The School of Mathematics

THE UNIVERSITY
of EDINBURGH

Conditional extreme value mixture models and their application to stationary time series

by

Ben Anson

Dissertation Presented for the Degree of
MSc in Statistics and Operational Research

August 2022

Supervised by
Dr. Ioannis Papastathopoulos

Abstract

Extreme value theory provides general results regarding the asymptotics of probability distribution functions that can be leveraged by statisticians to perform out-of-sample inference. This inference tends to be more difficult in a multivariate setting, but tools have been developed to combat the multivariate case, such as conditional extreme value models. In this dissertation, we explore the use of conditional extreme value mixture models (CEVMMs) for modelling tail switching behaviour of stationary time series. An EM algorithm for fitting CEVMMs is developed and its efficacy illustrated on data generated from an asymmetric logistic model. It is then shown that UK electricity imbalance prices exhibit tail switching behaviour and that CEVMMs are able to model this. Furthermore, it is demonstrated that CEVMMs are able to produce convincing one-step-ahead predictive distributions for imbalance price residuals.

Acknowledgements

Thank you to Ioannis, for the many hours spent explaining some of the finer details of the statistics of extremes, and of statistics and probability in general. I left every meeting with fresh problems, new ideas to try, and of course a thick wad of paper covered head to toe with scribbles.

Own Work Declaration

I declare that the following work is my own, except where otherwise stated.

Contents

1 Introduction **1**

2 Background **2**
2.1 Univariate extremes . . . 2
2.2 Multivariate extremes . . . 2
2.3 Conditional extreme value models . . . 3
2.4 Fitting conditional extreme value mixture models (CEVMMs) . . . 4
2.5 Tail switching . . . 5

3 Methods and results **6**
3.1 The 3-dimensional asymmetric logistic model . . . 6
3.1.1 Theoretical Extremal Dependence . . . 7
3.1.2 Fitting conditional extreme value mixture models — a worked example . . . 9
3.1.3 A look at logistic margins . . . 13
3.1.4 Regularization . . . 15
3.1.5 Quantifying error in CEVMM models with bootstrapped confidence intervals . 16
3.2 Application to Electricity Imbalance Prices . . . 20
3.2.1 Background, and obtaining electricity imbalance price data . . . 20
3.2.2 Cleaning imbalance price data . . . 21
3.2.3 Seasonalizing imbalance prices . . . 21
3.2.4 Fitting the CEVMM model . . . 25
3.2.5 Uncertainty in model parameters . . . 26
3.2.6 One-step-ahead residual prediction . . . 28
3.2.7 Conversion back to the original price scale . . . 33

4 Conclusions **35**

A Appendices **38**
A.1 EM algorithm for fitting conditional extremal K-mixture models with the normal assumption . . . 38
A.2 Constructing empirical CDFs in the CEVMM model . . . 39

Notation

Mathematical

$\mathbb{R}_+$	$\{x \in \mathbb{R} \,\vert\, x > 0\}$
$\mathcal{S}_K$	$\left\{x \in \mathbb{R}^K \,\middle\vert\, x \geq 0,\, \sum_{k=1}^K x_k = 1\right\}$, the $K-1$ dimensional simplex
$(x)_+$	$\max\{x, 0\}$
$x_{\{i_1,\ldots,i_m\}}$, where $x \in \mathbb{R}^d$	$(x_{i_1}, \ldots, x_{i_m})$
$x_{\{i_1,\ldots,i_m\},\ldots,\{j_1,\ldots,j_{m'}\}}$	$(x_{i_1}, \ldots, x_{i_m}, \ldots, x_{j_1}, \ldots, x_{j_{m'}})$
$x_{/i}$, where $x \in \mathbb{R}^d$	$(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_d)$
$\text{sigmoid}(x)$	$[1 + \exp(-x)]^{-1}$
$\text{logit}(x)$	$\log[x/(1-x)]$
$X_{1:t}$	$\{X_1, \ldots, X_t\}$
$\hat{F}_{\text{ecdf}}$	Empirical cumulative distribution function
$f(\cdot\,;\theta)$	$f : X \to Y$ with $f(x;\theta) = g(x,\theta)$ for some X, Y and $g : X \times \Theta \to Y$
$1(\cdot)$	Indicator function
$F_0(x)$	$\mathbb{P}(X_0 \leq x)$
$F_{1\vert 0}(x; x_0)$	$\mathbb{P}(X_1 \leq x \,\vert\, X_0 = x_0)$
$F_{2\vert 1,0}(x; x_1, x_0)$	$\mathbb{P}(X_2 \leq x \,\vert\, X_1 = x_1,\, X_0 = x_0)$

Nomenclature

LHS	Left hand side
RHS	Right hand side
CDF	Cumulative distribution function
ECDF	Empirical cumulative distribution function
MCMC	Markov chain Monte Carlo
MISE	Mean integrated square error
MIAE	Mean integrated absolute error
i.i.d.	Independent and identically distributed
r.v.	Random variable
EVT	Extreme value theory
GPD	Generalized Pareto distribution
CEVMM	Conditional extreme value mixture model

List of Tables

3.1 Parameter estimates and log-likelihoods, to 2 decimal places, obtained when fitting CEVMMs to asymmetric logistic data, with starting values initialized using different random seeds. 10
3.2 Mean integrated square errors ± 1 s.d. for Laplace and logistic marginals, and for different parameter combinations. MISEs are multiplied by 100 for readability. 15
3.3 Fitted parameters and log-likelihoods for CEVMM fits on the three datasets described above, for different levels of regularization γ. 16
3.4 Fitted parameters and log-likelihood for a CEVMM fit on $\mathcal{D}_1$, with $u = F^{-1}_{\text{Laplace}}(0.95)$, $\gamma = 0.05$. 17
3.5 Best CEVMM fit of 30 random initial values, with no regularization and $u = F^{-1}_{\text{Laplace}}(0.95)$. 26
3.6 Fitted parameter values and log likelihood for $K = 2$, $\gamma = 0.05$, $u = 2.8$. 26
3.7 Fitted parameter values and log-likelihood for $K = 3$, $\gamma = 0.05$, $u = 2.8$. 31

List of Figures

2.1 S&P 500 price and log returns on a Laplace scale, during a financially turbulent period in 2002. Tail switches across $u = F^{-1}_{\text{Laplace}}(0.95)$ (dashed line) denoted with $\star$. 5

3.1 X_{t+1} vs. X_t, on unit Laplace scale, for $\theta_{/012} = 0.3$ and different values of v (using the same random seed). The theoretical modes are shown in grey. 8

3.2 Data used for the worked example, in Laplace margins. The threshold u is shown as a dashed grey line on the left plot. 9

3.3 X_{t+1} vs. X_t, with each data point colour coded according to its more likely mode (a.k.a mixture component). The coloured lines show contours of the modelled normal mixture density for $X_{t+1} \,|\, X_t$, when viewed as a function of both X_{t+1} and X_t. 11

3.4 Probability densities of $X_{t+1}|X_t = x$ for 4 fixed values of x. The expected densities are shown in orange, and fitted densities in blue. These plots can be thought of as vertical cross sections of Figure 3.3 at $X_t = x$. 11

3.5 Effect of decreasing u (down from $u = 3$) on the fitted parameters. 12

3.6 Probability densities of $X_{t+1}|X_t = x$ for fixed values of x (Laplace margins). Expected densities are shown in orange. 10000 samples were drawn from the estimated distribution function; the associated Gaussian kernel density estimators for these samples are shown in blue, with histograms of the samples in green. 12

3.7 Probability densities of $X_{t+1}|X_t = x$ for fixed values of x (logistic margins). True densities are shown in orange, and fitted densities in blue. 14

3.8 Probability densities of $X_{t+1}|X_t = x$ for fixed values of x (logistic margins). True densities are shown in orange. 10000 samples were drawn from the estimated distribution function; the associated Gaussian kernel density estimators for these samples are shown in blue, with histograms of the samples in green. 14

3.9 The left plot shows a time series of length 15000 generated from the 3 dimensional asymmetric logistic process, placed in Laplace margins. The grey dashed line is u. The right plot shows $\mathcal{D}$. 17

3.10 Histograms of fitted values α_1, α_2, β_1 and β_2 on $M = 1000$ moving-block-bootstrapped time series with block size $B = 50$. The parameter values fitted on the full dataset are shown in red. 18

3.11 Estimates for $F_{1|0}(x; x_0)$, for two large x_0. The true distribution is shown in black, and the distribution estimated from the full data shown in red. The 99% confidence intervals, generated using $M = 1000$ bootstrap samples, are shown in blue. 19

3.12 Estimate for $F_{1|0}(x; F_0^{-1}(0.9999))$. The true distribution is shown in black, and the distribution estimated from the full data shown in red. The 99% confidence intervals, generated using $M = 1000$ bootstrap samples, are shown in blue. 20

3.13 Imbalance prices, and imbalance prices on a log scale, for 2021. Months labelled correspond to the first period of the month. 22

3.14 Plot of $y :=$ `log_price` $-$ $\widehat{\texttt{log\_price}}$ vs. period for model (3.10). Months labelled correspond to the first period of the month. 24

3.15 Plot of $y :=$ `log_price` $-$ $\widehat{\texttt{log\_price}}$ vs. period for model (3.11), chosen using stepwise selection. Months labelled correspond to the first period of the month. 24

3.16 Plots of y_t vs. y_{t-1} (Laplace margins). (a) shows all the data, with the dotted line denoting $y_{t-1} = u = 2.8$; (b) focuses on the region $y_{t-1} > u$. 25

3.17 Best CEVMM fit of 30 different models each fit with random initial values. Here, $\gamma = 0$ and $u = F^{-1}_{\text{Laplace}}(0.95)$. Data points are colour coded according to which normal mixture they most likely belong to. 26

3.18 Effect of threshold u on the fitted parameters, with $\gamma = 0.05$. 27

3.19 CEVMM fit, with regularization $\gamma = 0.05$ and $u = 2.8$. Data points are colour coded according to the normal mixture they most likely belong to. The coloured lines show contours of the modelled normal mixture density for $y_{t+1}|y_t$, when viewed as a function of both y_{t+1} and y_t. 27
3.20 Histograms of fitted values α_1, α_2, β_1 and β_2 on $M = 1000$ moving-block-bootstrapped time series with block size $B = 48$. The parameter values fitted on the full dataset are shown in red. 28
3.21 Imbalance price residuals for January 2022, with threshold $u = 2.8$ shown by the grey dashed line. 29
3.22 ($K = 2$) Histograms for simulated values of $y_{t+1}|y_t$ (here $y_{d,p} := y_{48d-1+p}$ denotes the residual on day d during period p in January 2022). The red line shows the actualized value of y_{t+1}, and the green line shows the value estimated by an ARIMA model (fit to the to the entire series of residuals). 30
3.23 Violin and box plots for likelihoods of the actualized residuals according to the estimated predictive distributions for the CEVMM models for $K = 2$ and $K = 3$. 31
3.24 ($K = 3$) Histograms for simulated values of $y_{t+1}|y_t$ (here $y_{d,p} := y_{48d-1+p}$ denotes the residual on day d during period p in January 2022). The red line shows the actualized value of y_{t+1}. 32
3.25 Histograms for simulated values of $P_{t+1}|y_t$ under the CEVMM model 3.6. Here $y_{d,p} := y_{48d-1+p}$ denotes the residual on day d during period p in January 2022, and $P_{t+1}|y_t$ represents the imbalance price (measured in GBP) at time $t + 1$, conditional on the previous period's imbalance residual. The red line shows the measured value of P_{t+1}. . 33

List of Algorithms

1 EM algorithm for fitting conditional extremal K-mixture models with the normal assumption 38
2 Sample residuals in the CEVMM model 39

Chapter 1

Introduction

Extreme value theory (EVT) is concerned with quantifying uncertainty around unlikely events. This is an intrinsically difficult task because, by definition, one usually has little data regarding an extreme event. Unfortunately, the naive approximation for the probability of an extreme event (namely 0) is sometimes insufficient. Accordingly, EVT has been an active area of research and it is commonly applied to areas such as finance (Embrechts et al., 1999; Gilli et al., 2006; Gkillas and Katsiampa, 2018) and the environmental sciences (Katz, 1999; Towler et al., 2010).

In this dissertation, we are interested in the methodologies that can be derived from EVT. In particular, we look at the development and application of an EM algorithm for conditional extremes of multivariate data. Conditional extreme value mixture models (CEVMMs) are a statistical construct, and require only a fairly mild assumption about the distribution F that generates the data being modelled. If a random d-vector X is distributed according to F, then we assume that

$$\mathbb{P}\left(\frac{X_{/j} - b(x)}{a(x)} \leq z \,\middle|\, X_j = x\right) \to G(z) \text{ as } x \to \infty,$$

where G is a non-degenerate probability distribution. This assumption originates from the seminal paper by Heffernan and Tawn (2004), where a methodology was developed for modelling the unknowns G, a and b. Their methodology has been extended and improved since then (Keef et al., 2013). The methodology is semi-parametric, in that parametric assumptions are made in order to find estimates of the functions a and b, and then nonparametric methods are used to model G. There is a subtle complication in that the functions a and b are not necessarily unique, but this naturally leads to use of mixture models where each mixture component accounts for a unique combination of the functions a and b. CEVMMs are based on this mixture model assumption.

In what is believed to be a novel development, we apply these mixture models to univariate stationary time series $\{X_t\}_{t\in\{1,\ldots,n\}}$. We do this by first viewing the time series as a set of consecutive pairs $\mathcal{D}' = \{(X_t, X_{t+1}) \,|\, t \in \{1, \ldots, n-1\}\}$. Then, assuming that the subset of pairs where the first element is large is a sample of i.i.d. draws from the same distribution, a CEVMM model can be fit to characterize the tail switching properties of X_t. Although an i.i.d. assumption is unreasonable in general, it is more palatable here since if n is large enough, other dependent variables will in effect be integrated out.

The structure of this dissertation proceeds as follows. In Section 2 a brief overview of EVT and two of its important results are given, as well as an explanation of the origins of the crucial assumptions and methods developed in Heffernan and Tawn (2004). The concept of tail switching is also introduced. Next, in Section 3, we show how CEVMMs can be fitted on simulated asymmetric logistic data viewed as a Markov chain, and we illustrate how uncertainty can be quantified. With the knowledge gained from working with simulated data, we apply CEVMMs to electricity imbalance price residuals. We show that CEVMMs are able to capture conditional extremal dependence of these residuals, and also offer convincing predictive distributions for the next residual. Finally, in Section 4, the results are summarised and suggestions for future work are put forward.

Chapter 2

Background

2.1 Univariate extremes

Crucial components of EVT are the Fisher-Tippet-Gnedenko Theorem (Coles, 2001), and the closely related Pickands-Balkema-de Haan Theorem (Balkema and de Haan, 1974; Pickands III, 1975). These asymptotic results give justification for extrapolating beyond data in a principled way, first pioneered in Davison and Smith (1990). This can be very powerful, and we approach the subject with this as our goal.

Theorem 2.1 (Pickands–Balkema–de Haan)
Let X be an r.v. with cumulative distribution F. Then if there exist $a(u) \in \mathbb{R}_+$, $b(u) \in \mathbb{R}$, and G non-degenerate such that

$$\mathbb{P}\left(\frac{X - b(u)}{a(u)} \leq z \,\middle|\, X > u\right) \to G(z) \text{ as } u \to \infty,$$

then G is a Generalized Pareto Distribution (GPD) function parameterized by $\xi \in \mathbb{R}$ such that

$$G(z) = G(z; \xi) = 1 - (1 + \xi z)_+^{-1/\xi}.$$

Theorem 2.1 allows for a natural application in distribution estimation. Given $\{X_i\}_{i=1}^n$ that are assumed to be i.i.d. drawn from an unknown distribution F, Theorem 2.1 motivates the following method for estimating F in its upper tail with a GPD distribution G

$$\hat{F}(z) = G((z - \hat{b})/\hat{a}\,;\, \hat{\xi}) \text{ for } z > v. \tag{2.1}$$

The parameters a, b, σ and ξ can be fit via some statistical procedure (e.g. maximum likelihood) on a set of exceedances $\{X_i \,|\, X_i > v,\, i \in \{1, \ldots, n\}\}$. A more standard estimator can be used to estimate $F(z)$ for $z < v$, such as the ECDF. Finding a suitable v can be troublesome as its selection corresponds to a bias-variance trade-off (Coles, 2001). However, given a sufficient amount of data it is generally possible to fit such a model and therefore allow for extrapolation in the tails. Relatively mature software packages exist for fitting these models and quantifying confidence intervals (see for example Belzile et al., 2022), so analyses of this kind should be considered approachable.

2.2 Multivariate extremes

In the univariate case, we have univariate extreme value distributions, which give rise to the useful Pickands-Balkema-de Haan Theorem (2.1). In the multivariate case, we have the following Theorem 2.2 concerning multivariate extreme value distributions, stated assuming Fréchet margins. Restricting to Fréchet margins is not limiting, since we can transform data to/from any margin via $Q \circ \hat{F}$, where Q is the relevant quantile function.

Theorem 2.2 (Multivariate Extremal Types)
Let $\{X^i\}_{i=1}^n = \{(X_1^i, \ldots, X_d^i)\}_{i=1}^n$ be independent random d-vectors with unit Fréchet marginals. Let $M^n \in \mathbb{R}^d$ be the vector of component-wise maxima, so that $M_j^n = \max_{i=1}^n X_j^i$. Take the following vector additions, multiplications, etc. to be component-wise, and read $a < b$, $a, b \in \mathbb{R}^d$ to mean that $a_i < b_i$, $\forall i \in \{1, \ldots, d\}$. Suppose there exist $a_n \in \mathbb{R}_+^d$, $b_n \in \mathbb{R}^d$ and non-degenerate G such that

$$\mathbb{P}\left(\frac{M^n - b_n}{a_n} \leq z\right) \to G(z) \text{ as } n \to \infty.$$

Then for any pair of norms $\|\cdot\|_1$, $\|\cdot\|_2$ on $\mathbb{R}^d$, G can be written in the following form

$$G(z) = \exp\left[-V(z)\right],$$
$$V(z) = \int_{\Xi} \max_{j \in \{1,\ldots,d\}} \left\{ \frac{\omega_j}{\|\omega\|_1} \frac{1}{z_j} \right\} S(\mathrm{d}\omega),$$

where $\Xi = \{\omega \in \mathbb{R}^d : \|\omega\|_2 = 1, \omega \geq 0\}$ and S is a measure over the Borel sets $\mathcal{B}(\Xi)$ satisfying $\int_{\Xi} \omega_j / \|\omega\|_1 \, S(\mathrm{d}\omega) = 1$ for all $j \in \{1, \ldots, d\}$.

If it were possible to establish V in Theorem 2.2, as well as the normalizing constants a_n, b_n for some multivariate data, then in principle it should be possible to form a multivariate equivalent of the distribution estimator in (2.1) for large n, by approximating F as follows

$$\begin{aligned} &\mathbb{P}((M^n - b_n)/a_n \leq z) \approx \exp\left[-V(z)\right] \\ &\quad \implies F(a_n z + b_n)^n \approx \exp\left[-V(z)\right] \text{ assuming } X_i \text{ i.i.d.} \\ &\quad\quad \implies F(z) \approx \exp\left[-\frac{1}{n} V\left(\frac{z - b_n}{a_n}\right)\right]. \end{aligned}$$

Notice, however, that V is characterized by a measure S, which makes inference very difficult. Parametric models (Tawn, 1990) as well as nonparametric models (Marcon et al., 2017) for V have been proposed, but ultimately there is no general and statistically practical multivariate equivalent of Theorem 2.1 that allows for multivariate density estimation/extrapolation in the tails.

