Generalized Structural Time Series Model

Abdelmadjid Djennad

A thesis submitted in partial fulfilment of the requirements of London Metropolitan University

for the degree of

Doctor of Philosophy

STORM, The Statistics, Operational Research and Mathematics Research Centre
London Metropolitan University

September, 2014

Acknowledgements

First and foremost, I would like to express my sincere gratitude to Allah who has given me the inspiration, the mental and physical strength to undertake my doctoral research.

I am very thankful to my supervisors Prof. Mikis Stasinopoulos, Dr Robert Rigby and Dr Vlasios Voudouris of the STORM Research Centre for their excellent supervision, assistance and guidance in my research. Their constant inspiration, educational guidance and suggestions have been remarkable during this period of my studies. They always created various opportunities to exchange ideas. Their timely discussion and organization have resulted in a well-defined and directed path for my PhD research.

In addition, I am very grateful to Dr Robert Rigby for all I have learned from him and for his continuous help and support in all stages of my thesis and research. I would like to thank again Dr Robert Rigby and Prof. Mikis Stasinopoulos for their time and patience.

I would like to take this opportunity to sincerely thank Prof. Robert Gilchrist, former director of the STORM Research Centre, for his assistance and positive encouragement.

I also take this opportunity to sincerely thank Prof. Paul Eilers, from the department of biostatistics, Erasmus University. I am very thankful to him for giving me excellent ideas at many stages of my Ph.D.

I would like to thank my family, especially my mother and my father for always believing in me, for their continuous love and supports in my decisions, and very special thanks to my wife, my brother and my sisters.

Thanks to my supportive friends, Khurram Majeed, Samson Habte, Dr Deepthi Ratnayake, and others who made the office a friendly environment for research.

Dedication

I dedicate this thesis to my father Layachi the son of Ali and Khadidja and to my mother Nedjma the daughter of Mohammed and Zahra with deepest gratitude and warmest affection and love.

Abstract

A new class of univariate time series models is developed, the Generalized Structural (GEST) time series model. The GEST model extends Gaussian structural time series models by allowing the distribution of the dependent variable to come from any parametric distribution, including highly skew and/or kurtotic distributions. Furthermore, the GEST model expands the systematic part of time series models to allow the explicit modelling of *any or all* of the *distribution parameters* as structural terms and (smoothed) functions of independent variables. The proposed GEST model primarily addresses the difficulty in modelling time-varying skewness and kurtosis (beyond location and dispersion time series models). The originality of the thesis starts from Chapter 6 and in particular Chapter 7 and Chapter 8, with applications of the GEST model in Chapter 9. Chapters 2 and 3 contain the literature review of non-Gaussian time series models, Chapter 4 is a reproduction of Chapter 17 in Pawitan (2001), which contains an alternative method for estimating the hyperparameters instead of using the Kalman filter, and Chapter 5 is an application of Chapter 4 to smoothing Gaussian structural time series models.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The movements in a time series may be described by formulating a time series model which is intended to be taken as a full description of the conditional distribution of an observation given the past. Many observable economic and financial variables are characterized by high skewness and heavy tails. For example, since the early work of Mandelbrot (1963) and Fama (1965), the failure of the Gaussian distribution to accurately model (high frequencies) financial returns has been extensively discussed in econometric and financial literature. The departure from normality constitutes an important issue in managing market risk since it means that extreme movements in the variables are more likely than a normal distribution would predict.

There are different approaches to capture the non-Gaussian movements of economic and financial observations (e.g., Engle, 1982; Bollerslev, 1986; Baillie *et al.*, 1996; Giraitis *et al.*, 2004; De Rossi and Harvey, 2009), in recognition of 'fat tail' events.

Non-Gaussian parameter-driven time series models that rely on parametric theoretical conditional distributions offer a way of modelling observable economic and financial observations such as the Standard and Poor 500 stock index. Previous non-Gaussian time series models are based on the structural model for the mean and stochastic volatility model for the variance. For example, West *et al.* (1985) approached the problem from a Bayesian perspective using Kalman filtering to model response observations from an exponential family distribution, while Kitagawa (1987) and Kitagawa (1996) presented a comprehensive treatment of both filtering and smoothing non-Gaussian data based on approximating non-Gaussian densities by Gaussian mixtures. Durbin and Koopman (2000) modelled the mean of an exponential family distribution with a state space model and separately modelled the variance as a stochastic volatility model.

1.1 Background

For many years stationary Gaussian time series models were used for much of the modelling of time series data (Box *et al.* (1994) and Brockwell and Davis (1996)). These stationary Gaussian models have proved useful for representing a range of data and have elegant properties. The Box-Jenkins methodology provided an important step in the development of time series. Prior to their work a variety of techniques were used and their work put the modelling of stochastic processes in a unified framework.

Two lines of approach, referred to as *observation-driven* and *parameter-driven* by Cox (1981), have been adopted for fitting Gaussian time series models. In the observation-driven approach the conditional distribution of Y_t, for $t \in T$, varies over time as a function of past observations, whilst in the parameter-driven approach the conditional distribution of Y_t, for $t \in T$, varies over time as a function of past parameters of the conditional distribution.

In addition, after the pioneering work of Kolmogorov (1941a,b), Wiener (1949), though still limited to stationary situations, Kalman (1960), and Kalman and Bucy (1961), in two seminal papers, achieved remarkable optimal procedures which covered non-stationary situations, known as the Kalman filter (in the discrete-time setting of state space models) and Kalman-Bucy filter (in the continuous-time setting of state space models).

State space models, also termed as dynamic models, are based on parameter-driven approach, where the observations y_t are related to unobserved "state", γ_t, by an observational model for y_t given γ_t. The states, which may be, e.g., unobserved trend and seasonal components or time-varying covariate effects, are assumed to follow a stochastic transition model.

State space models are very flexible class of models for dynamic phenomena, their applications range from engineering sciences, with aeronautics, electrical engineering, speech recognition, over automatic monitoring/surveillance systems with important applications in intensive care medicine, to genetics, with applications in gene sequencing, biology, to environmetrics and geo-statistics. In econometrics and finance, their application is used in the prediction of stock prices, modelling the stochastic volatility, option pricing and portfolio optimization.

Furthermore, state space models are very general models that subsume a whole class of special cases of interest in much the same way that linear regression does. Although the model was originally introduced as a method primarily for use in aerospace-related research in Kalman (1960) and Kalman and Bucy (1961), it has been applied to modeling data from economics (Harrison and Stevens, 1976; Harvey and Pierse, 1984; Harvey and Todd, 1983; Kitagawa and Gersch 1984, Shumway and Stoffer, 1982), medicine (Jones, 1984) and the soil sciences (Shumway, 1988).

An early work on state space modelling with non-Gaussian data is reviewed in chapter 8 of Anderson and Moore (1979). A more recent treatment of time series analysis based on the state space model is the text by Durbin and Koopman (2012), and Shumway and Stoffer (2012), chapter 6, with application in R statistical software.

This thesis focuses on structural models for modelling time series data, and generalizes and extends univariate structural models of Harvey (1989), to a non-Gaussian framework. Although the above papers use the Kalman filter/smoother for estimating the hyperparameters. This thesis uses a novel estimation method to estimate the hyperparameters of the non-Gaussian structural models, instead of the Kalman smoother.

1.2 Research motivation

Although non-Gaussian time series models relax the assumption of the conditional Gaussian distribution, they usually model the conditional mean and occasionally the conditional variance of the non-Gaussian distribution, but rarely both. Effectively, the systematic part of the model is limited to modelling explicitly the mean or variance which are usually two of the distribution parameters. However, economic and financial variables are usually characterized conditionally by high skewness and heavy tails. Hence, it seems reasonable that other features of the distribution (e.g. skewness and kurtosis) are conditional on past information and potentially on linear, nonlinear and smooth non-parametric functions of explanatory variables. Time series models that go *beyond* the mean and variance are still challenging.

The main motivations of this research are (i) the analysis of non-Gaussian response time series models, (ii) the generalization of time series models to structural models for all the parameters of the response distribution.

The author presents a new parameter-driven (rather than data-driven) approach to the modelling of the *conditional distribution* of observable economic and financial variables by explicitly modelling all the conditional distribution parameters denoted by μ_t, σ_t, ν_t and τ_t *stochastically* within a unified framework that relaxes the assumption of the *exponential family* distribution to allow the use of highly skewed and/or platykurtotic or leptokurtotic parametric distributions. We call the proposed model the **Ge**neralized **St**ructural time series (GEST) model.

Within the GEST model, the dependent (response) variable Y_t is assumed to come from a parametric distribution with probability (density) function $f_{Y_t}(y_t|\boldsymbol{\theta_t})$, where $\boldsymbol{\theta_t}$ is a vector of unknown distribution parameters at time t. The distribution $f_{Y_t}(y_t|\boldsymbol{\theta_t})$ can be any continuous or discrete distribution. The distribution parameter vector $\boldsymbol{\theta_t}$ is restricted in the current implementation to at most four parameters denoted $\boldsymbol{\theta_t} = (\theta_{1,t}, \theta_{2,t}, \theta_{3,t}, \theta_{4,t}) = (\mu_t, \sigma_t, \nu_t, \tau_t)$, where μ_t is in general a location parameter, σ_t a scale parameter, and ν_t and τ_t are shape parameters (often affecting the skewness and kurtosis respectively). Each of the distribution parameters $(\mu_t, \sigma_t, \nu_t, \tau_t)$ is modelled by a structural time series model and, if necessary, linear, non-linear and/or smooth non-parametric models to account for explanatory variables. Each structural model of μ_t, σ_t, ν_t, τ_t is a random walk or autoregressive model (not limited to order one), and/or a seasonal effect.

1.3 Framework

The proposed GEST model adapts the generalized additive model for location, scale and shape (GAMLSS) model of Rigby and Stasinopoulos (2005) to focus on time series modelling. Applications include modelling time series counts (e.g. discrete counts) using for example a negative binomial conditional distribution and including structural models for the location and/or scale of the distribution, and modelling continuous time series data such as the Standard and Poor 500 stock index (hereafter

S&P 500) using for example a skew t conditional distribution including structural models for the location, scale, skewness and kurtosis distribution parameters.

This research deals with univariate time series models, and in particular, with the analysis and modelling of non-Gaussian time series observations. In this context, the GAMLSS framework is used, where the response variable can have any distribution which may exhibit both positive or negative skewness and high or low kurtosis. One or more of the parameters of the distribution are modelled using structural models.

GAMLSS models are regression type models in which the response variable is assumed to come from a parametric distribution, $Y \sim f_Y(y|\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a vector of unknown parameters. The distribution $f_Y(y|\boldsymbol{\theta})$ can be any continuous, discrete or mixed distribution. (A "mixed" distribution is a continuous distribution with extra discrete points. e.g. a gamma distribution with possible values at zero).

In this thesis the author follows the R implementation of GAMLSS and restricts the parameter vector $\boldsymbol{\theta}$ to at most four parameters and denote the parameters as $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4) = (\mu, \sigma, \nu, \tau)$, where μ is in general a location parameter, σ a scale parameter, and ν and τ (if needed) are shape parameters. In several of the four parameter distributions implemented within the package `gamlss.dist` in R, ν and τ are parameters effecting the skewness and the kurtosis of the distribution respectively. Within the GAMLSS framework all the parameters (μ, σ, ν, τ) can be modelled as functions of the explanatory variables.

The GEST model contains two components, the first component considers the explanatory or regression variables, and the second component is the structural time series which contains the level, trend and seasonal effects across the time series. The importance of GEST framework is that it contains a regression component that is fixed and a time series component as a structural model that is allowed to change from time point to time point. The model is applied to each of the parameters of the response variable distribution.

1.4 Thesis outline

The plan of the thesis is as follows. The thesis contains ten chapters and four appendices. Chapter 2 provides a review of non-Gaussian data-driven time series models in a chronological order. Chapter 3 provides a review of non-Gaussian structural time series models in a chronological order.

Chapter 4 describes the estimation method of the random effects and hyperparameters for Gaussian random effect models, reproducing Chapter 17 of Pawitan (2001). The reasons for reproducing Chapter 17, is that this estimation method is an alternative method for estimating the hyperparameters in structural time series models instead of using the Kalman filter. In addition, the Chapter 4 is an introduction to Chapter 8, Section 8.3 and Section 8.4. If the reader is familiar with Gaussian random effects model they can skip this chapter.

Chapter 5 contains the application of Chapter 4 to smoothing Gaussian structural time series models with some examples using the data from Commandeur and Koopman (2007). The reason for this application is to test whether the random effect estimation method agrees with the results with the Kalman filter. Also this Chapter generalizes the estimation method of Pawitan (2001) Chapter 18, and Lee, Nelder and Pawitan (2006) Chapter 9 in fitting a local level model, and has new functions and extensions of the estimation method to a local level and trend, a local level and seasonal, a local level with trend and seasonal, and a local level with random coefficient of an explanatory variable. These new functions and extensions are part of the original work of the thesis.

Chapter 6 introduces new functions for simulating and fitting Gaussian structural time series models in R. The reasons for developing these functions is to compare between the simulated mean, μ_t, and the fitted mean, $\hat{\mu}_t$, for Gaussian structural time series observations, and to test whether the estimates of the hyperparameters of the fitted model agrees with the true hyperparameters of the simulation. These fitting functions have been used for fitting real data from Commandeur and Koopman (2007) in Chapter 5 and in this chapter they are used for fitting simulated data. These functions are part the originality of the thesis

Chapter 7 introduces the theory of a new stochastic process called the Generalized Structural (GEST) stochastic process, provides new simulated examples of the GEST process in R, fitting the non-Gaussian examples with the GEST model, and derives two theorems for the properties of the GEST process.

Chapter 8 introduces the statistical framework of the GEST model, defines the maximum likelihood estimation methods globally and locally, and gives the GEST algorithm. The GEST process and the GEST model are the main and original contributions of the thesis.

Chapter 9 provides the analysis of four data sets using the GEST model. Chapter 10 provides the conclusion with proposals for future developments.

Appendix A contains the derivations of Chapter 4. Appendix B contains the derivation of skew Student t distribution. Appendix C contains the proof of the GEST theorems. Appendix D provides the R commands for Chapter 5, 7 and 9. Note that the derivations of the GEST theorems, the proof of the theorems, and the derivation of the skew Student t distribution, have been contributed by Dr Robert Rigby (one the supervisors of the author of this thesis).

Chapter 2

Review of non-Gaussian data-driven time series models

Non-Gaussian time series modelling has developed rapidly during the last decade mainly because of the progressive development in computer softwares, allowing more complicated models and massive data to be fitted, often requiring non-Gaussian models. The size of the data has increased and so has the demand for analysing highly skewed and heavy tailed data.

This chapter reviews non-Gaussian data-driven time series models for the mean, volatility, skewness and kurtosis by describing the methodology employed and the conditional distribution (of the observation given the past information) used by the authors to extend the traditional Gaussian data-driven time series models. The review is illustrated in a chronological order.

2.1 ARMA models for the mean

The autoregressive moving average models with Gaussian innovations were developed by Box and Jenkins in 1970, as a stochastic process for modelling and forecasting the mean equation in time series.

Model

The ARMA(p,q) model is defined by

$$\begin{aligned} y_t|H_t &\sim NO(\mu_t, \sigma) \\ \mu_t &= \mu + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j e_{t-j} \end{aligned}$$

where $\phi^\top = (\phi_1, \ldots, \phi_p)$ are the autoregressive parameters and $\theta^\top = (\theta_1, \ldots, \theta_q)$ are the moving average parameters, $e_t = y_t - \mu_t$ are Gaussian innovations, and H_t past information.

Distribution

- Gaussian conditional distribution.

2.1.1 Zeger and Qaqish, 1988

Model

- Developed autoregressive generalized linear model, in particular autoregressive Poisson and Gamma models for modeling count time series data.

- The expected response variable at a given time depends on the covariates and on past outcomes.

Their model is defined by

$$
\begin{aligned}
y_t|H_t &\sim EF(\mu_t, \sigma) \\
g(\mu_t) &= \eta_t = x_t^\top \beta + \tau_t \\
\tau_t &= \sum_{j=1}^{p} \phi_j \{g(y^*_{t-j}) - x_{t-j}^\top \beta\}
\end{aligned} \tag{2.1}
$$

where g is a known 'link' function relating μ_t to predictor η_t. For certain functions g it may be necessary to replace y_{t-j} with y^*_{t-j} as in equation (2.1) to avoid the non-existence of $g(y_{t-j})$ for certain values of y_{t-j}. The form of y^*_{t-j} depends on the particular function g.

Distribution

- Exponential Family conditional distribution.

The conditional distribution of y_t given past information H_t is an Exponential Family (EF) distribution with conditional mean μ_t and constant scale σ (and hence dispersion σ^2), where

$$
\begin{aligned}
E(y_t|H_t) &= \mu_t \\
V(y_t|H_t) &= \sigma^2 v(\mu_t)
\end{aligned}
$$

where $v(\mu_t)$ is a known variance function of μ_t and $H_t = \{x_t, \ldots, x_1, y_{t-1}, \ldots, y_1\}$ represents the present and past covariates and past observations at time t.

2.1.2 Li, 1994

Model

- Developed moving average Exponential Family models based on Zeger and Qaqish, (1988).
- The conditional mean equation is similar to Zeger and Qaqish model, but the τ is a moving average generalized linear model.

The moving average generalized linear model is defined by

$$\begin{aligned} y_t|H_t &\sim EF(\mu_t, \sigma) \\ g(\mu_t) &= \eta_t = x_t^\top \beta + \tau_t \\ \tau_t &= \sum_{j=1}^{q} \theta_j \{g(y^*_{t-j}) - \eta_{t-j}\} \end{aligned} \tag{2.2}$$

where g is a known 'link' function relating μ_t to predictor η_t. Similarly, for certain functions g it may be necessary to replace y_{t-j} with y^*_{t-j} as in equation (2.2) to avoid the non-existence of $g(y_{t-j})$ for certain values of y_{t-j}. The form of y^*_{t-j} depends on the particular function g.

$H_t = \{x_t, \ldots, x_1, y_{t-1}, \ldots, y_1, \mu_{t-1}, \ldots, \mu_1\}$

Distribution

- Exponential Family conditional distribution.

2.1.3 Polasek and Pai, 1998

Model

- Derived the complete conditional densities for the parameters of the ARMA models with t innovations and the ARMA models with hyperbolic innovations.
- Used MCMC methods like the Metropolis-Hastings and Gibbs sampler for the Bayesian estimation.

Distribution

- Student t and hyperbolic conditional distributions.

2.1.4 Benjamin, Rigby and Stasinopoulos, 2003

Model

- Introduced the generalized autoregressive moving average (GARMA) model.
- Extended the Gaussian autoregressive moving average (ARMA) model to Exponential Family observations.
- Extended the work of Zeger and Qaqish (1988) and Li (1994).

The GARMA(p, q) model is defined by

$$
\begin{aligned}
y_t|H_t &\sim EF(\mu_t, \sigma) \\
g(\mu_t) &= \eta_t = x_t^\top \beta + \tau_t \\
\tau_t &= \sum_{j=1}^{p} \phi_j \{g(y_{t-j}^*) - x_{t-j}^\top \beta\} + \sum_{j=1}^{q} \theta_j \{g(y_{t-j}^*) - \eta_{t-j}\} \qquad (2.3)
\end{aligned}
$$

hence,

$$g(\mu_t) \;\; = \eta_t = \;\; x_t^\top \beta + \sum_{j=1}^{p} \phi_j \{g(y^*_{t-j}) - x_{t-j}^\top \beta\} + \sum_{j=1}^{q} \theta_j \{g(y^*_{t-j}) - \eta_{t-j}\}$$

for $H_t = \{x_t, \ldots, x_1, y_{t-1}, \ldots, y_1, \mu_{t-1}, \ldots, \mu_1\}$. If $\theta_j = 0$ for $j = 1, 2, \ldots, q$, this gives a model for counts from Zeger & Qaqish's (1988) and if $\phi_j = 0$ for $j = 1, 2, \ldots, p$, this gives Li's (1994). τ_t includes the autoregressive and moving-average terms, with $\phi^\top = (\phi_1, \ldots, \phi_q)$ and $\theta^\top = (\theta_1, \ldots, \theta_p)$ are the autoregressive and moving average parameters respectively. The moving-average error terms could for example be deviance residuals, Pearson residuals, residuals measured on the original scale (i.e. $y_t - \mu_t$) or residuals on the predictor scale (i.e. $g(y_t) - \eta_t$).

