-stable noise process for inter-day price shifts, we include as a special sub-case of our model the standard CVAR Gaussian models in Section 2.

The CVAR model we now consider incorporates a composite mixture of noise processes with $\boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \Sigma)$ for intra-day samples and $\boldsymbol{\epsilon}_t \sim \mathcal{S}_{\boldsymbol{a}}(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta})$ for inter-day observations. In this notation, the i-th asset has stable inter-day error model $\epsilon_t^{(i)} \sim \mathcal{S}_{a^{(i)}}(\beta^{(i)}, \gamma^{(i)}, \delta^{(i)})$. Here we utilize the notation of the majority of recent literature on α-stable models where $\mathcal{S}_{a^{(i)}}(\beta^{(i)}, \gamma^{(i)}, \delta^{(i)})$ is used for the class of stable laws, see Nolan (2012). Therefore, the resulting multivariate model we consider for innovation errors $\boldsymbol{\epsilon}_t$

at time t is given by dependent elements $\epsilon_t^{(i)}$,

$$\epsilon_t^{(i)} \sim N\left(0, \sigma^{(i)}\right) \mathbb{I}\left(t \notin \boldsymbol{\tau}\right) + \mathcal{S}_{a^{(i)}}\left(\beta^{(i)}, \gamma^{(i)}, \delta^{(i)}\right) \mathbb{I}\left(t \in \boldsymbol{\tau}\right), \tag{4}$$

where $\mathcal{S}_a\left(\beta, \gamma, \delta\right)$ denotes the α-stable distribution and $\boldsymbol{\tau}$ represents a vector of each of the first instants in time that both assets can be traded on their respective markets on each given day for the data series.

Under this α-stable composite model assumption we would like to utilize the innovation error structure provided in Equation (4) within a matrix variate Bayesian CVAR model. In the standard α-stable form, this model is intractable. We will first develop a non-standard SMiN representation for the matrix-variate CVAR likelihood. It will be specifically designed via a novel transformation to ensure that we can parameterize the model according to the same structure as presented in Section 2. The advantage of this is that we will then be able to derive conjugate models in our Bayesian framework. We achieve this by first considering a symmetric matrix variate representation, and then we generalize via ABC methods to the non-symmetric α-stable settings.

3.1 Matrix-variate SMiN CVAR Likelihood-model

In this section we develop a previously unpublished non-standard representation of a matrix-variate CVAR likelihood model. This involves utilizing a SMiN representation for the matrix-variate α-stable CVAR model inter-day errors combined with a Gaussian matrix-variate CVAR likelihood model for intra-day errors. This novel result is only achieved by our introduction of a specifically developed transformation of the intra and inter-day observation matrix. Under this transformation, we are able to demonstrate that we can combine these two likelihood models into a single Gaussian matrix-variate likelihood for the transformed data. In particular, the representation we develop is specifically designed to possess a covariance structure which will admit conjugacy in a Bayesian framework for the transformed parameters. In developing this transformation it is then important to clearly derive the properties of the resulting matrix-variate likelihood model.

Therefore, we then derive the properties of the resulting combined matrix-variate Gaussian likelihood, including relevant moments and the uniqueness of the inverse transform. This result for the inverse transform is important to ensure estimation of the model parameters can be obtained in the original untransformed model. The advantage of this representation is that it allowed us to obtain a conjugate family of matrix-variate

Bayesian models for the proposed CVAR model. We note that without the transformation we develop here in our non-standard SMiN representation it would not be possible to obtain the likelihood structure required for conjugacy.

Hence, summarizing this section we present Lemma 1 and Lemma 2 which are combined with Theorem 1 to demonstrate that under our specifically designed transformation for the vectorized matrix of observations, we can obtain a joint matrix-variate likelihood for the α-stable and Gaussian innovations mixture model. Next, we solve explicitly for the covariance matrix under this transformed representation. Lemma 3 and Theorem 2 then derive the form of the mean matrix for this matrix-variate likelihood, via a tensor product identity on vectorized transformed data. To achieve this we consider a special form of non-negative tensor factorization of our transformation matrix. Additionally we prove that the solution to the mean structure parameter matrix in the transformed model can be uniquely recovered under the transformation developed. Throughout this section we assume a lag $p = 1$, this can be extended trivially.

When the noise model in Equation (4) is strictly symmetric, i.e. the α-stable interday noise model is symmetric, it admits an exact SMiN representation given in Equation 5, see Godsill (2000),

$$\epsilon_t^{(i)} \sim N\left(0, \sigma^{(i)}\right)\mathbb{I}\left(t \notin \boldsymbol{\tau}\right) + N\left(\delta^{(i)}, \gamma^{(i)}\lambda^{(i)}\right)\mathbb{I}\left(t \in \boldsymbol{\tau}\right), \tag{5}$$

with auxiliary scale variables distributed as $\lambda^{(i)} \sim \mathcal{S}_{a^{(i)}/2}\left(0, 1, 1\right)$.

We denote by $\widetilde{Y}$ the matrix of observation differenced price vectors corresponding to intra-day prices with a total of $\widetilde{t}$ rows and $\widetilde{W}$ is the corresponding matrix for $\widetilde{Y}$, defined in Section 2. The definition of $W_{(T-\widetilde{t})}$ is the matrix for W corresponding to the observation vectors taken from the set of intra-day times when $t \in \boldsymbol{\tau}$. The vectors $\boldsymbol{\lambda} = \left[\widetilde{\lambda}^1\gamma^1, \ldots, \widetilde{\lambda}^n\gamma^n\right]$ are the scale parameters in the SMiN representation and $\boldsymbol{D_\lambda}$ is a diagonal matrix with each value of $\boldsymbol{\lambda}$ in the diagonal.

Lemma 1. *The combined grouped vectorized likelihood for intra-day and inter-day observation price vectors denoted* $Vec(Y_*) \sim N_{nT}\left(Vec(M_*), \Sigma_* \otimes \Psi_*\right)$ *is given by:*

$$\begin{aligned}
&L(\Sigma, B, \boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\delta}; Y_*) \\
&\quad = (2\pi)^{-0.5nT}|\Sigma_* \otimes \Psi_*|^{-0.5}\exp\left(-0.5Vec(Y_* - M_*)'(\Sigma_*^{-1} \otimes \Psi_*^{-1})Vec(Y_* - M_*)\right) \\
&\quad \propto |\Sigma_*|^{-0.5T}|\Psi_*|^{-0.5n}\exp\left(-0.5tr\left\{\Sigma_*^{-1}(Y_* - D_* - W_*B)'\Psi_*^{-1}(Y_* - D_* - W_*B)\right\}\right),
\end{aligned} \tag{6}$$

where we have ordered the intra and inter-day observation vectors according to

$$Y_* = y_{1:T} = [\boldsymbol{y}_1\, \boldsymbol{y}_2\, \ldots\, \boldsymbol{y}_{\tau_1 - 1}\, \boldsymbol{y}_{\tau_1 + 1}\, \ldots\, \boldsymbol{y}_T\, \boldsymbol{y}_{\tau_1}\, \boldsymbol{y}_{\tau_2}\, \ldots\, \boldsymbol{y}_{\tau_{i_D}}]',$$

and there are a total of i_D inter-day boundaries in the series. The relevant likelihood matrices are,

$$D_* = \begin{pmatrix} \mathbf{0} \\ \mathbf{1}_{i_D}\boldsymbol{\delta}^T \end{pmatrix}, W_* = \begin{pmatrix} \widetilde{W} \\ W_{(T-\widetilde{t})} \end{pmatrix}.$$

We consider a general covariance matrix structure $\Sigma_* \otimes \Psi_*$ (for Σ_* an $n \times n$ matrix and Ψ_* a $T \times T$ matrix)

Proof. The result in Lemma 1 follows by first utilizing the assumption of conditional independence of the observation vectors given model parameter matrices $\Sigma, B, \boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\delta}$ which states $\mathbb{E}[\boldsymbol{y}_s, \boldsymbol{y}_t] = \mathbb{E}[\boldsymbol{y}_s]\,\mathbb{E}[\boldsymbol{y}_t]$ for all $s, t\; s \neq t$. Followed by application of the theorems in Gupta and Nagar (1999) ([Theorem 2.2.1], [Theorem 2.3.11]) and the trace identity and determinant identities of Gupta and Nagar (1999) [Theorem 1.2.21 (v and x)] are applied. □

Remark 1: *To relate the matrix-variate Gaussian model in Lemma 1 to the original likelihood model in Equation (3) we need to find the relationship between the covariance matrices, Σ_*, Ψ_* and the original model parameters Σ, I_t. Under this reordered and repacked matrix-variate Gaussian, the independence of columns of the random matrix in the model in Equation (3) no longer holds, that is Ψ_* is only diagonal when $\boldsymbol{D_\lambda} = \Sigma$.*

Once we have developed the combined matrix-variate likelihood model for intra and inter-day observation vectors, we would like to exploit possible conjugacies. Conjugacy of the posterior for the standard matrix-variate Gaussian CVAR model is beneficial for inference and sampling. To achieve this under our α-stable mixture would require us to identify the sufficient statistics, (M_*, Σ_*, Ψ_*), for the grouped matrix-variate Gaussian model in Lemma 1, as $\Sigma_* = \Sigma$ and Ψ_* diagonal, conditional on parameters from the fitted α-stable SMiN intra-day noise model. Lemmas 2, 3 and Theorems 1 and 2 presented next allow us to identify the sufficient statistics and then transform the vectorized random observation matrix Y_* to recover required conjugacy properties.

Importantly, this will provide a significant dimension reduction in the posterior parameter space. Since, it allows us to specify a matrix-variate prior only on a matrix Σ_* which is $n \times n$ rather than on a multivariate covariance which is $nT \times nT$.