2.3 Conditional extreme value models

Suppose we have n i.i.d. random d-vectors $\{X^i\}_{i=1}^n = \{(X_1^i, \ldots, X_d^i)\}_{i=1}^n$. We drop the i superscript for notational convenience from here. We have established that it is difficult to estimate $\mathbb{P}(X > u+z \,|\, X > u)$ for some extreme u using the multivariate value distributions from Theorem 2.2. Heffernan and Tawn (2004) propose an alternative method for modelling the extremal dependence of multivariate data. Instead of considering $\mathbb{P}(X > u + z \,|\, X > u)$ directly, consider $\mathbb{P}(X_{/j} > u + z \,|\, X_j = x_j)$, for all $j \in \{1, \ldots, d\}$. To see why this is worthwhile, note that the latter can be used to find the former. For example, in the case $d = 2$,

$$\mathbb{P}(X > u + z \,|\, X > u) = \frac{1}{1 - F_2(u_2)} \int_{x_1 > u_1 + z_1} \mathbb{P}(X_2 > u_2 + z_2 \,|\, X_1 = x_1) f_1(x_1) \mathrm{d}x_1, \tag{2.2}$$

where F_2 is the distribution of X_2 and f_1 is the density of X_1. Assuming we have a model for $\mathbb{P}(X_2 > u_2 + z_2 \,|\, X_1 = x_1)$, then the RHS of (2.2) can be used to estimate the LHS. Whilst the denominator is also unknown, we can estimate it using univariate results such as (2.1), or with a simple empirical proportion. Similarly the integral is intractable, but can be estimated using Monte Carlo methods. All that remains is an approach to modelling such quantities as $\mathbb{P}(X_{/j} > u+z | X_j = x)$. We assume that there exist normalizing constants $a(x) > 0$, $b(x)$, and some non-degenerate G such that

$$\mathbb{P}(X_{/j} - b(x))/a(x) \leq z \,|\, X_j = x) \to G(z) \text{ as } x \to \infty. \tag{2.3}$$

This suggests that the following approximation is appropriate

$$\mathbb{P}\left(X_{/j} \leq y \middle| X_j = x\right) \approx G\left(\frac{y - b(x)}{a(x)}\right), \tag{2.4}$$

and indeed this is what Heffernan and Tawn (2004) studied. In practice, the functions a and b are not necessarily unique, and there can be several combinations of functions that yield non-degenerate limits in (2.3). This has been explored by Tendijck et al. (2021), and we can make (2.4) more general

with the following approximation

$$\mathbb{P}\left(X_{/j} \leq y | X_j = x\right) \approx \sum_{k=1}^{K} \pi_k G_k \left(\frac{y - b_k(x)}{a_k(x)}\right) \text{ for large } x, \text{ with } (\pi_1, \ldots, \pi_K) \in \mathcal{S}_K, \tag{2.5}$$

where K is the number of 'modes' (equivalently the number of unique combinations of suitable normalization functions), and $\mathcal{S}_K$ is the probability simplex with dimension $K-1$. We will focus heavily on fitting models of kind (2.5).

2.4 Fitting conditional extreme value mixture models (CEVMMs)

If all of K, a_k, b_k and G_k, for $k = 1, \ldots, K$, are known then fitting the model (2.5) is straightforward because the RHS becomes a simple mixture model of known distributions. Algorithmic options for fitting the mixture weights include

(a) EM — there are closed form solutions to the parameter updates for π, and EM is known to converge to the MLE for π (Redner and Walker, 1984);

(b) Bayesian methods — a Dirichlet prior for π yields closed form updates in a Gibbs sampler (Li et al., 2019) because it is a conjugate prior for the Multinomial distribution.

Unfortunately in practice, given real data, none of the parameters or distributions will be known. We will not even know the number of modes. There is a history of workarounds for this issue. In their original paper, Heffernan and Tawn (2004) proposed transforming data to Gumbel margins, and then using $a(x) = x^\beta$ and $b(x) = \alpha x + 1(\alpha = 0, \beta < 0) \cdot (\gamma - \delta \log(x))$ in (2.5). This parametric family was motivated by the fact that it fits with many types of dependence structures/copula. However, a similar, but simpler form for a and b was suggested in Keef et al. (2013). Using Laplace margins instead yields $a(x) = x^\beta$ and $b(x) = \alpha x$ for a large class of copula. Laplace margins have a further advantage over Gumbel margins in that an extreme value analysis can be performed easily in both the lower and the upper tails because the Laplace density is symmetric. Secondly, it has been common to use ECDF in place of the G_k's. It is not possible to fit α's, β's and π under this assumption, and so classically for the fitting step a false working assumption that the G's are normally distributed has been used. It is stressed that once estimates for the mixture weights and the normalization constants are found, the normal assumption is dropped. Putting these workarounds together yields the following model for the fitting step

$$\mathbb{P}\left(X_l \leq y | X_j = x\right) = \sum_{k=1}^{K} \pi_k \Phi\left(\frac{y - \alpha_k x}{x^{\beta_k}}; \mu_k, \sigma_k^2\right) \text{ for large } x,\ l \in \{1, \ldots, d\},\ l \neq j, \tag{2.6}$$

where Φ is the univariate normal CDF. For the $d = 2$ case, fitting (2.6) requires data $\mathcal{D}' = \{(x_i, y_i)\}_{i=1}^n$, where (x_i, y_i) are assumed to be i.i.d.. It is necessary to select a cut-off $u > 0$ and fit using $\mathcal{D} = \{(x, y) \in \mathcal{D}' \,|\, x > u\}$ so that the 'large x' constraint is satisfied. This is analogous to the choice of v in (2.1), and the selection of u also corresponds to a variance/bias trade-off. Once parameter estimates have been found, let $C_i \in \{1, \ldots, K\}$ be the classification of each pair $(x_i, y_i) \in \mathcal{D}$. By sampling C_i many times, we can construct sets of residuals $Z_1, \ldots, Z_K$, where $(y_i - \alpha_k x_i)/x_i^{\beta_k}$ is added to Z_k every time k is sampled from C_i. Finally, we can drop the normal assumption and construct ECDFs for each of these sets of residuals, yielding the final model

$$\mathbb{P}\left(X_l \leq y \,|\, X_j = x\right) \approx \sum_{k=1}^{K} \hat{\pi}_k \hat{F}^k_{\text{ecdf}}\left(\frac{y - \hat{\alpha}_k x}{x^{\hat{\beta}_k}}\right) \text{ for large } x. \tag{2.7}$$

The business of fitting model (2.6) is not simple — even though it is a mixture of Gaussians, it is not amenable to techniques for fitting Gaussian mixture models because the CDFs permit an argument that is non-linear in the data x. No closed form solution exists (nor are there a closed

form solutions to parameter updates if an EM algorithm is applied). Adaptive MCMC has been used previously (Tendijck et al., 2021), but here we consider an EM algorithm which is explained in depth in the appendix.

2.5 Tail switching

We briefly introduce the concept of tail switching, and discuss its relevance in terms of conditional extreme value models. We say that a time series $\{X_t\}_{t\in\mathbb{Z}}$ tail switches if for some fixed large threshold u, $X_t > u > X_{t+1}$. In other words, a series tail switches if it moves from the upper tail of its distribution to the main body. It is necessary to assume some degree of stationarity for X_t, as otherwise u may need to vary with t.

There are many examples of tail switching series in practice: see Figure 2.1 which shows several tail switches of S&P log returns (on the Laplace scale), along with the actual price. The log return transformation is necessary as it serves to residualize the data (so that it is stationary).

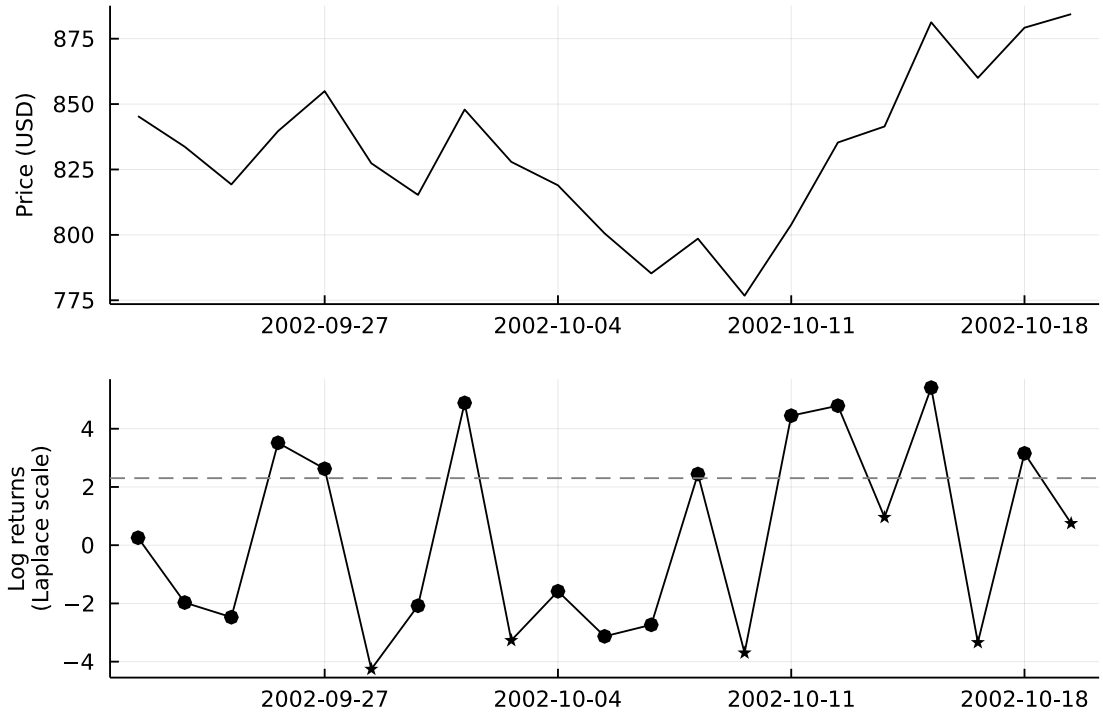

Figure 2.1: S&P 500 price and log returns on a Laplace scale, during a financially turbulent period in 2002. Tail switches across $u = F^{-1}_{\text{Laplace}}(0.95)$ (dashed line) denoted with $\star$.

Different aspects of tail switching have been studied. Bortot and Coles (2003) examine in particular not just the occurrences of tail switches, but tail switches from high extremes to low extremes (i.e. $X_t > u > 0 > u' > X_{t+1}$, where u' is small). This treatment is useful especially useful for finance, where such switches are common. Tail switching has also been explored in a multivariate setting, with multivariate Student-t models by Bernardi et al. (2013).

Crucially, conditional extreme value models can be applied to time series to characterize how they tail switch. The $K = 1$ case has previously been studied in the context of time series (Auld, 2021), and the $K > 1$ case has been studied in an environmental setting (Tendijck et al., 2021), but it is believed that an application of the $K > 1$ case in a time series context is novel. A model can be fit using (2.7) by constructing a dataset $\mathcal{D}' = \{(X_t, X_{t+1})\}_{t=1}^{n}$, and then fitting on $\mathcal{D} = \{(x, y) \in \mathcal{D}' | x > u\}$ for some large $u > 0$. It is assumed that $\{X_t\}$ is stationary, so that observations in $\mathcal{D}$ are (perhaps approximately) drawn i.i.d.. This assumption is an important ingredient so that model assumptions are met when fitting (2.7).

Chapter 3

Methods and results

We seek to investigate CEVMMs in the context of time series. In order to assess their efficacy, we begin by testing them on simulated data. We focus on data simulated from an asymmetric logistic distribution. The theoretical relevance of the asymmetric logistic distribution is that it is a multivariate extreme value distribution (in the sense that it the satisfies the conditions in Theorem 2.2). There are several further reasons for choosing this distribution: it permits several parameters, θ and v, allowing it to flexibly model a large range of possible copula (Coles and Tawn, 1991); it is workable analytically since its conditional distributions are known, and asymptotic properties relevant for our purposes have already been found.

By simulating asymmetric logistic data, we show the details of how to fit CEVMM models; consider the logistic distribution as an alternative marginal distribution; propose a method for regularizing the mixture likelihoods, and illustrate methods for quantifying uncertainty in CEVMM models with the moving block bootstrap.

Finally we use the methods and algorithms developed on the asymmetric logistic data to apply CEVMMs to UK electricity imbalance price data. Imbalance prices are seasonal, so we first perform a deseasonalization prior to model fitting. We show that there is an extremal dependence structure in the deseasonalized imbalance price, and demonstrate that CEVMM models can be useful for managing risk associated with the extremes of these residuals.

All the code written can be found on GitHub, see `https://github.com/lippirk/edi-diss`.

3.1 The 3-dimensional asymmetric logistic model

The 3-dimensional asymmetric logistic distribution with Fréchet margins is defined below via its cumulative distribution function.

$$\begin{aligned}
&F(x) = \exp\left(-V(x)\right),\ x \in \mathbb{R}^3_+,\\
&V(x) = \theta_0 x_0^{-1} + \theta_1 x_1^{-1} + \theta_2 x_2^{-1} + \theta_{01}\left\{(x_0^{-1/v_{01}} + x_1^{-1/v_{01}})^{v_{01}}\right\} + \theta_{02}\left\{(x_0^{-1/v_{02}} + x_2^{-1/v_{02}})^{v_{02}}\right\}\\
&\qquad + \theta_{12}\left\{(x_1^{-1/v_{12}} + x_2^{-1/v_{12}})^{v_{12}}\right\} + \theta_{012}\left\{(x_0^{-1/v_{012}} + x_1^{-1/v_{012}} + x_2^{-1/v_{012}})^{v_{012}}\right\},\\
&\text{subject to}\ \ \theta_0 + \theta_{01} + \theta_{02} + \theta_{012} = 1,\ \theta_1 + \theta_{01} + \theta_{12} + \theta_{012} = 1,\ \theta_2 + \theta_{01} + \theta_{02} + \theta_{012} = 1,\\
&\qquad 0 < v_{01}, v_{02}, v_{12}, v_{012} < 1,\ 0 < \theta_0, \theta_1, \theta_2, \theta_{01}, \theta_{02}, \theta_{12}, \theta_{012}.
\end{aligned} \tag{3.1}$$

When referring to 'θ' and 'v' we mean $(\theta_0, \theta_1, \theta_2, \theta_{01}, \theta_{02}, \theta_{12}, \theta_{012})$ and $(v_{01}, v_{12}, v_{012})$ respectively. In terms of notation, we will sometimes write, for example, $\theta_{/012}$ to mean $(\theta_0, \theta_1, \ldots, \theta_{12})$, i.e. all the θ's except θ_{012}.

The equality constraints on θ ensure that the marginals of F are unit Fréchet. Rather than working directly with (3.1), we use it to construct an order-2 Markov process (3.2), given by

$$\begin{aligned}
&X_0 \sim F_0,\\
&X_1 \,|\, X_0 = x_0 \sim F_{1|0}(\cdot; x_0),\\
&X_{t+2} \,|\, X_{t+1} = x_{t+1},\ X_t = x_t \sim F_{2|1,0}(\cdot; x_{t+1}, x_t),
\end{aligned} \tag{3.2}$$

where $F_{1|0}$ and $F_{2|1,0}$ are the conditionals of F. A stationary process can be specified by imposing the constraint $\mathbb{P}(X_t \le x, X_{t+1} \le y) = \mathbb{P}(X_{t+1} \le x, X_{t+2} \le y),\ \forall x, y > 0$, which implies $\theta_{01} = \theta_{12}$ and $v_{01} = v_{12}$.

Using inverse sampling, it is possible to draw from (3.2). Notice how decreasing the v's and increasing θ_{01}, θ_{02}, θ_{012} increases dependence between the r.v.s X_{t+2}, X_{t+1} and X_t.

3.1.1 Theoretical Extremal Dependence

We are interested in particular in the extremal dependence of X_0, X_1 and X_2. Since we know the conditionals $F_{1|0}$ and $F_{2|1,0}$ in (3.2), we can calculate the dependence directly. However, we also want the normalization constants necessary for limits such as (2.3) to exist, as then we can compare algorithmic results to what is expected asymptotically. These results exist already for $F_{1|0}$ and $F_{2|1,0}$, with X_i's in exponential margins (Papastathopoulos et al., 2017; Papastathopoulos and Tawn, 2019). These are summarised by

$$\begin{aligned}
&\mathbb{P}(X_2 \leq z | X_1 = x) \to_{x\to\infty} H_1(z),\\
&\mathbb{P}(X_2 \leq z | X_0 = x) \to_{x\to\infty} H_2(z),\\
&\mathbb{P}(X_2 \leq z | X_0 = x, X_1 = y) \to_{x\to\infty, y\to\infty} H_3(z),\\
&\mathbb{P}(X_2 - x \leq z | X_1 = x) \to_{x\to\infty} H_4(z),\\
&\mathbb{P}(X_2 - x \leq z | X_0 = x) \to_{x\to\infty} H_5(z) \text{ and}\\
&\mathbb{P}(X_2 + v_{012}\log(\exp(-x/v_{012}) + \exp(-y/v_{012})) \leq z | X_0 = x, X_1 = y) \to_{x\to\infty, y\to\infty} H_6(z).
\end{aligned} \tag{3.3}$$

The distributions H_i are not particularly important for our purposes — all that matters here is that they are non-degenerate. For consistency with the results from Keef et al. (2013), we would ideally work in Laplace margins. Fortunately the results from (3.3) carry over easily from exponential to Laplace margins. Note that

$$\left(F^{-1}_{\text{Laplace}} \circ F_{\text{exponential}}\right)(x) = x - \log 2, \text{ whenever } x > \log 2 = F^{-1}_{\text{exponential}}(0.5). \tag{3.4}$$

Equation (3.4) implies that the limits associated with $H_1, \ldots, H_6$ in expression (3.3) also hold in Laplace margins. This allows us to construct a CEVMM that captures the asymptotic extremal dependence between X_{t+1} and X_t with

$$\mathbb{P}\left(X_{t+1} \leq z | X_t = x\right) \approx \pi G_1(z - x) + (1 - \pi) G_2\left(z\right) \text{ for large } x. \tag{3.5}$$

We can also construct a CEVMM for the capturing the asymptotic extremal dependence between X_t, X_{t+1} and X_{t+2} with

$$\begin{aligned}
\mathbb{P}\left(X_{t+2} \leq z | X_t = x, X_{t+1} = y\right) \approx\ & \pi_1 G_1(z) + \pi_2 G_2\left(z\right) + \pi_3 G_3\left(z\right) + \pi_4 G_4(z - y) + \pi_5 G_5(z - x)\\
&+ \pi_6 G_6(z + v_{012}\log[\exp(-x/v_{012}) + \exp(-y/v_{012})]),
\end{aligned} \tag{3.6}$$

where $(\pi_1, \ldots, \pi_6) \in \mathcal{S}_6$, and at least one of x or y is large. The G_k's in (3.6) correspond to the H_k's in (3.3). In general the mixture probabilities π_k may depend on x and y. For example, if both x and y are very large, then we ought to have π_3, π_6 non zero, with the other mixtures negligible. Figure 3.1 depicts realizations of X_{t+1} and X_t, conditional on X_t large. Adopting notation from Keef et al. (2013) for the normalization functions ($a_k(x) = x^{\beta_k}$, $b_k(x) = \alpha_k x$) they are characterized by $(\alpha_1, \beta_1) = (1, 0)$ and $(\alpha_2, \beta_2) = (0, 0)$, and so we see straight lines. We will fit CEVMMs on this data. Notice that variance around the modes increases with v. This is due to v acting as a dependence parameter (high v meaning low dependence and vice versa). It follows that fitting CEVMMs will be easier to fit when relevant variables have high dependence, and at the opposite extreme, CEVMMs will not be a sensible model if v is close to 1, since in the $v = 1$ case the variables are independent.

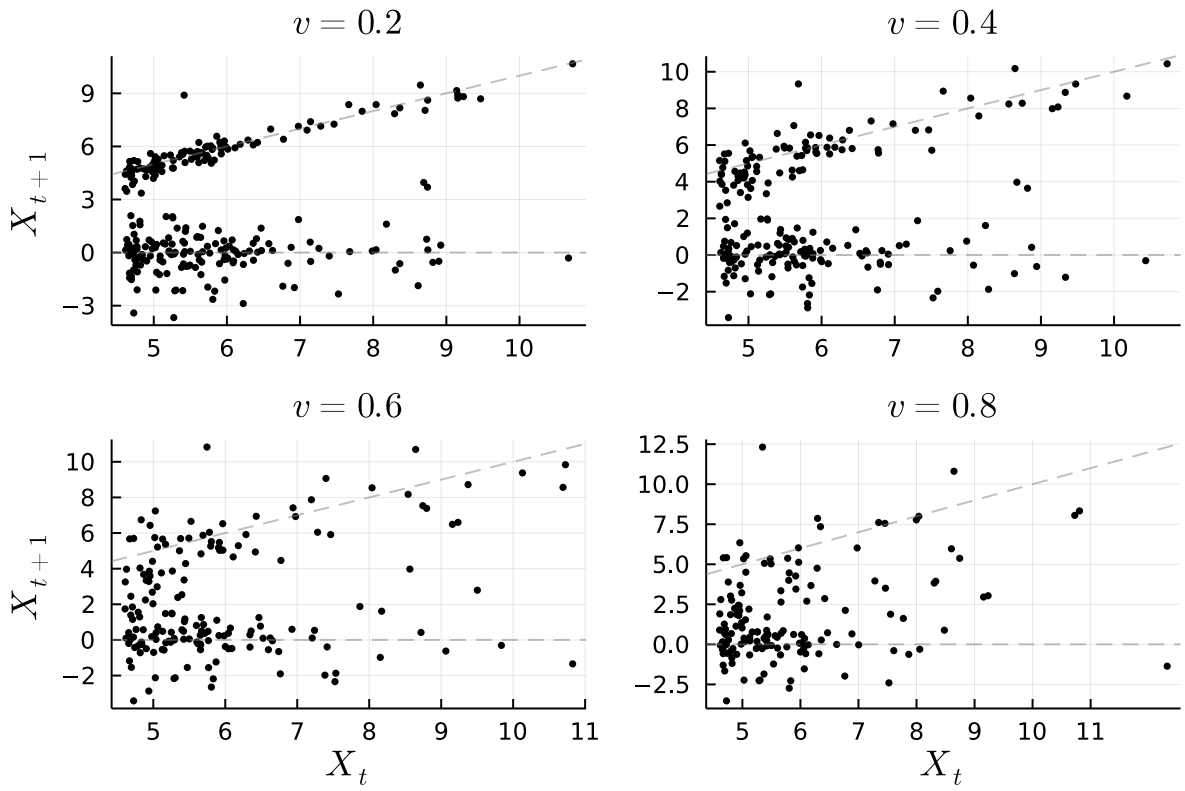

Figure 3.1: X_{t+1} vs. X_t, on unit Laplace scale, for $\theta_{/012} = 0.3$ and different values of v (using the same random seed). The theoretical modes are shown in grey.