The GARMA model can be used on a variety of time dependent responses which also have time dependent covariates. For example, count data with a conditional Poisson or Binomial distribution or continuous data with a conditional Gamma distribution (e.g. the volatility in a GARCH model). The GARMA model is flexible and parsimonious, it includes many well known special cases.

Distribution

GARMA-*Poisson*

Poisson GARMA(p,q) model with link g : log.

$$\begin{aligned} \eta_t &= \log(\mu_t) \\ &= \mathbf{x}_t^\top \boldsymbol{\beta} + \sum_{j=1}^{p} \phi_j [\log(y^*_{t-j}) - \mathbf{x}_{t-j}^\top \boldsymbol{\beta}] + \sum_{j=1}^{q} \theta_j [\log(y^*_{t-j}/\eta_{t-j})] \end{aligned}$$

where $y^*_{t-j} = \max(y_{t-j}, c)$, where $0 < c < 1$.

GARMA-*binomial*

Binomial GARMA(p,q) model with link g : logit.

$$\begin{aligned}\eta_t &= \text{logit}(\mu_t) \\ &= \mathbf{x}_t^\top \boldsymbol{\beta} + \sum_{j=1}^{p} \phi_j \left\{\text{logit}(y^*_{t-j}) - \mathbf{x}_{t-j}^\top \boldsymbol{\beta}\right\} + \sum_{j=1}^{q} \theta\{\text{logit}(y^*_{t-j}) - \eta_{t-j})\}\end{aligned}$$

where

$$y^*_t = \begin{cases} c & \text{if } y_t = 0 \\ y_t & \text{if } 1 \leq y_t \leq N_t - 1 \\ N_t - c & \text{if } y_t = N_t \end{cases}$$

GARMA-*gamma*

Gamma GARMA(p,q) model with link g: $\frac{1}{\mu_t}$.

$$\begin{aligned}\eta_t &= \frac{1}{\mu_t} \\ &= \mathbf{x}_t^\top \boldsymbol{\beta} + \sum_{j=1}^{p} \phi_j \left\{\frac{1}{y^*_{t-j}} - \mathbf{x}_{t-j}^\top \boldsymbol{\beta}\right\} + \sum_{j=1}^{q} \theta_j \left\{\frac{1}{y^*_{t-j}} - \eta_{t-j}\right\}\end{aligned}$$

GARMA-*GARCH*

Let ϵ_t be a Gaussian process with $\epsilon_t | H_t \sim N(0, h_t)$, and let $y_t = \epsilon_t^2$, where y_t has a conditional gamma distribution in GARMA(p,q) model with identity link since

$y_t | H_t \sim h_t \chi^2_1 \equiv Ga(\mu_t, 2)$

where $\mu_t = E(y_t|H_t) = h_t$ and

$$\begin{aligned}\mu_t &= \mathbf{x}_t^\top \boldsymbol{\beta} + \sum_{j=1}^{p} \phi_j (y_{t-j} - \mathbf{x}_{t-j}^\top \boldsymbol{\beta}) + \sum_{j=1}^{q} \theta_j (y_{t-j} - \mu_{t-j}) \\ h_t &= \beta_o + \sum_{j=1}^{p} \phi_j (y_{t-j} - \beta_o) + \sum_{j=1}^{q} \delta_j h_{t-j}\end{aligned}$$

2.1.5 Briet, Amerasinghe and Vounatsou, 2013

Model

- Extended the generalized autoregressive moving average (GARMA) model of Benjamin, Rigby and Stasinopoulos (2003), to generalized seasonal autoregressive integrated moving average (GSARIMA) models.
- Included seasonality in the GARMA model for modelling of non-Gaussian, non stationary and seasonal time series of count data.
- Implemented the package `gsarima` in R

Distribution

- Negative binomial conditional distribution.

2.2 GARCH, EGARCH and APARCH models for volatility

The ARMA model assumes a constant variance. The autoregressive conditional heteroscedastic (ARCH) model was introduced by Engle (1982) to model the changes in variance. Bollerslev (1986) extended the ARCH model to the generalized ARCH or GARCH model. A good review on ARCH and GARCH models and their extensions using a Gaussian distribution can be found in Poon and Granger (2002). This section focuses on the generalization of the ARCH models to non-Gaussian distributions.

The ARCH and GARCH models as they were first introduced by Engle and Bollerslev respectively, assumed a Gaussian distribution for the disturbances. The normality assumption in the models was insufficient to capture the observed returns in the tails and in the peak of the return series what is known as leptokurtosis.

The success of the ARCH/GARCH class of models is at capturing volatility clustering in financial markets.

According to the ARCH model the conditional variance is equal to a linear function of the past squared errors.

The ARCH(p) model is defined as

$$\begin{aligned} y_t &= \mu + e_t \\ E(e^2|H_{t-1}) &= \sigma_t^2 \\ &= \omega + \sum_{i=1}^{p} \alpha_i e_{t-i}^2 \end{aligned} \tag{2.4}$$

The GARCH(p,q) model is defined as

$$\begin{aligned} y_t &= \mu + e_t \\ E(e^2|H_{t-1}) &= \sigma_t^2 \\ &= \omega + \sum_{i=1}^{p} \alpha_i e_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \end{aligned} \tag{2.5}$$

Hence, there is a tendency for extreme values to be followed by other extreme values. However the assumption of conditional normally distributed errors made the ARCH/GARCH class incapable to model the excess kurtosis of the returns. Several alternative error distributions were proposed to this class of models, which are reviewed in this section

The reason for looking at the GARCH models with a non-Gaussian conditional distribution is because financial returns rarely have a normal conditional distribution and need a heavy tail distribution like the Student t distribution or skew Student t distribution.

2.2.1 Bollerslev, 1987

Model

- Extended the ARCH model with conditional t distribution errors by allowing the current conditional variance to be a function of past conditional variances and past squared errors of a t distribution.

The GARCH(p,q)-t model is defined as

$$\begin{aligned} y_t &= \mu + e_t \\ E(e^2|H_{t-1}) &= \sigma_t^2 \\ &= \omega + \sum_{i=1}^{p} \alpha_i e_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \end{aligned} \tag{2.6}$$

where $e_t = y_t - \mu \sim t_v$, and t_v is a t distribution with degrees of freedom parameter $v > 0$.

For a simple GARCH(1,1)-t model is

$$\begin{aligned} y_t &= \mu + e_t \\ \sigma_t^2 &= \omega + \alpha e_{t-1}^2 + \beta \sigma_{t-1}^2 \end{aligned} \tag{2.7}$$

Distribution

- Student t conditional distribution

2.2.2 Nelson, 1991

Model

- Introduced the exponential GARCH (EGARCH) model.
- Considered the family of generalized error distribution, GED, for the GARCH model.

Distribution

- Generalized error conditional distribution.

2.2.3 Forsberg and Bollerslev, 2002

Model

- Introduced a new parameterization of the normal inverse Gaussian distribution to build the GARCH-NIG model instead of the normal or Student t distribution.

Distribution

- Normal inverse Gaussian conditional distribution.

2.2.4 Mittnik, Paolella and Rachev, 2002

Model

- Presented a strict stationary GARCH model with stable Paretian innovations.
- Evaluated the GARCH processes driven by either stable Paretian or Student t innovations and compared in the context of prediction.

Distribution

- Stable Paretian and Student t conditional distributions.

2.2.5 Wurtz, Chalabi and Luksan, 2006

Methodology

- Implemented ARMA with GARCH/APARCH errors within the package `fGarch` in R statistical software using heavy tailed distributions.

Distribution

- Standardized Student t, skew standardized Student t, generalized error, and skew generalized error conditional distributions. The standardized Student t distribution has mean μ and variance exactly σ^2.

2.2.6 Ghalanos, 2012

Methodology

- Implemented GARCH, EGARCH, GJR-GARCH and APARCH within the package `rugarch` in R statistical software using heavy tailed distributions, with fitting, filtering, forecasting and simulation.

Distribution

- Standardized Student t, generalized error, generalized hyperbolic and sub-families, generalized hyperbolic skew Student t, skew generalized error, and Johnson's SU conditional distributions.

2.2.7 Broda, Haas, Krause, Paolella, and Steude, 2013

Model

- Introduced the Stable Mixture GARCH models.

- The new model nests several models or distributions for modelling asset returns, stable Paretian, mixtures of normals, normal-GARCH, stable-GARCH, and normal mixture GARCH.

Distribution

- Stable Paretian conditional distribution.

2.3 Models for skewness and kurtosis

One of the motivation behind modelling skewness and kurtosis in financial returns is that it has a great advantage for risk management and asset allocation, as well as a better description of the conditional distribution of the asset returns. Having a model for skewness and kurtosis if needed will improve the financial institutions' decisions in their asset allocations, in pricing and hedging the derivatives and in risk management.

2.3.1 Hansen, 1994

Model

- Developed a general model for autoregressive conditional density estimation using a skew Student t distribution with a GARCH-type dependence for four parameters of the conditional distribution, conditional mean, conditional variance, conditional skewness and conditional kurtosis.

- Extended Engle's ARCH model to permit parametric specifications for conditional dependence beyond the mean and variance.

- The suggestion is to model the conditional density with a small number of parameters, and then model these parameters as functions of the conditioning errors.

Distribution

- Student t and a new skewed Student t conditional distributions.

2.3.2 Harvey and Siddique, 1999

Model

- Proposed a model for time varying skewness based on skew Student t distribution but with a fixed kurtosis with a GARCH-type dependence of the third moment, from the motivation that some asset returns distribution appear to be negatively skewed implying a higher probability of negative returns than positive returns.

- Studied the conditional skewness of asset returns, and extended the traditional GARCH(1,1) model by explicitly modeling the conditional second and third moments jointly.

- Presented a framework for modeling and estimating time-varying volatility and skewness using a maximum likelihood approach assuming that the errors from the mean have a non-central conditional t distribution.

- Found a significant presence of conditional skewness and a significant impact of skewness on the estimated dynamics of conditional volatility.

- Suggested that conditional volatility is much less persistent after including conditional skewness in the modeling framework and asymmetric variance appears to disappear when skewness is included.

Distribution

- Skew Student t conditional distribution.

2.3.3 Rockinger and Jondeau, 2002

Model

- Proposed entropy densities with conditional skewness and conditional kurtosis, where the mean equal to zero, the variance equal to one and well defined skewness and kurtosis.
- Characterized the skewness and kurtosis domain over which entropy densities are well defined.
- They showed that their technique can be used to estimate GARCH model with time varying skewness and kurtosis.

Distribution

- Student t, generalized error distribution and skew Student-t (Hansen, 1994) conditional distributions

2.3.4 Brooks, Burke, Heravi and Persand, 2005

Model

- Developed a model for autoregressive conditional kurtosis, using a Student t distribution with a time varying degrees of freedom as a GARCH-type dependence. The variance and the degrees of freedom are modelled explicitly, by allowing the variance and degrees of freedom to vary over time.

Distribution

- Student t conditional distribution.

2.3.5 Lanne and Pentti, 2007

Model

- Introduced a GARCH-in-Mean (GARCH-M) model allowing for conditional skewness, using the z distribution of Barndorff-Nielsen *et al.* (1982) where a normal inverse Gaussian distribution belonging to the same family. Barndorff-Nielsen *et al.* (1982) showed that it can be represented as a variance-mean mixture of normal distributions.

- Found that the GARCH-M with z distribution is more accurate than using Student t distribution.

Distribution

- z conditional distribution.

2.3.6 Jondeau and Rockinger, 2003, 2009

Model

- Modeled the conditional volatility, skewness, and kurtosis of the returns as a GARCH type model

- Characterized the maximal range of skewness and kurtosis for which a density exists and showed that the skew Student t distribution of Hansen (1994) spans a large domain in the maximal set.

- Developed a graphical tool to summarize the impact of past shocks on the subsequent characteristics of the returns distribution.

- Introduced the concept of news impact curve (NIC) of skewness and kurtosis, which extends the well-known NIC of volatility developed by Engle and Ng (1993).

Distribution

- Skew Student t conditional distribution, Hansen (1994).

2.3.7 Wilhelmsson, 2009

Model

- Proposed a new model for financial returns with time varying variance, skewness and kurtosis based on the normal inverse Gaussian (NIG) distribution.

- Proposed a Value at Risk model with time varying variance, skewness and kurtosis using the normal inverse Gaussian (NIG) distribution.

Distribution

- Normal inverse Gaussian (NIG) conditional distribution.

Chapter 3

Review of non-Gaussian structural time series models

This chapter reviews non-Gaussian structural time series models for the mean, volatility, skewness and kurtosis by describing the methodology employed and the conditional distribution (of the observation given the past information) used by the authors to extend the traditional Gaussian state space time series models. The review is illustrated in a chronological order.

The introduction of state space models or dynamic linear models was pioneered by Kalman in 1960 and Kalman and Bucy in 1961, their methods of filtering, smoothing and forecasting were related primarily to aerospace related research for a Gaussian data. From 1974 until mid eighties the state space modeling were applied in economic related research by Harrison and Stevens, 1974; Shumway and Stoffer, 1982; Taylor, 1982; Harvey and Todd, 1983; Harvey and Pierse, 1984; Kitagawa and Gersch, 1984. However most of economic data are not Gaussian data therefore removing Gaussianity from state space models means the Kalman filter, smoother and predictor lose their optimality properties and does not apply for non-Gaussian dynamic models. For this particular problem with availability of fast computers several authors developed computer-intensive methods based on numerical integration to filter and smooth non-Gaussian state space models. Their techniques rely on approximating the non-Gaussian distribution by one or several Gaussian distributions. Markov chain Monte Carlo (MCMC) methods refer to Monte Carlo integration methods that use a Markovian updating scheme. For more discussion of these non-Gaussian state space models see Fahrmeir and Tutz (2000) chapter 8, Durbin and Koopman (2001, 2012) part 2, and chapter six of Shumway and Stoffer (2013).

3.1 Structural model for the mean

Most of state space models use Markov chain Monte Carlo (MCMC) methods, for example Gibbs sampling, importance sampling and the Metropolis algorithm, to compute the posterior distributions for the parameters of the states (hyperparamters). This method is very computer intensive and time consuming for parameter estimation. The problem with many early state space approaches is that the methods involved approximations of unknown hyperparameters of unobserved state variables, with convergence issues. For these reasons some authors later on attempted to solve these issues.

State space models contain two classes of variables: the unobserved state variables which describe the development over time of the underlying system and the observations. A simple Gaussian local level model is defined as:

$$\begin{aligned} y_t|H_t &\sim NO(\mu_t, \sigma) \\ y_t &= \mu_t + e_t \\ \mu_t &= \mu_{t-1} + \eta_t \end{aligned} \tag{3.1}$$

where the first equation is called the observation equation and the second equation is the state equation, and $e_t \sim NO(0, \sigma_e^2), \eta_t \sim NO(0, \sigma_\eta^2)$. The observation equation is a linear combination of unobserved variable or state and a white noise with Gaussian distribution, where the state equation is Markov process (random walk) with a white noise innovation. Both innovations are assumed normal in Gaussian structural time series models. However in non-Gaussian state space models, the observations innovations are assumed non-Gaussian, whereas the state in some models are assumed non-Gaussian and the others are assumed Gaussian. In the Gaussian structural model, the Kalman filter and smoother are applicable and provide optimal solutions to the process, but in the non-Gaussian model, the standard Kalman filter is not appropriate.

3.1.1 West, Harrison and Migon, 1985

Model

- Introduced state space models for exponential family observations with Gaussian state.
- Developed dynamic Bayesian models for application in nonlinear, non Gaussian time series and regression problems, provided dynamic extensions of standard generalized linear models.

- They used conjugate priors and at each update the posterior density was approximated to retain the conjugate structure, or assuming conjugate prior-posterior distributions for the natural parameter of the exponential family.

- Derived an approximate filter as an extension to Kalman filter for estimation of time-varying states or parameters.

Distribution

- Exponential Family conditional distribution.

3.1.2 Harvey and Durbin, 1986

Model

- modelled the effects of seat belt legislation on British road casualties over the period January 1969 to December 1984 using the Poisson distribution with a local linear trend and seasonality model for the mean equation.

- The mean equation has three state components, a random walk trend, seasonality, and intervention parameter which measures the effects of the seat belt law.

$$
\begin{aligned}
y_t &= \mu_t + \gamma_t + \sum_{j=1}^{k} \delta_j x_{jt} + \lambda \omega_t + e_t \\
\mu_t &= \mu_{t-1} + \beta_{t-1} + \eta_t \\
\beta_t &= \beta_{t-1} + \zeta_t \\
\gamma_t &= \sum_{j=1}^{s/2} \gamma_{jt} + w_t
\end{aligned}
\tag{3.2}
$$

where μ_t is the trend at time t, β_t is the slope at time t, x_{jt} is the value of the explanatory variable at time t and δ_j is its coefficient, ω_t is the intervention variable, and γ_t is seasonality. The innovations $e_t, \eta_t, \zeta_t, w_t$ are normal with zero expected values and unknown variances or hyperparameters.

Distribution

- Poisson conditional distribution

3.1.3 Kitagawa, 1987, 1989, 1990

Model

- Modelled non stationary time series data using non-Gaussian state-space approach, where, both the states and the observations are not Gaussian.
- Derived recursive formulas of prediction, filtering, and smoothing for the state estimation and identification of the non-Gaussian state-space model.
- Proposed a numerical method based on piecewise linear approximation.
- Approximated non-Gaussian densities by a mixture of normal distributions. At each update he collapsed the conditional density into a smaller number of components.

Distribution

- Mixture of normal distributions conditional distribution.

3.1.4 Fahrmeir, 1992

Model

- Introduced multivariate dynamic generalized linear models for time series analysis with exponential family observations and Gaussian state.
- Estimated the time-varying parameters by posterior modes to avoid a full Bayesian analysis, which are based on numerical integration and are computationally critical for higher dimensions.
- Proposed a generalization of the extended Kalman filter for conditionally Gaussian observations for approximate posterior mode filtering and smoothing. The recursions also can be interpreted as a simplified version of Fisher scoring for the posterior mode.
- Estimated the unknown hyperparameters by an empirical Bayes approach.

Distribution

- Exponential Family conditional distribution.

3.1.5 Shephard and Pitt, 1997

Methodology

- Used Markov chain Monte Carlo technique to carry out simulation smoothing and Bayesian posterior analysis of parameters.
- Used importance sampling to estimate the likelihood function for classical inference.

Distribution

- Exponential Family conditional distribution.

3.1.6 Durbin and Koopman, 2000

Model

- Considered non-Gaussian distribution for the state equation and non-Gaussian conditional distribution for the observations given the state.
- For the state, they used a heavy-tailed distribution to model structural shifts, and for the observations they considered both exponential family and heavy-tailed conditional distributions.
- Used importance sampling and antithetic variables simulation techniques.
- As an example they modeled van drivers killed in UK with a Poisson conditional distribution and modelled gas consumption in UK with a Student t conditional distribution.

Distribution

- Exponential Family and Student t conditional distributions.

3.1.7 Nakajima, Kunihama, Omori and Fruhwirth-Schnatter, 2012

Model

- Proposed a new state space approach to model the time-dependence in an extreme value process.

- Extended the generalized extreme value distribution to incorporate the time-dependence using a state space representation where the state variables either follow an autoregressive (AR) process or a moving average (MA) process with innovations arising from a Gumbel distribution.

- Proposed an efficient algorithm to implement the Markov chain Monte Carlo method where they exploited an accurate approximation of the Gumbel distribution by a ten-component mixture of normal distributions.

Distribution

- Ten-component mixture of normal conditional distribution.

3.2 Stochastic volatility model

Taylor (1982) introduced a stochastic volatility (SV) model for modelling time varying variance in a state space form with normal errors, but various extensions were proposed with non normal errors, as many empirical studies in finance and econometrics shown strong evidence of heavy tails for conditional mean errors in financial time series data (see for example Mandelbrot, 1963; Fama, 1965; Chib *et al.*, 2002; Jacquier *et al.*, 2004). SV models are alternative to the ARCH and GARCH models. Ghysels, Harvey and Renault (1996) and Shephard (1996) have a good review of the SV models with their applications.

Many generalizations of the standard SV models have been proposed, for example SV with leverage effects and SV with heavy tailed errors. The SV model with Student t and skew Student t are the most popular model to account for heavier tailed returns.