Lemma 2. *The mean and covariance of the vectorized observation matrix $Vec(Y_*)$ in terms of the original CVAR model matrices are given by,*

$$Cov(Vec(Y_*)) = \Sigma_* \otimes \Psi_* = \begin{pmatrix} \Sigma \otimes I_{\widetilde{t}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{D_\lambda} \otimes I_{(T-\widetilde{t})} \end{pmatrix}.$$

In addition we can obtain the covariance of $Vec(Y'_*)$ as

$$Cov(Vec(Y'_*)) = \Psi_* \otimes \Sigma_* = \begin{pmatrix} I_{\widetilde{t}} \otimes \Sigma & \mathbf{0} \\ \mathbf{0} & I_{(T-\widetilde{t})} \otimes \boldsymbol{D_\lambda} \end{pmatrix}.$$

Proof. Using [Definition 2.2.1] and [Theorem 2.2.1] of Gupta and Nagar (1999), the random vector $Vec(Y_*)$ is conditionally a multivariate Gaussian random vector of dimension $nT \times 1$. Using Lemma 1 and the SMiN CVAR model assumption of conditionally independent, but not identically distributed, Gaussian observation random vectors we can explicitly identify the mean and covariance structure of the vectorized observation matrix $Vec(Y_*)$ in terms of the original CVAR model matrices. □

Having identified the covariance structure for the vectorized reordered observation matrix, next we present Theorem 1 producing a likelihood structure that admits conjugacy under the priors presented in Section 4.1.

Theorem 1. *The transformed random vector denoted $Vec(Z_*) = Q_* Vec(Y_*)$ is multivariate Gaussian with $Vec(Z_*) = Q_* Vec(Y_*) \sim N(Q_* Vec(M_*), Q_*^T(\Sigma_* \otimes \Psi_*)Q_*)$. Under the specially designed transformation selected as,*

$$Q_* = \begin{pmatrix} I_n \otimes I_{\widetilde{t}} & \mathbf{0} \\ \mathbf{0} & Q \otimes I_{(T-\widetilde{t})}, \end{pmatrix},$$

one obtains $Z_* \sim N_{n,T}(\mu_*, \Sigma, I_T)$. In addition we can define

$$Q_* = \begin{pmatrix} I_{\widetilde{t}} \otimes I_n & \mathbf{0} \\ \mathbf{0} & I_{(T-\widetilde{t})} \otimes Q \end{pmatrix},$$

such that when it is used to transform $Q_* Vec(Y'_*)$ we obtain $Q_* Vec(Y'_*) \sim N_{n,T}(\mu_*, I_T, \Sigma)$ and we also have that $Z'_* = Q_* Vec(Y'_*)$.

Proof. Apply Lemma 1 and Lemma 2 which gives an $(nT \times 1)$ random vector $Vec(Y_*)$ conditionally distributed according to a multivariate Gaussian distribution, under a transformation by an $nT \times nT$ matrix Q_* to obtain $Vec(Z_*) = Q_* Vec(Y_*)$ which is also multivariate Gaussian. Using Luetkepohl (2005) [Proposition B.2] we obtain, for $Vec(Y_*) \sim N\left(Vec(M_*), \Sigma_* \otimes \Psi_*\right)$, a transformed random vector $Vec(Z_*) = Q_* Vec(Y_*) \sim N(Q_* Vec(M_*), Q_*^T(\Sigma_* \otimes \Psi_*)Q_*)$. Next, to prove the covariance structure of the transformed random vector under this particular transformation consider the new covariance structure for $Vec(Z_*)$ which will be given by,

$$
\begin{aligned}
&Cov(Vec(Z_*))\\
&= \begin{pmatrix} I_n \otimes I_{\widetilde{t}} & \mathbf{0} \\ \mathbf{0} & Q \otimes I_{(T-\widetilde{t})} \end{pmatrix}^T \begin{pmatrix} \Sigma \otimes I_{\widetilde{t}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{D_\lambda} \otimes I_{(T-\widetilde{t})} \end{pmatrix} \begin{pmatrix} I_n \otimes I_{\widetilde{t}} & \mathbf{0} \\ \mathbf{0} & Q \otimes I_{(T-\widetilde{t})} \end{pmatrix}\\
&= \begin{pmatrix} \Sigma \otimes I_{\widetilde{t}} & \mathbf{0} \\ \mathbf{0} & \left(Q^T \boldsymbol{D_\lambda} Q \otimes I_{(T-\widetilde{t})}\right) \end{pmatrix}.
\end{aligned}
$$

We can therefore obtain $Cov(Vec(Z_*)) = \Sigma \otimes I_T$ by solving the equation $Q^T \boldsymbol{D_\lambda} Q = \Sigma$ for matrix Q. We can make use of the fact that the $n \times n$ matrix $\boldsymbol{D_\lambda}$ is diagonal and the covariance matrix Σ is real and symmetric with an eigen decomposition $\Sigma = VFV^T$ with diagonal eigen values matrix F. Therefore if we select $Q = S^{\frac{1}{2}}U^T$ where $S^{\frac{1}{2}}$ is the diagonal matrix with the elements $S_{ii} = \sqrt{\frac{F_{ii}}{\boldsymbol{D_{\lambda,ii}}}}$ then the matrix U is the orthonormal matrix of eigen vectors for Σ, that is $U = V$. The proof for the transformation Q_* of $Vec(Y_*')$ follows trivially from this result. $\square$

Hence, we have transformed the observation vector $Vec(Y_*)$ via matrix Q_* to obtain a new random vector, which, when un-vectorized, produces a matrix-variate Gaussian with row dependence given by Σ and column dependence given by I_T. The significance of this new result is that it allows us to recover the conditional independence property of each vector observation whilst identifying under the transformation the identity $\Sigma_* = \Sigma$ and $\Psi_* = I_T$. Therefore the matrix-variate likelihood for transformed observations $\boldsymbol{z}_{1:T}$ is given by Lemma 3.

Lemma 3. *The likelihood of the transformed observations is given by,*

$$
\begin{aligned}
&L(\Sigma, B, \boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\delta}, Q; \boldsymbol{z}_{1:T})\\
&\quad \propto |\Sigma_* \otimes I_T|^{-0.5} \exp\left(-0.5\left(Vec(Z_*) - Q_* Vec(D_* - W_* B)\right)' \left(\Sigma_*^{-1} \otimes I_T^{-1}\right)\right.\\
&\quad\quad \left.\times \left(Vec(Z_*) - Q_* Vec(D_* - W_* B)\right)\right)\\
&\quad = |\Sigma_*|^{-0.5T} \exp\left(-\frac{1}{2} tr\left\{\Sigma_*^{-1}\left(\widehat{\widetilde{S}}_* + (\widetilde{B} - \widehat{\widetilde{B}}_*)'\widetilde{W}_*'\widetilde{W}_*(\widetilde{B} - \widehat{\widetilde{B}}_*)\right)\right\}\right),
\end{aligned}
$$

where we define $\widetilde{D}_* = HD_*G^T$, $\widetilde{W}_* = HW_*$, $\widetilde{B} = BG^T$ *and* $\widehat{\widetilde{B}}_* = \left(\widetilde{W}_*'\widetilde{W}_*\right)^{-1}\widetilde{W}_*'(Z_* - \widetilde{D}_*)$ *and* $\widehat{\widetilde{S}}_* = \left(Z_* - \widetilde{D}_* - \widetilde{W}_*\widehat{\widetilde{B}}_*\right)'\left(Z_* - \widetilde{D}_* - \widetilde{W}_*\widehat{\widetilde{B}}_*\right)$.

Proof. Utilize [Definition 2.2.1] and [Theorem 2.2.1] of Gupta and Nagar (1999), to obtain the likelihood of the transformed observations according to

$$
\begin{aligned}
&L(\Sigma, B, \boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\delta}, Q; \boldsymbol{z}_{1:T}) \\
&\quad \propto |\Sigma_* \otimes I_T|^{-0.5} \exp\left(-0.5\,(Vec(Z_*) - Q_* Vec(D_* - W_* B))'\,(\Sigma_*^{-1} \otimes I_T^{-1})\right. \\
&\qquad \left. \times (Vec(Z_*) - Q_* Vec(D_* - W_* B))\right).
\end{aligned}
$$

Then apply the identity in [Theorem 1.2.22] of Gupta and Nagar (1999), given by

$$(B' \otimes A)\, Vec(X) = Vec(AXB), \tag{7}$$

and rearrange the mean structure. Next, make an arbitrary choice of factorization of Q_* into the form $Q_* = G \otimes H$ with the only constraints that G is $(p \times n)$ and that H is $(q \times T)$ dimensions, with $pq = nT$. Hence, the rearranged mean structure gives,

$$
\begin{aligned}
&L(\Sigma, B, \boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\delta}, Q; \boldsymbol{z}_{1:T}) \\
&\quad \propto |\Sigma_* \otimes I_T|^{-0.5} \exp\left(-0.5\,(Vec(Z_*) - Q_* Vec(D_* - W_* B))'\,(\Sigma_*^{-1} \otimes I_T^{-1})\right. \\
&\quad \left. \times (Vec(Z_*) - Q_* Vec(D_* - W_* B))\right) \\
&\quad \propto |\Sigma_* \otimes I_T|^{-0.5} \exp\left(-0.5\left(Vec(Z_*) - Vec(\widetilde{D}_* - \widetilde{W}_* \widetilde{B})\right)'(\Sigma_*^{-1} \otimes I_T^{-1})\left(Vec(\widetilde{D}_* - \widetilde{W}_* \widetilde{B})\right)\right),
\end{aligned}
$$

with $\widetilde{D}_* = H D_* G^T$, $\widetilde{W}_* = H W_*$ and $\widetilde{B} = B G^T$. $\square$

We can now comment on the possible solutions to the tensor factorization utilized for $Q_* = G \otimes H$. Typically the basic Singular Value Decomposition is applied to perform a tensor factorization. This will be difficult in our setting as we are required to enforce the sub-matrix constraints that the first factored matrix must be $(p \times n)$ with n columns and the second $q \times T$ with T columns. There is a rich literature on alternative tensor factorizations such as the numerical algorithms for rank-k tensor approximations such as the orthogonal tensor decompositions (Higher-Order SVD) of De Lathauwer and Vandewalle (2004) or the Non-Negative Tensor Factorization (NTF) in Friedlandera and Hatzb (2008).

In Theorem 2 we provide a specific tensor factorization to satisfy the constraints required by the result in Lemma 3. It is important to obtain a specific factorization that decomposes the transformation matrix into a tensor factorization admitting a unique solution for the original mean estimate B.