3.1.2 Fitting conditional extreme value mixture models — a worked example

We begin with a worked example of how to fit CEVMMs to asymmetric logistic data. First, we generate data from (3.2). To avoid the need to generate copious amounts of data, we produce 100 independent realizations of the Markov process, each of length 30, crucially conditioning on $X_0 > q = F^{-1}_{\text{Fréchet}}(0.99)$ for each of the 100 chains to ensure several extreme values. Let X^j be the jth chain, then we form the datasets $\mathcal{D}' = \{(X^j_t, X^j_{t+1}) | j = 1, \dots, 100, t = 0, \dots, 28\}$, and $\mathcal{D} = \{(x, y) \in \mathcal{D}' | x > u\}$, where $u = 3 \approx F^{-1}_{\text{Laplace}}(0.975)$. The size of $\mathcal{D}$ is 460. Figure 3.2 shows the data generated, with $\mathcal{D}'$ on the left and $\mathcal{D}$ on the right. The parameters used were $\theta_0 = \theta_1 = \theta_2 = 0.1$, $\theta_{01} = \theta_{02} = \theta_{012} = 0.3$ and $v_{01} = v_{02} = v_{012} = 0.5$.

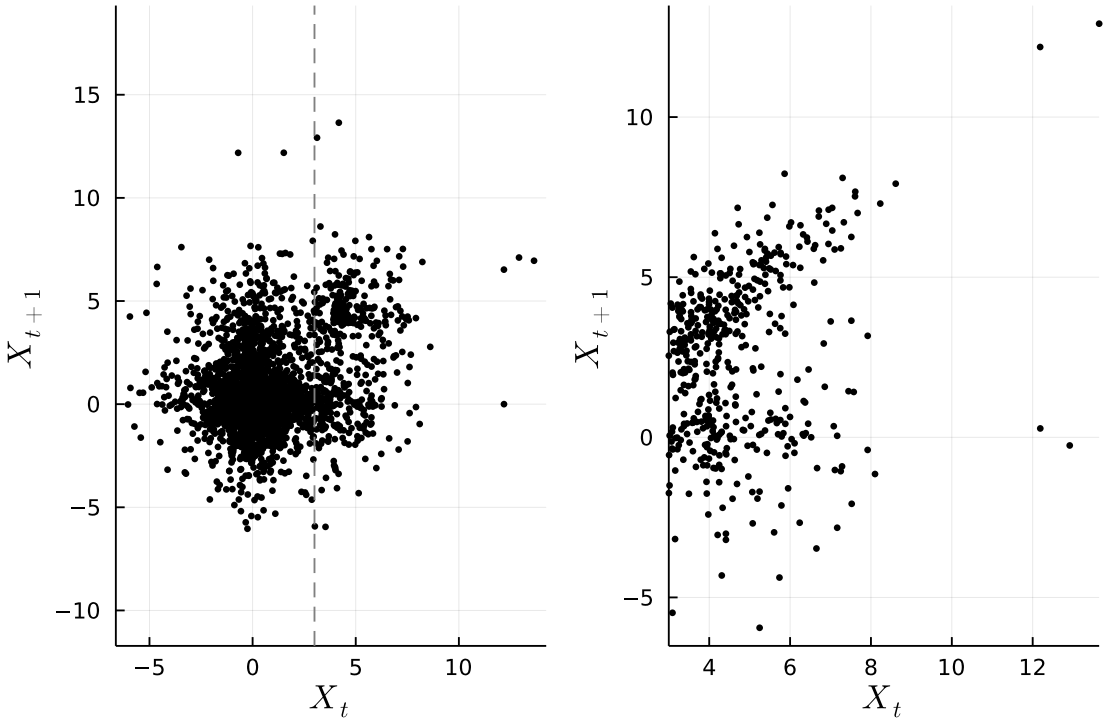

Figure 3.2: Data used for the worked example, in Laplace margins. The threshold u is shown as a dashed grey line on the left plot.

For fitting, we use the normal approximations as in (2.6), and pretend that the normalizing functions are unknown. The latter is both pedagogical and important if we wish to fit these models to general data. In other words, we are fitting $\theta = (\pi, \alpha_1, \alpha_2, \beta_1, \beta_2, \mu_1, \mu_2, \sigma_1, \sigma_2)$ in the model

$$\mathbb{P}\left(X_{t+1} \leq z | X_t = x\right) = \pi \Phi\left(\frac{z - \alpha_1 x}{x^{\beta_1}}; \mu_1, \sigma_1^2\right) + (1 - \pi)\Phi\left(\frac{z - \alpha_2 x}{x^{\beta_2}}; \mu_2, \sigma_2^2\right).$$

Note that we are making an assumption that each data point is i.i.d.. We use the EM algorithm A.1 to fit the parameters in this model. The parameter space is slightly awkward because of the constraints

$$0 \leq \pi \leq 1, -1 \leq \alpha_1 \leq \alpha_1 \leq 1, 0 \leq \beta_1, \beta_2 \leq 1, 0 < \sigma_1, \sigma_2. \tag{3.7}$$

The reason for constraints on π and σ should be clear. The bounds on the α's and β's are from Keef et al. (2013), but we impose $\alpha_1 \geq \alpha_2$ to ensure identifiability as in Tendijck et al. (2021). Numerically it is helpful to perform parameter transformations to allow us to work over $\mathbb{R}^9$, unconstrained. The

following transformations are proposed,

$$\begin{aligned}
\pi'_k &= \text{logit}(\pi_k),\ \pi_k = \text{sigmoid}(\pi'_k),\\
\beta'_k &= \text{logit}(\beta_k),\ \beta_k = \text{sigmoid}(\beta'_k),\\
\sigma'_k &= \log(\sigma_k),\ \sigma_k = \exp(\sigma'_k),\\
\alpha'_1 &= \text{logit}\left(\frac{\alpha_1+1}{2}\right),\ \alpha_1 = 2\text{sigmoid}(\alpha'_1) - 1,\\
\alpha'_2 &= \text{logit}\left(\frac{\alpha_2+1}{\alpha_1+1}\right),\ \alpha_2 = (\alpha_1+1)\text{sigmoid}(\alpha'_2) - 1.
\end{aligned} \tag{3.8}$$

In the transformed space, we take advantage of an implementation of the L-BFGS algorithm from Mogensen and Riseth (2018), which allows for fast model fitting.

We fit several models with EM using different (random) initial values for θ. Some results are shown in Table 3.1 (omitting the μ and σ values). We see that the initial value of θ greatly affects the fitting of α and β — this is likely because we are optimizing over a highly multimodal landscape. According to theory, we are expecting $(\alpha_1, \alpha_2, \beta_1, \beta_2) = (1, 0, 0, 0)$. We see some poor fits (e.g. seed 1), but thankfully the estimates with the highest likelihood are close to the expected values.

Seed	$\hat{\pi}$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	Log Likelihood
1	0.48	1.00	-1.00	0.89	0.95	-1002.45
2	0.51	0.93	-0.08	0.00	0.14	-985.77
3	0.03	1.00	-1.00	0.00	0.93	-1056.30
4	0.49	-0.12	-1.00	0.02	0.94	-995.32
5	0.49	-0.10	-0.10	0.00	0.89	-994.66
6	0.51	0.93	-0.08	0.00	0.14	-985.77
7	0.47	1.00	1.00	0.88	1.00	-1003.74
8	0.49	-0.12	-1.00	0.02	0.94	-995.32
9	0.51	0.93	-0.08	0.00	0.14	-985.77
10	0.51	1.00	-0.07	0.00	0.00	-987.19

Table 3.1: Parameter estimates and log-likelihoods, to 2 decimal places, obtained when fitting CEVMMs to asymmetric logistic data, with starting values initialized using different random seeds.

We continue our analysis with the parameters that induce the highest likelihood, i.e. seed 2 (though seeds 2, 6 and 9 yield equivalent results). We can examine the fit visually in Figures 3.3 and 3.4. Figure 3.3 shows the data used to fit the model, colour coded according to which mixture is more likely. Contours of the fitted normal mixture model are shown conditional on the value of X_t, such that taking a vertical cross-section yields a probability density. Figure 3.4 shows four of these cross sections, along with their expected probability densities. Here the expected density is calculated using $F_{1|0}$, by taking the appropriate partial derivative (using the chain rule to account for the change to Laplace margins). The model seems to have captured the mode corresponding to $\alpha = 1$ quite well, but performs poorly at the mode corresponding to $\alpha = 0$. One reason for this is that it is impossible for a normal distribution to properly approximate the discontinuity in the gradient of the true conditional density at 0.

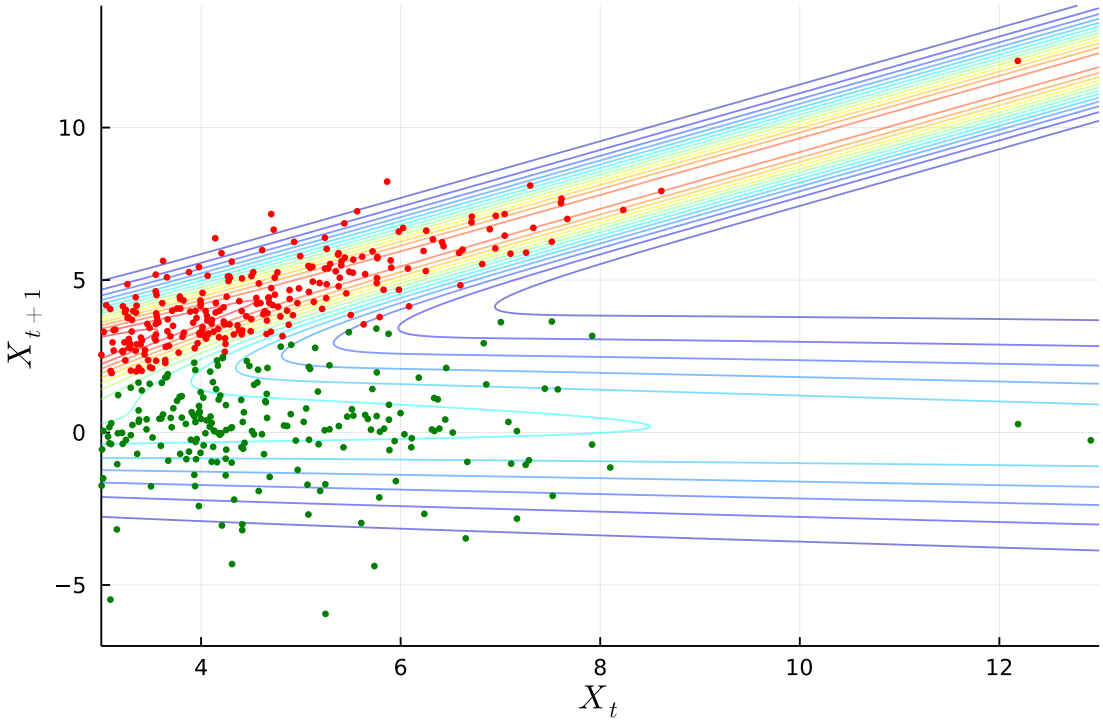

Figure 3.3: X_{t+1} vs. X_t, with each data point colour coded according to its more likely mode (a.k.a mixture component). The coloured lines show contours of the modelled normal mixture density for $X_{t+1} \,|\, X_t$, when viewed as a function of both X_{t+1} and X_t.

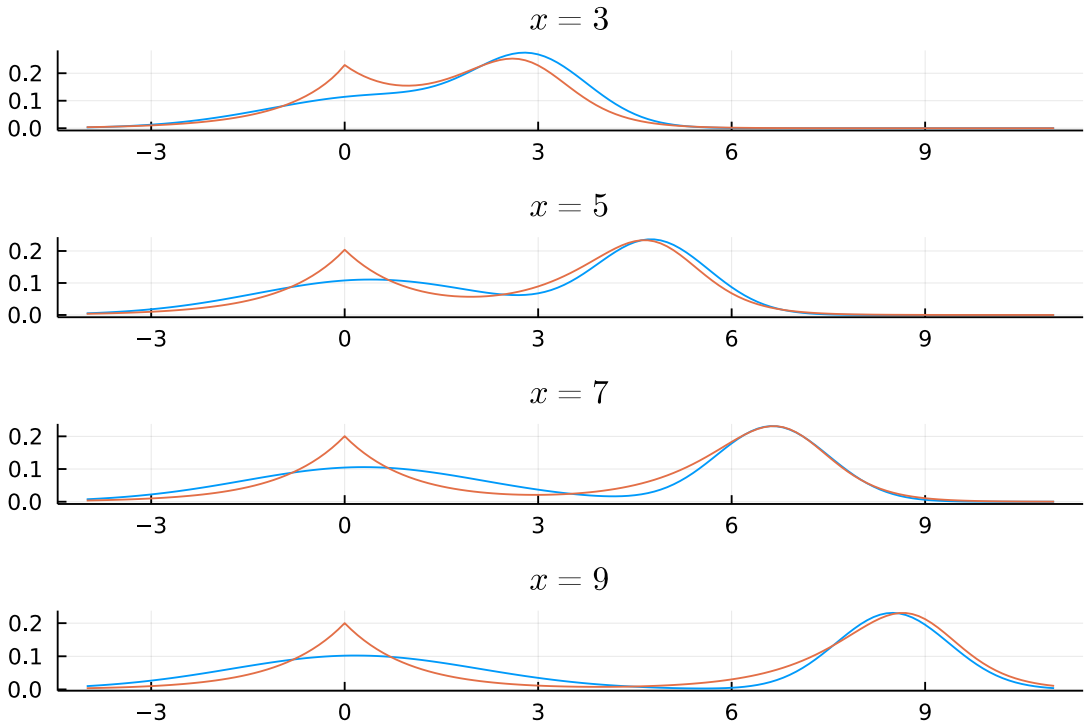

Figure 3.4: Probability densities of $X_{t+1}|X_t = x$ for 4 fixed values of x. The expected densities are shown in orange, and fitted densities in blue. These plots can be thought of as vertical cross sections of Figure 3.3 at $X_t = x$.

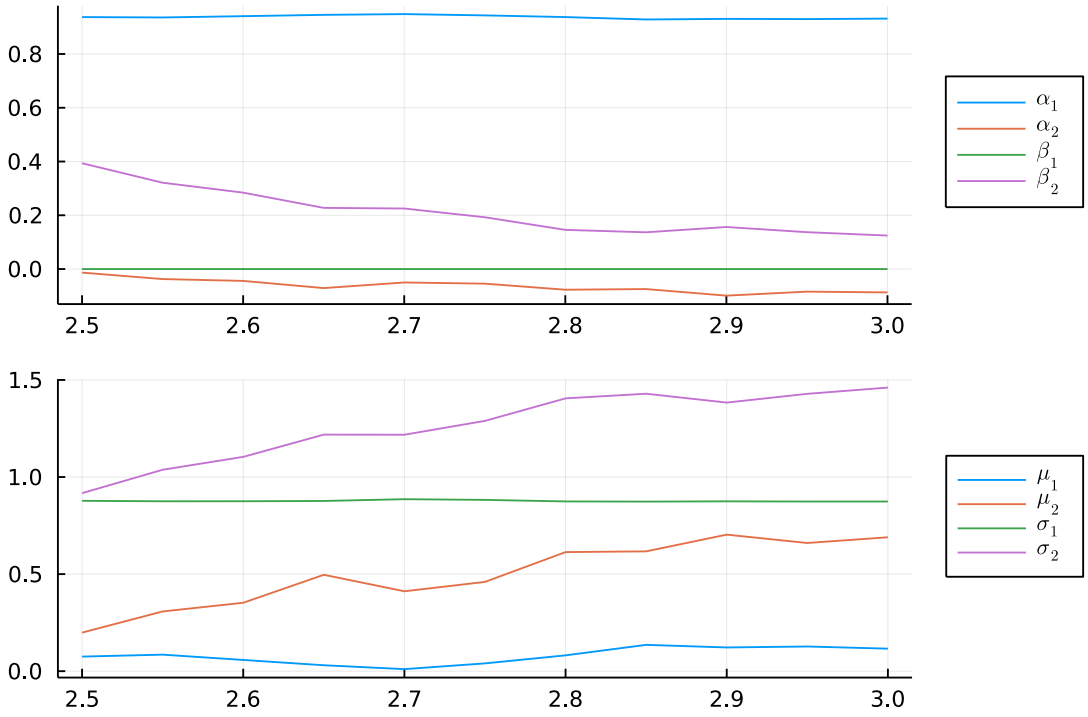

Figure 3.5: Effect of decreasing u (down from $u = 3$) on the fitted parameters.

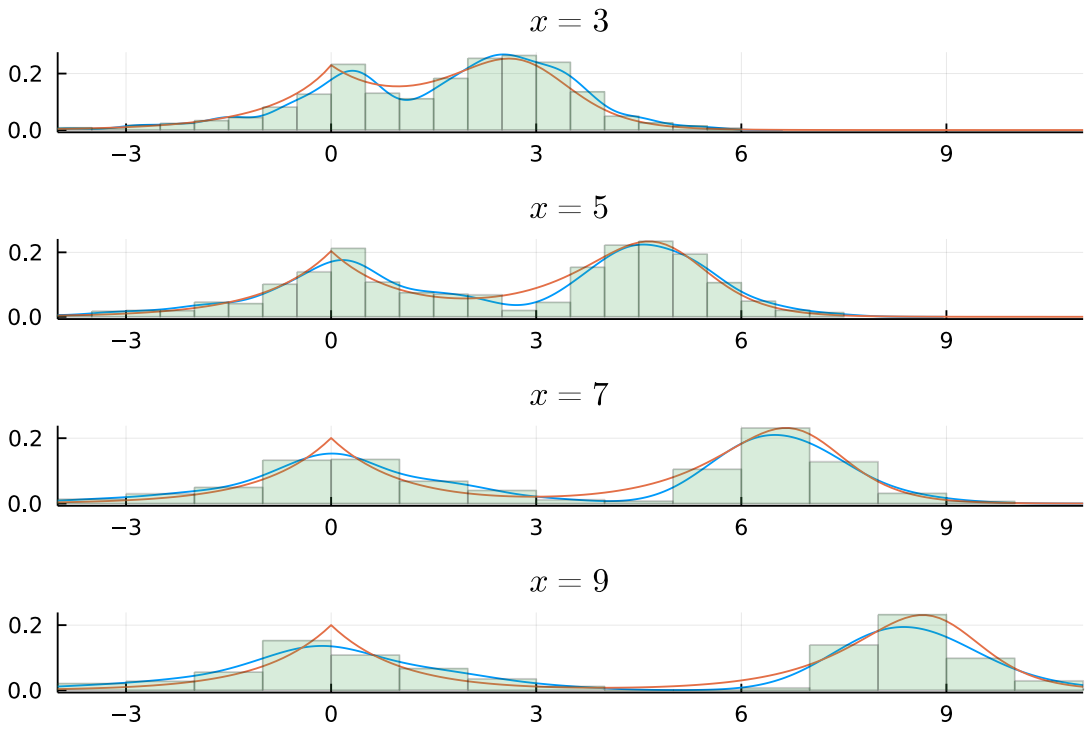

Figure 3.6: Probability densities of $X_{t+1}|X_t = x$ for fixed values of x (Laplace margins). Expected densities are shown in orange. 10000 samples were drawn from the estimated distribution function; the associated Gaussian kernel density estimators for these samples are shown in blue, with histograms of the samples in green.

It is expedient to reduce u as much as possible to reduce the variance of the ECDF estimators. This can be achieved by reducing u until the parameters $(\alpha_1, \alpha_2, \beta_1, \beta_2, \mu_1, \mu_2, \sigma_1, \sigma_2)$ begin to change (Tendijck et al., 2021). With the aid of Figure 3.5, which shows the effect of u on the fitted parameters, we choose $u = 2.9$. The rationale here is that both β_2, σ_2 and μ_2 begin to change for u below 2.9.