Let

$$y_t = \sigma_t e_t.$$

Taylor defined the stochastic volatility in a state space form by squaring the observations and taking the logarithms for the observations equations where the stochastic variance is considered as an unobserved state process,

$$\begin{aligned} \log y_t^2 &= \log \sigma_t^2 + \log e_t^2, \\ \log \sigma_t^2 &= \phi \log \sigma_{t-1}^2 + \eta_t. \end{aligned} \tag{3.3}$$

Letting $Y_t = \log y_t^2$ and $h_t = \log \sigma_t^2$ then

$$\begin{aligned} Y_t &= h_t + \log e_t^2, \\ h_t &= \phi h_{t-1} + \eta_t. \end{aligned} \tag{3.4}$$

Or alternatively,

$$y_t | \sigma_t^2 \sim N(0, \sigma_t^2),$$

where

$$\begin{aligned}\sigma_t^2 &= \sigma^2 \exp(h_t),\\ h_t &= \phi h_{t-1} + \eta_t,\\ \eta_t &\sim N(0, \sigma_\eta^2).\end{aligned} \tag{3.5}$$

Hence, the conditional variance is assumed to be lognormally distributed

$$\sigma_t^2 | H_{t-1} \sim LOGNO(\ln \sigma^2 + \phi h_{t-1}, \sigma_\eta^2).$$

Hence, the basic univariate stochastic volatility model specifies that conditional volatility follows a log-normal autoregressive model with innovations assumed to be independent of the innovations in the conditional mean equation.

If the innovations e_t^2 in 3.4 has a log normal distribution, then a stochastic volatility model for Y_t would have been a normal state space model, but $Y_t = \log y_t^2$ is not normal and assuming e_t is normal then e_t^2 is distributed as a chi-squared random variable with one degree of freedom and $\log e_t^2$ is distributed as the log chi-squared with one degree of freedom.

Various approaches have been proposed to the fitting of stochastic volatility either by Bayesian approaches using MCMC techniques or non-Bayesian approaches using quasi-maximum likelihood estimation algorithm. They approximate the log chi-squared density by a mixture of Gaussian densities to apply the Kalman filter techniques.

3.2.1 Shephard, 1994

Model

- Developed simulation techniques to extend the applicability of the usual Gaussian state space filtering and smoothing techniques to a class of non-Gaussian time series models.
- Fitted the stochastic volatility model with simulation.

Distribution

- Mixture of normal distributions conditional distribution.

3.2.2 Kim, Shephard and Chib, 1998

Model

- Proposed modelling $\log e_t^2$, a log of the chi-squared random variable with one degree of freedom, by a mixture of seven normal distributions to approximate the first four moments of the observational error distribution.
- Sampled all the unobserved volatilities at once using an approximating offset mixture model, followed by an importance reweighting procedure.
- Developed simulation-based MCMC methods for filtering, likelihood evaluation and model failure diagnostics.

Distribution

- Mixture of normal distributions conditional distribution.

3.2.3 Nagahara and Kitagawa, 1999

Model

- Proposed a non-Gaussian stochastic volatility model as an extension of the ordinary stochastic volatility model, assuming that the time series is distributed as a Pearson type-VII distribution, where the scale parameter of the distribution, which corresponds to the volatility of the process, is stochastic and is described by an autoregressive model with a constant term.

- Applied a non-Gaussian filter for estimating the parameters of the stochastic volatility model

- They suggested that the model can be further generalized to the case where the shape parameter of the Pearson type-VII distribution is also time-varying.

Distribution

- Pearson type-VII conditional distribution.

3.2.4 Chib, Nardari and Shephard, 2002

Model

- They estimated the SV model with jump and Student t errors but without leverage effects by extending Kim, Shephard and Chib (1998) model.

- Developed efficient and fast Bayesian Markov chain Monte Carlo (MCMC) estimation algorithm for estimating stochastic volatility models with Student t distribution.

Distribution

- Student t conditional distribution.

3.2.5 Nagahara, 2003

Model

- Proposed a fast and efficient algorithm by using an analytic approximation of successive conditional probability density functions for predicting, filtering and smoothing recursively using the Pearson type-VII distribution.

Distribution

- Pearson type-VII conditional distribution which includes the normal and Student t distributions.

3.2.6 Jacquier, Polson and Rossi, 2004

Model

- Developed a Bayesian MCMC method for estimating stochastic volatility with fat tailed and correlated errors, providing the first likelihood-based procedure for stochastic volatility with correlated errors.

- Extended the basic stochastic volatility model to allow for a leverage effect via correlation between the volatility and mean innovations, and for fat-tails in the mean equation innovation.

Distribution

- Student t conditional distribution.

3.2.7 Omori, Chib, Shephard and Nakajima, 2007

Model

- Extended method that was developed for SV models without leverage to models with leverage.
- Approximated the joint distribution of the outcome and volatility innovations by a suitably constructed ten-component mixture of bivariate normal distributions.
- The resulting posterior distribution is summarized by MCMC methods and the small approximation error in working with the mixture approximation is corrected by a reweighting procedure.
- They described some extensions of their method for superposition models (where the log-volatility is made up of a linear combination of heterogenous and independent autoregressions) and heavy-tailed error distributions (Student and log-normal).

Distribution

- Scale mixture of normal and Student t distributions conditional distribution.

3.2.8 Choy, Wan and Chan, 2008

Model

- Introduced the scale mixtures of uniform and the scale mixtures of normal representation to the Student t density and show that the setup of a Gibbs sampler for the SV-t model can be simplified, where the full conditional distribution of the log-volatilities has a truncated normal distribution which enables an efficient Gibbs sampling algorithm.

- Modeled a heavy-tailed stochastic volatility SV-t with leverage effect by expressing the bivariate Student t distribution as a scale mixture of bivariate normal distributions.

Distribution

- Student t and normal conditional distributions

3.2.9 Wang, Chan and Choy, 2011

Model

- Modeled a heavy-tailed stochastic volatility (SV) model with leverage effect, using a bivariate Student t distribution to model the error innovations of the return and volatility equations.

- Proposed an alternative formulation by first deriving a conditional Student t distribution for the return and a marginal Student t distribution for the log-volatility and then express these two Student t distributions as a scale mixture of normal (SMN) distributions.

- Their approach separated the sources of outliers and allows for distinguishing between outliers generated by the return process or by the volatility process, hence, improving Choy et al. (2008) approach.

- Expressed the Student t distribution as a SMN distribution at different stages, where the mixing parameters arose from the SMN representation play the role of identifying possible outliers.

Distribution

- Scale mixture of normal (i.e. Student t) conditional distribution.

3.2.10 Nakajima and Omori, 2009, 2012

Methodology

- In 2009 they proposed an efficient and fast Markov chain Monte Carlo estimation methods for the stochastic volatility model with leverage effects, heavy tailed errors and jump components, and for the stochastic volatility model with correlated jumps.
- In 2012 they provided a Bayesian analysis of stochastic volatility with leverage using generalized hyperbolic skew Student t distribution, and described an efficient Markov chain Monte Carlo estimation method that exploits a normal variance-mean mixture representation of the error distribution with an inverse gamma distribution as the mixing distribution.

Distribution

- Generalized hyperbolic (GH) skew Student t conditional distribution

3.2.11 Tsiotas, 2012

Model

- Introduced generalised asymmetric stochastic volatility (ASV) models that take account of volatility, asymmetry, skewness and excess kurtosis and using heavy-tailed conditional distributions.

- Used the Bayesian Markov chain Monte Carlo methods (Gibbs sampler and Metropolis-Hasting algorithm) for ASV estimation.

Distribution

- Skew normal, Student t and skew Student t conditional distributions of Azzalini (1985)

3.3 Models for skewness and kurtosis

All the above models are either modelling the structural mean or stochastic volatility, but not modelled jointly as in the data-driven approach. One of the aims of the author is to model skewness and kurtosis using the structural time series approach, and build a general structural time series framework for modelling all the parameters of the conditional distribution jointly and explicitly. This new class of model is the main contribution of the author to the current literature, where the structural time series models are extended by explicitly modelling the conditional mean, variance, skewness and kurtosis jointly. The author considers the modelling and estimation of non-Gaussian observational noise density by smoothing Gaussian state parameters.

Chapter 4

Random effect models

4.1 Introduction

This Chapter describes the estimation method of the random effects and hyperparameters for Gaussian random effect models, reproducing Chapter 17 of Pawitan (2001). The reasons for following Chapter 17 is that this estimation method is an alternative method for estimating the hyperparameters in structural time series models instead of using the Kalman filter. In addition, the Chapter 4 is an introduction to Chapter 8, Section 8.3 and Section 8.4. If the reader is familiar with Gaussian random effects model they can skip this chapter.

Random effect models provide a flexible framework for modelling Gaussian and non-Gaussian time series data. They provide a unified methodology for treating a wide range of problems in applied statistics and in particular in time series analysis. They relate time series observations, y_t, to a sequence of unknown or unobserved vector of random effects, γ_t, typically it includes a trend and/or seasonality.

A correlated random effects model assumes that the correlation arises among repeated measurements through time, they are also used in longitudinal studies, analysing the longitudinal relationships between the explanatory variables and the response variable.

Lee, Nelder and Pawitan (2006), p.148 and p.233, show that linear mixed models with correlated random effects over time can include many of state space models of Harvey (1989), and the extended likelihood in random effect models can be used instead of the Kalman filter for estimating the hyperparameters, (see Nelder 2000), where the underlying signal is assumed to be a random or unobserved state.

For Gaussian linear state space models, optimal conditional mean estimates are achieved by the linear Kalman filter and smoother. For nonlinear and non-Gaussian state space models, several extensions have been proposed, for example Sage and Melsa (1971), Anderson and Moore (1979), West, Harrison and Migon (1985), Kitagawa (1987), and Durbin and Koopman (2001). Although the Kalman filter was derived for filtering and smoothing Gaussian data, it has a major drawback in smoothing and filtering non-Gaussian data.

One general approach for modelling non-Gaussian state space models with non-Gaussian distributions is via the use of random effect models. Fahrmeir and Kaufmann (1991), Fahrmeir (1992) and Fahrmeir (1996) propose the same approach as an approximative method to model non-Gaussian state space models based on random effects models and generalized linear mixed models as in Breslow and Clayton (1993). Diggle, Liang and Zeger (1994) consider random effects models with first order autoregressive model for time series data. Zeger and Diggle (1994) propose a semiparametric model for longitudinal data, where the covariate entered parametrically and the time effect entered nonparametrically.

Generalized linear mixed models were proposed as a general framework by Breslow and Clayton (1993). They include an unobserved vector of random effects in a generalized linear model, assumed to arise from a normal distribution, and use an approximation of the marginal quasi-likelihood based on Laplace's method, leading to equations based on penalized quasi-likelihood. Generalized additive mixed models were introduced by Lin and Zhang (1999) which extended the generalized additive models of Hastie and Tibshirani (1990) to accommodate overdispersion and correlation in the data. Generalized additive models for location, scale and shape were introduced by Rigby and Stasinopoulos (2005) as a general unified framework for modelling all the parameters of the distribution with the random effects.

A simple random walk order 1 linear state space model is given by the observation equation, $y_t = \gamma_t + e_t$, where $\gamma_t = \gamma_{t-1} + b_t$ is the linear transition equation or state equation and e_t and b_t are Gaussian noises. Given the observations, $y_1, \ldots, y_T$, estimation (smoothing) of the unknown states γ_t is of primary interest to the author using the random effects models.

The estimation of Gaussian linear mixed models is described and analysed here, using the derivations from Pawitan (2001), chapter 17, with applications and illustrative examples to state space models in the next chapter, and generalizations that follow in the later chapters. The estimation of Gaussian linear mixed models is also discussed in Lee, Nelder and Pawitan (2006), chapter 5. The estimation of generalized liner mixed models is discussed in Fahrmeir and Tutz (2001), chapter 8, with application to state space models.

All state space model examples in the next chapter are estimated using the Q function derived from random effects estimation methods as an alternative to the Kalman smoother. The examples are taken from Commandeur and Koopman (2007) and the results are compared with their results.

4.2 Gaussian linear mixed models

Random effect models can be used for analyzing Gaussian and non-Gaussian data that are assumed to be clustered or correlated. The clustering can be due to repeated measurements over time. Moreover, the estimation techniques of random effect models can be used for Gaussian and non-Gaussian state space models. These estimation methods are discussed here using Pawitan (2001).

Consider a Gaussian liner mixed model

$$y = X\beta + Z\gamma + e \tag{4.1}$$

where y be an N-vector of outcome data, X and Z are (N x p) and (N x q) design matrices for the fixed effects parameter β and the random effects γ respectively, where $e \sim N(0, \Sigma)$, $\gamma \sim N(0, D)$, and γ and e are independent. The variance matrices $\Sigma = \sigma_e^2 I_N$ and $D = \sigma_\gamma^2 I_N$ are parameterized by an unknown variance component parameter $\theta = (\sigma_e^2, \sigma_\gamma^2)$.

Conditional on an unobserved random effects γ the outcome vector y is a multivariate normal with mean and variance:

$$\begin{aligned} E(y|\gamma) &= X\beta + Z\gamma \\ V(y|\gamma) &= \Sigma = \sigma_e^2 I_N. \end{aligned} \tag{4.2}$$

Hence,

$$
\begin{aligned}
y|\gamma &\sim N(X\beta + Z\gamma, \Sigma), \\
\gamma &\sim N(0, D),\ D = \sigma_\gamma^2 I_N,
\end{aligned}
\tag{4.3}
$$

and the marginal probability density of y is given by:

$$
\begin{aligned}
f(y) &= \int f(y|\gamma) f(\gamma) d\gamma \\
&= |2\pi V|^{-\frac{1}{2}} \exp\{-\frac{1}{2}(y - X\beta)^\top V^{-1}(y - X\beta)\}.
\end{aligned}
\tag{4.4}
$$

See Appendix A.1 for the derivation of the marginal probability density of y.

4.3 Estimation of the fixed parameters

The marginal log-likelihood of the fixed parameters (β, θ) is

$$
\log L(\beta, \theta) = -\frac{1}{2} \log |V| - \frac{1}{2}(y - X\beta)^\top V^{-1}(y - X\beta) \tag{4.5}
$$

where the parameter(s) $\theta = (\sigma_e^2, \sigma_\gamma^2)$ enters through the marginal variance V.

The *profile of the marginal log-likelihood* of the variance parameter θ is

$$
pl(\theta) = -\frac{1}{2} \log |V| - \frac{1}{2}(y - X\hat{\beta})^\top V^{-1}(y - X\hat{\beta}) \tag{4.6}
$$

where $\hat{\beta}$ are the fitted values computed by the weighted least square formula

$$\hat{\beta} = \left(X^\top V^{-1} X\right)^{-1} X^\top V^{-1} y \tag{4.7}$$

Patterson and Thompson (1971) and Harville (1974) derived a restricted maximum likelihood (REML) adjusted so it takes into account the estimation of β, by integrating out both β and γ from $f(y|\beta,\gamma)f(\gamma)$.

The *modified profile of the marginal log-likelihood* for θ is given by

$$pl_m(\theta) = -\frac{1}{2}\log|V| - \frac{1}{2}\log|X^\top V^{-1} X| - \frac{1}{2}(y - X\hat{\beta})^\top V^{-1}(y - X\hat{\beta}) \tag{4.8}$$

where $-\frac{1}{2}\log|X^\top V^{-1} X|$ is the extra REML adjustment.

See Appendix A.2.

4.4 Estimation of the random effects

The joint (or extended) likelihood of **all** the parameters is based on the **joint** density distribution of (y, γ):

$$L(\beta, \theta, \gamma) = p(y|\beta, \gamma)p(\gamma|\theta) \tag{4.9}$$

The log-likelihood of the random effects γ, by omitting the constant which does not depend on the parameters, is given by

$$\begin{aligned} \log L(\beta,\theta,\gamma) &= -\frac{1}{2}\log|\Sigma| - \frac{1}{2}(y - X\beta - Z\gamma)^\top \Sigma^{-1}(y - X\beta - Z\gamma) \\ &\quad -\frac{1}{2}\log|D| - \frac{1}{2}\gamma^\top D^{-1}\gamma \end{aligned} \tag{4.10}$$

The estimate $\hat{\gamma}$ is the solution of

$$\hat{\gamma} = (Z^\top \Sigma^{-1} Z + D^{-1})^{-1} Z^\top \Sigma^{-1}(y - X\beta) \tag{4.11}$$

See Appendix A.3.

4.5 Computing the hyperparameters via $\hat{\beta}$ and $\hat{\gamma}$

The profile of the marginal log-likelihood (4.6) for the variance component parameter $\theta = (\sigma_e^2, \sigma_\gamma^2)$ is not computationally desirable due to the terms involving V^{-1} and the following alternative is more desirable because it is easier to compute.

$$\begin{aligned} pl(\theta) &= -\frac{1}{2}\log|\Sigma| - \frac{1}{2}(y - X\hat{\beta} - Z\hat{\gamma})^\top \Sigma^{-1}(y - X\hat{\beta} - Z\hat{\gamma}) \\ &\quad -\frac{1}{2}\log|D| - \frac{1}{2}\hat{\gamma}^\top D^{-1}\hat{\gamma} - \frac{1}{2}\log|Z^\top \Sigma^{-1} Z + D^{-1}| \\ &= \log L(\hat{\beta},\theta,\hat{\gamma}) - \frac{1}{2}\log|Z^\top \Sigma^{-1} Z + D^{-1}| \end{aligned} \tag{4.12}$$

where θ enters the function through Σ, D, $\hat{\beta}$ and $\hat{\gamma}$. Pawitan (2001) proves that $pl(\theta)$ in (4.6) is equal to $pl(\theta)$ in (4.12). The profile of the marginal log-likelihood (4.6) or (4.12) is the joint (or extended) log likelihood evaluated at $(\hat{\beta}, \hat{\gamma})$ with the extra term

$$-\frac{1}{2}\log|Z^\top\Sigma^{-1}Z+D^{-1}|,$$

where the matrix $(Z^\top\Sigma^{-1}Z+D^{-1})$ is the Fisher information of γ. The profile likelihood (4.12) is simpler than (4.6), since the matrices involved are simpler than V^{-1}.

See Appendix A.4.

4.6 Estimation procedure

Computationally the whole estimation of θ, β and γ can be done through maximising Q as an objective function, where

$$\begin{aligned} Q &= -\frac{1}{2}\log|\Sigma| - \frac{1}{2}(y-X\beta-Z\gamma)^\top\Sigma^{-1}(y-X\beta-Z\gamma) \\ &\quad -\frac{1}{2}\log|D| - \frac{1}{2}\gamma^\top D^{-1}\gamma - \frac{1}{2}\log|Z^\top\Sigma^{-1}Z+D^{-1}| \end{aligned}$$

The score equations for β and γ yield the usual formulae for $\hat{\beta}$ and $\hat{\gamma}$ at fixed θ.

The following algorithm can be applied for estimating θ, β and γ, by starting with an initial estimate for θ, then:

Numerically maximize $Q(\theta, \hat{\beta}_\theta, \hat{\gamma}_\theta)$ over θ until convergence, where within the procedure the $Q(\theta, \hat{\beta}_\theta, \hat{\gamma}_\theta)$ is calculated by

(1) Given θ, compute $\hat{\beta}_\theta$ and $\hat{\gamma}_\theta$ using:

$$\begin{aligned} \hat{\beta}_\theta &= (X^\top V^{-1}X)^{-1}X^\top V^{-1}y \\ \hat{\gamma}_\theta &= (Z^\top\Sigma^{-1}Z+D^{-1})^{-1}Z^\top\Sigma^{-1}(y-X\beta) \end{aligned}$$

(2) Fixing β and γ at the values $\hat{\beta}_\theta$ and $\hat{\gamma}_\theta$, calculate $Q(\theta, \hat{\beta}_\theta, \hat{\gamma}_\theta)$, where

$$\begin{aligned} Q(\theta, \hat{\beta}_\theta, \hat{\gamma}_\theta) &= -\frac{1}{2}\log|\Sigma| - \frac{1}{2}(y - X\hat{\beta}_\theta - Z\hat{\gamma}_\theta)^\top \Sigma^{-1}(y - X\hat{\beta}_\theta - Z\hat{\gamma}_\theta) \\ &\quad -\frac{1}{2}\log|D| - \frac{1}{2}\hat{\gamma}_\theta^\top D^{-1}\hat{\gamma}_\theta - \frac{1}{2}\log|Z^\top\Sigma^{-1}Z + D^{-1}|. \end{aligned}$$

See Appendix A.5 for an alternative method of estimation.