Theorem 2. *Given transformed observations, Z_*', an analytic tensor factorization for Q_* satisfying the dimensionality constraints on each tensor factor in Lemma 3 is* *given by,*

$$Q_* = \sum_{i=1}^{T} \sum_{j=1}^{T} U_{ij} \otimes Q_{ij},$$

where Q_{ij} represent the (i,j)-th sub-block of dimension $n \times n$ in the $nT \times nT$ transform matrix Q_* and U_{ij} represent the $(T \times T)$ matrix whose ij-th element is 1 and whose remaining elements are 0. Therefore the mean structure of the CVAR likelihood model in Lemma 3 is given by,

$$\begin{aligned}\mathbb{E}\left[Vec(Z_*')\right] &= Q_* Vec(D_*' - B'W_*') \\ &= \sum_{i=1}^{T} Vec(Q_{ii}(D_*' - B'W_*')U_{ii}').\end{aligned}$$

This allows us to make explicit the mean structure of the matrix-variate transformed data likelihood of Lemma 3 by identifying the following elements $\widetilde{D}_*' = \sum_{i=1}^{T} Q_{ii} D_*' U_{ii}'$, $\widetilde{W}_*' = \sum_{i=1}^{T} W_*' U_{ii}$ and $\widetilde{B}' = \sum_{i=1}^{T} Q_{ii} B'$.

Proof. Using identity [(1.29) p. 343] in Harville (2008) we can exploit the fact that the transformation matrix Q_* we have selected is a square $nT \times nT$ matrix which has an $n \times n$ block diagonal structure. Hence we will consider the following structure in Q_*,

$$\left(\begin{array}{cccc} Q_{11} & Q_{12} & \cdots & Q_{1T} \\ \vdots & \vdots & & \vdots \\ Q_{T1} & Q_{T2} & \cdots & Q_{TT} \end{array} \right)^T,$$

with each sub matrix Q_{ij} being selected as an $(n \times n)$ matrix. Then obtain the tensor factorization, using the fact that all Q_{ij} matrices will be comprised of 0 elements other than those with $i = j$ giving a sparse representation

$$Q_* = \sum_{i=1}^{T} \sum_{j=1}^{T} U_{ij} \otimes Q_{ij} = \sum_{j=1}^{T} U_{ii} \otimes Q_{ii}.$$

Under this factorization the mean structure we obtain in the likelihood model in Theorem 1 with application of the identity in [Theorem 1.2.22] of Gupta and Nagar (1999) shown in Equation (7), is given by,

$$\begin{aligned}\mathbb{E}\left[Q_* Vec(Y_*')\right] &= Q_* Vec(D_*' - B'W_*') \\ &= \sum_{i=1}^{T}\sum_{j=1}^{T}\left(U_{ij}\otimes Q_{ij}\right)Vec(D_*' - B'W_*') \\ &= \sum_{i=1}^{T} Vec(Q_{ii}D_*'U_{ii}' - Q_{ii}B'W_*'U_{ii}')).\end{aligned}$$

This allows us to make explicit the mean structure of the matrix-variate transformed data likelihood by identifying the following elements $\widetilde{D}_*' = \sum_{i=1}^{T} Q_{ii}D_*'U_{ii}'$, $\widetilde{W}_*' = \sum_{i=1}^{T} W_*'U_{ii}'$ and $\widetilde{B}' = \sum_{i=1}^{T} QiiB'$.
Finally, we note that we can uniquely solve the system

$$\widetilde{B}' = \sum_{i=1}^{T} Q_{ii}B',$$

for B' given $\widetilde{B}'$. This is due to the fact that the matrices Q_{ii} for $i < T$ are constructed from identity matrices and the case of $i = T$ is constructed in our transform as a real matrix of eigen vectors of covariance matrix Σ, which is therefore invertible. We can therefore obtain the unique solution for B' as

$$B' = \widetilde{B}'\left((T-1)I_n + Q_{TT}\right)^{-1}.$$

□

Hence, we have shown that this particular choice of factorization for Q_* ensures that a unique solution to B' is attainable given $\widetilde{B}$. This result is important for the conjugate Bayesian model derivation developed in Section 4.

4 Bayesian CVAR and Approximate Bayesian Computation

In this section we consider the composite noise model developed in Section 3 and derive Bayesian models in the presence of the fitted α-stable inter-day noise. In general, conjugacy is lost for the general asymmetric noise model in Equation (4). In these cases we resort to ABC methodology. However, in the case of symmetric α-stable noise we derive two novel conjugate models under the transformation Q_*. This is achieved utilizing the scale mixture of Normals (SMiN) representation developed in Section 3.1.

The two conjugate SMiN models are based around two classes of prior structure considered in the econometrics literature. The choice of prior in CVAR models is an important consideration, see discussions in Strachan and Inder (2004) and Kleibergen and Paap (2002). We do not consider in detail the issue of prior distortions illustrated by Kleibergen and Van Dijk (2009). Instead we select two popular prior choices from the literature. The first is based on a prior specification on the cointegration space, rather than the actual parameters, as proposed in Strachan and Inder (2004). The second prior choice is based on linear identification constraints on the cointegration vectors developed in Geweke (1996) and utilized in Peters et al. (2010a).

4.1 Conjugate CVAR models via a symmetric α-stable SMiN representation

The prior model which we consider is hierarchical and produces conjugate posterior distributions for matrix-variate parameters Σ and B.

- $\Sigma \sim IW(S,h)$ where $IW(S,h)$ is the Inverse Wishart distribution with h degrees of freedom and S is an $(n \times n)$ positive definite matrix.
- $B'|\Sigma \sim N(P', \Sigma \otimes A^{-1})$ where $N(P, \Sigma \otimes A^{-1})$ is the matrix-variate Gaussian distribution with h degrees of freedom and S is an $(n \times n)$ positive definite matrix.

The first prior we consider for the cointegration matrix β is based on the proposal developed in Villani (2005). As discussed in Strachan and Inder (2004) there are several approaches to prior specifications in CVAR models, either directly on the cointegration vector parameters $\boldsymbol{\beta}$ or on the cointegration space, which produces an induced prior on the parameters. A uniform prior on the cointegration space Grassman manifold is proposed in Villani (2005) as relevant when considering models conditional on knowledge of the rank. We follow the uniform prior specification on the Grassman manifold given in Lemma 3.4 of Villani (2005), which results in a prior on the parameters given by

- $\boldsymbol{\beta}' \sim t_{d-r\times r}(0, \mathbb{I}_{d-r}, \mathbb{I}_r, 1)$ where this $t_{d-r\times r}$ represents a matrix-variate t-distribution as defined in Definition 4.2.1 of Gupta and Nagar (1999).

The second prior we consider is based on the approach adopted in Geweke (1996) and is identical to the choice of Geweke (1996), Peters et al. (2010a) and Sugita (2002),

- $\boldsymbol{\beta}' \sim N(\bar{\boldsymbol{\beta}}', Q \otimes H^{-1})$ where $N(\bar{\boldsymbol{\beta}}, Q \otimes H^{-1})$ is the matrix-variate Gaussian distribution with prior mean $\bar{\boldsymbol{\beta}}$, Q is an $(r \times r)$ positive definite matrix, H an $(n \times n)$ matrix.

In this case we make identical model assumptions and restrictions for the Bayesian CVAR model as in Peters et al. (2010a). In particular, for any non-singular matrix A, the matrix of long run multipliers $\Pi = \boldsymbol{\alpha}\boldsymbol{\beta}'$ is indistinguishable from $\Pi = \boldsymbol{\alpha} A A^{-1} \boldsymbol{\beta}'$, see Koop et al. (2006). As proposed in Sugita (2002), we remove this problem by incorporating an identification constraint which imposes the required r^2 restrictions as follows $\boldsymbol{\beta} = [I_r, \boldsymbol{\beta}_*']'$, where I_r denotes the $r \times r$ identity matrix. The choice of possible constraints is not unique, though the choice we select is computationally convenient, see discussions in Kleibergen and Van Dijk (2009).

Here we derive the conjugate model for the matrix-variate parameters of the posterior under the transformation Q_* developed in Theorem 1. By working under this transformation, we not only obtain conjugacy but also reduce the posterior dimension significantly from a parameterization of the posterior covariance matrix in dimension $nT \times nT$ to parameterization in $n \times n$ dimensions.

Given parameter estimates of the multivariate α-stable statistical model, $\mathcal{S}_{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta})$, fitted to historical price series inter-day level shifts for each asset in the CVAR model, the following posterior conjugacy properties are satisfied for the prior choices presented.

1. Σ **Conditional:** Conditional on the re-arranged un-transformed subset of observation vectors from intra-day prices matrix $\widetilde{Y}$ we obtain an Inverse Wishart distribution for
$$p(\Sigma|\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\alpha}, \widetilde{Y}) \propto |S_{\widetilde{Y}}|^{(t+h)/2} |\Sigma|^{-(t+h+n+1)/2} \exp\left(-0.5 tr(\Sigma^{-1} S_{\widetilde{Y}})\right);$$
where $S_{\widetilde{Y}}$ is defined to be given by
$$S_{\widetilde{Y}} = S + \widehat{S} + (P - \widehat{B})' \left[A^{-1} + (W'W)^{-1}\right]^{-1} \left(P - \widehat{B}\right).$$

2. $\widetilde{B}$ **Conditional:** Under the SMiN model and conditional on the re-arranged transformed complete vector of observations for intra and inter-days, $Vec(Z_*) = Q_* Vec(Y_*)$ we obtain a Matrix-variate Gaussian for
$$p(\widetilde{B}|\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\alpha}, \Sigma, Z_*, Q_*) \propto |A_{Z_*}|^{n/2} |\Sigma|^{-k/2} \exp\left(-0.5 tr\left(\Sigma^{-1} (\widetilde{B} - B_{Z_*})' A_\star (\widetilde{B} - B_{Z_*})\right)\right)$$
where $A_{Z_*} = \widetilde{A} + \widetilde{W}_*' \widetilde{W}_*$ and $B_{Z_*} = \left(\widetilde{A} + \widetilde{W}_*' \widetilde{W}_*\right)^{-1} \left(\widetilde{A}\widetilde{P} + \widetilde{W}_*' \widetilde{W}_* \widehat{\widetilde{B}}_*\right)$.