Finally, we sample residuals (this is explained in Appendix A.2, here we sample each classification $M = 500$ times) to obtain the semi-parametric model in the form (2.7). Comparing the semi-parametric model fit (Figure 3.6) to the fit of the model with normal assumptions (Figure 3.4), we see that moving to use residuals has somewhat remedied the issue of poor fit at the lower mode, since an ECDF function is more easily able to capture the discontinuity in the gradient of the expected density. There is, however, still a bias, perhaps induced by the fact that the theoretical values for α and β were not obtained exactly during the fitting process.

From the above, there are a few things that should be taken into account when fitting these models. The EM algorithm is sensitive to initial values. Here we fit several models, and selected the one with the highest likelihood, but in general this may lead to overfitting. Sensitivity may be further exacerbated if we tried to model $K \geq 3$ modes. We work around this issue practically later when establishing uncertainty by adding regularization (see Section 3.1.4). Also, the best fitted model in terms of log-likelihood found biased values for the normalizing function constants, which possibly induces the slight bias in the modes we see in Figure 3.6. It is unclear exactly the cause of this bias, but potential reasons include an insufficiently large threshold or overfitting.

Overall, we can expect the EM algorithm to perform a reasonable fit, since $(\hat{\alpha}_1, \hat{\beta}_1) = (0.93, -0.08)$ and $(\hat{\alpha}_2, \hat{\beta}_2) = (0.00, 0.14)$ are reasonably close to the true values.

3.1.3 A look at logistic margins

The cause of this undesirable 'kink' in the true conditional density in Figure 3.4 is due to definition of the Laplace density. The logistic distribution has similar asymptotic properties to the Laplace distribution in the upper and lower tails (both tails are symmetric and are exponential). In particular, the results from (3.4) hold in logistic margins. Crucially, however, the logistic density is smooth, meaning that the conditionals are more likely to be easily modelled by a mixture of Gaussians. Repeating the above with the same data, but with logistic margins and starting with $u = 3.65 \approx F^{-1}_{\text{logistic}}(0.975)$ (for fair comparison), yields similar parameter estimates $(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\beta}_1, \hat{\beta}_2) = (0.93, -0.08, 0.00, 0.17)$, but more convincing looking normal mixture densities (see Figure 3.7), especially at the lower mode. However, it is not clear whether moving to logistic margins actually improves the semi-parametric model, since Figures 3.8 and 3.6 are very similar. Notice that the logistic version suffers from the same bias problem as the Laplace version.

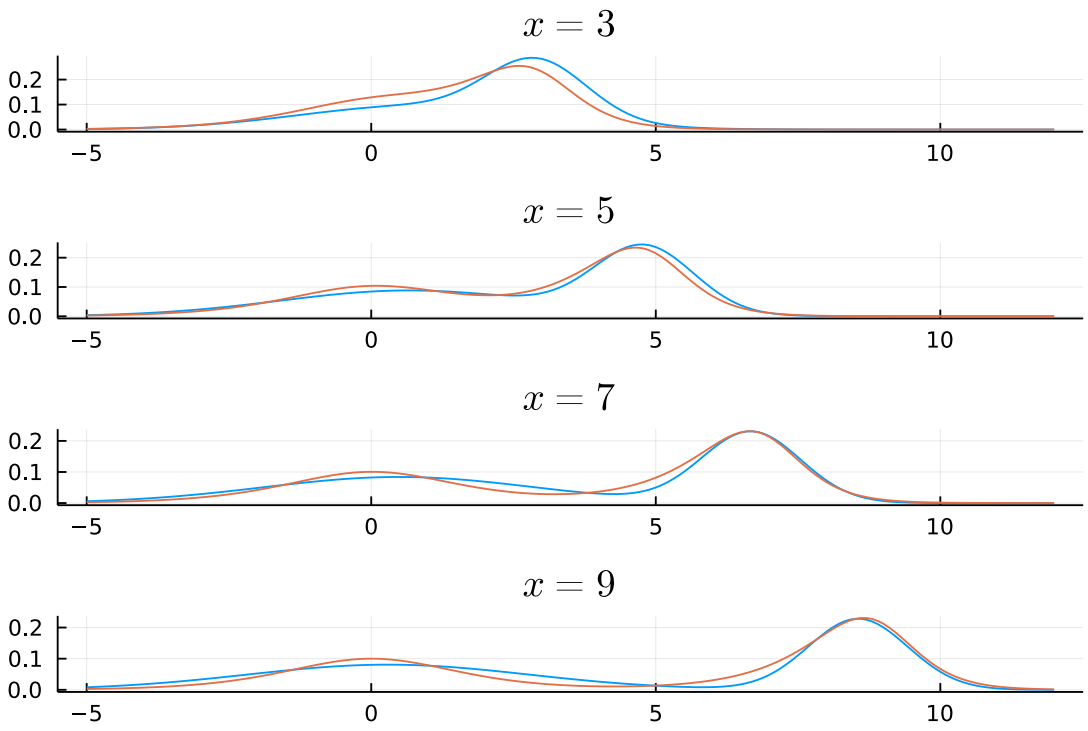

Figure 3.7: Probability densities of $X_{t+1}|X_t = x$ for fixed values of x (logistic margins). True densities are shown in orange, and fitted densities in blue.

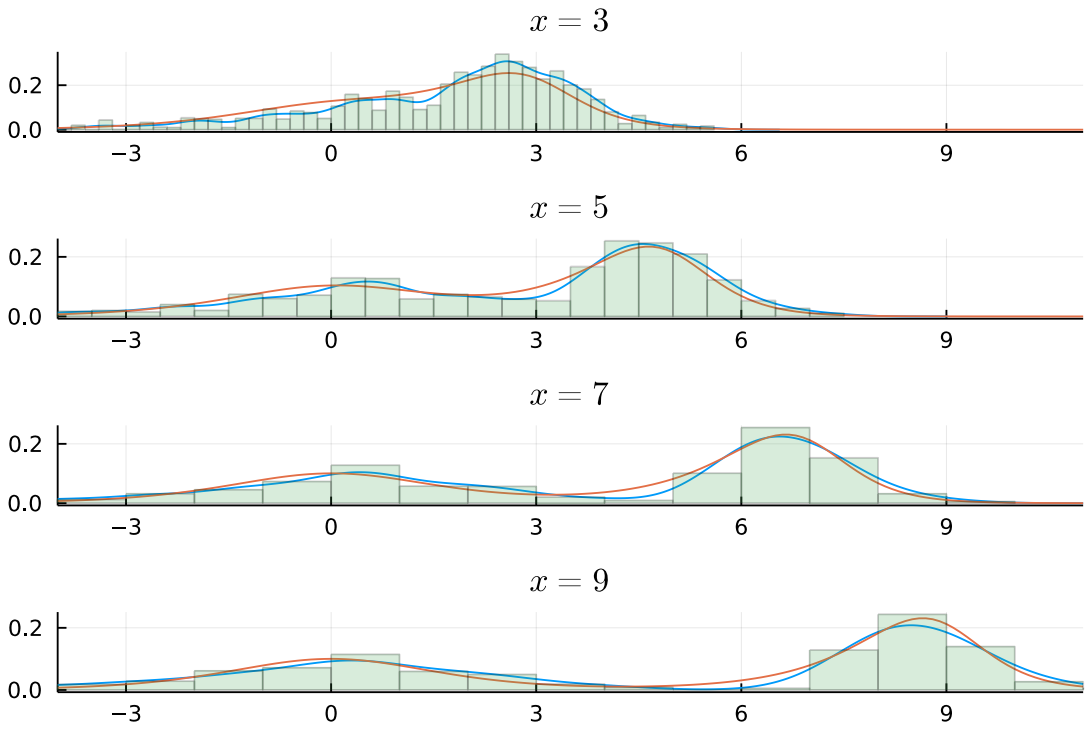

Figure 3.8: Probability densities of $X_{t+1}|X_t = x$ for fixed values of x (logistic margins). True densities are shown in orange. 10000 samples were drawn from the estimated distribution function; the associated Gaussian kernel density estimators for these samples are shown in blue, with histograms of the samples in green.

Simulation study to compare Laplace and logistic margins

To quantitatively examine the effect of the choice of marginal distribution on the estimated conditional $\hat{F}_{1|0}$, we perform a simulation study comparing Laplace and logistic margins. Specifically we generate 100 datasets (each) for different combinations of u, v and θ, fitting CEVMMs and recording integrated absolute and square errors with respect to the expected distribution. Mean errors (as well as their standard errors) can then be calculated by averaging over the 100 datasets. Formulae for these metrics are given by

$$\begin{aligned} \text{MISE}_{x_0} &= \frac{1}{100}\sum_{d=1}^{100}\int_{\mathbb{R}}\left(F_{1|0}(x;x_0) - \hat{F}_{1|0}^{(d)}(x;x_0)\right)^2 \mathrm{d}x, \\ \text{MIAE}_{x_0} &= \frac{1}{100}\sum_{d=1}^{100}\int_{\mathbb{R}}\left|F_{1|0}(x;x_0) - \hat{F}_{1|0}^{(d)}(x;x_0)\right| \mathrm{d}x, \end{aligned} \tag{3.9}$$

where $\hat{F}_{1|0}^{(d)}$ is the estimated conditional distribution on the dth dataset (out of 100), and the quantities MISE_{x_0} and MIAE_{x_0} also implicitly depend on v, u and θ. By varying x_0 in (3.9), we test the estimators' efficacy conditional on different values. Varying the parameters v and θ, allows us to test across different types of datasets. We vary the level of dependence in the data (high, medium and low via v: $v = 0.3$; $v = 0.5$; $v = 0.7$, and low and high via θ: $\theta_{\{0\},\{1\},\{2\}} = 0.1, \theta_{/\{0\},\{1\},\{2\}} = 0.3$; $\theta_{/012} = 0.3$), and also the threshold for computing the dataset $\mathcal{D}$ via $u = F_0^{-1}(q_u)$. All models were fit using the same initial parameter estimates. A subset of the mean integrated square error results, with low dependence via θ, are shown in Table 3.2.

(v, q_u)	Marginal	$100\text{MISE}_{0.97}$	$100\text{MISE}_{0.98}$	$100\text{MISE}_{0.99}$	$100\text{MISE}_{0.999}$	$100\text{MISE}_{0.9999}$
(0.3, 0.95)	Laplace	$\mathbf{0.4 \pm 0.2}$	$\mathbf{0.4 \pm 0.2}$	0.4 ± 0.3	0.5 ± 0.5	0.8 ± 0.9
	Logistic	0.4 ± 0.2	0.4 ± 0.2	$\mathbf{0.4 \pm 0.2}$	$\mathbf{0.4 \pm 0.3}$	$\mathbf{0.6 \pm 0.6}$
(0.3, 0.98)	Laplace	$\mathbf{0.4 \pm 0.2}$	$\mathbf{0.4 \pm 0.3}$	0.4 ± 0.4	0.6 ± 0.7	1.0 ± 1.1
	Logistic	0.4 ± 0.3	0.4 ± 0.3	$\mathbf{0.4 \pm 0.3}$	$\mathbf{0.5 \pm 0.6}$	$\mathbf{0.8 \pm 1.1}$
(0.5, 0.95)	Laplace	$\mathbf{0.7 \pm 0.4}$	$\mathbf{1.0 \pm 0.5}$	$\mathbf{1.7 \pm 0.9}$	$\mathbf{4.7 \pm 2.9}$	$\mathbf{8.1 \pm 5.5}$
	Logistic	1.1 ± 0.7	1.5 ± 1.0	2.4 ± 1.8	5.8 ± 5.2	9.7 ± 9.1
(0.5, 0.98)	Laplace	$\mathbf{0.7 \pm 0.3}$	$\mathbf{1.1 \pm 0.4}$	$\mathbf{1.9 \pm 0.9}$	$\mathbf{5.5 \pm 3.2}$	$\mathbf{9.6 \pm 6.2}$
	Logistic	1.1 ± 0.6	1.6 ± 0.9	2.7 ± 1.7	6.9 ± 5.1	11.7 ± 9.0
(0.7, 0.95)	Laplace	$\mathbf{0.4 \pm 0.3}$	$\mathbf{0.6 \pm 0.4}$	$\mathbf{1.1 \pm 0.6}$	$\mathbf{4.0 \pm 2.3}$	$\mathbf{9.2 \pm 5.4}$
	Logistic	0.9 ± 0.4	1.6 ± 0.6	3.0 ± 1.1	11.1 ± 4.0	22.5 ± 8.3
(0.7, 0.98)	Laplace	$\mathbf{0.4 \pm 0.2}$	$\mathbf{0.6 \pm 0.3}$	$\mathbf{1.1 \pm 0.6}$	$\mathbf{4.5 \pm 2.8}$	$\mathbf{10.4 \pm 6.4}$
	Logistic	0.9 ± 0.3	1.5 ± 0.5	2.9 ± 1.1	10.9 ± 4.3	22.4 ± 8.8

Table 3.2: Mean integrated square errors ± 1 s.d. for Laplace and logistic marginals, and for different parameter combinations. MISEs are multiplied by 100 for readability.

The result with the better MISE (Laplace vs. logistic) for a given $v, q_u = F_0(u)$ is shown in bold. The estimators fit in Laplace margins appear to perform better than the estimators fit in logistic margins for lower x_0, but the performance gap decreases for larger x_0. The logistic margin's improvement for higher x_0 may be due to the fact that the exponential tail approximation improves for larger x_0. The MIAE results yield similar conclusions.

Overall, Laplace margins perform better. They have better mean errors (MISE and MIAE) 82.5% of the time (averaged over all variables).

3.1.4 Regularization

Changing initial parameter estimates can drastically affect the resulting model fit of a CEVMM (e.g. see Table 3.1). This is expected if initial parameter estimates are far from the true values, but we want to avoid this instability otherwise. We may also encounter instability when fitting if the data

changes slightly, which is in effect what we do in Section 3.1.5 when bootstrapping new datasets from an existing one. Regularization is the proposed solution to avoid such instability. It is possible to regularize any and every parameter in a model, but it was found by experimentation on asymmetric logistic datasets that regularization in the σ parameters is sufficient to allow for stable model fitting. Without this, at least in the $K = 2$ case, the fitting algorithm sometimes has a tendency to either remove one of the mixture components (by sending π_1 or π_2 to zero, and making the corresponding variance large), or alternatively to assign a very small variance to one of the modes. Neither of these are desirable when we specifically want 2 mixtures. As such the following adjusted log-likelihood is proposed,

$$l_{\text{regularized}}(\theta) = l(\theta) - \gamma \sum_{k=1}^{K} (\sigma_k^{-2} + \sigma_k^2/C), \ \gamma > 0, \ C > 0,$$

with the σ_k^{-2} term designed to discourage small variances, and the σ_k^2/C term to discourage variances from becoming too large. In our implementation, we fixed $C = 20$.

We also generally preferred $\gamma = 0.05$, motivated by a quick experiment. Three datasets were generated, similar to the one described in Section 3.1.2, but each has a different level of dependence ($v = 0.3$, $v = 0.5$ or $v = 0.7$, but $\theta = (0.1, 0.1, 0.1, 0.3, 0.3, 0.3)$ held constant throughout). The threshold was fixed as $u = F^{-1}_{\text{Laplace}}(0.95)$. CEVMM models were then fit to each dataset, indexed by v for different values of $\gamma \in \{0, 0.005, 0.05, 0.5, 5\}$. The results in Table 3.3 show that $\gamma = 0.05$ tended to perform better, both in terms of fitted parameters matching their expected theoretical values, and in terms of log-likelihood (high is better).

v	γ	$\hat{\pi}$	$\hat{\alpha_1}$	$\hat{\alpha_2}$	$\hat{\beta}_1$	$\hat{\beta}_2$	Log likelihood
0.3	0.000	0.51	1.00	-1.00	0.71	0.85	-1789.00
0.3	0.005	0.52	0.96	-1.00	0.62	0.82	-1788.52
0.3	0.050	0.57	1.00	-0.13	0.00	0.00	-1497.89
0.3	0.500	0.60	1.00	-0.12	0.00	0.00	-1612.62
0.3	5.000	0.63	1.00	-0.34	0.00	0.00	-1752.90
0.5	0.000	0.60	1.00	-1.00	0.00	0.97	-1280.03
0.5	0.005	0.62	1.00	-0.01	0.00	0.00	-1263.10
0.5	0.050	0.62	1.00	-0.02	0.00	0.00	-1264.42
0.5	0.500	0.62	1.00	-0.23	0.00	0.00	-1306.99
0.5	5.000	0.66	1.00	-0.54	0.00	0.00	-1354.92
0.7	0.000	0.49	1.00	-1.00	0.68	0.90	-831.60
0.7	0.005	0.49	1.00	-1.00	0.55	0.83	-831.59
0.7	0.050	0.47	1.00	-0.16	0.00	0.00	-816.93
0.7	0.500	0.47	1.00	-0.22	0.00	0.00	-820.90
0.7	5.000	0.49	1.00	-0.26	0.00	0.00	-827.02

Table 3.3: Fitted parameters and log-likelihoods for CEVMM fits on the three datasets described above, for different levels of regularization γ.

3.1.5 Quantifying error in CEVMM models with bootstrapped confidence intervals

We briefly show that it is possible to produce confidence intervals for a CEVMM's parameters and its predictions. Due to the constraints on the parameter space for CEVMMs (see (3.7) for the constraints in the $K = 2$ case), we cannot construct asymptotic confidence intervals. The reason for this is that the usual regularity conditions, sufficient for asymptotic normality of maximum likelihood estimates, do not necessarily hold — in particular our estimates may lie on the boundary of the parameter space. In order to establish confidence intervals for a CEVMM model we must try to quantify sources of error directly. These sources include error in the parameter estimates $\theta = (\pi, \alpha, \beta, \mu, \sigma)$; error due to

the choice of threshold u (is u large enough for the asymptotic approximation to be valid but small enough such that variance of the model parameters is low?), and error due to false assumptions (e.g. the i.i.d. assumption not being met).

It is possible to estimate uncertainty due to error in the parameter estimates using the bootstrap (Efron and Tibshirani, 1994). It would be difficult to estimate the error due to u; it may be possible via a combination of the bootstrap and varying u, but this is likely to be very computationally intensive, and also it would hard to decouple from the estimated error in the parameters. Estimating error due to a false assumption is incredibly difficult, since if an assumption has been broken it is likely that the model is invalid.

As such, we concentrate on only uncertainty due to the model parameters. From a high level this means that we run Algorithms 1 and 2 on many bootstrapped datasets to obtain many models. Then we can construct empirical distributions, and hence confidence intervals, for a quantity of interest by computing it with each model. Note that there is unfortunately no guarantee that any resulting empirical distribution will be accurate. We demonstrate the procedure for a dataset generated using the asymmetric logistic model. The dataset $\mathcal{D}_1$ is a single time series of length 15000, generated using (3.2) having set $\theta = (0.1, 0.1, 0.1, 0.3, 0.3, 0.3)$, $v = (0.5, 0.5, 0.5)$. After moving the time series to Laplace margins, we set threshold $u = F^{-1}_{\text{Laplace}}(0.95)$. See Figure 3.9 for a visual representation of the data, and Table 3.4 for the fitted parameters, with regularization $\gamma = 0.05$.

$\hat{\pi}$	$\hat{\alpha_1}$	$\hat{\alpha_2}$	$\hat{\beta_1}$	$\hat{\beta_2}$	$\hat{\mu_1}$	$\hat{\mu_2}$	$\hat{\sigma_1}$	$\hat{\sigma_2}$	Log likelihood
0.5066	1.0000	0.0002	0.0000	0.0000	-0.5491	0.4895	1.1040	1.4649	-1410.2723

Table 3.4: Fitted parameters and log-likelihood for a CEVMM fit on $\mathcal{D}_1$, with $u = F^{-1}_{\text{Laplace}}(0.95)$, $\gamma = 0.05$.