4.7 Several random effects

Pawitan (2001), page 452, extends the estimation procedure of one random factor to two random factors, and Rigby and Stasinopoulos (2013) extends the estimation procedure furthermore, of one random factor to N random factors.

The extension to more than one random factor has a similar computational procedure as one random factor with some extension on the estimations; if we have two random factors γ_1 and γ_2, the conditional mean of y will be

$$E(y|\gamma_1, \gamma_2) = X\beta + Z_1\gamma_1 + Z_2\gamma_2$$

and variance Σ, where X, Z_1, Z_2 are appropriate design matrices and β is the fixed effects parameter. γ_1 and γ_2 are independent normal with mean zero and variance D_1 and D_2 respectively.

See Appendix A.6 for the derivations the Q function and the hyperparameters.

Chapter 5

Smoothing for Gaussian structural time series models

5.1 Introduction

This Chapter contains the application of Chapter 4 to smoothing Gaussian structural time series models with some examples using the data from Commandeur and Koopman (2007). The reason for this application is to test whether the random effect estimation method agrees with the results with the Kalman filter. In addition, this Chapter generalizes the estimation method of Pawitan (2001) Chapter 18, and Lee, Nelder and Pawitan (2006) Chapter 9 in fitting a local level model, and has new functions and extensions of the estimation method to a local level and trend, a local level and seasonal, a local level with trend and seasonal, and a local level with random coefficient of an explanatory variable.

State space models allow a natural interpretation of a time series as the result of several components, such as trend, seasonal or regressive components. At the same time, they have an elegant and powerful probabilistic structure. They can be used for modeling univariate or multivariate time series, also in presence of non-stationarity, structural changes, irregular patterns. The state space model or dynamic linear model was introduced in Kalman (1960) and Kalman and Bucy (1961). Although the model was originally introduced as a method primarily for use in aerospace-related research, it has been applied to modeling data from economics (Harrison and Stevens, 1976; Harvey and Pierse, 1984; Kitagawa and Gersch, 1984; Durbin and Koopman, 2001; Shumway and Stoffer, 2013). In last two decades there has been an increasing interest for the application of non Gaussian state-space models in time series analysis; see for example West and Harrison (1997), Durbin and Koopman (2001), the overviews by Kunsch (2001) and Migon *et al.* (2005).

5.2 Local level model

A basic example of the state space model is the local level model, which is a simple Gaussian signal plus noise model.

Let $\mathbf{y}^\top = (y_1, y_2, \ldots, y_n)$ be the observed vector of response variable. The random walk process of the local level model is defined as:

$$\begin{aligned} y_t &= \gamma_t + e_t \\ \gamma_t &= \gamma_{t-1} + b_t \end{aligned} \tag{5.1}$$

where e_t and b_t are two independent Gaussian white noise series. $e_t \sim N(0, \sigma_e^2)$, $b_t \sim N(0, \sigma_b^2)$.

The irregular and signal disturbances, e_t and b_t respectively, are mutually independent with mean zero and variances σ_e^2 and σ_b^2 respectively. The signal to noise ratio, $\lambda = \sigma_b^2/\sigma_e^2$ plays the key role in determining how observations should be weighted for prediction and signal extraction.

Note that γ_t is a random walk which is not directly observable, and y_t is the observed data with observational noise e_t. The dynamic dependance of y_t is governed by the hidden state γ_t. The first equation is called the observation or measurement equation, while the second equation is called the state equation. In the state equation, time dependencies in the observed time series are dealt with by letting the state at time $t+1$ be a function of the state at time t. In state space models, the unknown estimates are known as hyperparameters (variances). Unlike classical regression analysis, when a state space model contains two or more hyperparameters, the maximum likelihood estimation of these variances requires an iterative procedure. The iterations aim to maximise the log-likelihood with respect to the hyperparamters. Numerical optimization methods are employed (`nlminb` and `optim` functions from `R` statistical software) and they are based on an iterative search process to find the maximum in a numerically efficient way.

Random walk order (1),

$$\begin{aligned} y_t &= \gamma_t + e_t \\ b_t &= \gamma_t - \gamma_{t-1} = \mathbf{D}\gamma_t, \end{aligned}$$

in a random walk order (2),

$$\begin{aligned} y_t &= \gamma_t + e_t \\ \gamma_t &= 2\gamma_{t-1} - \gamma_{t-2} + b_t \\ b_t &= \gamma_t - 2\gamma_{t-1} + \gamma_{t-2} = \mathbf{D}_2\gamma_t \end{aligned}$$

where $e_t \sim N(0, \sigma_e^2)$ and $b_t \sim N(0, \sigma_b^2)$, this can be generalized to random walk order J, i.e. rw(J).

Autoregressive order (J)

The autoregressive process of the local level model is defined as

$$
\begin{aligned}
y_t &= \gamma_t + e_t \\
\gamma_t &= \sum_{j=1}^{J} \phi_j \gamma_{t-1} + b_t
\end{aligned}
\tag{5.2}
$$

The smoothing matrices $\mathbf{D}$ and $\mathbf{D}^\top\mathbf{D}$ for a random walk of order (1) are defined as

$$
\mathbf{D} = \begin{pmatrix}
-1 & 1 & & & 0 \\
0 & -1 & 1 & & \\
 & .. & .. & .. & \\
 & .. & .. & .. & \\
0 & & & -1 & 1
\end{pmatrix}
$$

$$
\mathbf{D}^\top\mathbf{D} = \begin{pmatrix}
1 & -1 & & & 0 \\
-1 & 2 & -1 & & \\
 & .. & .. & .. & \\
 & & -1 & 2 & -1 \\
0 & & & -1 & 1
\end{pmatrix}
$$

The smoothing matrices $\mathbf{D}_2$ and $\mathbf{D}_2^\top \mathbf{D}_2$ for a random walk of order (2) are defined as

$$\mathbf{D}_2 = \begin{pmatrix} 1 & -2 & 1 & & & & & 0 \\ 0 & 1 & -2 & 1 & & & & \\ & .. & .. & .. & .. & .. & .. & \\ & & & & 1 & -2 & 1 & 0 \\ 0 & & & & & 1 & -2 & 1 \end{pmatrix}$$

$$\mathbf{D}_2^\top \mathbf{D}_2 = \begin{pmatrix} 1 & -2 & 1 & & & & & & 0 \\ -2 & 5 & -4 & 1 & & & & & \\ 1 & -4 & 6 & -4 & 1 & & & & \\ & .. & .. & .. & .. & .. & .. & & \\ & & & & 1 & -4 & 6 & -4 & 1 \\ & & & & & 1 & -4 & 5 & -2 \\ 0 & & & & & & 1 & -2 & 1 \end{pmatrix}$$

The smoothing matrices $\mathbf{D}$ and $\mathbf{D}^\top\mathbf{D}$ for an autoregressive of order (1) are defined as:

$$\mathbf{D} = \begin{pmatrix} -\phi & 1 & & & 0 \\ 0 & -\phi & 1 & & \\ & .. & .. & .. & \\ & .. & .. & .. & \\ 0 & & & -\phi & 1 \end{pmatrix}$$

$$\mathbf{D}^\top\mathbf{D} = \begin{pmatrix} \phi^2 & -\phi & 0 & & & & & 0 \\ -\phi & 1+\phi^2 & -\phi & 0 & & & & \\ 0 & -\phi & 1+\phi^2 & -\phi & 0 & & & \\ & .. & .. & .. & .. & .. & .. & \\ & .. & .. & .. & .. & .. & .. & \\ & & & & 0 & -\phi & 1+\phi^2 & -\phi \\ 0 & & & & & 0 & -\phi & \phi^2 \end{pmatrix}$$

5.2.1 Examples of local level model

Norwegian road fatalities

Applying the local level model (i.e. a random walk order 1) to the log of the annual number of road traffic fatalities in Norway from 1970 to 2003. The maximum likelihood estimates of the hyperparameters are given in Table 5.1, and the fitted random walk local level is plotted with the data in Figure 5.1. The R fitting commands and output are given in Appendix D.

Table 5.1: The estimated hyperparameters of the Norwegian fatalities

Data	Local level model	σ_e^2	σ_b^2	LogLik/T	LogLik
Norwegian	The Q function	0.00326821	0.0047030	0.8468622	28.7933
fatalities	Kalman filter	0.00326838	0.0047026	0.8468622	28.7933

Reference to data source: http://www.ssfpack.com/CKbook.html
and Commandeur and Koopman (2007), Appendix B.
Reference to Kalman filter results: Commandeur and Koopman (2007) p:18,19.

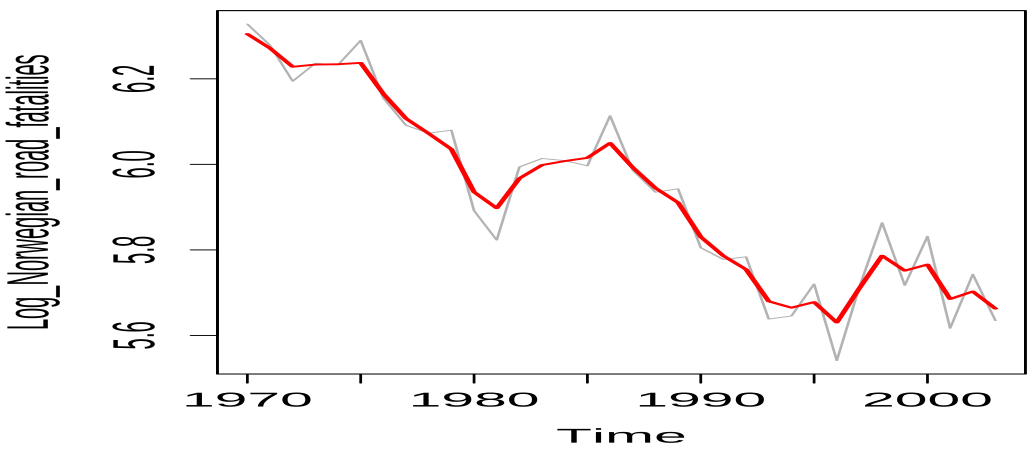


Figure 5.1: Observed log Norwegian road fatalities in gray and the fitted local level in red.

UK drivers killed or seriously injured

The local level model (i.e. a random walk order 1) is applied to the log of the monthly number of drivers killed or seriously injured (KSI) in the UK, from January 1969 to December 1984. The maximum likelihood estimates of the hyperparameters are given in Table 5.2, and the fitted local level is plotted with the data in Figure 5.2. The R fitting commands and output are given in Appendix D.

Table 5.2: The estimated hyperparameters for UK drivers KSI

Data	Local level model	σ_e^2	σ_b^2	LogLik/T	LogLik
UK drivers	The Q function	0.00222155	0.011866	0.6451960	123.8776
KSI	Kalman Filter	0.00222157	0.011866	0.6451960	123.8776

Reference to data source: http://www.ssfpack.com/CKbook.html
and Commandeur and Koopman (2007), Appendix A.
Reference to Kalman filter results: Commandeur and Koopman (2007) p.16.

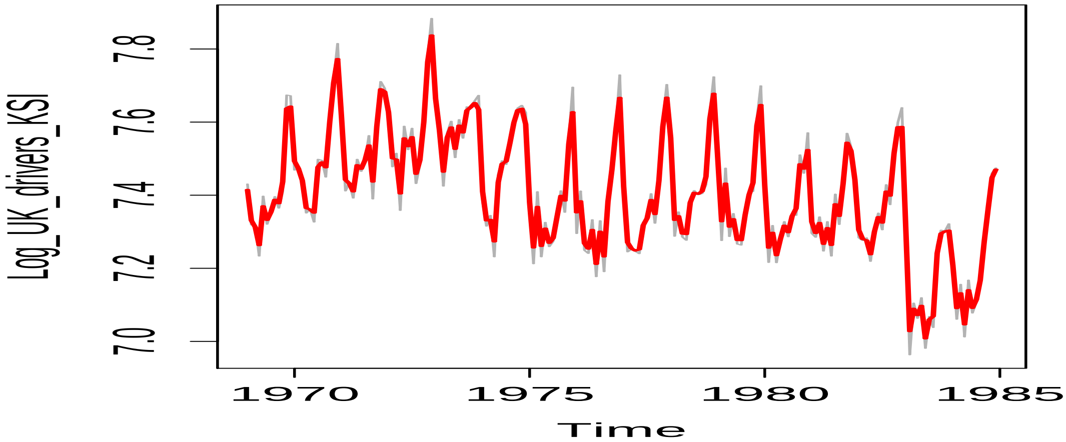


Figure 5.2: Observed log UK drivers KSI in gray and the fitted local level in red.

5.3 Local level and trend model

The local linear trend model is obtained by adding a random slope to the local level model. The random walk process of the local level and trend (or local linear trend) model is defined as

$$\begin{aligned} y_t &= \gamma_t + e_t \\ \gamma_t &= \gamma_{t-1} + \psi_t + b_t \\ \psi_t &= \psi_{t-1} + d_t \end{aligned} \tag{5.3}$$

where ψ_t is a random walk slope and where $e_t \sim N(0, \sigma_e^2)$, $b_t \sim N(0, \sigma_b^2)$, and $d_t \sim N(0, \sigma_d^2)$.

The autoregressive process of the local level and trend model is defined as

$$\begin{aligned} y_t &= \gamma_t + e_t \\ \gamma_t &= \sum_{j=1}^{J} \phi_j \gamma_{t-j} + \psi_t + b_t \\ \psi_t &= \sum_{l=1}^{L} \rho_l \psi_{t-l} + d_t \end{aligned} \tag{5.4}$$

where ψ_t is an autoregressive trend (or slope) and where $e_t \sim N(0, \sigma_e^2)$, $b_t \sim N(0, \sigma_b^2)$ and $d_t \sim N(0, \sigma_d^2)$.

The local linear trend model contains two state equations: one for modelling the level, and one for modelling the slope. The slope is also referred to as the drift in time series.

5.3.1 Examples of local level and trend model

Finnish traffic fatalities

In this example the local level and trend model is applied to the log of the annual number of road traffic fatalities in Finland as observed from 1970 to 2003, allowing both the level and the slope to vary over time. The estimated hyperparameters are given in Table 5.3.

Figure 5.3 shows the log of the annual number of road traffic fatalities in Finland from 1970 to 2003, and the decomposition of the fitted local level and the fitted slope. The stochastic slope is shown separately at the bottom of Figure 5.3. When the slope is positive and increasing, the level of fatalities in Finland was increasing as shown is Figure 5.3, from 1983 to 1987. On the other hand, when the slope is negative and decreasing, the level of fatalities in Finland was decreasing too, as shown is Figure 5.3, from 1971 to 1976, 1989 to 1992, and from 2000 to 2003. Figure 5.4 shows the log of the annual number of road traffic fatalities in Finland from 1970 to 2003, and the fitted local linear trend (local level and trend).

The R fitting commands and output are given in Appendix D.

Table 5.3: The estimated hyperparameters for Finnish fatalities

Data	local linear trend	σ_e^2	σ_b^2	σ_d^2	LogLik/T	LogLik
Finn.	The Q function	0.00320085	1.00024e-08	0.00153302	0.7864746	26.740
fatal.	Kalman filter	0.00320083	9.69606E-26	0.00153314	0.7864746	26.740

Reference to data source: http://www.ssfpack.com/CKbook.html
and Commandeur and Koopman (2007), Appendix B.
Reference to Kalman filter results: Commandeur and Koopman (2007) p.28.

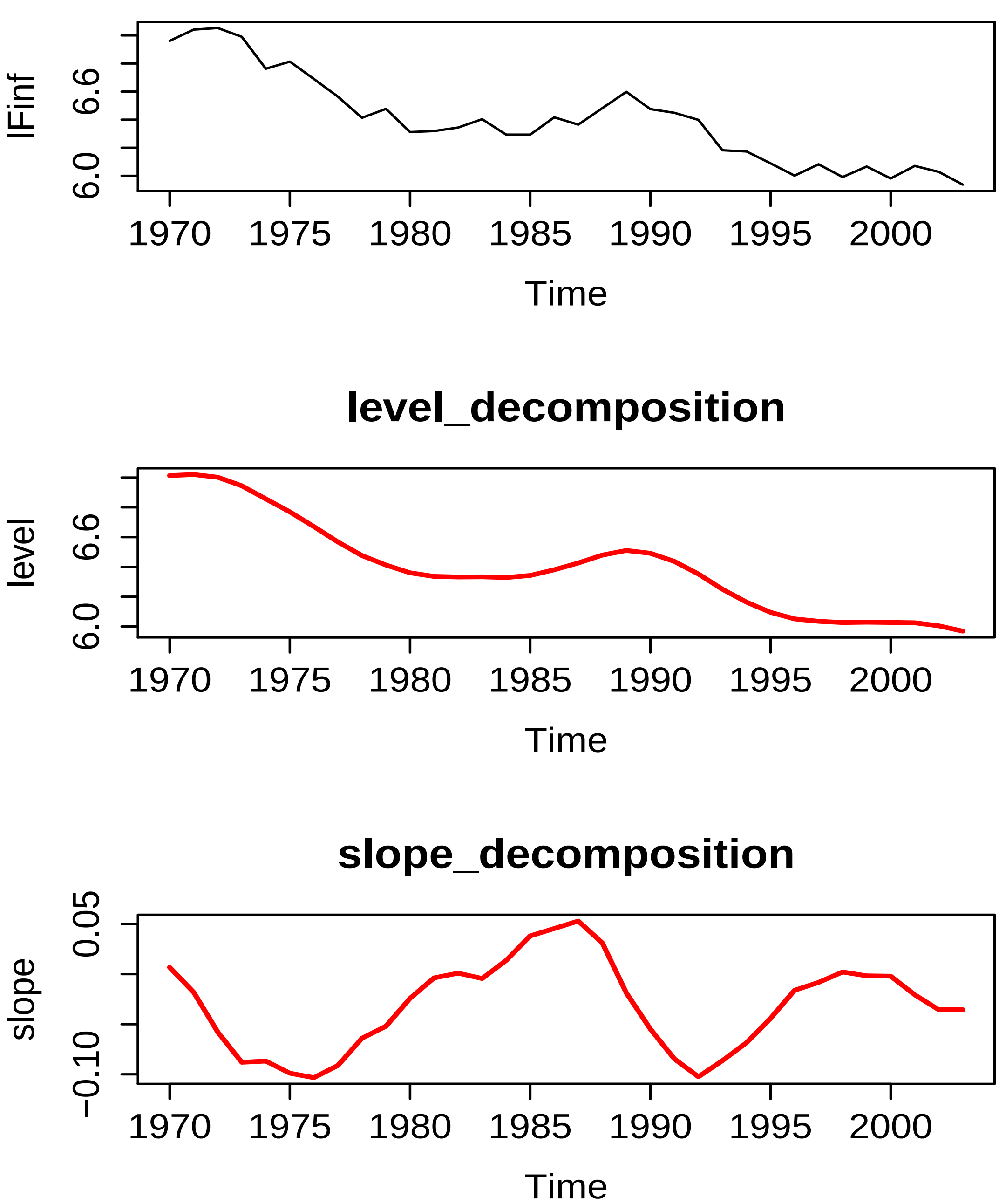


Figure 5.3: Observed log Finnish fatalities and the fitted local level and trend decomposed.