3. $\boldsymbol{\beta}$ **Conditional:** Under the SMiN model and conditional on the re-arranged transformed complete vector of observations for intra and inter-days, $Vec(Z_*) = Q_* Vec(Y_*)$ we obtain the marginal matrix-variate posterior for the cointegration vectors, $\boldsymbol{\beta}$ given by,

$$p(\boldsymbol{\beta}|\boldsymbol{\lambda}, \boldsymbol{\alpha}, Z_*, Q_*) \propto p(\boldsymbol{\beta})|S_{Z_*}|^{-(t+h+1)/2}|A_{Z_*}|^{-n/2},$$

for

$$S_{Z_*} = S + \widehat{\widehat{S}}_* + (P - \widehat{\widehat{B}}_*)' \left[\widetilde{A}^{-1} + (\widetilde{W}_*'\widetilde{W}_*)^{-1}\right]^{-1} \left(P - \widehat{\widehat{B}}_*\right),$$

and A_{Z_*} defined in Conditional 2. The choice of prior for the cointegration vectors, given either by Geweke (1996) or Strachan and Inder (2004) can then be substituted for $p(\boldsymbol{\beta})$.

4. λ **Conditional:** Under the SMiN model we obtain the marginal distribution for each random variable λ^i in the $n \times 1$ random vector $\boldsymbol{\lambda}$ given by,

$$p(\lambda_i|\boldsymbol{\alpha}, \chi, \widetilde{B}, Q_*, \boldsymbol{\beta}) \propto \prod_{t \in \tau} N\left(\epsilon_t^i; 0, \lambda_i \gamma_i\right) \times S_{a_i/2}\left(\lambda_i; 0, 1, 1\right),$$

where for all $t \in \tau$ we define $\epsilon_t^i = \chi_{i,t} - \left[W_{(T-\widetilde{t})}B\right]_{i,t}$ and $\boldsymbol{\chi} = Y_{-\widetilde{Y}} - \mathbf{1}_k \boldsymbol{\delta}^T$ is the inter-day observation matrix of differenced price vectors not including rows for $\widetilde{Y}$ after subtracting the location parameters for each α-stable fit, given by $\boldsymbol{\delta} = \left[\delta^{(1)}, \ldots, \delta^{(n)}\right]'$.

The conjugacy for Conditional 1 and Conditional 2 are provided in Sugita (2002) [Section 2.2, Equations (10) and (11)] as a direct consequence of Theorem 1 and Theorem 2 and the transformation developed and conjugate prior choices. The derivation of Conditional 3 also follows from Sugita (2002) [Section 2.2, Equation (14)]. The proof for Conditional 4 is presented in Godsill (2000) [Section 2 Equation (4)].

4.2 Asymmetric α-stable Approximate Bayesian Computation CVAR

When considering the asymmetric α-stable noise model, presented in Equation (4), we have an intractable matrix-variate likelihood, which can not be evaluated pointwise even up to a normalizing constant. To overcome this problem we formulate a novel approximate Bayesian computation (ABC) solution. ABC modeling is a new class of statistical estimation techniques specifically designed for situations in which the likelihood and thus the posterior distribution is intractable. These have now been studied and applied in a range of settings, see Peters et al. (2010c) and Peters and Sisson (2006) for ABC modeling for financial risk and insurance contexts. In addition, there are now several methodological papers and reviews available for this new class of modeling technique, see Peters et al. (2008), Tavaré et al. (1997), Fearnhead and Prangle (2010), Beaumont et al. (2009), Del Moral et al. (2011) and the review of Sisson and Fan (2010).

In this section we develop an ABC model and associated Markov chain Monte Carlo (MCMC-ABC) sampler to perform estimation in this general composite asymmetric α-stable and Gaussian noise CVAR model setting. MCMC-ABC samplers are actively studied in the statistical literature since Tavaré et al. (1997), see a review chapter in Sisson and Fan (2010).

The notation we will adopt to represent our ABC approximation is that developed in the recent book chapter [Section 1.2.1] of Sisson and Fan (2010). ABC inference adopts the approach of augmenting the target posterior distribution from the intractable "True" model, denoted $p(\Sigma, B, \boldsymbol{\beta}|Y) \propto p(Y|\Sigma, B, \boldsymbol{\beta})p(\Sigma, B, \boldsymbol{\beta})$, into an augmented parameter target posterior distribution. The ABC posterior model approximation, denoted $p_{ABC}(\Sigma, B, \boldsymbol{\beta}, Y_S|Y)$, is defined according to the representation of Reeves and Pettitt (2005) and Wilkinson (2008) by,

$$p_{ABC}(\Sigma, B, \boldsymbol{\beta}, Y_S|Y) = p(Y|Y_S, \Sigma, B, \boldsymbol{\beta})p(Y_S|\Sigma, B, \boldsymbol{\beta})p(\Sigma, B, \boldsymbol{\beta}), \tag{8}$$

where the auxiliary parameters "synthetic observation" matrix Y_S are a (simulated) dataset or set of data sets from $p(Y|\Sigma, B, \boldsymbol{\beta})$, on the same space as Y.

The function $p(Y|Y_S, \Sigma, B, \boldsymbol{\beta})$ is chosen to weight the posterior $p(\Sigma, B, \boldsymbol{\beta}|Y)$ with high values in regions where Y_S and Y are similar. There are many choices for this function, discussed and studied in Peters et al. (2010c), Sisson and Fan (2010) and Grelaud et al. (2009). Generally, the weighting function $p(Y|Y_S, \Sigma, B, \boldsymbol{\beta})$ is simplified in two important ways, the first involves replacing the observation and synthetic data vector / matrix with summary statistics and the second involves making a kernel approximation to the weighting function. Therefore we obtain a kernel representation of the form

$$p_\epsilon(Y|Y_S, \Sigma, B, \boldsymbol{\beta}) = \frac{1}{\epsilon} K\left(\frac{|\boldsymbol{S}(Y_S) - \boldsymbol{S}(Y)|}{\epsilon}\right), \tag{9}$$

see Peters et al. (2010c), Ratmann et al. (2009) and Beaumont et al. (2009). In this simplification the data matrix Y is replaced with summary statistics (ideally sufficient statistics) vector or matrix denoted $\boldsymbol{S}(Y)$ of significantly lower dimension than Y. When sufficient statistics are not available, then summary statistics are utilized at the cost of bias (as $\epsilon \to 0$), see recent discussion in Fearnhead and Prangle (2010).

In this paper we consider a popular kernel weighting function (uniform kernel) with Euclidean distance measure between summary statistics on vectorized observation matrices $Vec(Y)$ and $Vec(Y_S)$ given by

$$p_\epsilon(Y|Y_S, \Sigma, B, \boldsymbol{\beta}) = \begin{cases} 1 & \text{if } ||\boldsymbol{S}(Vec(Y)) - \boldsymbol{S}(Vec(Y_S))|| \leq \epsilon \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

This kernel has been used successfully in several ABC studies, see discussions in Peters et al. (2010b), Toni et al. (2009), Sisson and Fan (2010) and Fearnhead and Prangle (2010). We note that the ABC approximation we develop can utilize any choice of kernel.

Finally, given this ABC approximation, it is a natural statistical question to ask in what sense is ABC approximating the intractable target posterior distribution. This has been studied in several different contexts, see Beaumont et al. (2009), Del Moral et al. (2011), Peters et al. (2008) and Fearnhead and Prangle (2010) for details.

Here we briefly summarize, for our ABC approximation, the discussion presented in Sisson and Fan (2010) which describes the basic relationship between the ABC estimate and the intractable posterior. When one assumes that the weighting function is constant at the point $Y_S = Y$ with respect to parameters $\Sigma, B, \boldsymbol{\beta}$ this results in $p(Y|Y_S, \Sigma, B, \boldsymbol{\beta}) = c$, for some constant $c > 0$. The result of this is that the target posterior is recovered exactly at $Y_S = Y$, that is $p_{ABC}\left(\Sigma, B, \boldsymbol{\beta}, Y_S|Y\right) = p(\Sigma, B, \boldsymbol{\beta}|Y)$.

We also mention that, given the augmented ABC posterior distribution $p_{ABC}\left(\Sigma, B, \boldsymbol{\beta}, Y_S|Y\right)$ generally inference involves the marginal posterior,

$$p\left(\Sigma, B, \boldsymbol{\beta}|Y\right) \propto p(\Sigma, B, \boldsymbol{\beta}) \int p(Y|Y_S, \Sigma, B, \boldsymbol{\beta}) p(Y_S|\Sigma, B, \boldsymbol{\beta}) dY_S, \tag{11}$$

obtained by integrating out the auxiliary dataset. The ABC distribution $p_{ABC}\left(\Sigma, B, \boldsymbol{\beta}|Y\right)$ then acts as an approximation to $p\left(\Sigma, B, \boldsymbol{\beta}|Y\right)$ and is obtained in practice by discarding realizations of the auxiliary dataset from the output of any sampler targeting the joint posterior $p_{ABC}\left(\Sigma, B, \boldsymbol{\beta}, Y_S|Y\right)$. In this paper we consider the MCMC-ABC sampler approach.

The algorithm considered in Section 5 demonstrates how to combine the SMiN and ABC CVAR models developed. In particular we provide a general adaptive MCMC based sampling algorithm for matrix-variate α-stable CVAR posterior distributions in the ABC setting. This involves use of the conjugate models derived under the SMiN assumption being used to reduce the required dimension of the adaptive proposal kernel.

5 Sampling and Estimation

Here we present (Algorithm 1) the sampling methodology used for the posterior $p_{ABC}(\Sigma, B, \boldsymbol{\beta}, Y_S|Y)$. Our approach involves several of the posterior matrix variables $(\Sigma,\widetilde{B})$ being sampled via the conjugate model derived, in Section 4.1, for the symmetric α-stable case. However, since we are considering the general case of asymmetric α-stable noise, these conjugate posterior distributions, which are derived for the symmetric α-stable noise, can therefore be considered "proposals" for the MCMC-ABC Algorithm 1. Since, having sampled these matrices $(\Sigma,\widetilde{B})$ from the conjugate model "proposals", we still perform an accept-reject step in the spirit of the classical Metropolis-Hastings algorithm. The remaining matrix posterior parameters $(\boldsymbol{\beta},\boldsymbol{\lambda})$ are sampled via an adaptive Metropolis and adaptive Rejection Sampling framework. The proposals are combined into the hybrid adaptive ABC methodology as presented in Algorithm 1.