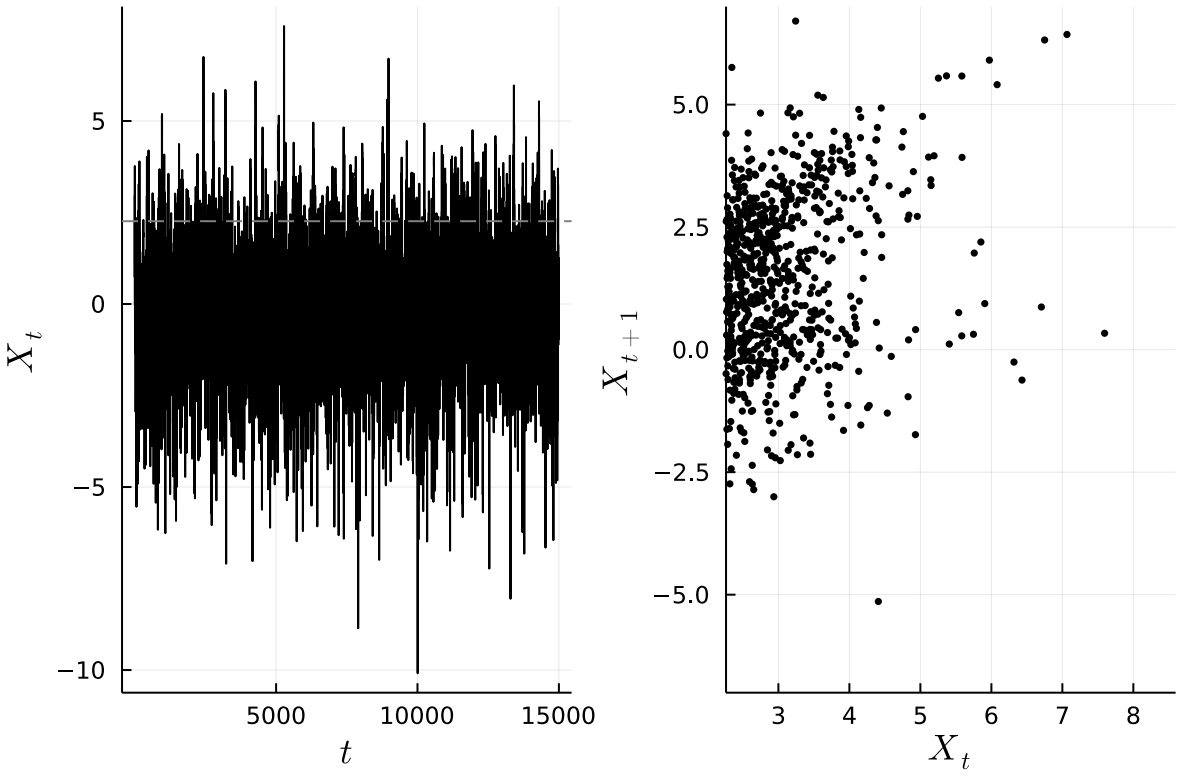

Figure 3.9: The left plot shows a time series of length 15000 generated from the 3 dimensional asymmetric logistic process, placed in Laplace margins. The grey dashed line is u. The right plot shows $\mathcal{D}$.

Regarding the bootstrap, we reach in particular for the moving block bootstrap (Kunsch, 1989), which can be applied to bootstrap stationary time series. To explain how this can be applied to a

CEVMM model, suppose we have a time series $\{X_t\}_{t=1}^n$ and that an appropriate threshold u has been established. We form bootstrapped series by sampling with replacement, n/B times, blocks of size B and then concatenating these sampled blocks together to form bootstrapped series. Repeating this M times yields M bootstrapped series $\{\tilde{X}_{1:n}^j\}_{j=1}^M$. A CEVMM can then be fit on each of the bootstrapped datasets $\tilde{\mathcal{D}}^j := \{(\tilde{X}_t^j, \tilde{X}_{t+1}^j) \,|\, t = 1, \ldots, n-1, \tilde{X}_t^j > u\}$, which allows us to construct empirical confidence intervals on estimators of interest. The 'moving block' aspect comes from the fact that the set to be sampled from is $\{X_{1:B}, X_{2:B+1}, \ldots, X_{n-B+1:n}\}$. A crucial assumption of the moving block bootstrap is that the original series $\{X_t\}$ from which blocks are formed is stationary; since otherwise the resulting bootstrapped series will not be comparable to the original series and any statistical analysis will therefore be meaningless.

Figure 3.10 shows empirical distributions of fitted parameters, generated using the moving block bootstrap. It is promising that these empirical distributions corroborate with the theoretically expected values of α_1, α_2, β_1 and β_2. We see, however, that the distribution for α_2 is slightly biased, and that the bootstrap distribution for α_1 places a small amount of mass near 0.8.

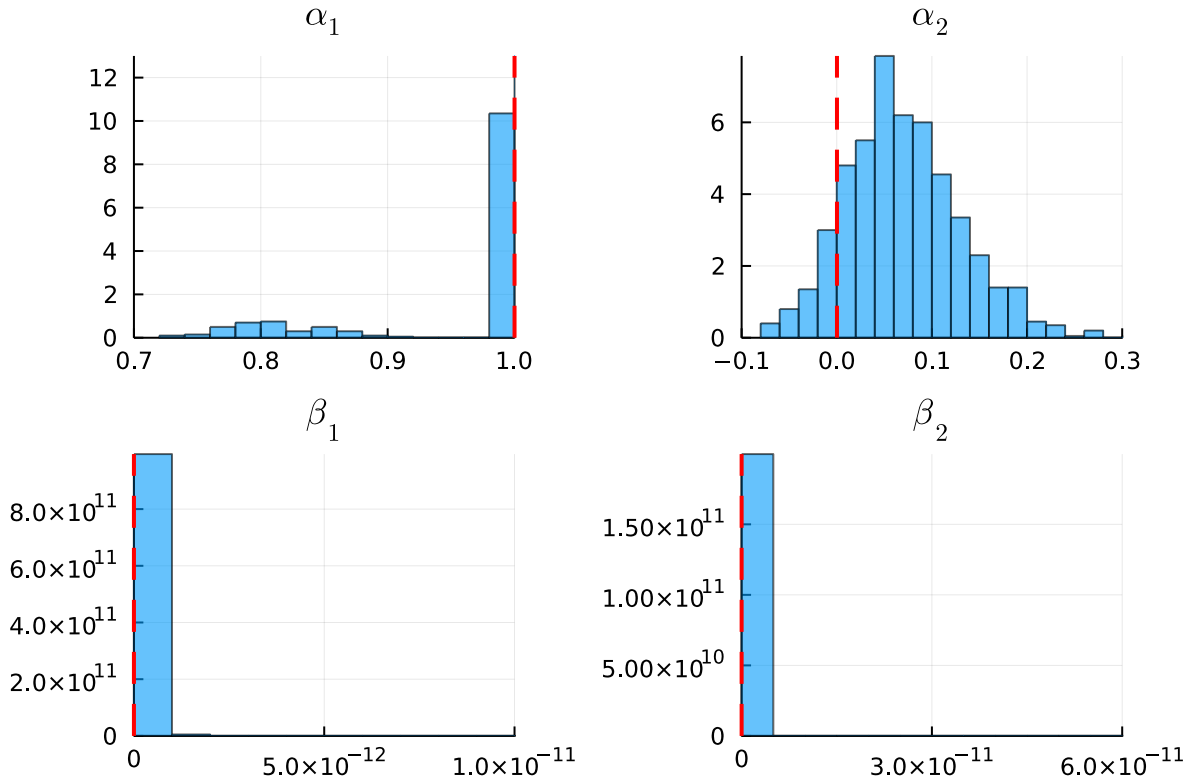

Figure 3.10: Histograms of fitted values α_1, α_2, β_1 and β_2 on $M = 1000$ moving-block-bootstrapped time series with block size $B = 50$. The parameter values fitted on the full dataset are shown in red.

Now that we have bootstrapped $M = 1000$ models, we can establish, as an example, confidence intervals for $\hat{F}_{1|0}(\cdot; x_0)$. These have been calculated and appear to work reasonably well for $x_0 \leq F_0^{-1}(0.99)$ (Figure 3.11), even if the confidence regions are perhaps a bit narrow for 99% confidence. Notice that there is no guarantee that estimators derived from the model fit on the full data will actually lie in the bootstrapped confidence interval. This may be problematic. If this happens, it suggests that the chosen u is not suitable, or the model assumptions do not hold. As we take x_0 larger in $\hat{F}_{1|0}(\cdot; x_0)$, we begin to see a bias (Figure 3.12). The bias is more obvious when x_0 is large because the modes of the distribution are further apart. Any error in the probability assigned to each mode leads to a discrepancy where the CDF flatlines.

From what we have seen, these bootstrap confidence intervals for $\hat{F}_{1|0}$ should be taken lightly. This is not a huge surprise, since we have not shown any underlying theory to support their use. However, they may still be of practical use, in particular when conditioning on smaller values of x_0. Having said this, the bootstrap distributions for the model parameters seem more convincing. It would be

interesting to implement a Bayesian CEVMM model and compare associated confidence regions to these bootstrap confidence intervals, but this is an area for future work.

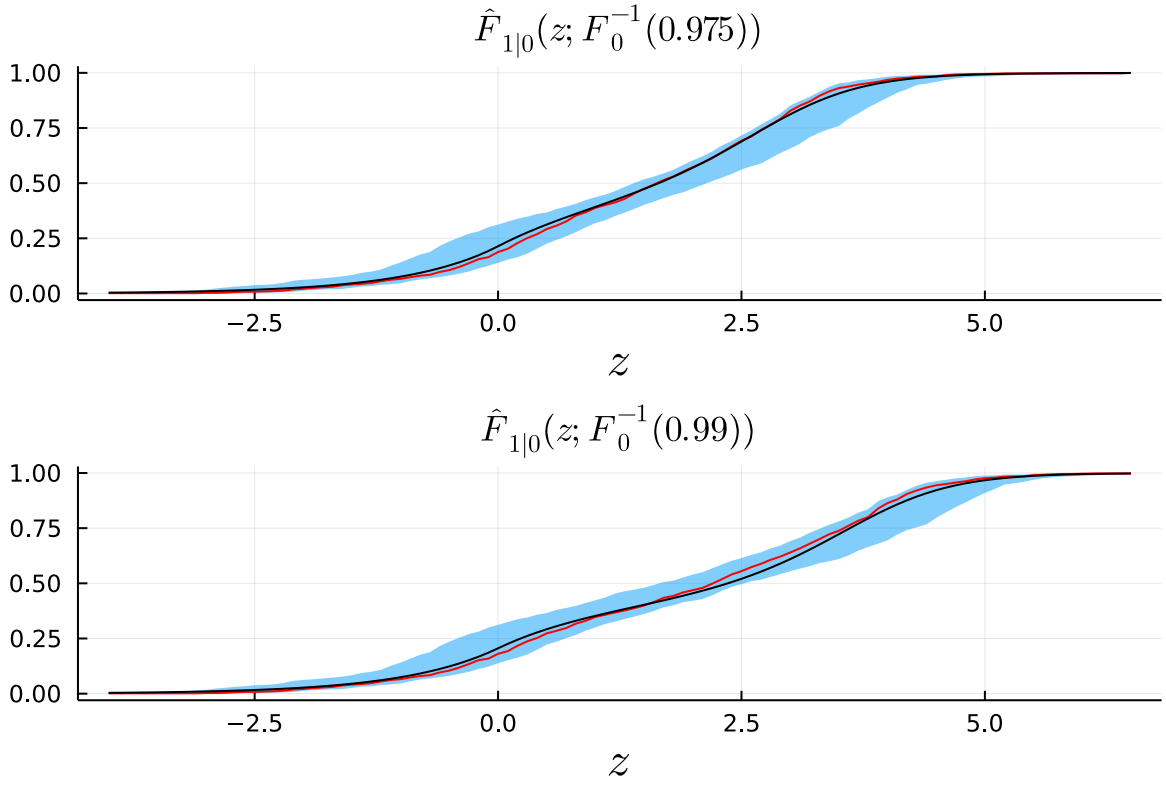

Figure 3.11: Estimates for $F_{1|0}(x; x_0)$, for two large x_0. The true distribution is shown in black, and the distribution estimated from the full data shown in red. The 99% confidence intervals, generated using $M = 1000$ bootstrap samples, are shown in blue.

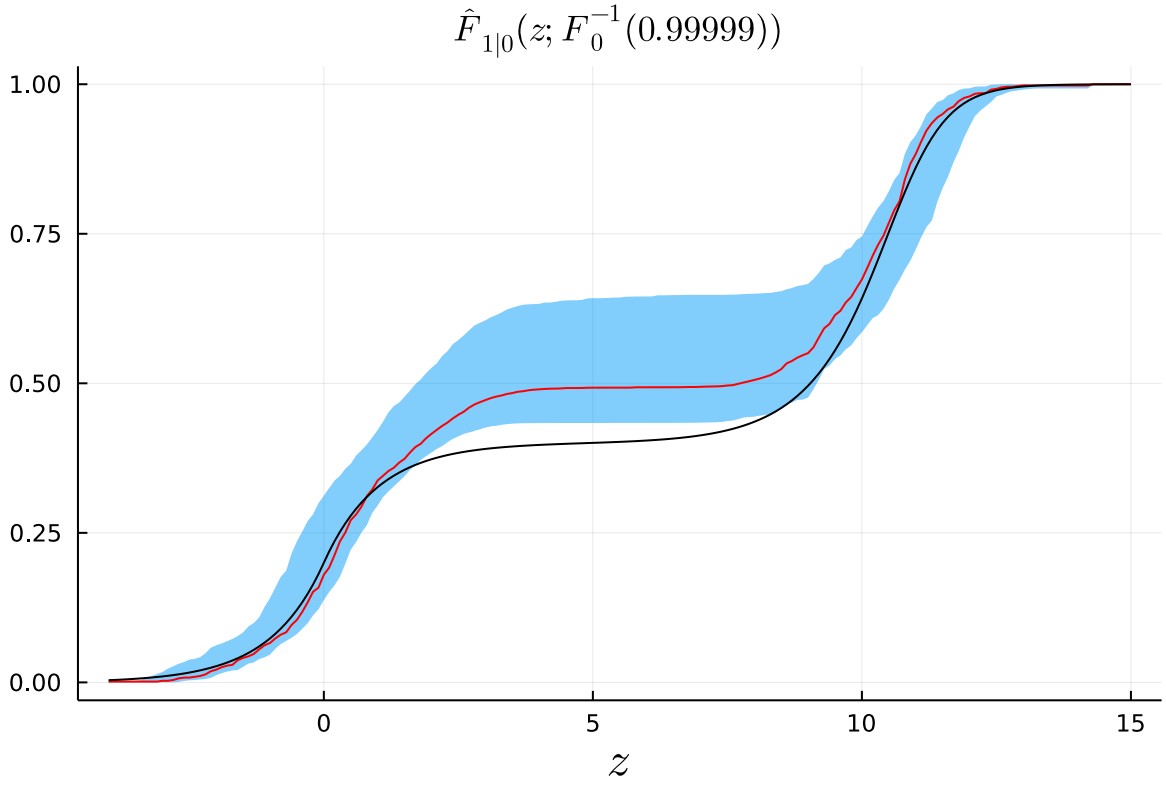

Figure 3.12: Estimate for $F_{1|0}(x; F_0^{-1}(0.9999))$. The true distribution is shown in black, and the distribution estimated from the full data shown in red. The 99% confidence intervals, generated using $M = 1000$ bootstrap samples, are shown in blue.

3.2 Application to Electricity Imbalance Prices

We apply the CEVMM methodology discussed above to electricity imbalance prices. Because they can be very volatile, imbalance prices are a natural place to apply extreme value statistics. We first give a short background on the UK electricity markets, as this is necessary to understand the structure of imbalance price data. We then perform a deseasonalization step to obtain stationary-like residuals for model fitting. Finally, we fit a CEVMM on the residuals and analyse the results.

3.2.1 Background, and obtaining electricity imbalance price data

It is in the interest of electricity suppliers and generators to strike contracts with each other in order for them to supply electricity to their customers, and to profit from electricity generation. In the UK, there exist markets to trade these electricity contracts (Liu et al., 2022) and in fact organisations that are neither suppliers nor generators can participate (Elexon).

Contracts will usually stipulate that the parties involved either consume or produce a certain volume of electricity. In reality, parties will likely fail to meet the contract requirements — for example, a supplier may incorrectly predict the demand of its customers, or the gales blowing on the eastern coast of Scotland may suddenly turn into light breezes. The volume of electricity that does not balance in these contracts, the imbalance, must be bought or sold on the central electricity grid at a certain price. This price, typically measured in £ per MWh, is determined half-hourly by Elexon (`https://elexon.co.uk`), and is called the *imbalance price*. Intuitively, we can expect the imbalance price to be high when there is a high demand for electricity, but there is little supply on the central grid, incentivising electricity generators to produce more. Conversely, we expect a low imbalance price when there is excess power on the central grid – it is even possible for imbalance prices to be negative.

Imbalance price data is not available to download directly. However, it is possible to obtain it via Elexon's API. There is a Python library (`https://github.com/MichaelKavanagh/elexon`) to assist

with this. Imbalance volume data is also collected for each period — this is the total imbalance over all contracts in a particular period.

3.2.2 Cleaning imbalance price data

A graphical representation of the raw 2021 data can be found in Figure 3.13. The year 2021 had 365 days, and we have 48 periods per 24 hours, so we expect to work with an imbalance price time series of length 17520.

Unfortunately, price data is missing from 2021-06-23 for periods 34 and 37 (there were no other missing data). Mean imputation was used to impute these missing values (the mean prices were calculated using data from the same month, June, and from the same periods). While multiple imputation is generally preferable to mean imputation, it is not expected that these values will actually affect the final CEVMM model since from Figure 3.13 we see that it is unlikely June will have extreme values in the upper tail.

We later model the imbalance prices in order to extract residuals (see (3.11) for the final model used). In particular we use $\texttt{period} \in \{1, \dots, 48\}$ as a covariate, which is complicated by the UK switching time zones to/from UTC from/to BST for daylight savings. The upshot of this is that 2021-03-28 has 2 fewer periods and 2021-10-31 has 2 more periods than other days. To work around this, the missing values in March are mean imputed, and the data corresponding to periods 49 and 50 in October are dropped. Imbalance volumes are imputed in a similar way in March.

To avoid massive outliers in the imbalance price, we perform a log transformation on the price, namely $\texttt{log\_price'} := \log(\texttt{price} + a)$, where price denotes the imbalance price. The adjustment term a in the argument to the logarithm is to ensure that the argument is positive. Though we are fitting models with 2021 data, for the purposes of compatibility with 2022 data (e.g. for prediction), we select $a = 91 = -\min\lceil\texttt{price}\rceil$, where the min is taken over all periods between 2021-01-01 and 2022-08-01. Finally, we standardize the volumes and prices to have mean 0 and standard deviation 1.

3.2.3 Seasonalizing imbalance prices

Imbalance prices are seasonal. This is because demand for electricity fluctuates throughout the day and throughout the year. In order for i.i.d. assumptions to be valid when fitting CEVMM models, we first fit a model on the imbalance prices in the hope of recovering stationary residuals; we call this step *deseasonalization*. We will then fit a CEVMM model on these residuals. We begin with a simple linear model with the factors weekday, month and period, given by

$$\begin{aligned}
\texttt{log_price} &= \beta_0 + \beta_1 \cdot \texttt{weekday} + \beta_2 \cdot \texttt{month} + \beta_3 \cdot \texttt{period} + \varepsilon \\
\varepsilon &\sim N(0, \sigma^2) \\
\texttt{weekday} &\in \{\texttt{Monday}, \dots, \texttt{Sunday}\} \\
\texttt{period} &\in \{1, \dots, 48\} \\
\texttt{month} &\in \{\texttt{January}, \dots, \texttt{December}\} \\
&\text{where} \\
\texttt{log_price} &= \frac{\texttt{log_price'} - \text{mean}(\texttt{log_price'})}{\text{std}(\texttt{log_price'})} \\
\texttt{log_price'} &= \log(\texttt{price} + a).
\end{aligned} \tag{3.10}$$

The log and standardization transforms are as discussed in Section 3.2.2. A deseasonalization can then be obtained with $y := \texttt{log\_price} - \widehat{\texttt{log\_price}}$, of which a plot can be found in Figure 3.14. It is clear that large values of y correspond with high imbalance prices, and similarly small values of y correspond with low imbalance prices.