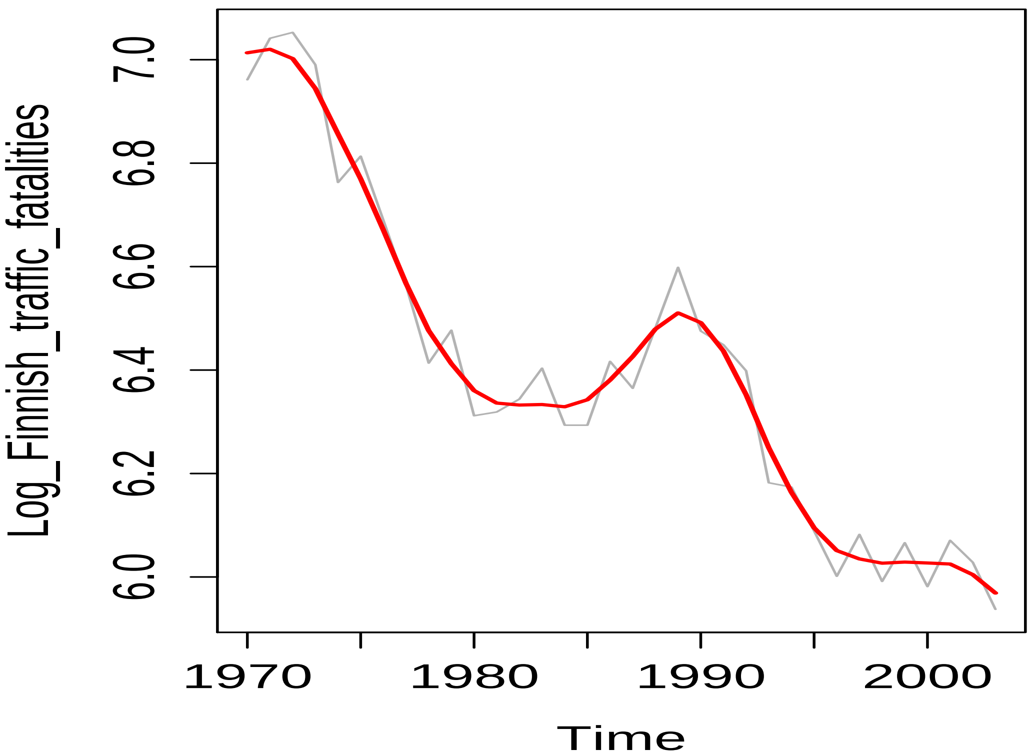


Figure 5.4: Observed Finnish fatalities with the fitted local level and trend.

UK drivers killed or seriously injured

The local level and trend model is applied to the log of the monthly number of drivers killed or seriously injured (KSI) in the UK in the period January 1969 to December 1984. The estimated variances are given in Table 5.4,

Figure 5.5 shows the log of the monthly number of drivers killed or seriously injured (KSI) in the UK, from January 1969 to December 1984, and the decomposition of the fitted local level and the fitted random walk slope.

The fitted slope is constant as shown in Figure 5.5, the variance of the slope is almost zero, hence the fluctuation in the slope is almost negligible. This means the addition of a random slope to the local level model is not effective in analysing the observed log of the monthly number of drivers killed or seriously injured (KSI) in the UK. Hence, the random slope is redundant in this case.

The R fitting commands and output are given in Appendix D.

Table 5.4: The estimated hyperparameters for UK drivers KSI

Data	local linear trend	σ_e^2	σ_b^2	σ_d^2	LogLik/T	LogLik
UK	The Q function	0.0021181	0.012128	1.0E-11	0.6247934	119.9603
KSI	Kalman filter	0.0021181	0.012128	1.5E-11	0.6247935	119.9604

Reference to data source: http://www.ssfpack.com/CKbook.html
and Commandeur and Koopman (2007), Appendix A.
Reference to Kalman filter results: Commandeur and Koopman (2007) p.27.

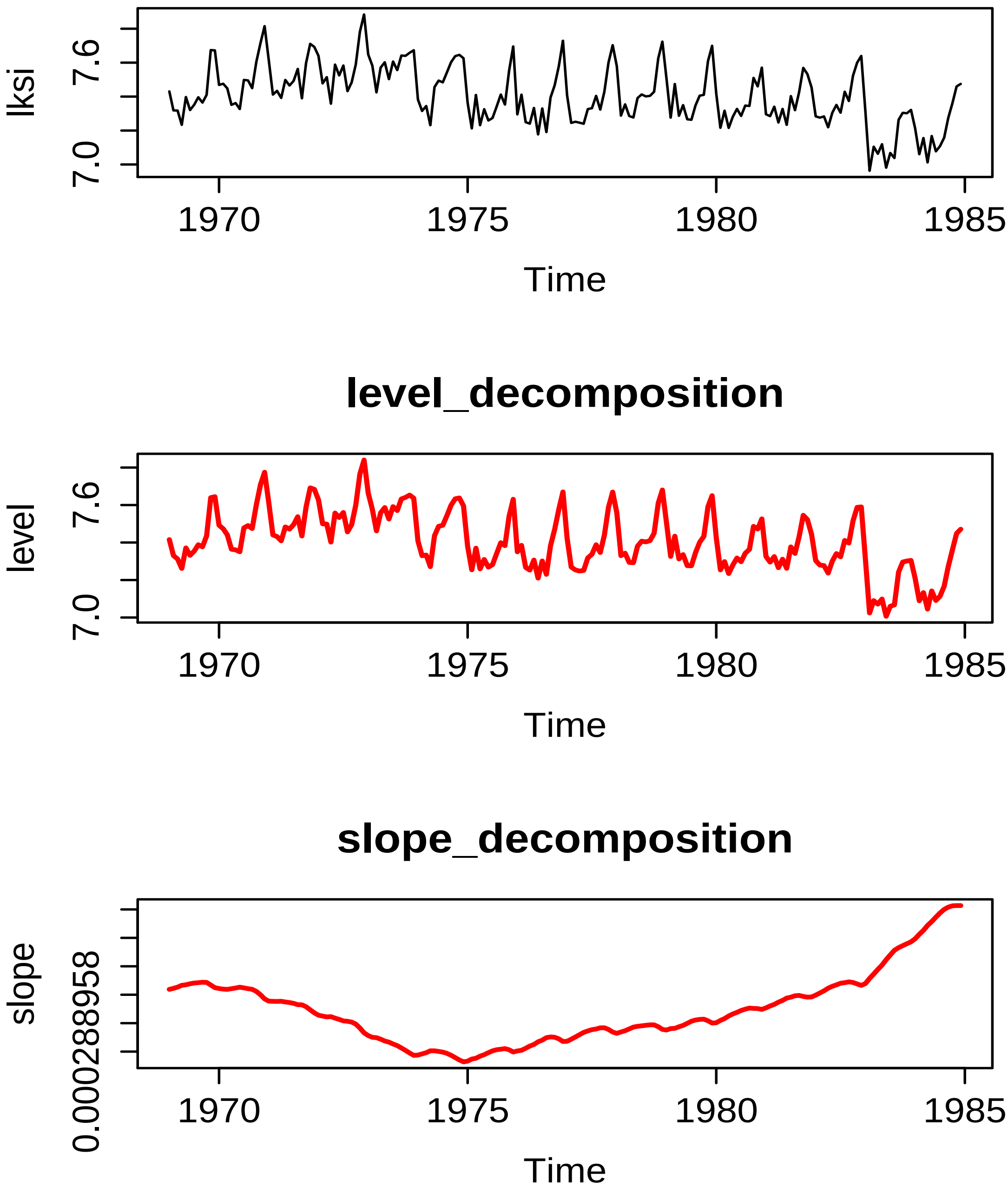


Figure 5.5: Observed log UK drivers KSI with the fitted local level and trend decomposed.

5.4 Local level and seasonal model

The random walk process of the local level and seasonal is defined as

$$
\begin{aligned}
y_t &= \gamma_t + s_t + e_t \\
\gamma_t &= \gamma_{t-1} + b_t \\
s_t &= -\sum_{m=1}^{M-1} s_{t-m} + w_t
\end{aligned}
\tag{5.5}
$$

where $e_t \sim N(0, \sigma_e^2)$, $b_t \sim N(0, \sigma_b^2)$ and $w_t \sim N(0, \sigma_w^2)$.

The random walk process of the local level with trend and seasonal is defined as

$$
\begin{aligned}
y_t &= \gamma_t + s_t + e_t \\
\gamma_t &= \gamma_{t-1} + \psi_t + b_t \\
\psi_t &= \psi_{t-1} + d_t \\
s_t &= -\sum_{m=1}^{M-1} s_{t-m} + w_t
\end{aligned}
\tag{5.6}
$$

where $e_t \sim N(0, \sigma_e^2)$, $b_t \sim N(0, \sigma_b^2)$, $d_t \sim N(0, \sigma_d^2)$ and $w_t \sim N(0, \sigma_w^2)$.

The autoregressive process of the local level and seasonal is defined as

$$
\begin{aligned}
y_t &= \gamma_t + s_t + e_t \\
\gamma_t &= \sum_{j=1}^{J} \phi_j \gamma_{t-1} + b_t \\
s_t &= -\sum_{m=1}^{M-1} s_{t-m} + w_t
\end{aligned}
\tag{5.7}
$$

where $e_t \sim N(0, \sigma_e^2)$, $b_t \sim N(0, \sigma_b^2)$ and $w_t \sim N(0, \sigma_w^2)$.

The autoregressive process of the local level with trend and seasonal is defined as

$$\begin{aligned} y_t &= \gamma_t + s_t + e_t \\ \gamma_t &= \sum_{j=1}^{J} \phi_j \gamma_{t-j} + \psi_t + b_t \\ \psi_t &= \sum_{l=1}^{L} \rho_l \psi_{t-l} + d_t \\ s_t &= -\sum_{m=1}^{M-1} s_{t-m} + w_t \end{aligned} \tag{5.8}$$

where $e_t \sim N(0, \sigma_e^2)$, $b_t \sim N(0, \sigma_b^2)$, $d_t \sim N(0, \sigma_d^2)$ and $w_t \sim N(0, \sigma_w^2)$.

5.4.1 Examples of local level and seasonal model

UK inflation

The local level and seasonal model is applied to UK inflation, as measured on a quarterly basis for the years 1950-2001. The estimated variances are given in Table 5.5, while the stochastic level and stochastic seasonal are displayed separately in the Figure 5.7. The R fitting commands and output are given in Appendix D.

Table 5.5: The estimated hyperparameters for quarterly UK inflation.

Data	local level & seasonal	σ_e^2	σ_b^2	σ_w^2	LogLik/T	LogLik
UK	The Q function	3.3713e-05	2.1241e-05	4.34e-07	3.201381	665.8873
inf.	Kalman filter	3.3717e-05	2.1197e-05	1.09e-07	3.198464	665.2805

Reference to data source: http://www.ssfpack.com/CKbook.html
and Commandeur and Koopman (2007), Appendix D.
Reference to Kalman filter results: Commandeur and Koopman (2007) p.44.

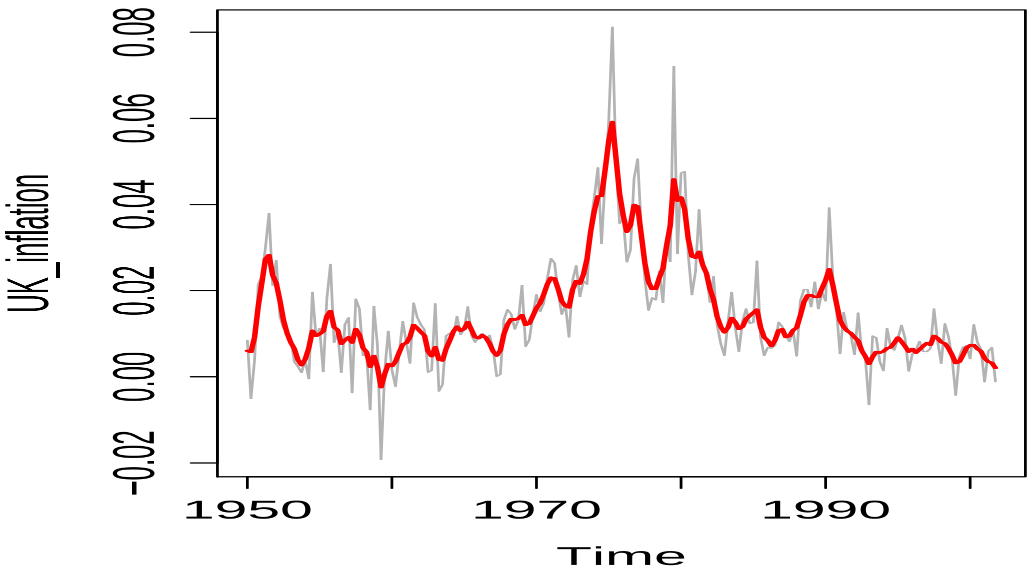


Figure 5.6: The observed UK quarterly inflation with the fitted local level.

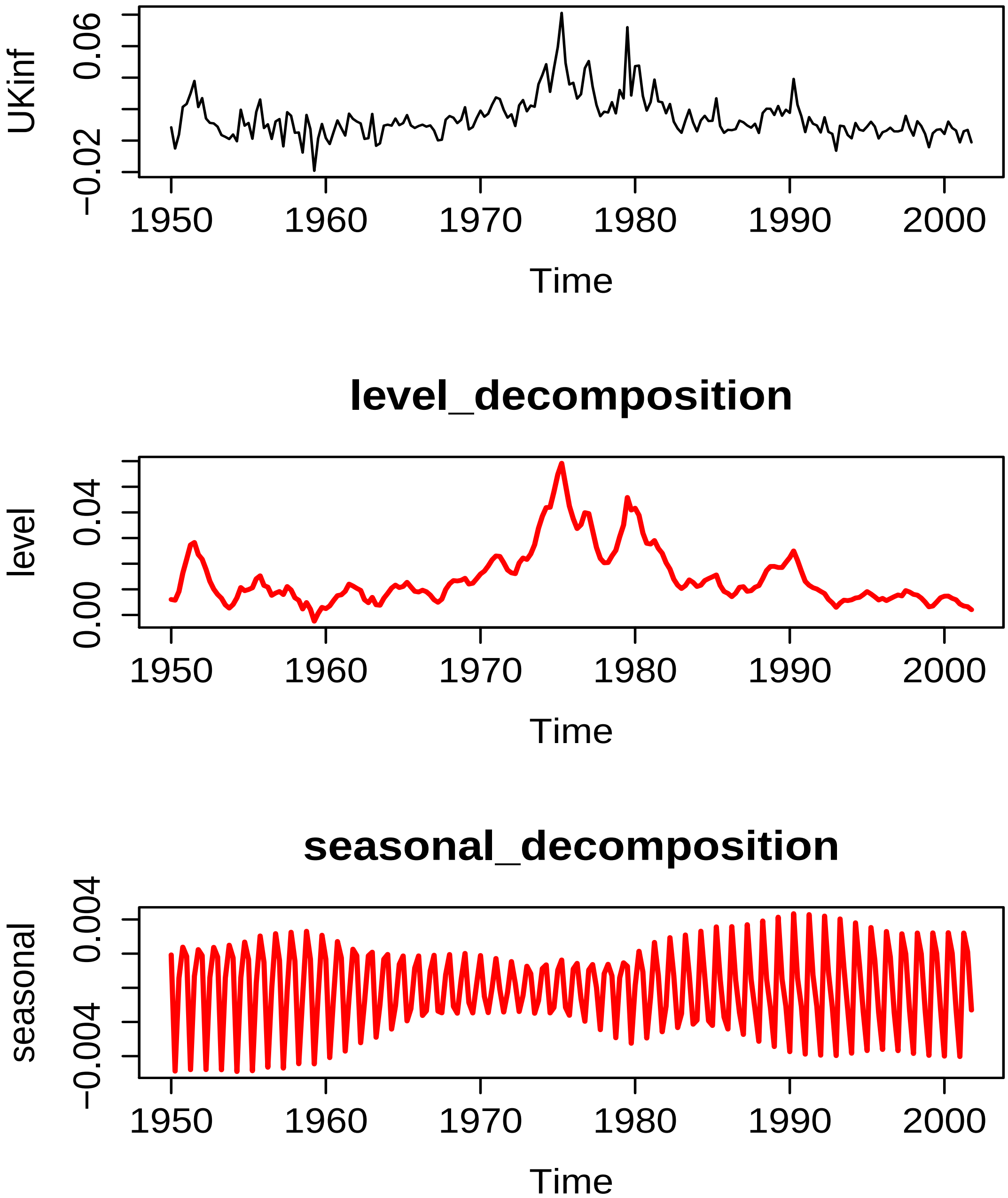


Figure 5.7: Observed UK quarterly inflation with fitted stochastic level and stochastic seasonal.

UK drivers killed or seriously injured

The local level and seasonal is applied to the log of the monthly number of drivers killed or seriously injured (KSI) in the UK in the period January 1969 to December 1984. The estimated variances are given in Table 5.6, while the stochastic level and the stochastic seasonal are displayed separately in the Figure 5.9. The R fitting commands and output are given in Appendix D.

Table 5.6: The estimated hyperparameters for UK derivers KSI

Data	local level & seasonal	σ_e^2	σ_b^2	σ_w^2	LogLik/T	LogLik
UK	The Q function	0.00351382	0.000945617	0.1E-9	0.9829963	188.735
KSI	Kalman filter	0.00341592	0.000935947	0.5E-6	0.9369063	179.886

Reference to data source: http://www.ssfpack.com/CKbook.html
and Commandeur and Koopman (2007), Appendix A.
Reference to Kalman filter results: Commandeur and Koopman (2007) p.38.

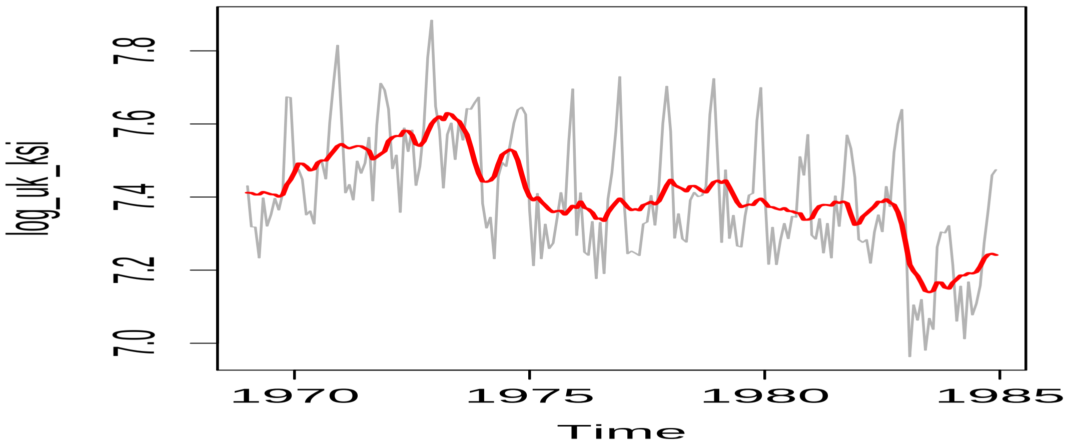


Figure 5.8: Observed UK drivers KSI with fitted local level.

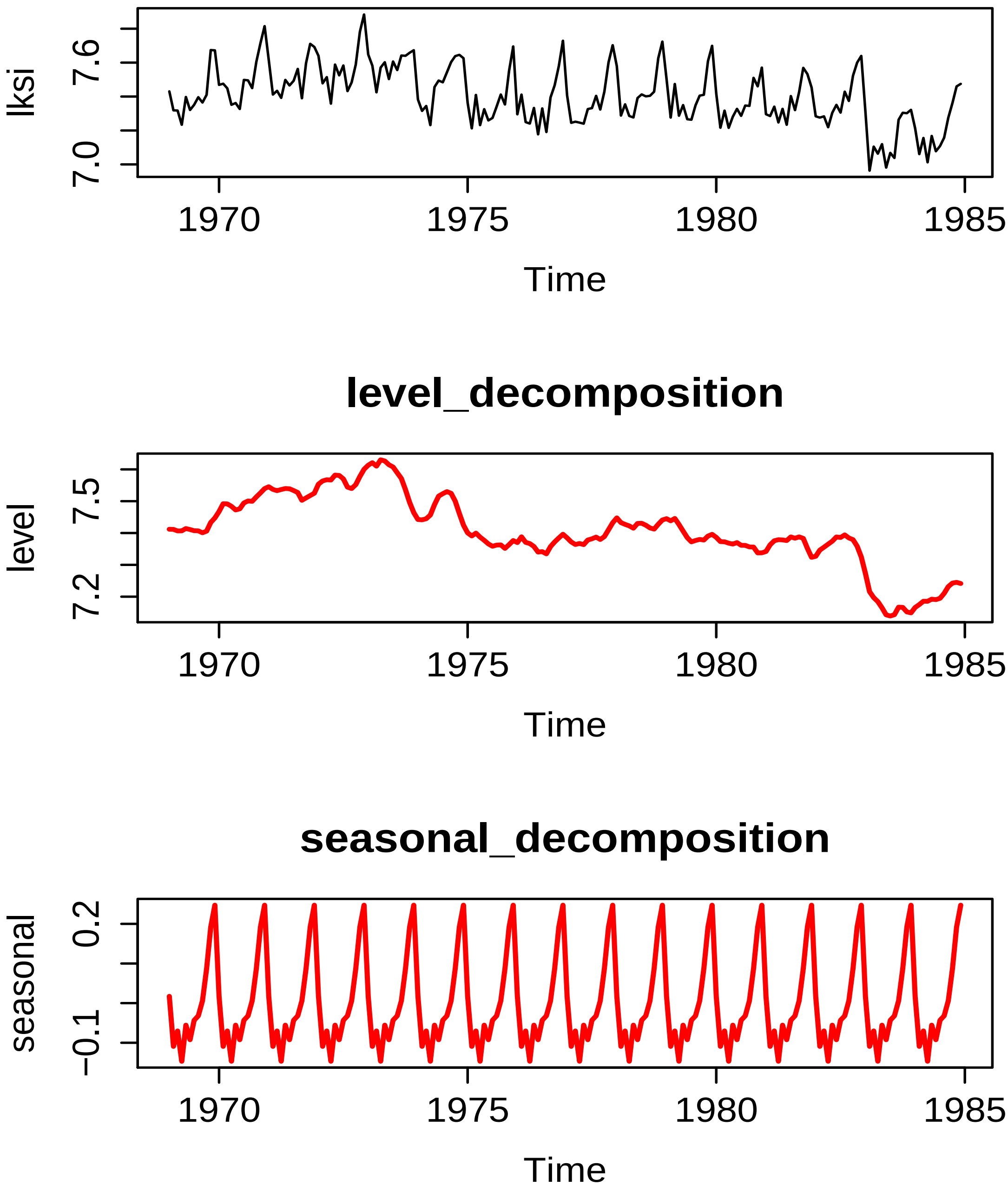


Figure 5.9: Observed log UK drivers KSI with fitted local level and seasonal decomposed.