The version of the HAdMCMC-ABC algorithm we present updates at each iteration of the Markov chain all matrix parameters, however block Metropolis-within-Gibbs frameworks are also possible. Proposing to update the matrix-variate Markov chain parameters from iteration $j-1$ to iteration j involves sampling proposal $\{\Sigma, B, \boldsymbol{\beta}, \lambda\}'$ given Markov chain state $\{\Sigma, B, \boldsymbol{\beta}, \lambda\}[j-1]$ according to the proposal,

$$
\begin{aligned}
& q\left(\{\Sigma, B, \boldsymbol{\beta}, \boldsymbol{\lambda}\}[j-1]; \{\Sigma, B, \boldsymbol{\beta}, \boldsymbol{\lambda}\}'\right) \\
& \quad = p(\Sigma|\{\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\alpha}\}', \widetilde{Y}) p(\widetilde{B}|\{\boldsymbol{\beta}, \boldsymbol{\lambda}\}', \{\boldsymbol{\alpha}, \Sigma\}[j-1], Z_*, Q_*) \\
& \quad \times p\left(\boldsymbol{\lambda}|\{\boldsymbol{\beta}\}', \left\{\boldsymbol{\alpha}, \widetilde{B}\right\}[j-1], Z_*, Q_*\right) q(\boldsymbol{\beta}[j-1]; \boldsymbol{\beta}).
\end{aligned}
$$

The first three distributions are given by the conjugate models derived under the symmetric α-stable intra-day assumption, allowing them to be sampled exactly and $q(\boldsymbol{\beta}[j-1]; \boldsymbol{\beta})$ is given by the adaptive Metropolis proposal developed in Peters et al. (2010a) [Algorithm 2] for the cointegration modelling framework. In general several adaption strategies are possible, see discussions in Roberts and Rosenthal (2009), Atchadé and Rosenthal (2005), Haario et al. (2001) and Andrieu and Moulines (2006). The adaptive Metropolis scheme involves a proposal distribution given by,

$$
\begin{aligned}
q_j\left(\boldsymbol{\beta}[j-1],\cdot\right) &= w_1 N\left(\boldsymbol{\beta};\boldsymbol{\beta}[j-1],\frac{(2.38)^2}{d}\Phi_j\right) \\
&+ \left(1-w_1\right) N\left(\boldsymbol{\beta};\boldsymbol{\beta}[j-1],\frac{(0.1)^2}{d}I_{d,d}\right). \qquad (12)
\end{aligned}
$$

Here, Φ_j is the current empirical estimate of the covariance between the parameters of $\boldsymbol{\beta}$ estimated using samples from the Markov chain up to iteration j. The theoretical motivation for the choices of scale factors 2.38, 0.1 and dimension d are all provided in Roberts and Rosenthal (2009). We note that the update of the covariance matrix can be done recursively online via the following recursion,

$$
\begin{aligned}
&\boldsymbol{m}_{j+1} = \boldsymbol{m}_j + \frac{1}{j+1}\left(\boldsymbol{\beta}[j-1]-\boldsymbol{m}_j\right) \\
&\Phi_{j+1} = \Phi_j + \frac{1}{j+1}\left(\left(\boldsymbol{\beta}[j-1]-\boldsymbol{m}_j\right)\left(\boldsymbol{\beta}[j-1]-\boldsymbol{m}_j\right)'-\Phi_j\right).
\end{aligned} \qquad (13)
$$

Some code associated with this paper may be made available upon request to the corresponding author.

6 Results and Analysis

In the first study we fit univariate α-stable models to historical price series data to assess if there is evidence for modeling inter-day level shifts via an α-stable distribution in the differenced price series. If the series indicates substantial deviation away from the standard CVAR model assumption of a Gaussian error model ($\alpha = 2, \beta = 0$), then a composite mixture model for the errors proposed in Equation (4) becomes tenable. Otherwise, since the Gaussian distribution is also contained in the stable family, the model we propose reduces to the standard CVAR cointegration Bayesian model in Peters et al. (2010a).

6.1 Real Data - α-stable Empirical Assessment

Before fitting the Bayesian CVAR model, via the MCMC-ABC sampler developed, we first estimate the α-stable noise model parameters $(\boldsymbol{\alpha},\boldsymbol{\gamma},\boldsymbol{\delta})$. These give us the model $\mathcal{S}_\alpha(\beta,\gamma,\delta)$ used for the (inter-day) day boundary level shifts in each asset that comprise the α-stable noise component in our composite noise model in Equation 5.

Algorithm 1: Hybrid Adaptive Markov Chain Monte Carlo Approximate Bayesian Computation (HAdMCMC-ABC).

Input: Initialized Markov chain matrix-variate states
$\boldsymbol{\theta}^{(0)} = \left(\Sigma^{(0)}, \widetilde{B}^{(0)}, \boldsymbol{\beta}^{(0)}, \boldsymbol{\lambda}^{(0)}\right)$.

Output: Markov chain samples
$\{\boldsymbol{\theta}^{(j)}\}_{j=1:J} = \{\Sigma^{(j)}, B^{(j)}, \boldsymbol{\beta}^{(j)}\}_{j=1:J} \sim p_{ABC}\left(\Sigma, B, \boldsymbol{\beta}, Y_S|Y\right)$.

begin

1a. Set ABC tolerance level ϵ (note annealing of the tolerance can be utilized).

1b. Evaluate summary statistic vector for observed price series $\boldsymbol{S}(Vec(Y))$. (We use a vector of quantile estimates - several choices possible - see Fearnhead and Prangle (2010))

repeat

2. Sample conjugate proposals for matrix parameters $(\Sigma,\widetilde{B})$:

2a. Sample proposed matrix state Σ^* via inversion from posterior $p(\Sigma|\boldsymbol{\beta}^{(j-1)}, \boldsymbol{\lambda}^{(j-1)}, \boldsymbol{\alpha}^{(j-1)}, \widetilde{Y})$, [Conditional 1].

2b. Evaluate transformation matrix Q_*^* based on proposed Σ^* and obtain transformed observation matrix Z_*, [Lemma 3].

2c. Sample proposed matrix state $\widetilde{\boldsymbol{B}}^*$ via inversion from $p(\widetilde{B}|\boldsymbol{\beta}^{(j-1)}, \boldsymbol{\lambda}^{(j-1)}, \boldsymbol{\alpha}^{(j-1)}, \Sigma^*, Z_*, Q_*^*)$, [Conditional 2].

3. Sample adaptive proposals for matrix parameters $(\boldsymbol{\beta},\boldsymbol{\lambda})$:

3a. Sample components of proposed vector $\boldsymbol{\lambda}^*$ from $p(\lambda_i|\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \chi, \widetilde{B}, Q_*, \boldsymbol{\beta})$, in [Conditional 4] via single component adaptive rejection in Godsill (2000) [Section 3.1.1., p.2].

3b. Sample proposed unconstrained elements of matrix $\boldsymbol{\beta}$ from adaptive Metropolis proposal in Peters et al. (2010a) [Alg. 2, p.12].

4. Generate synthetic data set Y_S given proposal $\{\Sigma, B, \beta, \boldsymbol{\lambda}\}'$ and fitted intra-day model $S_{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta})$, evaluate summary statistic vector $\boldsymbol{S}(Vec(Y_S))$ and calculate weighting function in Equation (10).

5. Calculate ABC - Metropolis Hastings acceptance probability according to the general specification in Sisson and Fan (2010) [Equation (1.3.2)] for joint proposal $\boldsymbol{\theta} = (\Sigma, B, \beta, \boldsymbol{\lambda})$:

$$A\left(\boldsymbol{\theta}^{(j-1)}, \boldsymbol{\theta}^*\right) = \frac{p_{ABC}\left(\boldsymbol{\theta}^*|Y\right) q\left(\boldsymbol{\theta}^* \rightarrow \boldsymbol{\theta}^{(j-1)}\right)}{p_{ABC}\left(\boldsymbol{\theta}^{(j-1)}|Y\right) q\left(\boldsymbol{\theta}^{(j-1)} \rightarrow \boldsymbol{\theta}^*\right)}$$

Accept $\boldsymbol{\theta}^{(j)} = \boldsymbol{\theta}^*$ via rejection using A, otherwise $\boldsymbol{\theta}^{(j)} = \boldsymbol{\theta}^{(j-1)}$.
Set $j = j + 1$.

until $j = J$

end

We analyze inter-day price shifts by first extracting 'daily' close/open differenced price series for each asset pairs inter-day price shifts. 'Daily' here refers to the times when both markets for the pairs are first jointly open, or when the first market closes. Data consists of 10 minute interval price data. The assets considered are AUD as Australian Dollars, CD as Canadian Dollars, FV as a US five year note (bond), NQ as the NASDAQ mini-index and TU as a US two year note. In total each asset pair considers 30 contract segments, with varying numbers of days present and consecutive segment periods in time (a segment ends when a contract rolls over for one of the assets). Figure 1 shows each asset's differenced price series $\triangle x_t = x_t - x_{t-1}$ from open to close of market each day, including the associated level shifts at the close/open day boundaries in the currency in which the asset is traded.

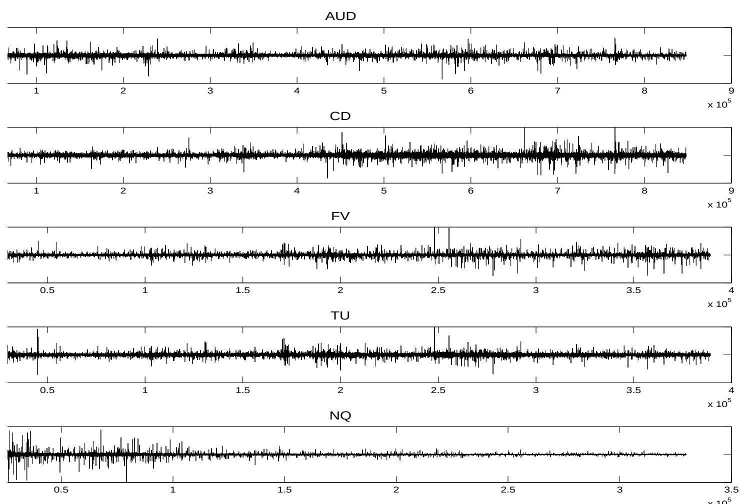

Figure 1: Plots of observations for differenced price series over 30 contract segments.