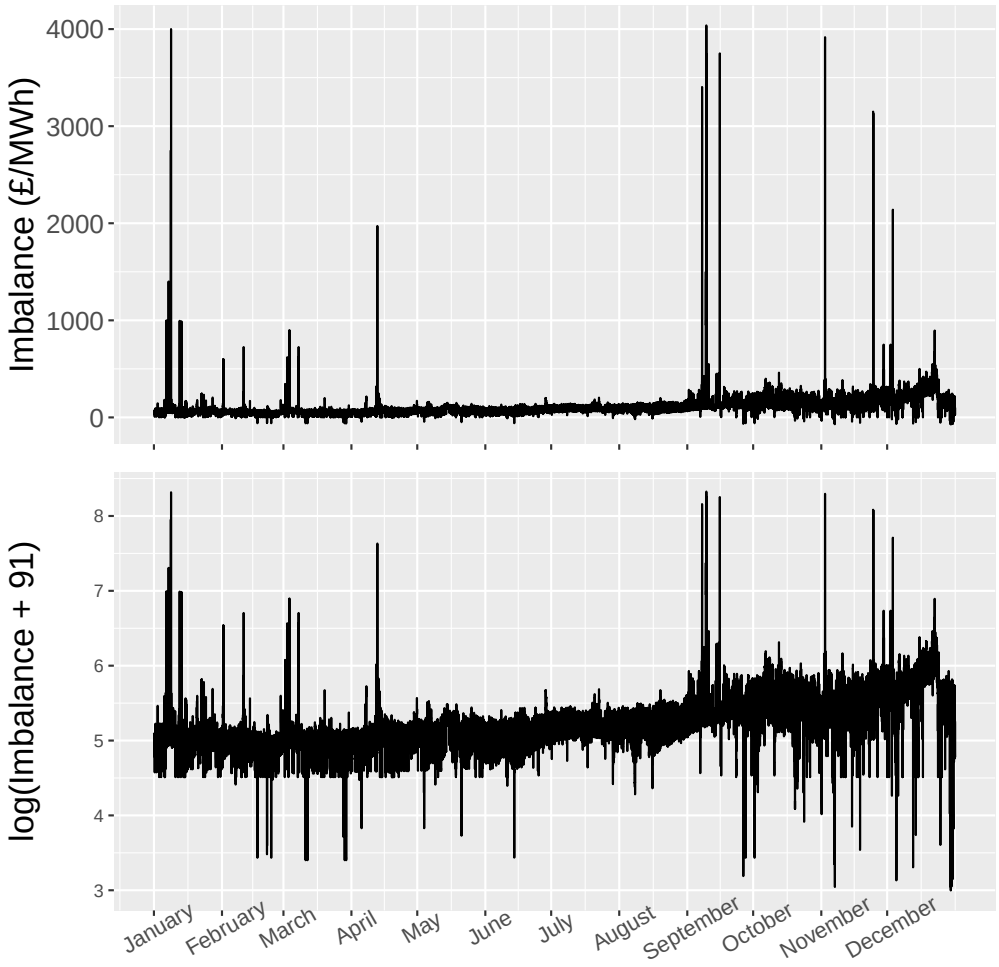

Figure 3.13: Imbalance prices, and imbalance prices on a log scale, for 2021. Months labelled correspond to the first period of the month.

One issue, at least visually, with the residuals in Figure 3.14 is the uptick towards the end of the series, and then the sudden jump down. Some experimentation finds that autoregressive terms help capture more variance. In particular, we consider the following covariates

(a) `last_log_price1` and `last_log_price2`, the `log_price` of the previous and second to last period respectively;

(b) `last_quantity1` and `last_quantity2`, the imbalance quantities of the previous and second to last period respectively;

(c) `yday_log_price` and `yday_quantity`, the `log_price` and imbalance quantity of the same period yesterday.

Stepwise selection was used (with the AIC criterion) to find which terms to include, with the selected model shown in (3.11). Note that by introducing `yday_quantity` and `yday_log_price`, we have no residuals for 2021-01-01, meaning that there are 17472 residuals in total.

$$\texttt{log\_price} = \begin{bmatrix} \texttt{yday\_log\_price} \\ \texttt{last\_log\_price1} \\ \texttt{last\_quantity1} \\ \texttt{last\_log\_price2} \\ \texttt{last\_quantity2} \\ \texttt{period} \\ \texttt{month} \\ \texttt{weekday} \\ \texttt{last\_log\_price2} \cdot \texttt{last\_quantity2} \end{bmatrix}^T \beta + \varepsilon,\ \varepsilon\ \sim N(0, \sigma^2) \tag{3.11}$$

Visually, the residuals for model (3.11) look more convincingly stationary than the model defined by (3.10) (Figure 3.15 vs. Figure 3.14). Note that in neither model is the normal assumption reasonable given the huge spikes seen in the figures, but this is the behaviour we aim to capture later with the CEVMM model.

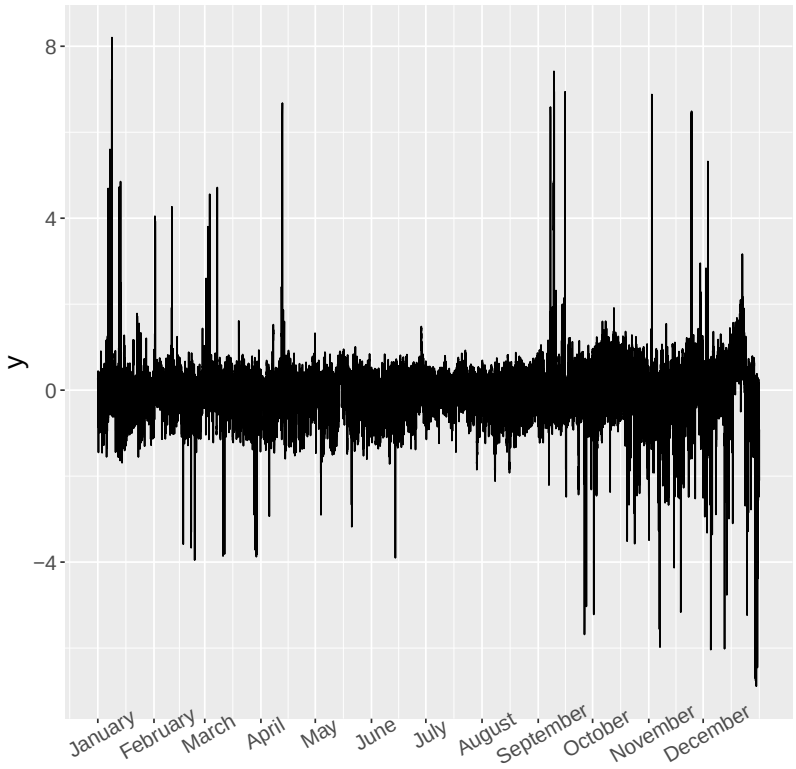

Figure 3.14: Plot of $y :=$ `log_price` $-$ $\widehat{\texttt{log\_price}}$ vs. period for model (3.10). Months labelled correspond to the first period of the month.

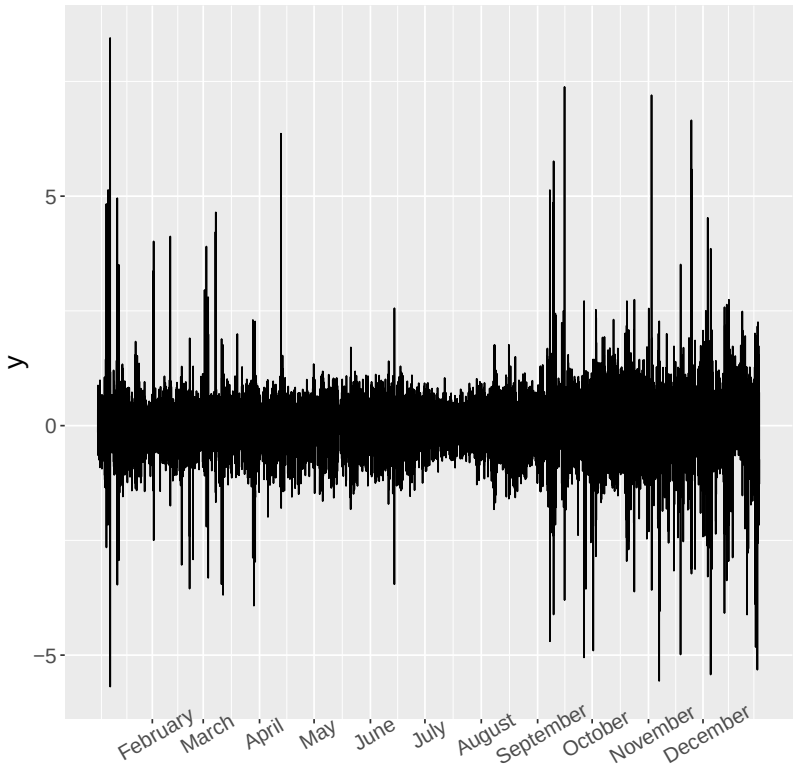

Figure 3.15: Plot of $y :=$ `log_price` $-$ $\widehat{\texttt{log\_price}}$ vs. period for model (3.11), chosen using stepwise selection. Months labelled correspond to the first period of the month.

To quantitatively check that y actually resembles *something* like a deseasonalization, we run a KPSS test to check for level stationarity.

```
KPSS Test for Level Stationarity

data:  resids
KPSS Level = 0.065412, Truncation lag parameter = 14, p-value = 0.1
```

Model (3.11) is not perfect as there is not overwhelming evidence for stationarity of its residuals. However, there is also no evidence against stationarity by the KPSS test above. Since we are mostly interested in applying a CEVMM model, we consider this model sufficient.

As a final note on deseasonalization, we point out that there is one clear disadvantage of moving from the simpler model (3.10) to (3.11). The residuals of the latter model no longer have a simple interpretation in terms of the price imbalance because the residuals are now coupled to the imbalance volume, as well as the imbalance price. This, unfortunately, reduces interpretability.

3.2.4 Fitting the CEVMM model

Given deseasonalized residuals of the imbalance data $\{y_t\}_{t=1}^{17472} = \{\texttt{log\_price} - \widehat{\texttt{log\_price}}\}$, we first move them to Laplace margins so that we can fit a CEVMM model. We choose Laplace margins over logistic margins as a result of the experiments of Section 3.1.3. Since the distribution of the residuals is unknown, we apply the transformation $y_t \mapsto F^{-1}_{\text{Laplace}}[n/(n+1)\hat{F}_{\text{ecdf}}(y_t)]$ to map to approximate Laplace margins. The factor $n/(n+1)$ is necessary to avoid the largest residual being mapped to the upper end point $F^{-1}_{\text{Laplace}}(1) = \infty$.

Figure 3.16 shows the raw data, where it is possible to see at least 2 modes for $y_t|y_{t-1} > u$. One mode corresponds to positive dependence, $\text{corr}(y_{t-1}, y_t) > 0$, and the other mode corresponds to negative dependence. This leads us to assume that there are $K = 2$ modes.

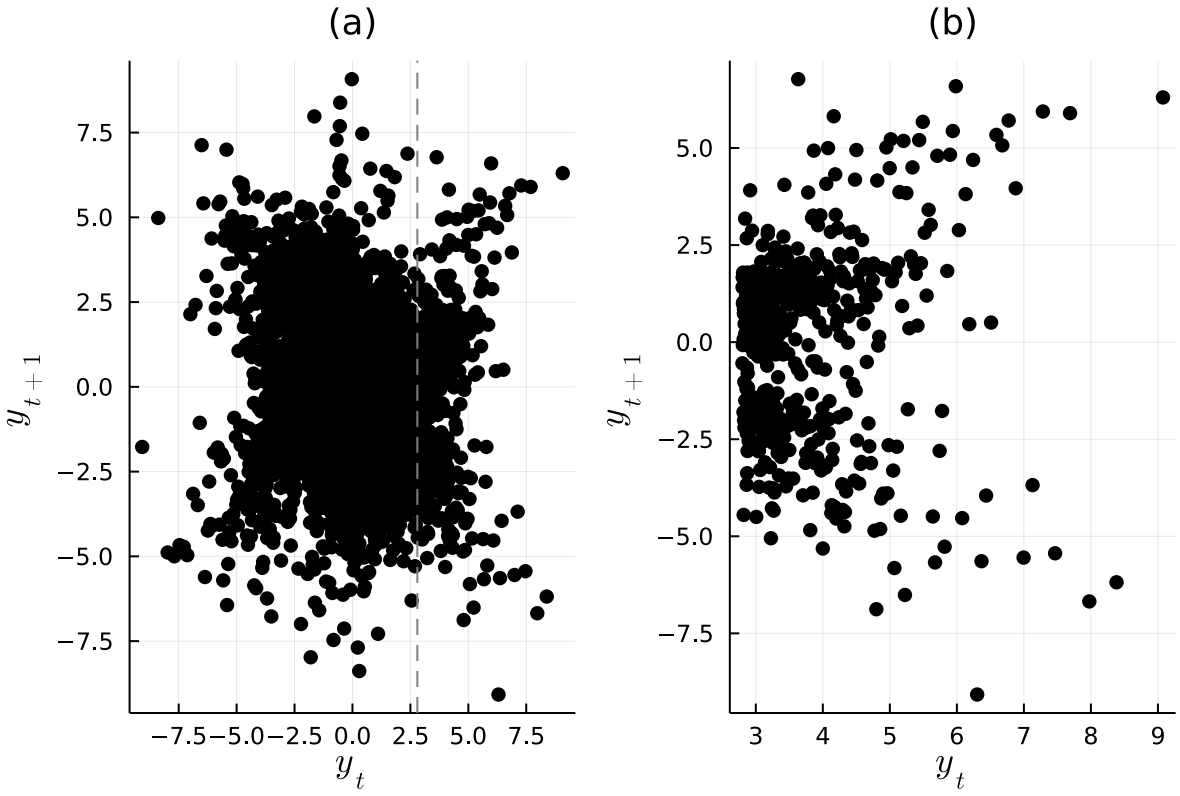

Figure 3.16: Plots of y_t vs. y_{t-1} (Laplace margins). (a) shows all the data, with the dotted line denoting $y_{t-1} = u = 2.8$; (b) focuses on the region $y_{t-1} > u$.

We first set the regularization parameter to $\gamma = 0.05 > 0$ as $\gamma = 0$ leads to overfitting ($\gamma = 0.05$ in particular was chosen because it was found to have worked well for asymmetric logistic data). See Figure 3.17 and the corresponding parameter estimates in Table 3.5 as an example of this overfitting. We see that the data has not been clustered naturally, and the table shows that the fitted variances are perhaps artificially small.

$\hat{\pi}$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	Log likelihood
0.4374	1.0000	-0.9994	0.6515	1.0000	-0.9555	0.7594	0.2632	0.6768	-1674.3615

Table 3.5: Best CEVMM fit of 30 random initial values, with no regularization and $u = F^{-1}_{\text{Laplace}}(0.95)$.

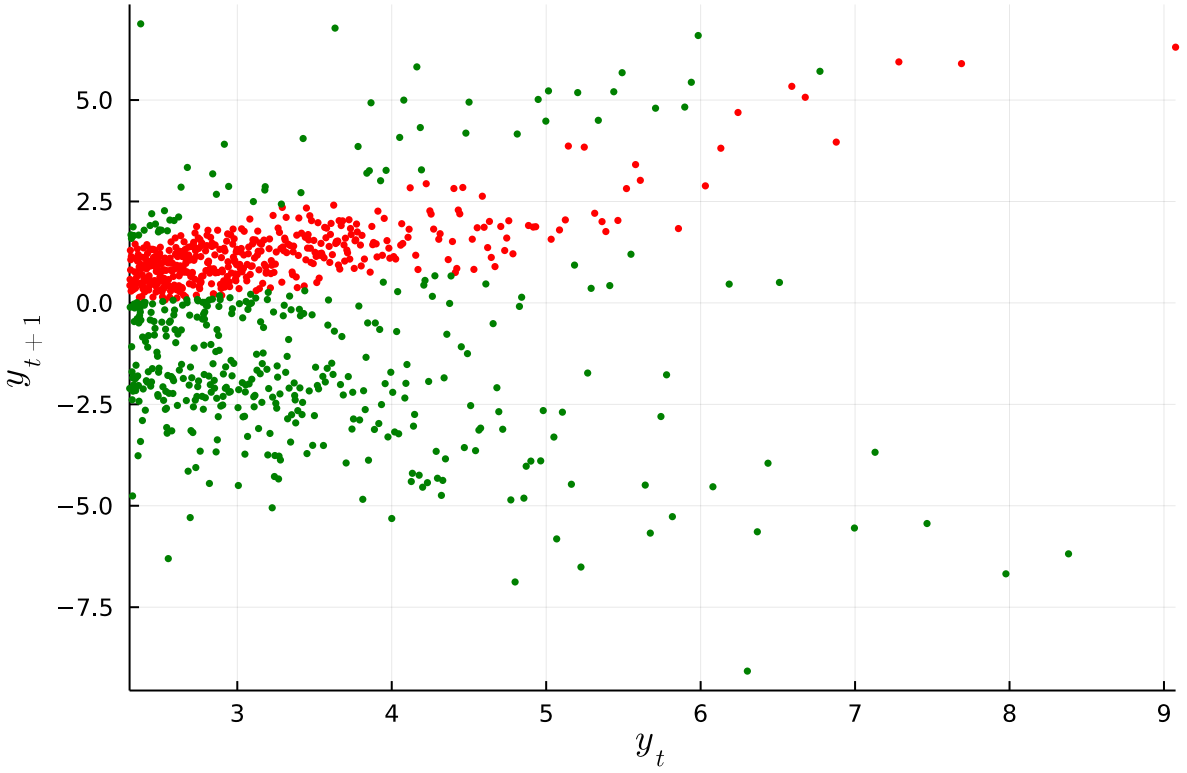

Figure 3.17: Best CEVMM fit of 30 different models each fit with random initial values. Here, $\gamma = 0$ and $u = F^{-1}_{\text{Laplace}}(0.95)$. Data points are colour coded according to which normal mixture they most likely belong to.

It is not obvious which value should be chosen for the threshold u. We proceed by fitting CEVMMs for several $u \in [1.6, 3.9] \approx F^{-1}_{\text{Laplace}}([1.6, 3.9])$. Figure 3.18 shows how the fitted values change with u — we select $u = 2.8$ since for $u < 2.8$ we see α_1 steadily decreasing and μ_2 steadily increasing.

Figure 3.19 and Table 3.6 show the fitted model with the selected parameters. We see a much more natural separation of the modes, and most data points are being clustered to the mode that one might expect.

$\hat{\pi}$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	Log likelihood
0.6702	1.0000	-1.0000	0.0000	0.0000	-2.3069	1.1855	1.2633	1.4060	-1126.2791

Table 3.6: Fitted parameter values and log likelihood for $K = 2$, $\gamma = 0.05$, $u = 2.8$.

3.2.5 Uncertainty in model parameters

Using the moving block bootstrap once again, as in Section 3.1.5, we can estimate uncertainty in the fitted parameters. Figure 3.20 shows estimated distributions for $M = 1000$ bootstrap replications.

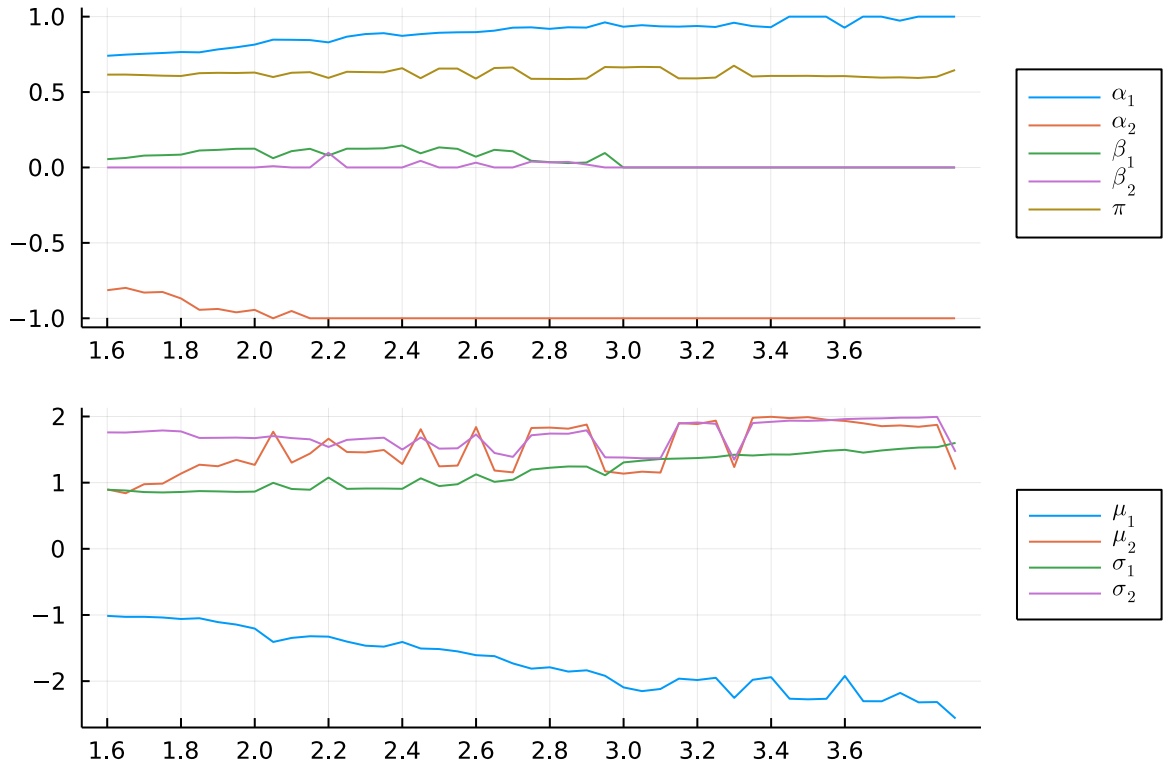

Figure 3.18: Effect of threshold u on the fitted parameters, with $\gamma = 0.05$.