Johnson & Johnson quarterly earnings

The quarterly earnings series from the U.S company Johnson & Johnson is highly nonstationary as given in Figure 5.10. There is both a trend signal that is gradually increasing over time and a seasonal component that cycles every four quarters or once per year. The autoregressive order 1 local level and seasonal model is fitted to Johnson & Johnson earnings for comparison with the results of Shumway and Stoffer (2011), then to log Johnson & Johnson earnings. The estimated hyperparameters for the ar local level and seasonal model are given in Table 5.7. On the normal scale, the estimated growth was $\phi_1 = 1.035$, corresponding to exponential growth with inflation at about 3.5% per year. The fitted values are displayed separately in Figure 5.10. On the log scale, the estimated growth was $\phi_1 = 1.019$, corresponding to 1.9% growth with inflation per year. The estimated hyperparameters are given in Table 5.7, while the fitted values are displayed separately in Figure 5.11. The fitted values of seasonality is more illustrative and realistic on the log scale than on the normal scale. The R fitting commands and output are given in Appendix D.

Table 5.7: The estimated hyperparameters for Johnson & Johnson quarterly earnings

Data	local level & seasonal	σ_e^2	σ_b^2	σ_w^2	ϕ_1
q.jnj	The Q function	3.8e-06	0.01963217	0.05032152	1.035
	Kalman filter	2.5e-07	0.01951609	0.04879681	1.035
log(q.jnj)	The Q function	2.1e-09	0.00358159	0.00099529	1.019

Reference to data source: the data are available in R as `q.jnj` in package `FinTS`.
Reference to Kalman filter results: Shumway and Stoffer (2011), p.351,352.

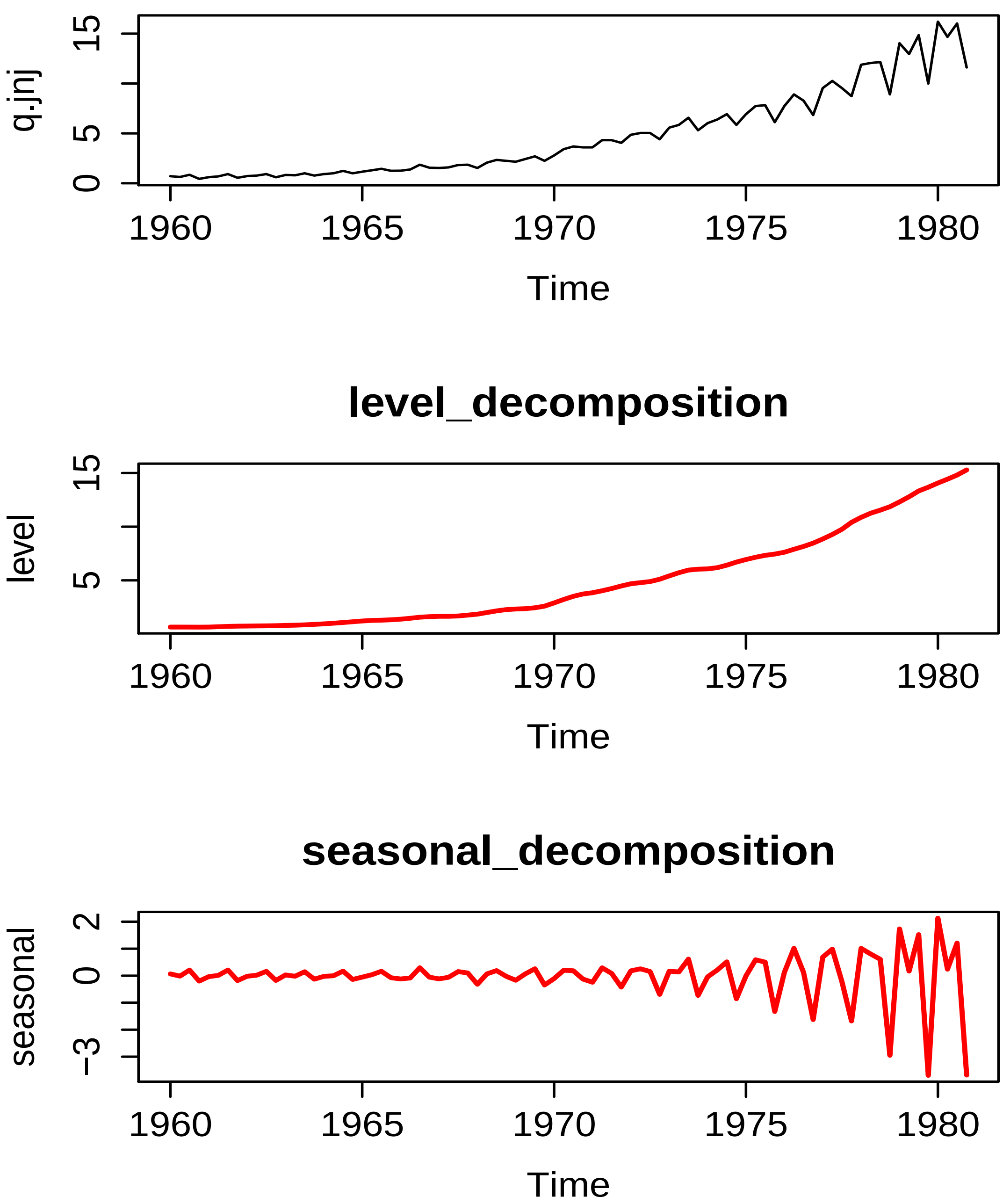


Figure 5.10: Johnson & Johnson quarterly earnings with the fitted AR(1) local level and seasonal.

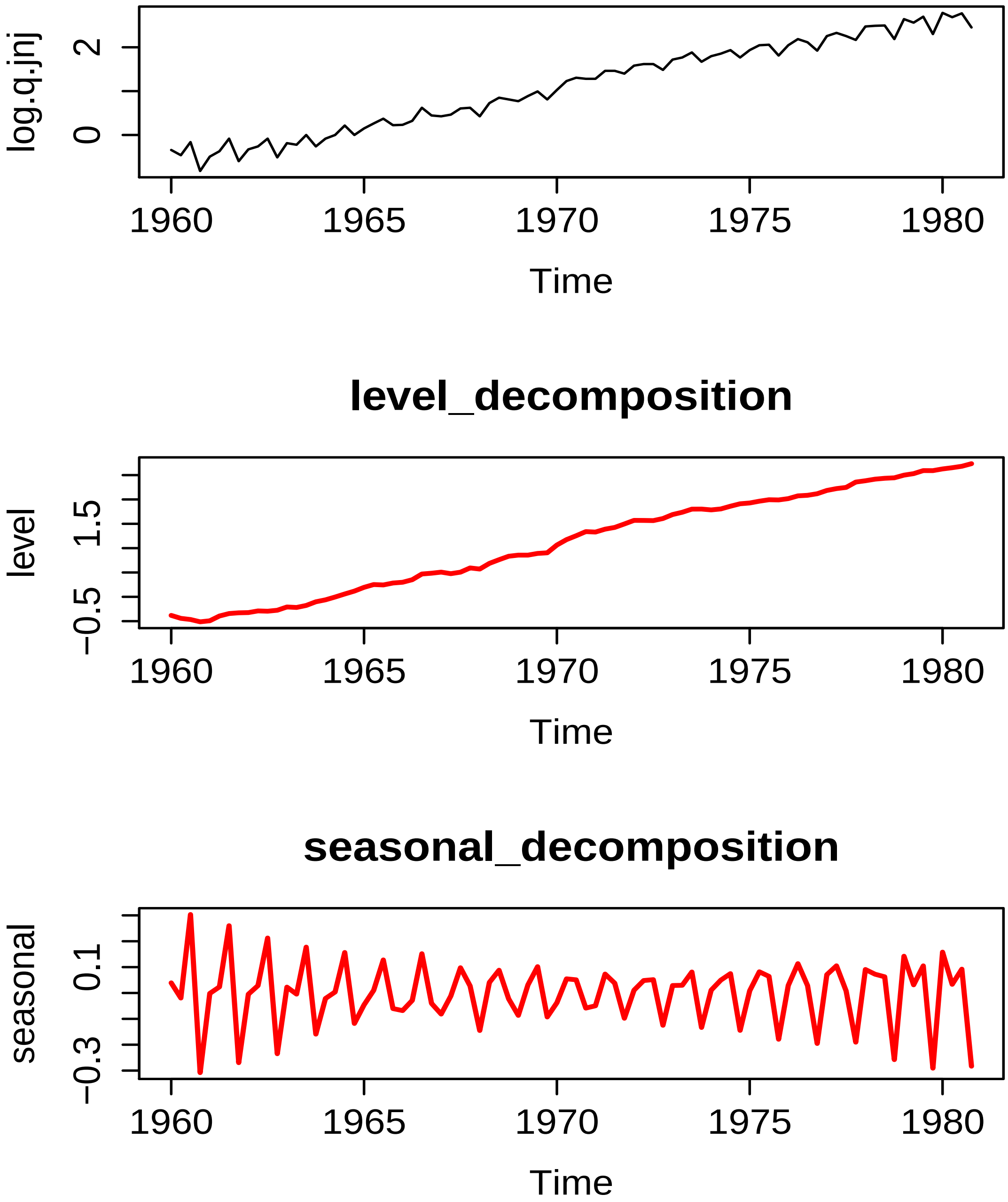


Figure 5.11: log of Johnson & Johnson quarterly earnings with the fitted AR(1) local level and seasonal.

5.5 Local level with random coefficient of an explanatory variable model

The local level with a random coefficient of an explanatory variable is defined as

$$\begin{aligned} y_t &= \gamma_t + \beta_t x_t + e_t \\ \gamma_t &= \sum_{j=1}^{J} \phi_j \gamma_{t-j} + b_t \\ \beta_t &= \beta_{t-1} + v_t \end{aligned} \tag{5.9}$$

where $e_t \sim NO(0, \sigma_e^2)$, $b_t \sim NO(0, \sigma_b^2)$, and $v_t \sim NO(0, \sigma_v^2)$, where γ_t is the autoregressive local level and x_t is the explanatory variable.

5.5.1 Example of local level with explanatory variable model

UK drivers killed or seriously injured

Random walk local level with a random coefficient of an explanatory variable is applied to the log of the monthly number of drivers killed or seriously injured (KSI) in the UK, from January 1969 to December 1984. The explanatory variable is the log of the monthly petrol prices in the UK from January 1969 to December 1984. The estimated variances are given in Table 5.8, while the fitted values for the local level with a random coefficient of an explanatory variable are displayed separately in the Figure 5.12. The R fitting commands and output are given in Appendix D.

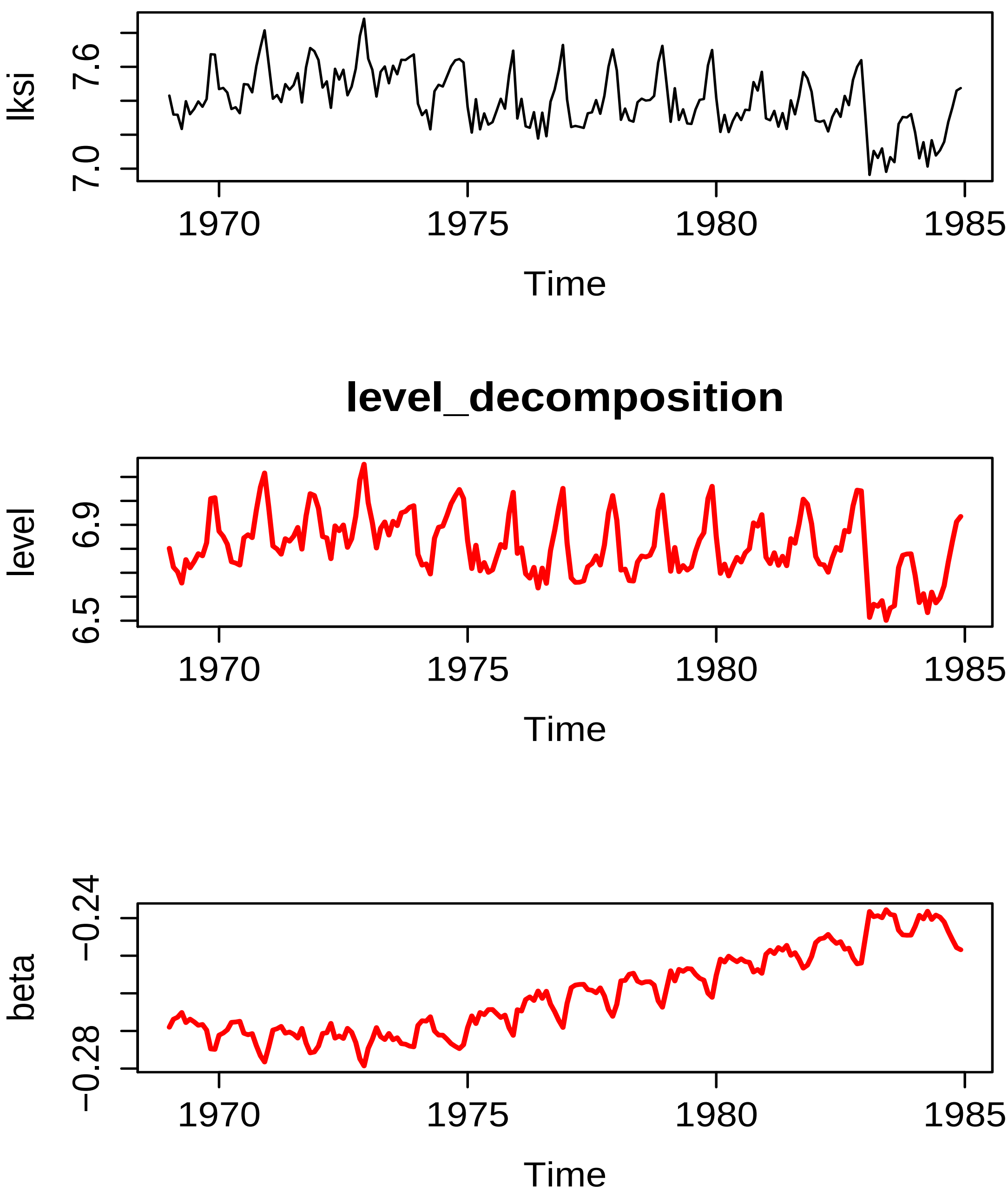


Figure 5.12: Observed log UK drivers KSI with fitted random walk local level with a random coefficient of monthly UK log petrol prices.

Table 5.8: The estimated hyperparameters for UK derivers KSI

Data	level & r. coef.	σ_e^2	σ_b^2	σ_v^2	$\hat{\beta}_1$	LogLik/T	LogLik
UK	Q function	0.00235670	0.0109746	1.302e-4	-0.26900	0.655216	125.801
KSI	Kalman filter	0.00234791	0.0116673	*	-0.26105	0.645636	123.962

Reference to data source: http://www.ssfpack.com/CKbook.html
and Commandeur and Koopman (2007), Appendix A.
Reference to Kalman filter results: Commandeur and Koopman (2007) p.52.
(*) not available in Commandeur and Koopman (2007).

5.6 Maximum likelihood estimation

Here the most general model, the autoregressive process of the local level with trend and seasonal (6.3), is fitted with the following maximum likelihood estimation.

$$y_t = \gamma_t + s_t + e_t \tag{5.10}$$

where γ_t, ψ_t and s_t are given by (6.3) assumptions. [The other structural models given earlier are all special cases of this model.]

Hence,

$$\begin{aligned}
e_t &= y_t - \gamma_t - s_t \\
b_t &= \gamma_t - \sum_{j=1}^{J} \phi_j \gamma_{t-j} - \psi_{t-1} \\
d_t &= \psi_t - \sum_{l=1}^{L} \rho_l \psi_{t-l} \\
w_t &= \sum_{m=1}^{M} s_{t-m+1}
\end{aligned} \tag{5.11}$$

where $\mathbf{e} \sim N_T(0, \sigma_e^2 \mathbf{W}^{-1})$, $\mathbf{b} \sim N_{T-J}(0, \sigma_b^2 \mathbf{I}_{T-J})$, $\mathbf{d} \sim N_{T-L}(0, \sigma_d^2 \mathbf{I}_{T-L})$, $\boldsymbol{\omega} \sim N_{T-M+1}(0, \sigma_w^2 \mathbf{I}_{T-M+1})$. Note $\mathbf{e} = (e_1, e_2, \ldots, e_T)^\top$, $\mathbf{b} = (b_{J+1}, b_{J+2}, \ldots, b_T)^\top$, $\mathbf{d} = (d_{L+1}, d_{L+2}, \ldots, d_T)^\top$, $\boldsymbol{\omega} = (w_M, w_{M+1}, \ldots, w_T)^\top$.

Algorithm for estimating $(\sigma_e^2, \sigma_b^2, \sigma_d^2, \sigma_w^2, \boldsymbol{\phi}, \boldsymbol{\rho})$

1. Given starting values for $(\sigma_e^2, \sigma_b^2, \sigma_d^2, \sigma_w^2, \boldsymbol{\phi}, \boldsymbol{\rho})$, estimate initial values for $(\boldsymbol{\gamma}, \boldsymbol{\psi}, s)$.

2. Having estimated initial values for $(\boldsymbol{\gamma}, \boldsymbol{\psi}, s)$, estimate initial values for $(\sigma_e^2, \sigma_b^2, \sigma_d^2, \sigma_w^2)$.

3. Maximize Q over $(\sigma_e^2, \sigma_b^2, \sigma_d^2, \sigma_w^2, \boldsymbol{\phi}, \boldsymbol{\rho})$ using a numerical algorithm, where $(\boldsymbol{\gamma}, \boldsymbol{\psi}, s)$ given $(\sigma_e^2, \sigma_b^2, \sigma_d^2, \sigma_w^2, \boldsymbol{\phi}, \boldsymbol{\rho})$ is obtained before calculating Q in the function evaluating Q.

4. Having the maximum values for $(\sigma_e^2, \sigma_b^2, \sigma_d^2, \sigma_w^2, \boldsymbol{\phi}, \boldsymbol{\rho})$ estimate the maximum values for $(\boldsymbol{\gamma}, \boldsymbol{\psi}, s)$.