From the data in Figure 1 we extract inter-day differenced level shifts and fit them independently for each asset with an α-stable model. This first step in the modeling, prior to the Bayesian CVAR estimation, is achieved via a simple maximum likelihood based numerical approach of Nolan (1997), using open source software available at URL (`http://academic2.american.edu/~jpnolan/stable/stable.html`) and detailed in [Section VII] of Alder et al. (1998). The results of this analysis for the S0 α-stable model, comprised of level shift data for inter-day boundaries, in the 30 segments, are

Asset i.d.	# days	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\gamma}$	$\widehat{\delta}$
AUD	1535	1.833 (0.07)	0.019 (0.34)	195.365 (9)	5.1510 (17)
CD	1535	1.666 (0.08)	0.028 (0.20)	97.344 (5)	-4.699 (8)
FV	960	1.855 (0.08)	-0.551 (0.42)	105.134 (6)	15.922 (11)
NQ	1054	1.254 (0.09)	0.009 (0.14)	313.678 (23)	1.673 (31)
TU	960	1.807 (0.09)	-0.059(0.37)	88.119 (5)	-0.088 (10)

Table 1: Maximum Likelihood estimates and in brackets half the width of the symmetric 95% CI. Dates of analysis for each asset: AUD - 05/09/99 - 30/11/05; CD - 05/09/99 - 30/11/05; FV - 05/09/99 - 18/08/03; NQ - 05/09/99 - 02/12/03; TU - 05/09/99 - 18/08/03

provided in Table 1. The approach we propose here is flexible and can involve fitting the stable model to any sub segment of data required, with different stable parameter estimates per data segment.

The analysis shows that, for each of the assets, the α-stable shape parameter has 95% confidence intervals which do not contain the Gaussian case $\alpha = 2$, even with large historical data sets. Furethermore, in the case of the Canadian dollar and the NASDAQ mini index, the value of α obtained implies a signifcantly heavy-tailed model is appropriate. Additionally, several series demonstrate asymmetry ($\beta \neq 0$) with estimated 95% CI not precluding skewness, violating the assumptions of Gaussianity at these inter-day boundary points and also demonstrating that the symmetric simplification proposed in Chen and Hsiao (2010) can be invalid in many real data settings. Hence, this analysis suggests that it is clearly suitable to consider modeling the inter-day level shifts seperately from Gaussian intra-day data, verifying the appropriateness of our model assumption in Equation 5.

6.2 ABC Matrix-variate α-stable and Gaussian CVAR Analysis

In this section we perform three studies. The first and second studies involve analysis of the algorithms developed to sample from the matrix-variate posterior distribution on synthetic data sets generated with known parameters. The first study considers a mixture noise model (Equation (4)) with very heavy-tailed symmetric α-stable inter-day noise ($\alpha = 1.3$). In this symmetric α-stable case the SMiN CVAR matrix-variate likelihood derived in Section 3.1 and the resulting conjugate posterior models, developed in Section 4.1, can be utilized exactly and one does not require ABC methods.

The second study considers asymmetric heavy tailed α-stable ($\alpha = 1.3$, $\beta = 0.5$) inter-day noise. In this case we can not obtain analytic expressions for the Bayesian CVAR model as the SMiN representation does not apply, therefore we utilize the ABC estimation developed. Hence we compare the ABC model and MCMC results from Algorithm 1 to the case in which intra-day level shifts are ignored in the "Gaussian" case and sampling occurs as in Peters et al. (2010a). In the third study we consider a real data set analysis via our general MCMC-ABC sampler in Algorithm 1, for a pair of assets, observed in practice to have a cointegration relationship with rank $r = 1$, with α-stable fits from Table 1 for $AUD - CD$.

In the results section we present several different sets of simulations. The estimated results, denoted "Gaussian", are from a Bayesian CVAR model with Gaussian likelihood as presented in Section 2. This is the standard Gaussian CVAR model in the literature which ignores the fact that inter-day level shifts data is best modeled with a composite error model in Equation 5. We sample this model using the adaptive MCMC sampler of Peters et al. (2010a) and Sugita (2009), to assess the bias in parameter estimates if intra-day level shifts are not modeled explicitly. The estimated results, denoted by "Mixture Exact", are obtained under a composite error model in Equation 5, involving the assumption of symmetric α-stable noise in which the conjugate SMiN Bayesian CVAR model of Section 4.1 is sampled. The estimated results denoted by "Mixture ABC" are obtained under a composite error model in Equation 5 involving general asymmetric α-stable noise in which the ABC approximate CVAR model of Section 4.2 is sampled via Algorithm 1. In all cases we are particularly interested in the estimated cointegration basis vectors β, which directly affect portfolio weights in Bayesian pairs trading.

In all studies we consider pairs data, with a cointegration rank of $r = 1$. We ran samplers with 10,000 burn-in and 20,000 actual samples. In studies one and two we perform analysis on 20 independently generated pairs of price data sets, with each price series of length 500 samples and every 50-th sample modeled with an α-stable innovation. In the real data analysis we take the series described in Section 4.1. This corresponds to considering 10,000 observations sampled at 10 min intervals, which is equivalent to around 200 days of data.

	Gaussian model	Mixture Gaussian and α-stable intra-day		
Parameter Estimates	**Gaussian**	**Mixture ABC**	**Mixture Exact**	**Truth**
Ave. MMSE $\beta_{1,2}$	-0.02 (0.21)	0.39 (0.27)	0.42 (0.25)	0.5
Ave. Stdev. $\beta_{1,2}$	0.28 (0.08)	0.31 (0.12)	0.35 (0.09)	-
Ave. MMSE tr (Σ)	3.17 (2.03)	2.61 (2.12)	2.23 (1.91)	2
Ave. Stdev. tr (Σ)	0.16 (0.12)	0.21 (0.16)	0.19 (0.21)	-
Ave. MMSE μ_1	-0.03 (0.08)	-0.01 (0.03)	0.05 (0.01)	-
Ave. Stdev. μ_1	0.06 (0.03)	0.08 (0.02)	0.07 (0.02)	-
Ave. MMSE μ_2	4.0E-3 (0.01)	7E-3 (0.03)	6E-3 (0.01)	0.1
Ave. Stdev. μ_2	0.05 (0.01)	0.07 (0.02)	0.09 (0.03)	-
Ave. MMSE $\alpha_{1,1}$	-0.06 (0.02)	0.05 (0.02)	0.08 (0.04)	0.1
Ave. Stdev. $\alpha_{1,1}$	0.02 (2E-3)	0.03 (4E-3)	0.05 (3E-3)	-
Ave. MMSE $\alpha_{1,2}$	3E-3 (0.02)	-0.19 (0.01)	-0.21 (0.02)	-0.3
Ave. Stdev. $\alpha_{1,2}$	0.02 (0.01)	0.02 (0.01)	0.04 (0.02)	-
Ave. Mean accept. prob.	0.37	0.21	1	-

Table 2: Results for the first study with composite noise model (Equation 5) considering symmetric α-stable inter-day innovations. **Sampler Analysis:** Ave. MMSE and Stdev is averaged posterior mean or variances obtained from posterior estimation of the parameters from 20 independently generated data sets. In $(\cdot)$ are the standard error in estimates.

Synthetic Data Analysis - Symmetric α-stable heavy tailed.

The model used for this synthetic study considers parameter settings $\beta = [1, 0.5]$, $\alpha = [0.1, -0.3]$, $\Sigma = \mathbb{I}_2$ $\mu = [0, 0]$ and ($\alpha = 1.3$, $\beta = 0$, $\gamma = 1$, $\delta = 0$). The prior settings for the Bayesian model are those specified in Peters et al. (2010a). We considered several tolerance ranges ϵ in the ABC approximation, the result used for the final simulations was $\epsilon = 0.1$. This provided a reasonable trade-off between accuracy of the estimation results in the synthetic studies and mixing of the MCMC-ABC sampler. In Table 2 we present the results, comparing the performance of the estimation of the parameters for the resulting Bayesian posterior model in Theorem 3. The results demonstrate that the effect of ignoring the inter-day level shifts when fitting the Bayesian model has a significant effect on the estimation of the cointegration vector $\boldsymbol{\beta}$. In addition, it is clear that in this symmetric case, the estimates, obtained via the exact MCMC sampler and the ABC approximation, are similar. However, as expected, the computational cost for the ABC approach is significantly higher than the non-ABC approach. We also see that the estimation of the remaining parameters is accurate.

	Gaussian model	**Mixture Gaussian and α-stable intra-day**	
Parameter Estimates	**Gaussian**	**Mixture ABC**	**Truth**
Ave. MMSE $\beta_{1,2}$	-0.01 (0.21)	0.36 (0.32)	0.5
Ave. Stdev. $\beta_{1,2}$	0.28 (0.08)	0.41 (0.16)	-
Ave. MMSE $\mathrm{tr}(\Sigma)$	2.92 (1.32)	3.0 (1.49)	2
Ave. Stdev. $\mathrm{tr}(\Sigma)$	0.14 (0.07)	0.21 (0.12)	-
Ave. MMSE μ_1	-0.02 (0.07)	-0.01 (0.09)	0.1
Ave. Stdev. μ_1	0.06 (0.02)	0.10 (0.03)	-
Ave. MMSE μ_2	-3.0E-3 (0.01)	4E-3 (0.03)	0.1
Ave. Stdev. μ_2	0.05 (0.01)	0.09 (0.03)	-
Ave. MMSE $\alpha_{1,1}$	-0.06 (0.01)	0.06 (0.03)	0.1
Ave. Stdev. $\alpha_{1,1}$	0.01 (2E-3)	0.03 (8E-3)	-
Ave. MMSE $\alpha_{1,2}$	2E-3 (0.02)	1E-3 (8E-3)	-0.3
Ave. Stdev. $\alpha_{1,2}$	0.02 (0.01)	0.03 (0.01)	-
Ave. Mean accept. prob.	0.42	0.28	-

Table 3: Results for the second study with composite noise model (Equation 5) considering asymmetric α-stable inter-day innovations. **Sampler Analysis:** Ave. MMSE or Stdev is averaged posterior mean or variances obtained from estimation of the posterior parameters from 20 independently generated data sets. In $(\cdot)$ are the standard error in estimates.

Synthetic Data Analysis - Asymmetric α-stable heavy tailed

The model, used for this synthetic study, considers identical parameter settings and prior settings for the CVAR model as the previous study, with the asymmetric inter-day noise model with α-stable parameters ($\alpha = 1.3$, $\beta = 0.5$, $\gamma = 1$, $\delta = 0$). In the asymmetric case we must work with the ABC model. In Table 3 we present the results, comparing the performance of the estimation of the parameters for the resulting ABC posterior versus the basic Gaussian conjugate Bayesian model. Table 3 demonstrates that significantly more accurate results for the estimation of the cointegration vectors occur when inter-day noise modeling is incorporated.