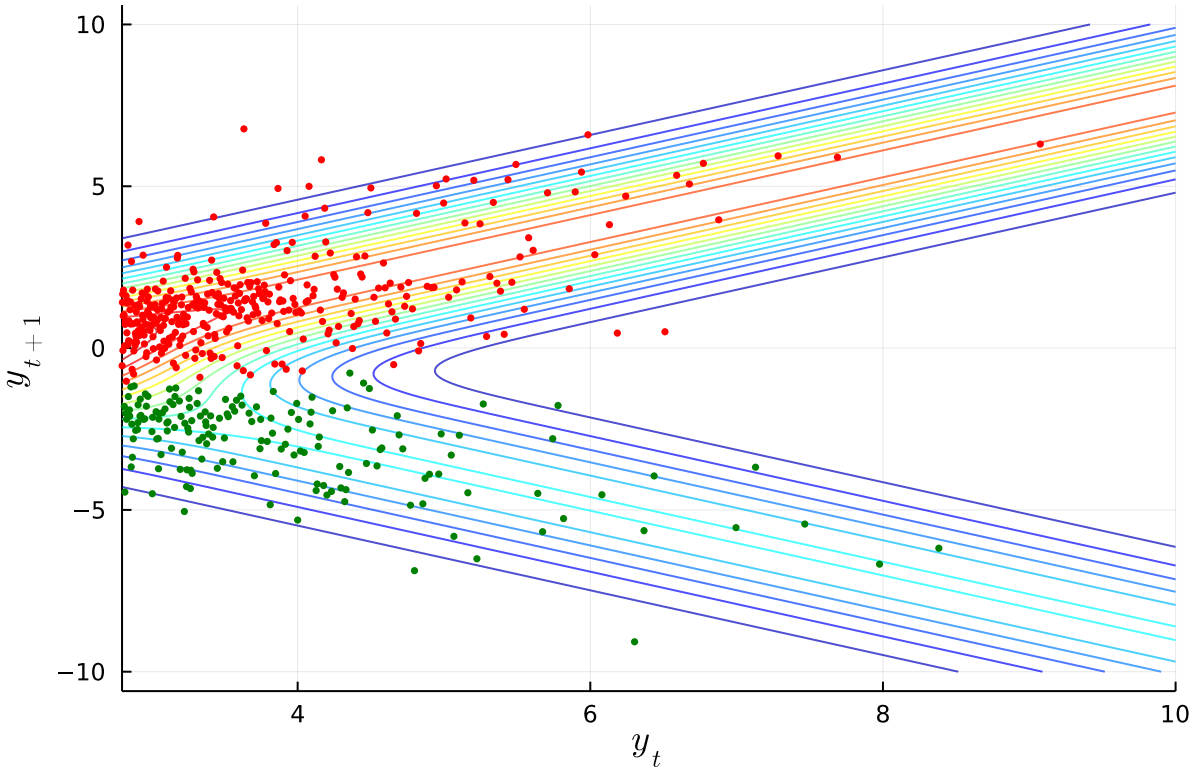

Figure 3.19: CEVMM fit, with regularization $\gamma = 0.05$ and $u = 2.8$. Data points are colour coded according to the normal mixture they most likely belong to. The coloured lines show contours of the modelled normal mixture density for $y_{t+1}|y_t$, when viewed as a function of both y_{t+1} and y_t.

A block size $B = 48$ was used for convenience, as 48 divides the total length of the time series $\{y_t\}$ evenly. The samples clearly corroborate with the values fitted on the full dataset and there is little uncertainty. However, the small islands of mass away from the parameter boundaries are suspect. We saw similar behaviour with the asymmetric logistic data in Figure 3.10. It may be that there are a few data points in the dataset that are influential in encouraging the parameters to the boundary, and sometimes they are not included by the bootstrap.

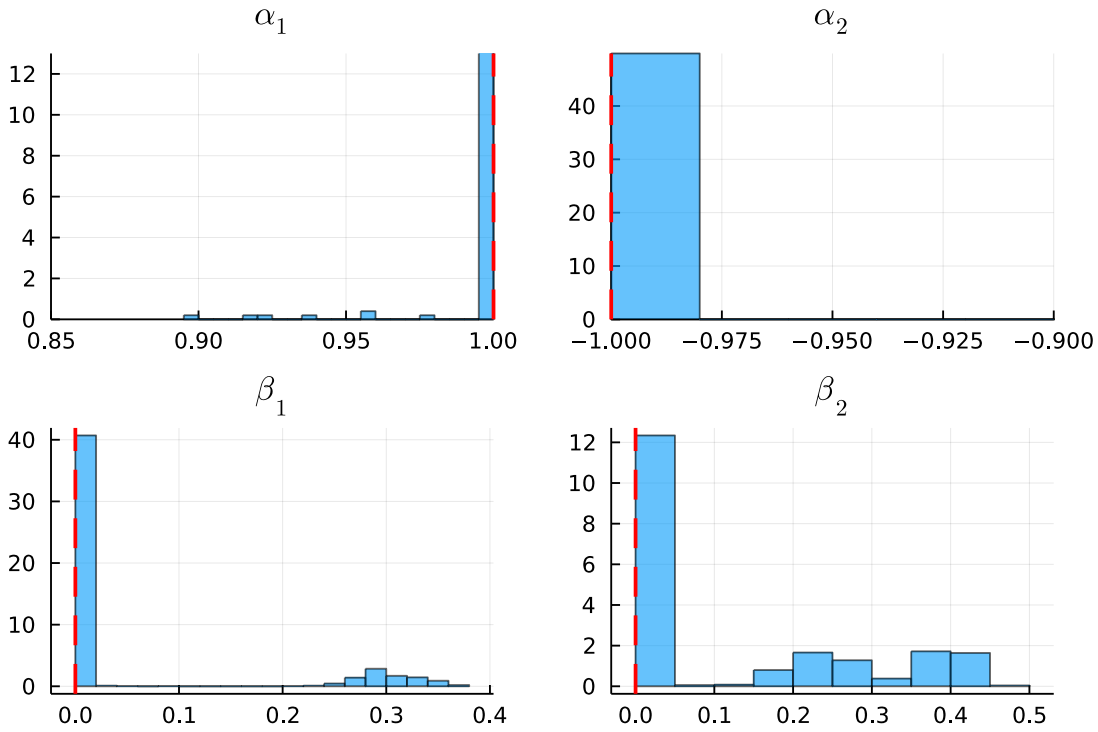

Figure 3.20: Histograms of fitted values α_1, α_2, β_1 and β_2 on $M = 1000$ moving-block-bootstrapped time series with block size $B = 48$. The parameter values fitted on the full dataset are shown in red.

3.2.6 One-step-ahead residual prediction

If the current price's residual $y_t = y$ is large, then it is simple to use a CEVMM to find a one-step-ahead predictive distribution for the imbalance price residuals. We demonstrate this with the above CEVMM model (defined by parameters in Table 3.6) by finding predictive distributions for January 2022 data.

Concretely, one-step-ahead prediction means sampling from $\hat{F}_{1|0}$, as we assume that $y_{t+1}|y_t \sim F_{1|0}(\cdot\,; y_t)$. Note that the procedure does not take into account uncertainty in the CEVMM model's parameters. This is a good test case because the data is out-of-sample for the CEVMM model, and additionally the true residuals are known, so it is possible to assess the performance of our predictions.

Imbalance price residuals for January 2022 (after the usual transformation to Laplace margins) are shown in Figure 3.21 (of which there are $48 \cdot 31 = 1488$), and the results of sampling predictions are shown in Figure 3.22 for the first 16 exceedances (out of 45).

We can see that when predicting one step ahead, the CEVMM assigns a larger probability to the upper mode than the lower mode. See Figure 3.22; sometimes the upper mode is a good predictor (e.g. row 1, column 3); sometimes the lower mode is a good predictor (e.g. row 2, column 2), and sometimes neither are good estimators (e.g. row 3, column 4). As such, the CEVMM model is not going to be useful for constructing point estimates, but more for managing risk. The ARIMA predictions are plotted merely out of interest, but it illustrates that more standard time series model tend not to predict such large residuals as the CEVMM model, even when they are likely to happen. The estimate from the ARIMA model tends to be more conservative than the CEVMM model.

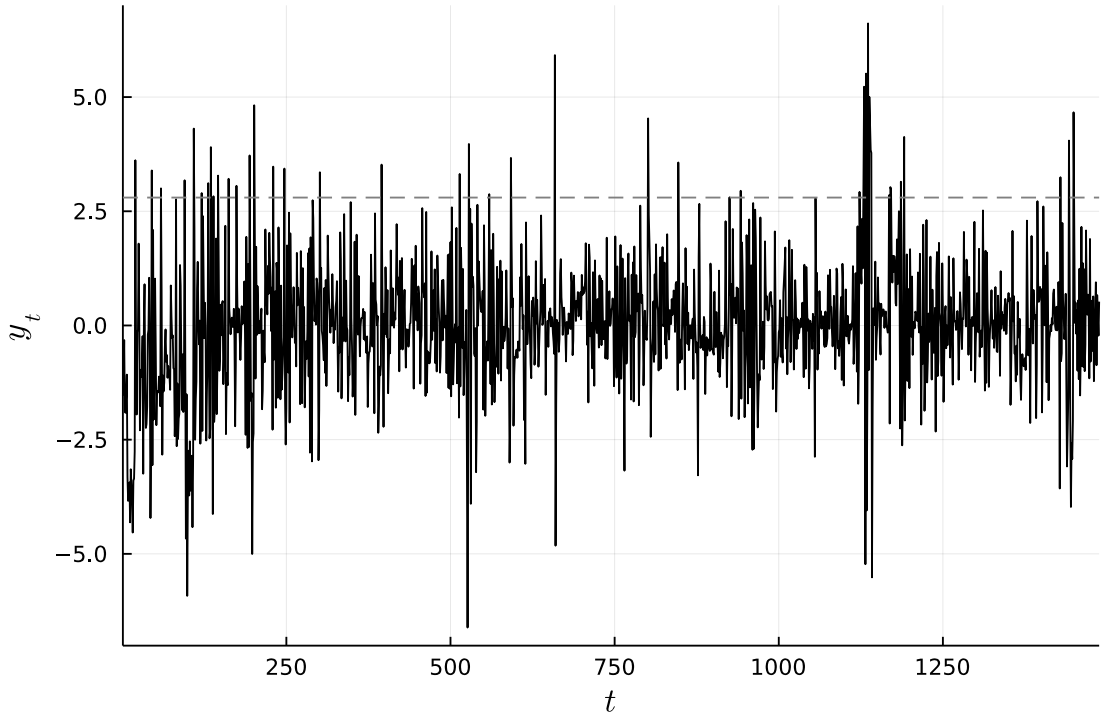

Figure 3.21: Imbalance price residuals for January 2022, with threshold $u = 2.8$ shown by the grey dashed line.

We repeat the above for $K = 3$; the model used has parameters detailed in Table 3.7. The results are shown in Figure 3.24. We compare the two values of K by looking at their predictive capabilities. The proxy we use is likelihood of the data, where the likelihood is determined by a Gaussian kernel density estimator; that is, we fit density estimator on the simulated residuals of $y_{t+1}|y_t$, for each of the 45 exceedances in January 2022, and calculate likelihoods of the actual values. We compare the distributions of these likelihoods in Figure 3.23. The distribution of likelihoods looks similar, but $K = 2$ has a slightly higher mean, and also has a higher 'best likelihood'. Overall, there is no reason to prefer $K = 3$.

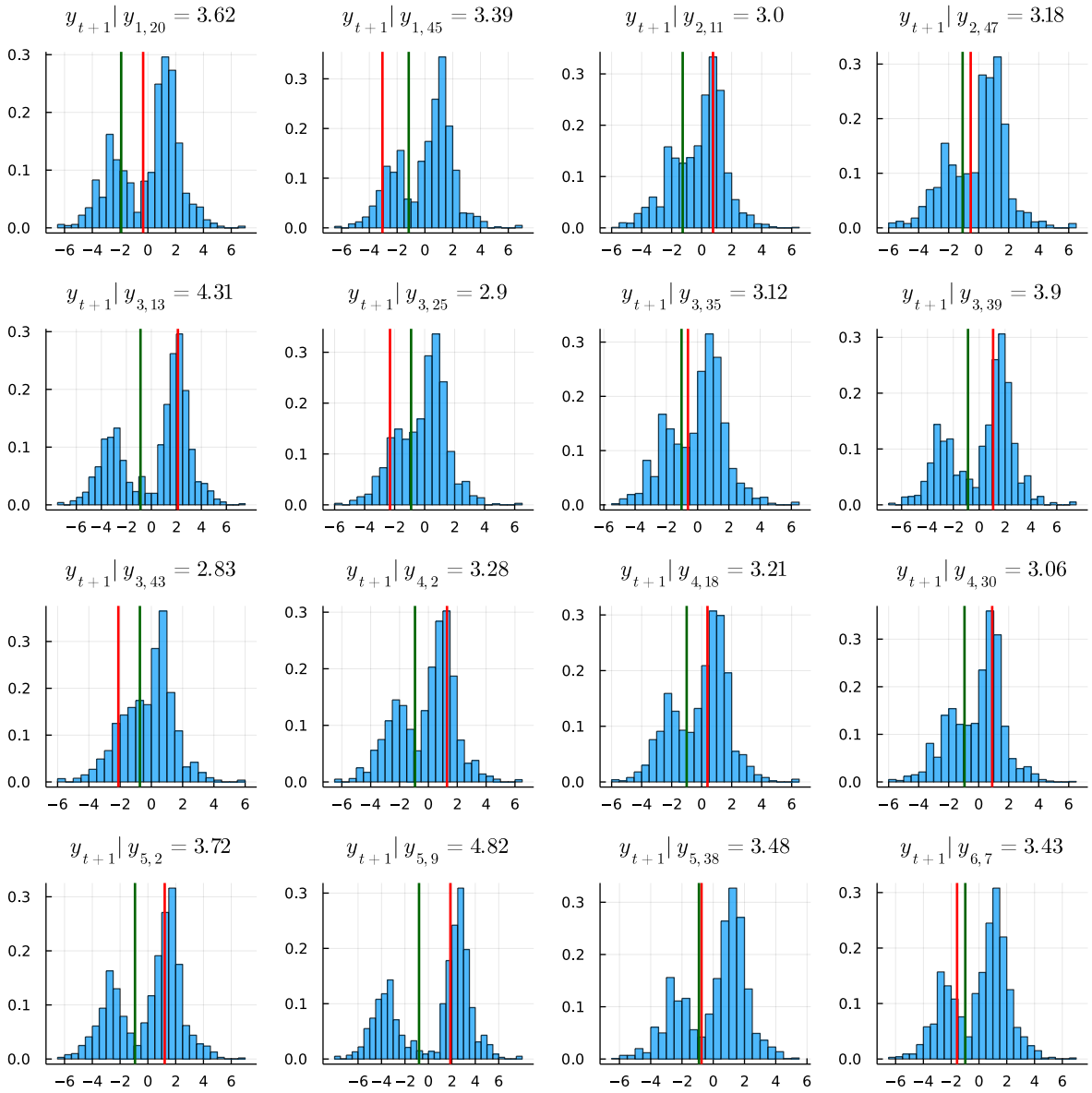

Figure 3.22: ($K = 2$) Histograms for simulated values of $y_{t+1}|y_t$ (here $y_{d,p} := y_{48d-1+p}$ denotes the residual on day d during period p in January 2022). The red line shows the actualized value of y_{t+1}, and the green line shows the value estimated by an ARIMA model (fit to the to the entire series of residuals).

$\hat{\pi}_1$	$\hat{\pi}_2$	$\hat{\pi}_3$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
0.4886	0.1799	0.3314	1.0000	0.1951	-1.0000	0.0000	0.2061	0.0000

$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\mu}_3$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\sigma}_3$	Log likelihood
-1.8380	-0.2785	1.3205	1.2349	1.2609	1.6148	-1133.6306

Table 3.7: Fitted parameter values and log-likelihood for $K = 3$, $\gamma = 0.05$, $u = 2.8$.

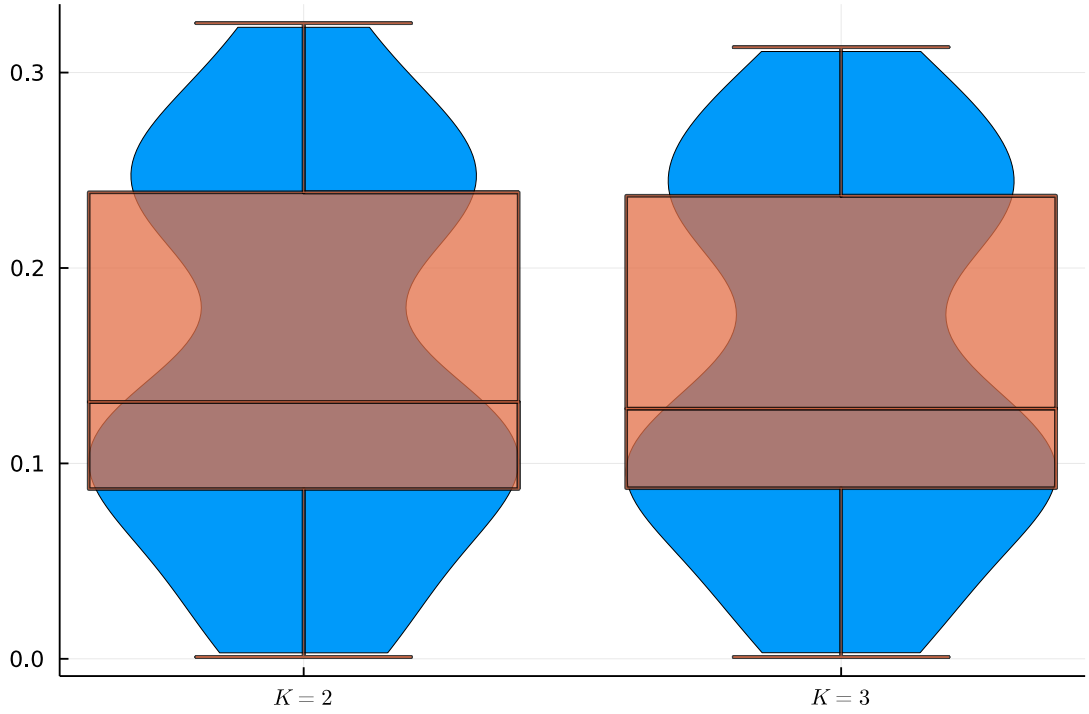

Figure 3.23: Violin and box plots for likelihoods of the actualized residuals according to the estimated predictive distributions for the CEVMM models for $K = 2$ and $K = 3$.