Let Q be given by

$$\begin{aligned}
Q &= \log f(\mathbf{y}|\boldsymbol{\gamma}, \mathbf{s}) + \log f(\boldsymbol{\gamma}, \boldsymbol{\psi}, \mathbf{s}) - \frac{1}{2} \log \left| \mathbf{A} + \mathbf{D}^\top \mathbf{M}^{-1} \mathbf{D} \right| + \frac{3T}{2} \log 2\pi \\
\log f(\mathbf{y}|\boldsymbol{\gamma}, \mathbf{s}) &= -\frac{1}{2} \log \left| 2\pi \sigma_e^2 \mathbf{W}^{-1} \right| - \frac{1}{2} (\mathbf{y} - \boldsymbol{\gamma} - \mathbf{s})^\top \sigma_e^{-2} \mathbf{W} (\mathbf{y} - \boldsymbol{\gamma} - \mathbf{s}) \\
\log f(\boldsymbol{\gamma}, \boldsymbol{\psi}, \mathbf{s}) &= -\frac{1}{2} \log |2\pi \mathbf{M}| - \frac{1}{2} \left(\boldsymbol{\gamma}^\top \boldsymbol{\psi}^\top \mathbf{s}^\top \right) \mathbf{D}^\top \mathbf{M}^{-1} \mathbf{D} \left(\boldsymbol{\gamma} \; \boldsymbol{\psi} \; \mathbf{s} \right)^\top
\end{aligned}$$

where $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_T)^\top$, $\boldsymbol{\psi} = (\psi_1, \psi_2, \ldots, \psi_T)^\top$, $\mathbf{s} = (s_1, s_2, \ldots, s_T)^\top$, $\mathbf{M} =$ matrix diagonal $(\sigma_b^2 \mathbf{I}_{T-J}, \sigma_v^2 \mathbf{I}_{T-L}, \sigma_w^2 \mathbf{I}_{T-M+1})$

$$\mathbf{D} = \begin{pmatrix} D_\gamma & D_{\gamma\psi} & 0 \\ 0 & D_\psi & 0 \\ 0 & 0 & D_s \end{pmatrix}$$

$$\mathbf{D}_\gamma = \begin{pmatrix} -\phi_J & -\phi_{J-1} & \dots & \dots & -\phi_1 & 1 & 0 & 0 & \dots & \dots & 0 & 0 \\ 0 & -\phi_J & -\phi_{J-1} & \dots & \dots & -\phi_1 & 1 & 0 & \dots & \dots & 0 & 0 \\ \\ \\ 0 & 0 & 0 & \dots & \dots & 0 & -\phi_J & -\phi_{J-1} & \dots & \dots & -\phi_1 & 1 \end{pmatrix}$$

$$D_\psi = \begin{pmatrix} -\rho_L & -\rho_{L-1} & \dots & \dots & -\rho_1 & 1 & 0 & 0 & \dots & \dots & 0 & 0 \\ 0 & -\rho_L & -\rho_{L-1} & \dots & \dots & -\rho_1 & 1 & 0 & \dots & \dots & 0 & 0 \\ \\ \\ 0 & 0 & 0 & \dots & \dots & 0 & -\rho_L & -\rho_{L-1} & \dots & \dots & -\rho_1 & 1 \end{pmatrix}$$

$$D_s = \begin{pmatrix} 1 & 1 & \dots & \dots & \dots & \dots & 1 & 0 & 0 & \dots & \dots & 0 & 0 \\ 0 & 1 & 1 & \dots & \dots & \dots & \dots & 1 & 0 & \dots & \dots & 0 & 0 \\ \\ \\ 0 & 0 & 0 & \dots & \dots & 0 & 1 & 1 & \dots & \dots & \dots & \dots & 1 \end{pmatrix}$$

$$D_{\gamma\psi} = (-\mathbf{I}_{T-1}\ \mathbf{0})$$

with the first $(J-1)$ rows removed from $D_{\gamma\psi}$

$$\mathbf{A} = \begin{pmatrix} \Sigma^{-1} & 0 & \Sigma^{-1} \\ 0 & 0 & 0 \\ \Sigma^{-1} & 0 & \Sigma^{-1} \end{pmatrix}$$

Note that $\mathbf{D}_\gamma$ is on $(T-J)$ x T matrix, $\mathbf{D}_\psi$ is $(T-L)$ x T, $\mathbf{D}_{\gamma\psi}$ is $(T-J)$ x T, $\mathbf{D}_s$ is $(T-M+1)$ x T, Σ^{-1} is $(T$ x $T)$ and M is $(3T$ x $3T)$ matrix.

Chapter 6

R functions for simulating and fitting Gaussian structural time series models

This Chapter introduces new functions for simulating and fitting Gaussian structural time series models in R. The reasons for developing these functions is to compare between the simulated mean, μ_t, and the fitted mean, $\hat{\mu}_t$, for Gaussian structural time series, and to test whether the estimates of the hyperparameters of the fitted model agrees with the true hyperparameters of the simulation. These fitting functions have been used for fitting real data from Commandeur and Koopman (2007) in Chapter 5, and in this chapter they are used for fitting simulated data.

Gaussian structural time series models are implemented on CRAN, the Comprehensive R Archive Network, statistical software through the packages: `dse` (Gilbert, 2009), `dlm` (Petris *et al.*, 2009), and `KFAS` (Helske, 2010), whereas exponential family state space models are implemented in `sspir` package (Dethlefsen and Lundbye-Christensen, 2006). A comparative review of the tools available in R for state space analysis is available in Tusell (2011).

In addition, `StructTS()` (Riply, 2002) is an R function included in the base package `stats` for fitting Gaussian structural time series models. It has three options, `level`, `trend`, and `BSM` for fitting local level, local level and trend, local level and seasonality (basic structural) models respectively. Petris, Petrone and Campagnoli (2009) provide good analyses, concepts and techniques of modeling and forecasting with dynamic linear models in a Baysian approach using the R package `dlm` for their practical implementation. To the best of the author's knowledge all these packages apply for model fitting only not for a simulation purpose.

Sections 6.1 and 6.2 provide new simulation functions for simulating Gaussian structural time series models. Sections 6.3 and 6.4 provide fitting functions for fitting Gaussian structural time series models. The reason for developing the simulation functions, is for testing the fitting functions which have been used for smoothing Gaussian structural time series models in Chapter 5. The simulation and fitting functions are implemented in R and the commands are given in Appendix D.

The simulation generates random values for variables in a specified probability distribution by using the following two steps:

- Random number generation: generate a sequence of uniform random numbers in [0,1]

- Random variate generation: transform a uniform random sequence to produce a sequence with the desired probability distribution.

The default random number generator in R is the Mersenne-Twister, (Matsumoto and Nishimura, 1998), which is a twisted generalised feedback shift register with period $2^{19937}-1$ and equidistribution in 623 consecutive dimensions. The algorithm is based on Mersenne primes, which take their names from the mathematician Marin Mersenne who studied the prime numbers in the early 17th century.

6.1 Local level simulation functions

6.1.1 Random walk local level order 1

In random walk local level model, the observations are assumed to be dependent on an unobserved state vector that is generated by a random walk process, and on Gaussian measurement error that is independent of the state vector. In local level models, the state vector is the mean of the observations.

Let

$$\begin{aligned} y_t &= \gamma_t + e_t \\ \gamma_t &= \gamma_{t-1} + b_t \end{aligned}$$

i.e.

$$\Delta(\gamma_t) = \gamma_t - \gamma_{t-1} = b_t,$$

where e_t and b_t are two independent Gaussian white noise series, $e_t \sim N(0, \sigma_e^2)$, $b_t \sim N(0, \sigma_b^2)$, for $t = 1, 2, \ldots, T$. This model can be simulated by the following function,

`mrwAll.sim(N=4000, mu=0, sige=3, sigb=1, order=1, plot=TRUE).`

This produces a simulation of a random walk local level model with different orders for the mean, i.e. random walk order 1, random walk order 2, random walk order 3, etc. It has six arguments which takes different inputs for different simulations, `N`: the number of observations, `mu`: is the starting value of the mean, i.e. γ_1, `sige`: is the true value of the standard deviation of the observations measurement error, `sigb`: is the true value of the standard deviation of the state vector innovations, `order`: is the number of differences in the random walk, and `plot=TRUE`: is for plotting both the observations and the unobserved random walk state vector (mean). Note that by increasing the orders in the random walk, the mean becomes more smooth, also note that the simulations uses standard deviations of the innovations and not variances.

Here a sequence of 4000 observations were simulated using the function `mrwAll.sim` with $\gamma_1 = 0$ and hyperparameters $\sigma_e = 3$, $\sigma_b = 1$ and order 1 (`order = 1`). The simulated observations and random walk mean are plotted in Figure 1.1. The simulated data is a gray color and the mean in a dark color.

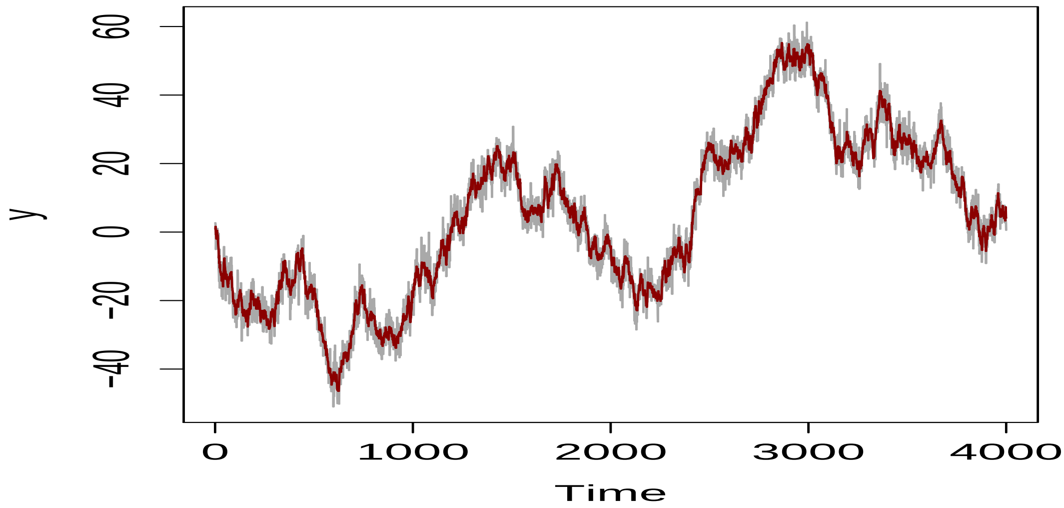


Figure 6.1: Simulation of a random walk local level with order 1, $\sigma_e = 3$ and $\sigma_b = 1$.

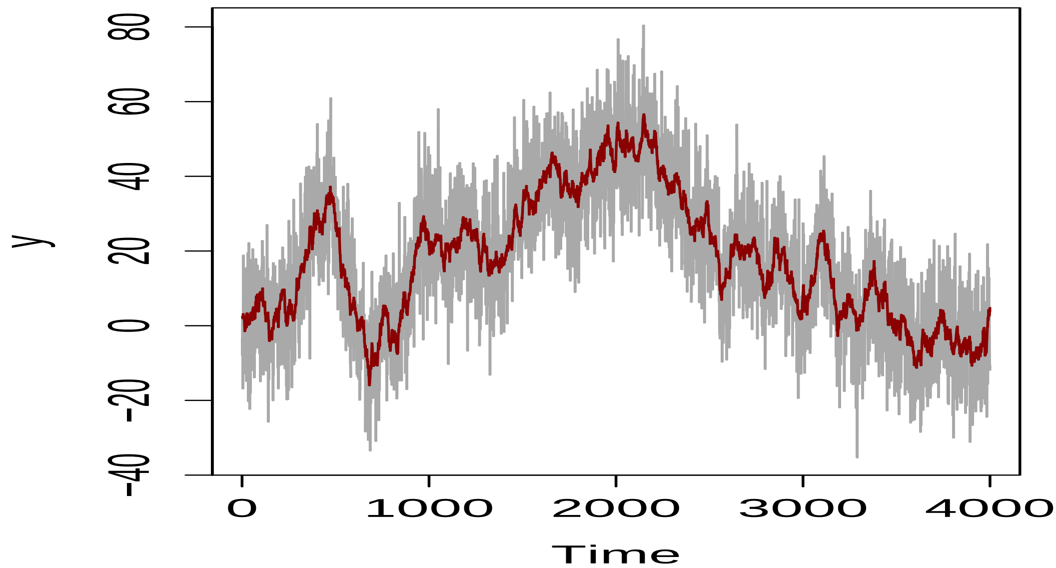


Figure 6.2: Simulation of a random walk local level with order 1, $\sigma_e = 10$ and $\sigma_b = 1$.

6.1.2 Random walk local level order 2

Here we have an example for a random walk local level model with order 2, where the simulated function is exactly the same with a different input (`order=2`),

```
mrwAll.sim(N=4000, mu=0, sige=1, sigb=.0001, order=2, plot=T).
```

This produces a random walk local level simulation of order 2. If the order is increased the signal to noise ratio, $\lambda = \sigma_b^2/\sigma_e^2$, should be small or the variance of the mean vector should be small.

The simulation model is given by

$$
\begin{aligned}
y_t &= \gamma_t + e_t \\
\gamma_t &= 2\gamma_{t-1} - \gamma_{t-2} + b_t \\
\Delta^2(\gamma_t) &= \Delta(\Delta\gamma_t) = b_t,
\end{aligned}
$$

where Δ is the difference operator.

Figure 6.3 gives an example simulation using the above command. A random walk order 2 is smoother then random walk order 1, and becomes much smoother as the order is increased.

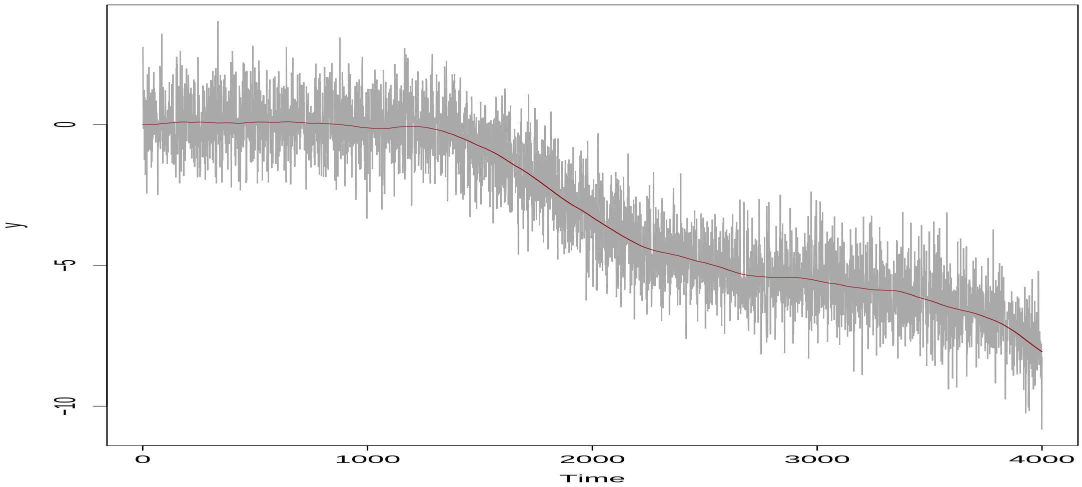


Figure 6.3: Simulation of a random walk local level with order 2, $\sigma_e = 1$ and $\sigma_b = .0001$.

6.1.3 Random walk local level order d

The simulation model is given by

$$\begin{aligned} y_t &= \gamma_t + e_t \\ \Delta^d(\gamma_t) &= b_t, \end{aligned}$$

where d is the order of the random walk, which can take any positive value. This is simulated by specifying the order in the `mrwAll.sim()` function.

6.1.4 Random walk local level and trend

Random walk local level and trend also referred to as local linear trend model is obtained by adding a random slope or a drift to local level model. Local linear trend contains two state equations, one equation for modelling the mean level, and the other equation for modelling the slope. In regression analysis the slope is fixed, whereas in state space models the slope is stochastic, allowed to change over time.

The model is given by

$$\begin{aligned} y_t &= \gamma_t + e_t \\ \gamma_t &= \gamma_{t-1} + \psi_t + b_t \\ \psi_t &= \psi_{t-1} + d_t \end{aligned}$$

where $e_t \sim NO(0, \sigma_e^2), b_t \sim NO(0, \sigma_b^2), d_t \sim NO(0, \sigma_d^2)$. The following function is used to simulate the local linear trend model,

```
mrwrd.sim(N=4000, mu=0, d=1, sige=50, sigb=.01, sigd=.09, plot=T).
```

The function `mrwrd.sim()` produces a simulation of a local level and trend model.

It has the same arguments as `mrwAll.sim()` in addition to two input arguments, `d`: for the starting value of the drift, i.e. ψ_1, and `sigd`: for the true value of the standard deviation of the drift.

In this example a sequence of 4000 observations were simulated using the function `mrwrd.sim()` with initial value for the mean equal to 0, i.e. $\gamma_1 = 0$, and initial value for the drift is equal to 1, i.e. $\psi_1 = 1$, and initial values for the hyperparameters $\sigma_e = 50$, $\sigma_b = .01$, $\sigma_d = .09$. The simulation has two plots, one plot for the observations with local mean and slope altogether, and a second plot for the slope only, as shown in Figure 6.4.

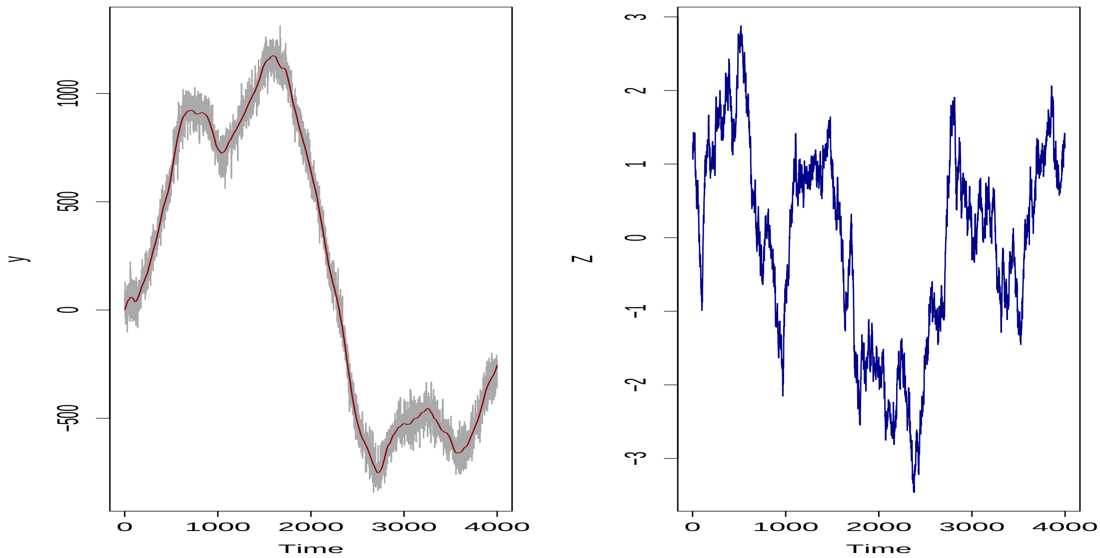


Figure 6.4: Simulation of random walk local level and trend.

6.1.5 Autoregressive local level

In autoregressive local level model the observations are driven by an unobserved autoregressive state vector instead of a random walk vector. This model has a big advantage in extracting a stationary signal for the mean from stationary observations.

The model is given by

$$\begin{aligned} y_t &= \gamma_t + e_t \\ \gamma_t &= \sum_{j=1}^{J} \phi_j \gamma_{t-j} + b_t \end{aligned}$$

where $e_t \sim NO(0, \sigma_e^2)$ and $b_t \sim NO(0, \sigma_b^2)$.

The following functions simulate the autoregressive local level models of order 1 and 2 respectively,

```
mar.sim(N=1000, mu=0, sige=2, sigb=.5, phi=c(.5), plot=T)
mar.sim(N=1000, mu=0, sige=2, sigb=.5, phi=c(.5,.4), plot=T).
```

This produces two autoregressive local level simulations for a number of observations. The function `mar.sim()` has the same arguments as `mrw.sim()` with an extra argument for the ar parameter (`phi`), for the mean which can take more then one value depending on the order of the autoregressive model.

In the above examples two series of 1000 observations were simulated using the function `mar.sim()` with $\gamma_1 = 0$ and hyperparameters $\sigma_e = 2$, $\sigma_b = .5$, $\phi_1 = .5$ for an autoregressive local level with order 1, and $\gamma_1 = 0$ and $\sigma_e = 2$, $\sigma_b = .5$, $\phi_1 = .5$, $\phi_2 = .4$ for an autoregressive local level with order 2.