The results are summarized for the estimated MMSE of the cointegration vector $\boldsymbol{\beta}$ in Figure 2. The left panel in Figure 2 demonstrates the first study with a symmetric α-stable composite noise model, comparing samples from the "Gaussian" case and the "Mixture Exact" case. The right panel in Figure 2 demonstrates the second study with an asymmetric α-stable composite noise model, comparing samples from the "Mixture Exact" and the "Mixture ABC" case. The "Mixture Exact" case is sub-optimal in this study since it is assuming symmetric α-stable noise, while the "Mixture ABC" case appropriately assumes asymmetric α-stable noise. In all cases we are particularly interested in the estimated cointegration basis vectors β which directly affect portfolio weights in Bayesian pairs trading.

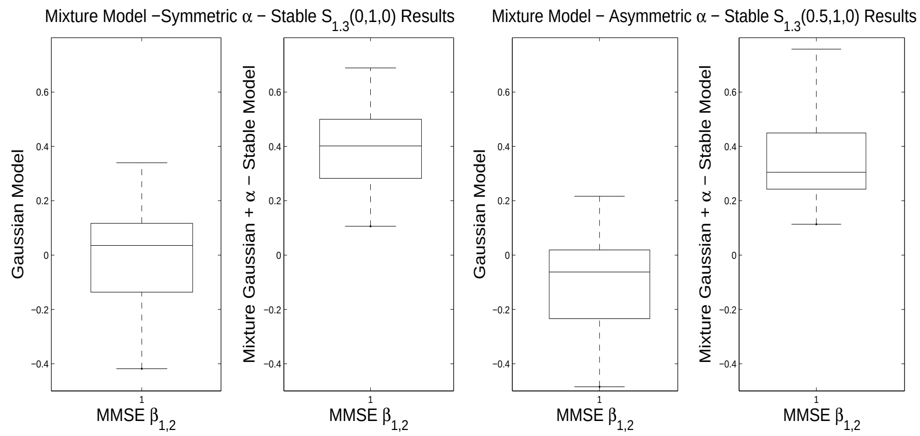

Figure 2: Estimated cointegration vector $\boldsymbol{\beta}$ in the first and second study.

Real Data Analysis

In this section we focus on the accuracy of estimation for vectors $\boldsymbol{\beta}$. These are important to the design of algorithmic trading strategies, since they are the basis for projection of the raw price series to obtain a stationary deviation series to consider trading analysis. In addition we provide estimation results for the reversion rate of the stochastic trends to stationarity as denoted by the matrix $\boldsymbol{\alpha}$. We analyze the performance of the basic "Gaussian" posterior model of Peters et al. (2010a) and Sugita (2009) in the presence of inter-day price series level shifts versus the estimation of the "Mixture ABC" model via Algorithm 1.

We present data analysis for two pairs of assets, AUD/CD and FV/TU. The first example demonstrates an asset pair with approximate symmetry though quite heavy tailed level shift statistical model. The second example is the opposite with statistically significant asymmetry and a not so heavy tailed statistical model fit for the level shift components in each asset TU and FV. In the first case where approximate symmetry is reasonable we demonstrate that results compare favorably between the 'Mixture Exact' and the 'Mixture ABC' results. In the second case the asymmetry requires that we utilize the mixture ABC approach.

	Gaussian model	Gaussian and α-stable intra-day	
Parameter Estimates	**Gaussian**	**Mixture Exact**	**Mixture ABC**
Ave. MMSE $\beta_{1,2}$	-0.31 (0.25)	0.24 (0.19)	0.18 (0.21)
Ave. Var. $\beta_{1,2}$	0.20 (0.04)	0.76 (0.12)	0.83 (0.08)
Ave. MMSE $\alpha_{1,1}$	-0.02 (1.36E-3)	-0.04 (2.1E-3)	-0.01 (3.8E-3)
Ave. Var. $\alpha_{1,1}$	3.90E-5 (4.09E-6)	4.2E-5 (9.21E-6)	5.3E-5 (2.5E-5)
Ave. MMSE $\alpha_{1,2}$	1.24E-3(1.20E-3)	-2.9E-4 (2.2E-3)	-6.3E-4 (1.7E-3)
Ave. Var. $\alpha_{1,2}$	2.18E-5(3.07E-6)	3.1E-5 (9.01E-4)	1.7E-5 (1.0E-3)

Table 4: **Sampler Analysis:** In $(\cdot)$ are the standard error estimates obtained from 20 batches of MCMC samples each of length 1,000, averaged over each of the sets of 2 days of data.

First we assess the price series for AUD / CD with base currency in AUD sampled at 10min intervals during the joint open market hours. Analysis is performed for the first contract in Table 1, starting from 05/09/99, containing 60 days worth of market data, producing a time series of prices of length 29,621 samples. The data was transformed by translation of each series by the median and scaled by the standard deviation. The analysis performed considers 30 batches of 2 days of data, giving on average 489 data samples per batch, and the posterior parameter estimates are averaged over samplers analysis of each data set and presented in Table 4. These results demonstrate that since this asset pair has approximately symmetric, though heavy tailed inter-day level shifts, the parameter estimates from the Mixture Exact (assuming $\beta = 0$) and Mixture ABC ($\beta \approx 0$) are similar. Furthermore, they demonstrate that failing to account for the inter-day level shifts observed can significantly affect the estimation of the cointegration vectors and reversion rates as demonstrated in the comparison in Table 4 and in Figure 4.

Next we assess the price series for FV / TU sampled at 10min intervals during the jointly open market hours. Analysis is performed for the first contract in Table 1, starting from 05/09/99, containing 53 days worth of market data, producing a time series of prices of length 19,935 samples. The data was transformed by translation of each series by the median and scaled by the standard deviation. The analysis performed considers 26 batches of 2 days of data and the posterior parameter estimates are averaged over samplers analysis of each data set and presented in Table 5. Again, these results demonstrate that failing to account for the inter-day level shifts observed can significantly affect the estimation of the cointegration vectors and reversion rates, this

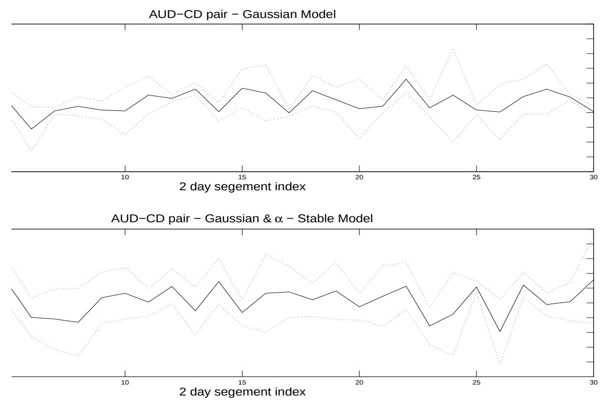

Figure 3: Estimated cointegration vector $\boldsymbol{\beta}$ for AUD-CD pair for 2 day segments at 10min samples. TOP: Gaussian model; Bottom: Mixture ABC model; Solid line is estimated MMSE and dashed line is posterior 95% Confidence Interval.

time for the asymmetric case.

7 Conclusions

We studied the impact of price series level shifts on statistical estimation of matrix-variate parameters in CVAR models utilized in algorithmic trading. In particular, we first demonstrated the significant impact on estimation of CVAR models when failing to appropriately model observed level shifts in price series.

We developed a composite noise model, comprised of Gaussian and α-stable innovation noise, for the CVAR model in the presence of price series level shifts. The example, that we illustrated this model on, involved the situation that occurs at deterministic times each trading day, at inter-day market boundaries. However, we point out that our methodology is general and extends also to settings in which the level shift times are unknown *a priori*.

Working under this composite noise model of Gaussian and α-stable CVAR innovations, we developed a novel conjugate Bayesian model under transformation, allowing for exact MCMC sampling frameworks to be developed in the symmetric heavy tailed α-stable scenario. In the asymmetric skewed noise setting, a non-standard approximate Bayesian computation model was developed and an advanced, adaptive MCMC

	Gaussian model	**Gaussian and α-stable intra-day**
Parameter Estimates	**Gaussian**	**Mixture ABC**
Ave. MMSE $\beta_{1,2}$	-1.01 (0.11)	-5.55E-2 (0.12)
Ave. Var. $\beta_{1,2}$	0.07 (9.5E-3)	0.05 (4.5E-3)
Ave. MMSE $\alpha_{1,1}$	-0.01 (2.4E-3)	-0.52(5.89E-2)
Ave. Var. $\alpha_{1,1}$	1.89E-4 (1.94E-5)	7.1E-3 (1.3E-3)
Ave. MMSE $\alpha_{1,2}$	4.96E-2 (1.8E-3)	4.48E-2 (5.65E-2)
Ave. Var. $\alpha_{1,2}$	1.28E-4 (1.77E-5)	4.8E-3 (1.6E-3)

Table 5: **Sampler Analysis:** In $(\cdot)$ are the standard error estimates obtained from 20 batches of MCMC samples each of length 1,000, averaged over each of the sets of 2 days of data.

algorithm was utilized to sample this ABC posterior.

We were able to demonstrate and verify on synthetic data sets under both symmetric and asymmetric α-stable models, that the sampling methodology we developed for estimation of the MMSE for the matrix-variate posterior parameters is accurate. We then compared the performance of our model and sampler to the standard Gaussian Bayesian CVAR model on real financial pairs, demonstrating a marked difference in the estimated CVAR model parameters, hence justifying the applicability of such a model in applied financial models for trading.

The model developed in this paper assumes that the underlying model for the price series pair is appropriately modeled by the basic CVAR model presented in Section 2. This differs significantly from the underlying assumption of Chen and Hsiao (2010). Alternative approaches, that could be developed in future work, include the use of a Markov switching regime model, see for example Krolzig (1997). Under such a model the CVAR parameters may vary depending on a latent regime state variable, see Sugita (2008) for details.