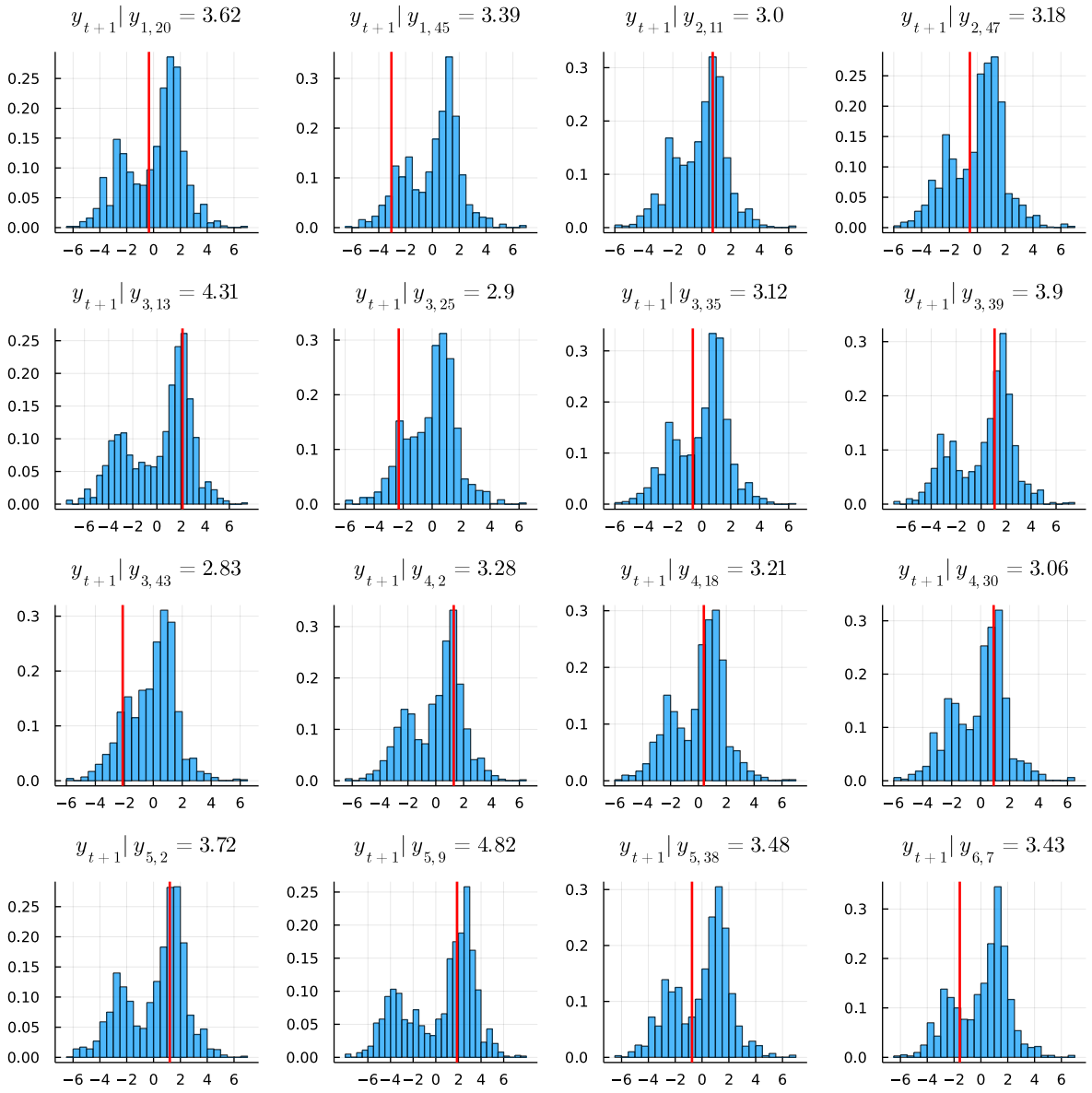

Figure 3.24: ($K = 3$) Histograms for simulated values of $y_{t+1}|y_t$ (here $y_{d,p} := y_{48d-1+p}$ denotes the residual on day d during period p in January 2022). The red line shows the actualized value of y_{t+1}.

3.2.7 Conversion back to the original price scale

While we have fitted the CEVMM on imbalance price residuals, this only provides utility to electricity suppliers and producers if it is possible to frame the results using GBP. In terms of one-step-ahead price prediction, this can be done by first moving residuals from unit Laplace scale to their original scale, and then applying (3.12).

$$\begin{aligned} &\texttt{log\_price}_{t+1} = \widehat{\texttt{log\_price}}_{t+1} + y_{t+1} \\ \implies &\texttt{price}_{t+1} = \exp\left(\text{sd}(\texttt{log\_price'})(\widehat{\texttt{log\_price}}_{t+1} + y_{t+1}) + \text{mean}(\texttt{log\_price'})\right) - a \end{aligned} \tag{3.12}$$

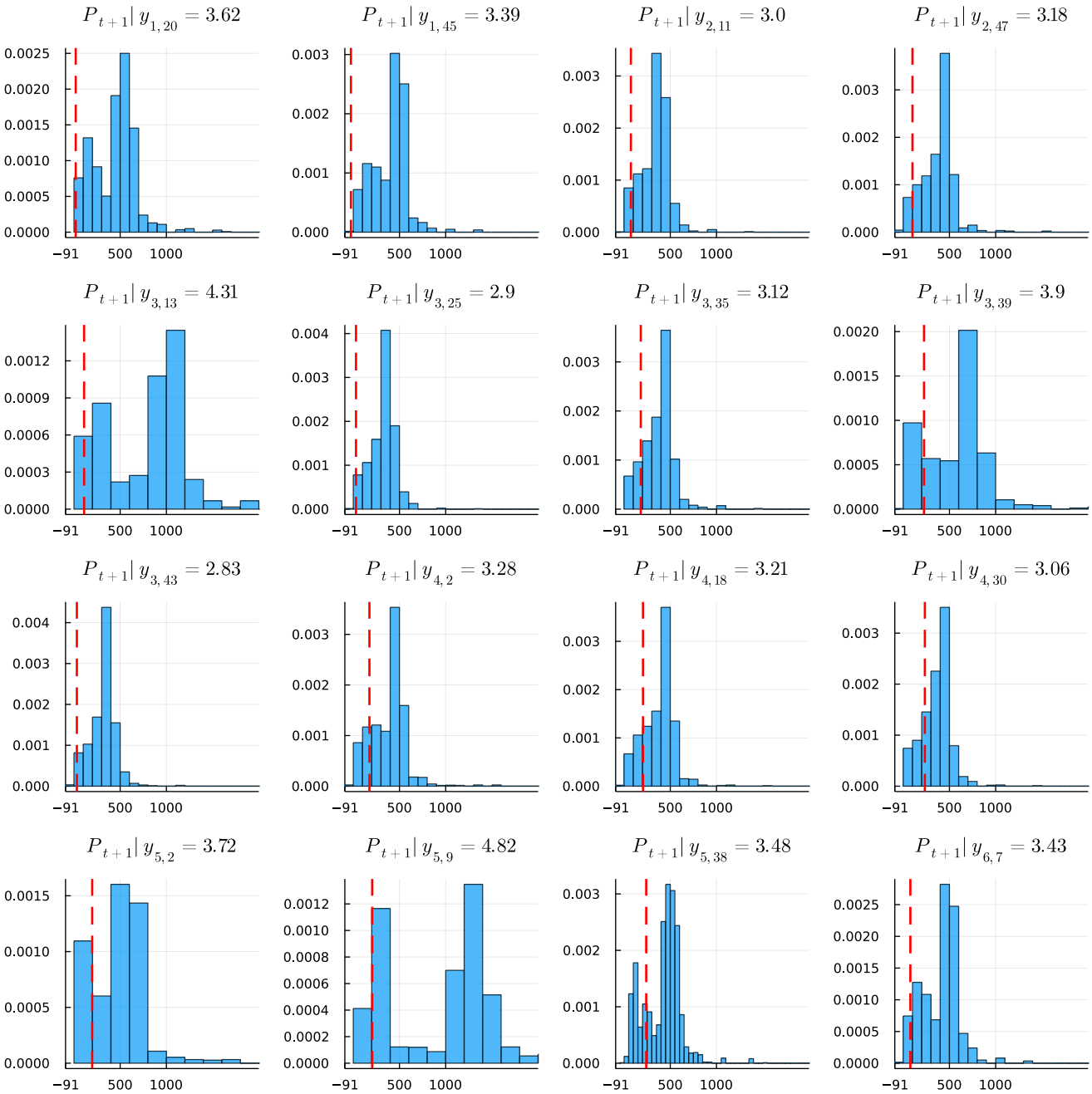

Figure 3.25: Histograms for simulated values of $P_{t+1}|y_t$ under the CEVMM model 3.6. Here $y_{d,p} := y_{48d-1+p}$ denotes the residual on day d during period p in January 2022, and $P_{t+1}|y_t$ represents the imbalance price (measured in GBP) at time $t+1$, conditional on the previous period's imbalance residual. The red line shows the measured value of P_{t+1}.

This reverse transform was applied to the predictive distributions of Figure 3.22, and the results can be found in Figure 3.25. The results are underwhelming, most notably because the predictive distributions suggest that high prices, around £500 are likely, but we never see these actualized in the figure. This may be because the model (3.11) used to deseasonalize the imbalance prices is too simplistic. It is also likely that a more sophisticated model is required to perform accurate price prediction; one suggestion is to use CEVMM models as part of a larger ensemble of models.

Chapter 4

Conclusions

We have developed and implemented the CEVMM model based on the EM algorithm, and used it to characterize bivariate data. Models of this type have been applied in the past, but the application here to stationary time series is believed to be novel. Simulated data from a 3-dimensional asymmetric logistic model was used to illustrate methods for fitting CEVMM models, assessing uncertainty, and to validate the implementation. CEVMMs were easily able to fit parameters that align closely with theoretical parameter values, and we saw that the fitted distributions also aligned reasonably well, though there was perhaps some bias. We then considered electricity imbalance price residuals, where there was evidence of extremal dependence. A CEVMM model was fitted, and used to perform one-step-ahead residual prediction with some success.

Extreme events are rare by definition but can be catastrophic. Most statistical and machine learning models are trained to fit to existing data, and are therefore unlikely to be able to take these unlikely events into account when making decisions. CEVMMs however have the ability to shine in this domain. By taking advantage of the theory of extremes that has been developed over several decades, CEVMMs have the potential to play a crucial role in managing this risk of extremes. In their presented form, the CEVMM models discussed can be applied to any bivariate dataset, under the mild assumption (2.3), and also the assumption that exceedances are i.i.d.. The latter is stringent in general, but is not unreasonable if we consider datasets of consecutive pairs of realizations from stationary time series. It is therefore expected that CEVMM models will be broadly applicable, especially in fields where stationary time series are prevalent.

While CEVMMs can be applied in the case of extremes, they do not apply generally, and so to perform wholistic inference it would be necessary to use them in tandem with other models, which has not yet been investigated. Additionally, whilst the methods proposed are relatively easy to use, CEVMMs may rarely apply because they are conditional on extreme values. This is not a problem directly, but it makes model validation difficult as any test data is, by definition, rare.

There are several avenues for future work. Firstly, if CEVMMs were to be used in tandem with other methods for wholistically modelling, it is not obvious how this ought to be done. However, there is potential here. For example, we saw that sometimes the CEVMM model was able to perform imbalance residual prediction better than others; perhaps in the context of a larger system, it would be possible to identify the times when the signal from the CEVMM model is more reliable. Secondly, we have only demonstrated how CEVMMs can be applied to bivariate data. It should be possible to extend the model to the multivariate case (for example by repeating the methods shown to each extra dimension, or even by adopting a multivariate normal model in (2.6)). This could allow the development of extremal models that take into account data from multiple stationary time series, or that simply increase the number of consecutive data points used for fitting beyond 2. Finally, there are many aspects regarding the implementation that could be tested or improved. For example, construction of confidence intervals was achieved with a moving block bootstrap, but it is not clear if this is the best approach. Additionally, it is not understood how violation of the i.i.d. assumption affects results; since this assumption can be unrealistic for real data, it may be useful for practitioners to quantify the tolerances here.

Bibliography

Graeme Auld. *Statistical Modelling of Environmental Extremes*. PhD thesis, University of Edinburgh, 2021.

August A Balkema and Laurens de Haan. Residual life time at great age. *The Annals of Probability*, 2(5):792–804, 1974.

Leo Belzile et al. *mev: Modelling Extreme Values*, 2022. URL https://CRAN.R-project.org/package=mev. R package version 1.14.

Mauro Bernardi, Antonello Maruotti, and Lea Petrella. Multivariate Markov-switching models and tail risk interdependence. *arXiv preprint arXiv:1312.6407*, 2013.

Paola Bortot and Stuart Coles. Extremes of Markov chains with tail switching potential. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(4):851–867, 2003.

Stuart Coles. *An Introduction to Statistical Modeling of Extreme Values*, volume 208. Springer, 2001.

Stuart Coles and Jonathan A Tawn. Modelling extreme multivariate events. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(2):377–392, 1991.

Anthony C Davison and Richard L Smith. Models for exceedances over high thresholds. *Journal of the Royal Statistical Society: Series B (Methodological)*, 52(3):393–425, 1990.

Bradley Efron and Robert J Tibshirani. *An Introduction to the Bootstrap*. CRC press, 1994.

Elexon. BSC explained. https://www.elexon.co.uk/about/bsc-explained/.

Paul Embrechts, Sidney I Resnick, and Gennady Samorodnitsky. Extreme value theory as a risk management tool. *North American Actuarial Journal*, 3(2):30–41, 1999.

Manfred Gilli et al. An application of extreme value theory for measuring financial risk. *Computational Economics*, 27(2):207–228, 2006.

Konstantinos Gkillas and Paraskevi Katsiampa. An application of extreme value theory to cryptocurrencies. *Economics Letters*, 164:109–111, 2018.

Janet E Heffernan and Jonathan A Tawn. A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66 (3):497–546, 2004.

Rick W Katz. Extreme value theory for precipitation: sensitivity analysis for climate change. *Advances in Water Resources*, 23(2):133–139, 1999.

Caroline Keef, Ioannis Papastathopoulos, and Jonathan A Tawn. Estimation of the conditional distribution of a multivariate variable given that one of its components is large: Additional constraints for the Heffernan and Tawn model. *Journal of Multivariate Analysis*, 115:396–404, 2013.

Hans R Kunsch. The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, pages 1217–1241, 1989.

Yuelin Li, Elizabeth Schofield, and Mithat Gönen. A tutorial on Dirichlet process mixture modeling. *Journal of Mathematical Psychology*, 91:128–144, 2019.

Jinqi Liu, Jihong Wang, and Joel Cardinal. Evolution and reform of UK electricity market. *Renewable and Sustainable Energy Reviews*, 161:112317, 2022.

Giulia Marcon, SA Padoan, Philippe Naveau, Pietro Muliere, and Johan Segers. Multivariate nonparametric estimation of the Pickands dependence function using Bernstein polynomials. *Journal of Statistical Planning and Inference*, 183:1–17, 2017.

Patrick Kofod Mogensen and Asbjørn Nilsen Riseth. Optim: A mathematical optimization package for Julia. *Journal of Open Source Software*, 3(24), 2018.

Ioannis Papastathopoulos and Jonathan A. Tawn. Hidden tail chains and recurrence equations for dependence parameters associated with extremes of higher-order Markov chains, 2019. URL https://arxiv.org/abs/1903.04059.

Ioannis Papastathopoulos, Kirstin Strokorb, Jonathan A Tawn, and Adam Butler. Extreme events of Markov chains. *Advances in Applied Probability*, 49(1):134–161, 2017.

James Pickands III. Statistical inference using extreme order statistics. *the Annals of Statistics*, pages 119–131, 1975.

Richard A Redner and Homer F Walker. Mixture densities, maximum likelihood and the EM algorithm. *SIAM review*, 26(2):195–239, 1984.

Jonathan A Tawn. Modelling multivariate extreme value distributions. *Biometrika*, 77(2):245–253, 1990.

Stan Tendijck, Emma Eastoe, Jonathan A Tawn, David Randell, and Philip Jonathan. Modeling the extremes of bivariate mixture distributions with application to oceanographic data. *Journal of the American Statistical Association*, pages 1–12, 2021.

Erin Towler, Balaji Rajagopalan, Eric Gilleland, R Scott Summers, David Yates, and Richard W Katz. Modeling hydrologic and water quality extremes in a changing climate: A statistical approach based on extreme value theory. *Water Resources Research*, 46(11), 2010.

Chapter A

Appendices

A.1 EM algorithm for fitting conditional extremal K-mixture models with the normal assumption

Let $\Phi(\cdot\,;\mu,\sigma^2)$ and $\phi(\cdot\,;\mu,\sigma^2)$ be the normal CDF and PDF respectively. We are concerned with fitting the model

$$\mathbb{P}\left(Y \leq y | X = x\right) = \sum_{k=1}^{K} \pi_k \Phi_k \left(\frac{y - \alpha_1 x}{x^{\beta_k}}; \mu_k, \sigma_k^2 \right). \tag{A.1}$$

We fit on a dataset $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^n$ where $x_i > u$ for some preferably large u ($u > 0$ is strictly necessary for us to be able to evaluate fractional powers of x_i), and it is assumed that elements of $\mathcal{D}$ are i.i.d. draws. An EM algorithm for finding π, α, β, μ, and σ is given in Algorithm 1.

Algorithm 1 EM algorithm for fitting conditional extremal K-mixture models with the normal assumption

Initialize a probability vector $\pi^{(0)} \in \mathcal{S}_K$
Initialize $\alpha_k^{(0)}, \beta_k^{(0)} \in [0,1]$, $k = 1, \ldots, K$.
Initialize $\mu_k^{(0)} \in \mathbb{R}$, $k = 1, \ldots, K$.
Initialize $\sigma_k^{(0)} \in \mathbb{R}_+$, $k = 1, \ldots, K$.
$t \leftarrow 0$
while $\pi^{(t)}, \alpha^{(t)}, \beta^{(t)}, \mu^{(t)}, (\sigma^2)^{(t)}$ not all converged **do**
  for $i \in 1, \ldots, n$ **do** ▷ E-step
    for $k \in 1, \ldots, K$ **do**

$$\tilde{\pi}_{ik} \leftarrow \frac{\pi_k^{(t)} \phi_k\left((y_i - \alpha_k^{(t)} x_i)/x_i^{\beta_k^{(t)}}; \mu_k, \sigma_k^2 \right)}{\sum_{j=1}^{K} \pi_j^{(t)} \phi_j\left((y_i - \alpha_j^{(t)})/x_i^{\beta_j^{(t)}}; \mu_j, \sigma_j^2 \right)}$$

    end for
  end for
  for $k \in 1, \ldots, K$ **do** ▷ M-step
    $\pi_k^{(t+1)} \leftarrow \frac{1}{n} \sum_{i=1}^n \tilde{\pi}_{ik}$
  end for
  $\alpha^{(t+1)}, \beta^{(t+1)}, \mu^{(t+1)}, \sigma^{(t+1)} \leftarrow$
    $\operatorname{argmax}_{\alpha,\beta,\mu,\sigma} \sum_{i=1}^n \sum_{k=1}^K \tilde{\pi}_{ik} \phi\left((y_i - \alpha_k x_i)/x_i^{\beta_k}; \mu_k; \sigma_k^2 \right)$ ▷ evaluated numerically
  $t \leftarrow t + 1$
end while
return $\pi^{(t)}, \alpha^{(t)}, \beta^{(t)}, \mu^{(t)}, \sigma^{(t)}$

The method for calculating the argmax in Algorithm 1 is described in Section 3.1.2. In particular, it is worth emphasising that the reparameterization (3.8) is used (it is easily generalized to the general $K \in \mathbb{N}$ case). We also clarify what we mean by convergence in Algorithm 1. We say $x^{(t)}$ has converged if

$$\left| x^{(t)} - x^{(t-1)} \right| < \max\left\{ \texttt{atol}, \max\left\{ \left| x^{(t)} \right|, \left| x^{(t-1)} \right| \right\} \texttt{rtol} \right\}$$

with $\texttt{rtol} = \texttt{atol} = 10^{-6}$.

A.2 Constructing empirical CDFs in the CEVMM model

This steps assumes that a normal K-mixture model (A.1) has already been fit and that fitting (A.2) is desired.

$$\mathbb{P}\left(Y \leq y | X = x\right) = \sum_{k=1}^{K} \pi_k F_{\text{ecdf}}^{k}\left(\frac{y - \alpha_1 x}{x^{\beta_k}}\right) \tag{A.2}$$

We take π_k, α_k and β_k to be the fitted values from Algorithm 1, and then use Algorithm 2, as described in Tendijck et al. (2021), to sample residuals $Z_1, \ldots, Z_K$.

Algorithm 2 Sample residuals in the CEVMM model

Assumes $\hat{\alpha}, \hat{\beta} \in [0,1]^K$, $\hat{\pi} \in \mathcal{S}_K$, $\hat{\mu} \in \mathbb{R}^K$, $\hat{\sigma} \in \mathbb{R}_+^K$.
Initialize M ▷ # of samples
Initialize empty arrays $Z_1, \ldots, Z_K$ ▷ Residuals
for $i \in 1, \ldots, n$ **do** ▷ Calculate mixture probabilities for each data point
  for $k \in 1, \ldots, K$ **do**
    $\tilde{\pi}_{ik} \leftarrow \frac{\hat{\pi}_k \phi_k\left((y_i - \hat{\alpha}_k x_i)/x_i^{\hat{\beta}_k}; \hat{\mu}_k, \hat{\sigma}_k^2\right)}{\sum_{j=1}^{K} \hat{\pi}_j \phi_j\left((y_i - \hat{\alpha}_j)/x_i^{\hat{\beta}_j}; \hat{\mu}_j, \hat{\sigma}_j^2\right)}$
  end for
end for
for $m \in 1, \ldots, M$ **do**
  for $i \in 1, \ldots, n$ **do**
    $C \leftarrow$ sample from Categorical $\left((\pi_{ik})_{k \in \{1, \ldots, K\}}\right)$
    Append $\frac{y_i - \hat{\alpha}_C x_i}{x_i^{\hat{\beta}_C}}$ onto array Z_C
  end for
end for
return $Z_1, \ldots, Z_K$

After running Algorithm 2, the estimators become

$$\hat{F}_{\text{ecdf}}^{k}(x) = \frac{1}{|Z_k|} \sum_{j=1}^{|Z_k|} 1\left((Z_k)_j \leq x\right),$$

where $1(\cdot)$ is the indicator function. The normal assumption, $\hat{\mu}$ and $\hat{\sigma}$ are then discarded.