References

Ackert, L. and Racine, M. (1999). "Stochastic trends and cointegration in the market for equities." *Journal of Economics and Business*, 51(2): 133–143. 756

Alder, R., Feldman, R., and Taqqu, M. S. (1998). *A practical guide to heavy-tails: Statistical techniques for analysing heavy-tailed distributions*. Birkhäuser, Boston.

778

Andrieu, C. and Moulines, É. (2006). "On the ergodicity properties of some adaptive MCMC algorithms." *The Annals of Applied Probability*, 16(3): 1462–1505. 775

Atchadé, Y. and Rosenthal, J. (2005). "On adaptive Markov chain Monte Carlo algorithms." *Bernoulli*, 11(5): 815–828. 775

Bauwens, L. and Lubrano, M. (1994). "Identification restrictions and posterior densities in cointegrated Gaussian VAR systems." *Universite catholique de Louvain, Center for Operations Research and Econometrics Discussion Papers, Report - 1994018.*. 756

Beaumont, M., Cornuet, J., Marin, J., and Robert, C. (2009). "Adaptivity for ABC algorithms: the ABC-PMC scheme." *Biometrika*, 96(4): 983–990. 757, 773, 774

Bock, M. and Mestel, R. (2009). "A Regime-Switching Relative Value Arbitrage Rule." *Operations Research Proceedings 2008*, 9–14. 756

Chambers, J., Mallows, C., and Stuck, B. (1976). "A method for simulating stable random variables." *Journal of the American Statistical Association*, 71: 340–334. Correction (1987), *82*, 704. 761

Chen, P. and Hsiao, C. (2010). "Subsampling the Johansen test with stable innovations." *Australian & New Zealand Journal of Statistics*, 52(1): 61–73. 779, 786

De Lathauwer, L. and Vandewalle, J. (2004). "Dimensionality reduction in higher-order signal processing and rank-(R1, R2,..., RN) reduction in multilinear algebra." *Linear Algebra and its Applications*, 391: 31–55. 767

Del Moral, P., Doucet, A., and Jasra, A. (2011). "An adaptive sequential Monte Carlo method for approximate Bayesian computation." *Statistics and Computing - to appear*. 757, 773, 774

Engle, R. and Granger, C. (1987). "Co-integration and error correction: representation, estimation, and testing." *Econometrica: Journal of the Econometric Society*, 55(2): 251–276. 756

Fama, E. and Roll, R. (1968). "Some properties of symmetric stable distributions." *Journal of the American Statistical Association*, 63: 817–83. 761

Fearnhead, P. and Prangle, D. (2010). "Semi-automatic Approximate Bayesian Computation." *Arxiv preprint arXiv:1004.1112v2 [stat.ME]*. 757, 773, 774, 777

Friedlandera, M. and Hatzb, K. (2008). "Computing non-negative tensor factorizations." *Optimization Methods and Software*, 23(4): 631–647. 767

Gatev, E., Goetzmann, W., and Rouwenhorst, K. (2006). "Pairs trading: Performance of a relative-value arbitrage rule." *Review of Financial Studies*, 19(3): 797. 756

Geweke, J. (1996). "Bayesian reduced rank regression in econometrics* 1." *Journal of Econometrics*, 75(1): 121–146. 756, 770, 772

Godsill, S. (2000). "Inference in symmetric alpha-stable noise using MCMC and the slice sampler." In *Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing*, volume VI, 3806–3809. 761, 763, 772

Granger, C. and Hyung, N. (2004). "Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns." *Journal of Empirical Finance*, 11(3): 399–421. 756

Granger, C. and Weiss, A. (2001). "Time series analysis of error correction models." *Spectral analysis, seasonality, nonlinearity, methodology and forecasting: collected papers of Clive WJ Granger*, 129. 756, 759

Grelaud, A., Robert, C., and Marin, J. (2009). "ABC methods for model choice in Gibbs random fields." *Comptes Rendus Mathematique*, 347(3-4): 205–210. 773

Gupta, A. and Nagar, D. (1999). *Matrix variate distributions*. Monograph and Surveys in Pure and Applied Mathematics. Chapman Hall CRC. 760, 764, 765, 766, 767, 768, 770

Haario, H., Saksman, E., and Tamminen, J. (2001). "An adaptive Metropolis algorithm." *Bernoulli*, 7(2): 223–242. 775

Harville, D. (2008). *Matrix algebra from a statistician's perspective*. Springer Verlag, New York. 768

Johansen, S. (1988). "Statistical analysis of cointegration vectors." *Journal of Economic Dynamics and Control*, 12(2-3): 231–254. 758

Kleibergen, F. and Paap, R. (2002). "Priors, posteriors and Bayes factors for a Bayesian analysis of cointegration." *Journal of Econometrics*, 111(2): 223–249. 770

Kleibergen, F. and Van Dijk, H. (2009). "On the shape of the likelihood/posterior in cointegration models." *Econometric Theory*, 10(3-4): 514–551. 756, 770, 771

Koop, G., Strachan, R., Van Dijk, H., and Villani, M. (2006). "Bayesian approaches to cointegration." *Palgrave Handbook on Econometrics*, 1: 871–898. 756, 759, 771

Krolzig, H. (1997). "Statistical analysis of cointegrated VAR processes with Markovian regime shifts." *unpublished, Nuffield College, Oxford*. 786

Luetkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer-Verlag, Berlin. 765

Mills, T. and Markellos, R. (2008). *The econometric modelling of financial time series*. Cambridge University Press. 756

Neslehova, J., Embrechts, P., and Chavez-Demoulin, V. (2006). "Infinite mean models and the LDA for operational risk." *Journal of Operational Risk*, 1(1): 3–25. 761

Nolan, J. P. (1997). "Numerical computation of stable densities and distributions." *Communications in Statistics, Stochastic Models*, 13: 759–774. 757, 778

— (2012). "Stable Distributions - Models for Heavy Tailed Data." In progress, Chapter 1 online at `http://academic2.american.edu/~jpnolan`. 761

Peters, G., Kannan, B., Lasscock, B., and Mellen, C. (2010a). "Model Selection and Adaptive Markov chain Monte Carlo for Bayesian Cointegrated VAR model." *Bayesian Analysis*, 5(3): 465–492. 756, 758, 761, 770, 771, 775, 776, 780, 781, 783

Peters, G., Nevat, I., Sisson, S., Fan, Y., and Yuan, J. (2010b). "Bayesian symbol detection in wireless relay networks via likelihood-free inference." *Signal Processing, IEEE Transactions on*, 58(10): 5206–5218. 774

Peters, G. and Sisson, S. (2006). "Bayesian inference, Monte Carlo sampling and operational risk." *Journal of Operational Risk*, 1(3): 27–50. 773

Peters, G., Wuethrich, M., and Shevchenko, P. (2010c). "Chain ladder method: Bayesian bootstrap versus classical bootstrap." *Insurance: Mathematics and Economics*, 47(1): 36–51. 773

Peters, G. W., Fan, Y., and Sisson, S. A. (2008). "On sequential Monte Carlo, partial rejection control and approximate Bayesian computation." Technical report, Tech. rep. UNSW. 757, 773, 774

Qiou, Z. and Ravishanker, N. (1998). "Bayesian inference for time series with infinite variance stable innovations." *Journal of Time Series Analysis*, 19: 235–249. 757

Ratmann, O., Andrieu, C., Hinkley, T., Wiuf, C., and Richardson, S. (2009). "Model criticism based on likelihood-free inference, with an application to protein network evolution." *Proceedings of the National Academy of Science USA*, 106: 10576–10581. 773

Reeves, R. and Pettitt, A. (2005). "A theoretical framework for approximate Bayesian computation." In *Statistical Solutions to Modern Problems: Proceedings of the 20th International Workshop on Statistical Modelling, Sydney, Australia, July 10–15, 2005*, 393–396. 773

Roberts, G. and Rosenthal, J. (2009). "Examples of adaptive MCMC." *Journal of Computational and Graphical Statistics*, 18(2): 349–367. 775, 776

Samorodnitsky, G. and Taqqu, M. S. (1994). *Stable non-Gaussian random processes: Stochastic models with infinite variance*. Chapman and Hall/CRC. 757, 761

Sisson, S. and Fan, Y. (2010). "Likelihood-free Markov chain Monte Carlo." *Arxiv preprint arXiv:1001.2058v1 [stat.ME]*. 757, 773, 774

Strachan, R. (2003). "Valid Bayesian estimation of the cointegrating error correction model." *Journal of Business and Economic Statistics*, 21(1): 185–195. 756

Strachan, R. and Inder, B. (2004). "Bayesian analysis of the error correction model." *Journal of Econometrics*, 123(2): 307–325. 756, 759, 770, 772

Sugita, K. (2002). "Testing for cointegration rank using Bayes factors." In *Royal Economic Society Annual Conference*. Royal Economic Society. 756, 770, 771, 772

— (2008). "Bayesian analysis of a Markov switching temporal cointegration model." *Japan and the World Economy*, 20(2): 257–274. 786

— (2009). "A Monte Carlo comparison of Bayesian testing for cointegration rank." *Economics Bulletin*, 29(3): 2145–2151. 756, 780, 783

Tavaré, S., Balding, D. J., Griffiths, R. C., and Donnelly, P. (1997). "Inferring coalescence times from DNA sequence data." *Genetics*, 145: 505–518. 757, 773

Toni, T., Welch, D., Strelkowa, N., Ipsen, A., and Stumpf, M. P. H. (2009). "Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems." *Journal of the Royal Society Interface*, 6: 187–202. 774

Villani, M. (2005). "Bayesian reference analysis of cointegration." *Econometric Theory*, 21(02): 326–357. 770

Wang, J. and Zivot, E. (2000). "A Bayesian time series model of multiple structural changes in level, trend, and variance." *Journal of Business & Economic Statistics*, 18(3): 374–386. 756

Wilkinson, R. (2008). "Approximate Bayesian computation (ABC) gives exact results under the assumption of model error." *Arxiv preprint arXiv:0811.3355v1 [stat.CO]*. 773

Zolotarev, V. M. (1986). *One-Dimensional Stable Distributions*. Translations of Mathematical Monographs. American Mathematical Society, Providence, Rhode Island. 757, 761

Acknowledgments

We thank Prof. Arnaud Doucet and Prof. Richard Gerlach for comments and suggestions during this research. In addition, we thank the Institute of Statistical Mathematics, Tokyo, Japan, for hosting GWP whilst performing aspects of this research and Boronia for funding.