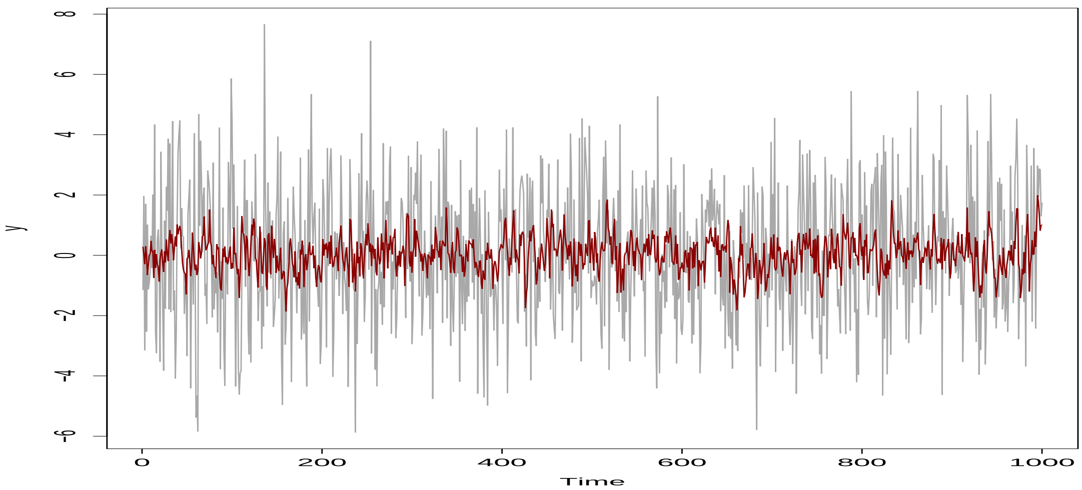


Figure 6.5: Simulation of an AR(1) model, $\phi_1 = .5$

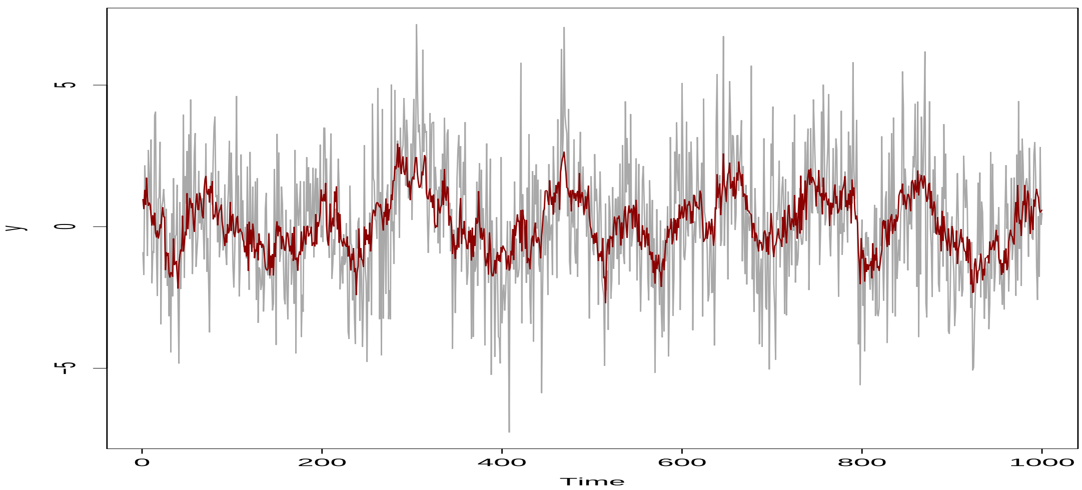


Figure 6.6: Simulation of an AR(2) model, $\phi_1 = .5, \phi_2 = .4$

6.2 Seasonality simulation functions

Seasonality models applies when there is a recurring pattern in time series data, either daily, monthly or quarterly patterns. This regular pattern is refereed to as a seasonal effect. In structural time series models, the seasonal effect can be modelled or extracted by adding a seasonal state vector to local level model.

6.2.1 Seasonality

The seasonality model is defined as

$$\begin{aligned} y_t &= s_t + e_t \\ s_t &= -\sum_{m=1}^{M-1} s_{t-m} + w_t \end{aligned}$$

where $e_t \sim NO(0, \sigma_e^2)$, $w_t \sim NO(0, \sigma_w^2)$, and s_t is the seasonal effect.

The following function simulates the seasonality model,

`mseas.sim(N=240, mu=0, sige=.1, sigw=.1, init=NULL, sigI=1, frequency=12, plot=T).`

This function produces a simulation of the seasonality model. The inputs are the initial values for the seasonal effects, i.e. $s_1, s_2, \ldots, s_{M-1}$ were, by default (since `init=NULL`), randomly generated from a normal distribution with mean 0 and standard deviation 1 (since `sigI=1`). `sige`: is the true value of the standard deviation of the observations measurement error, `sigw`: is the true value of the standard deviation of the seasonal vector innovations. `Frequency=4` for a quarterly data. For monthly data, the input is `Frequency=12`.

In example above a sequence of 240 monthly observations were simulated using the function `mseas.sim()` with two hyperparameters $\sigma_e = .1$, and $\sigma_w = .1$. The

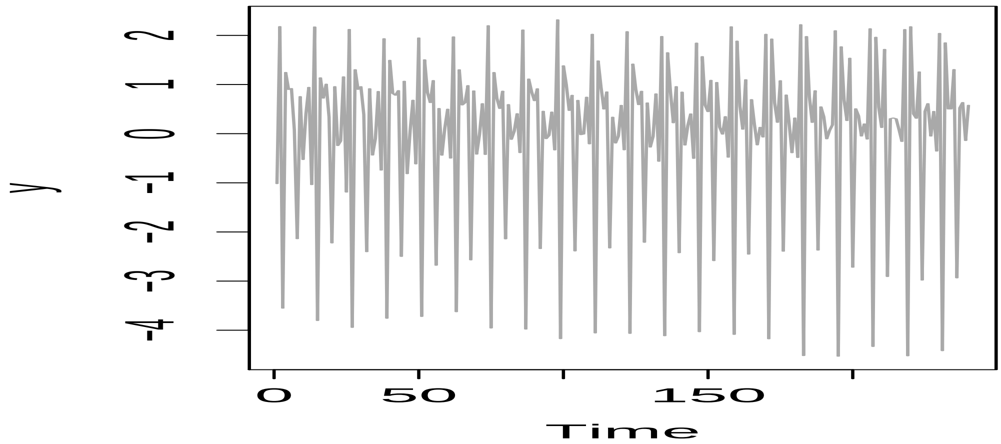


Figure 6.7: Simulation of seasonality of monthly observations.

simulation function returns a plot with y: the seasonality, as shown in Figure 6.7

6.2.2 Random walk local level and seasonal

The model is given by

$$
\begin{aligned}
y_t &= \gamma_t + s_t + e_t \\
\gamma_t &= \gamma_{t-1} + b_t \\
s_t &= -\sum_{m=1}^{M-1} s_{t-m} + w_t
\end{aligned}
$$

where $e_t \sim NO(0, \sigma_e^2)$, $b_t \sim NO(0, \sigma_b^2)$, and $w_t \sim NO(0, \sigma_w^2)$, where γ_t is the random walk local level and s_t is the seasonal effects.

The following function simulates the random walk local level and seasonal model, `mrw.seas.sim(N=240, mu=1, sige=1, sigb=.4, sigw=.04, init=NULL, sigI=1, frequency=4, plot=T)`.

This function produces a simulation of the random walk local level and seasonal.

The inputs are the same as `mrwAll.sim()`, with additional inputs for the random walk local level and seasonal, which are the standard deviation of the seasonal effects, `sigw`= σ_w, the initial values for the seasonal effects, i.e. The initial values for the seasonal effects, i.e. $s_1, s_2, \ldots, s_{M-1}$ were, by default (since `init=NULL`), randomly generated from a normal distribution with mean 0 and standard deviation 1 (since `sigI=1`). `Frequency=4` for a quarterly data. For monthly data, the input is `Frequency=12`.

In example above a sequence of 240 quarterly observations were simulated using the function `mrw.seas.sim()` with $\gamma_1 = 1$ and three hyperparameters $\sigma_e = 1$, $\sigma_b = .4$ and $\sigma_w = .04$. The simulation function returns a plot with `y`: the observations, `T`: random walk local level, `S`: seasonality, `e`: irregular component of the observations, as shown in Figure 6.8

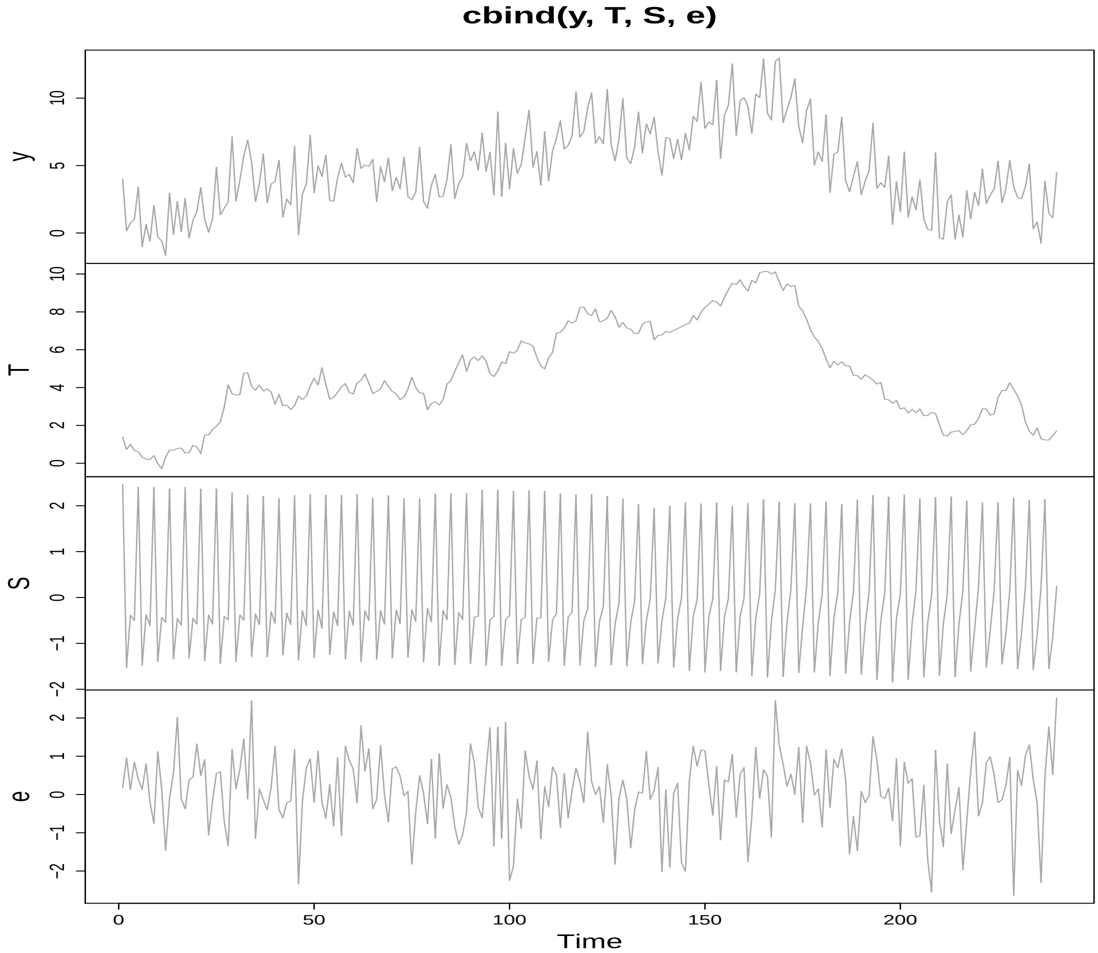


Figure 6.8: Simulation of random walk and seasonality of quarterly observations.

6.2.3 Autoregressive local level and seasonal

The model is given by

$$
\begin{aligned}
y_t &= \gamma_t + s_t + e_t \\
\gamma_t &= \sum_{j=1}^{J} \phi_j \gamma_{t-j} + b_t \\
s_t &= -\sum_{m=1}^{M-1} s_{t-m} + w_t
\end{aligned}
$$

where $e_t \sim NO(0, \sigma_e^2)$, $b_t \sim NO(0, \sigma_b^2)$, and $w_t \sim NO(0, \sigma_w^2)$, where γ_t is the autoregressive local level and s_t is the seasonal effects.

The following function simulates the autoregressive local level and seasonal model,

```
mar.seas.sim(N=240, mu=0, sige=1, sigb=.2, sigw=.05, phi=c(.5),
init=NULL, sigI=1, frequency=4, plot=T)
```

This gives the autoregressive and seasonality simulation. This function is the same as the random walk and seasonality simulation function `mrw.seas.sim()`, but with an extra input parameter `phi` for the ar model. The current function `mar.seas.sim()` allows order up to and including three.

In this example a sequence of 240 quarterly observations were simulated using the function `mrw.seas.sim` with $\gamma_1 = 0$ and three hyperparameters $\sigma_e = 1$, $\sigma_b = .2$, $\sigma_w = .05$ and $\phi = 0.5$. The initial values for the seasonal effects, i.e. $s_1, s_2, \ldots, s_{M-1}$ were, by default (since `init=NULL`), randomly generated from a normal distribution with mean 0 and standard deviation 1 (since `sigI=1`).

The simulation function returns a plot with `y`: the observations, `T`: random walk local level, `S`: seasonality, `e`: irregular component of the observations, as shown in Figure 6.9

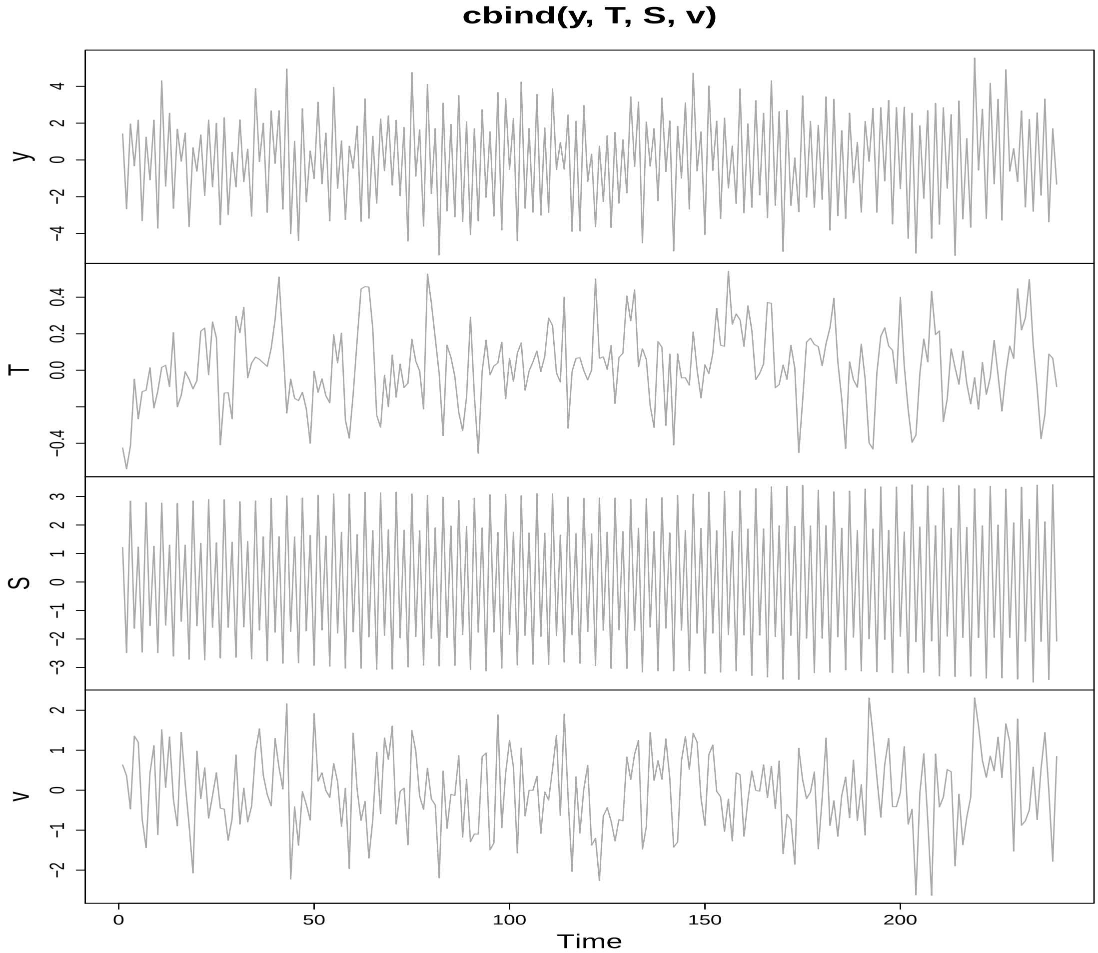


Figure 6.9: Simulation of autoregressive and seasonality quarterly observations.

6.3 Local level fitting functions

6.3.1 Random walk local level

The function to fit the random walk local level model in R is

`RW(y, weights=rep(1,length(y)), order=1, sig2e=1, sig2b=1, plot=FALSE, sig2e.fix = FALSE, sig2b.fix=FALSE, penalty=FALSE, delta=c(0.01, 0.01), shift=c(0,0))`.

The model fitted is given by

$$\begin{aligned} y_t &= \gamma_t + e_t \\ \Delta^d(\gamma_t) &= b_t, \end{aligned}$$

where $e_t \sim NO(0, \sigma_e^2)$, $b_t \sim NO(0, \sigma_b^2)$, and d is the order of the random walk, which can take any positive value.

The `RW()` function fits a random walk local level model, returns and also plots the fitted values of the mean (i.e. $\hat{\gamma}_t$ for $t = 1, 2, \ldots, T$) for the fitted model and estimates the hyperparameters (σ_e^2 and σ_b^2). It returns the estimated hyperparameters, degrees of freedom, global deviance, AIC, SBC, and marginal deviance. By calling `plot(m1)` where m1 is the fitted model, it gives the summary of the randomised quantile residuals along with the QQ plot.

The inputs of the function are: `y` a vector of observations; and if `plot=TRUE`, it plots the data along with the fitted mean. The other arguments are set by default but they can be changed, `weights`: prior weights for the observations, `order`: the order of the random walk, i.e. the number of differences of γ_t in the random walk, `sig2e`: the initial value for the variance in the observations error measurements (σ_e^2), `sig2b`:

the initial value for the variance in the state vector innovations (σ_b^2), `sig2e.fix`, `sig2b.fix`: are used when the variances are fixed and not estimated. The optional penalty arguments, `delta` and `shift`, are penalties on the hyperparameters, and used when we maximize the Q function, for avoiding local optima incase the surface of the maximum likelihood of the Q function is flat. Both penalties can be used for tilting the surface of the maximum likelihood of the Q function. These options are mainly used for non-Gaussian structural time series observations.

The following function:

```
RW.s(y, weights=rep(1,length(y)), order=1, sig2e=1, sig2b=1,
sige.fix=FALSE, plot=FALSE)
```

is a special case of the `RW()` function, without penalties on the hyperparameters. It fits a random walk local level model, and returns the fitted values and estimated hyperparameters. If the fit is just a basic fit without options (penalty, shift and delta), then calling the function RW.s(y,plot=T) is adequate.

Example of fitting the simulated data in Figure 6.1

In this example the simulated data in Figure 6.1 is fitted with the `RW()` function and the fitted hyperparameters are compared with the true hyperparameters.

The R commands are:

```
set.seed(11111)
y1 <- mrwAll.sim(N=4000, mu=0, sige=3, sigb=1, order=1, plot=TRUE)
m1 <- RW(y1,plot=T)
```

The fitted mean of model `m1` is identical to the simulated true mean in Figure 6.1, as shown in Figure 6.10, and the fitted hyperparameters of model `m1` agrees with the true hyperparameters, as shown in Table 6.1.

Table 6.1: The true hyperparameters and fitted hyperparameters of model `m1`.

Data	Random walk local level	σ_e	σ_b
y1	true hyperparameters	3	1
	fitted hyperparameters	3.0028	1.019

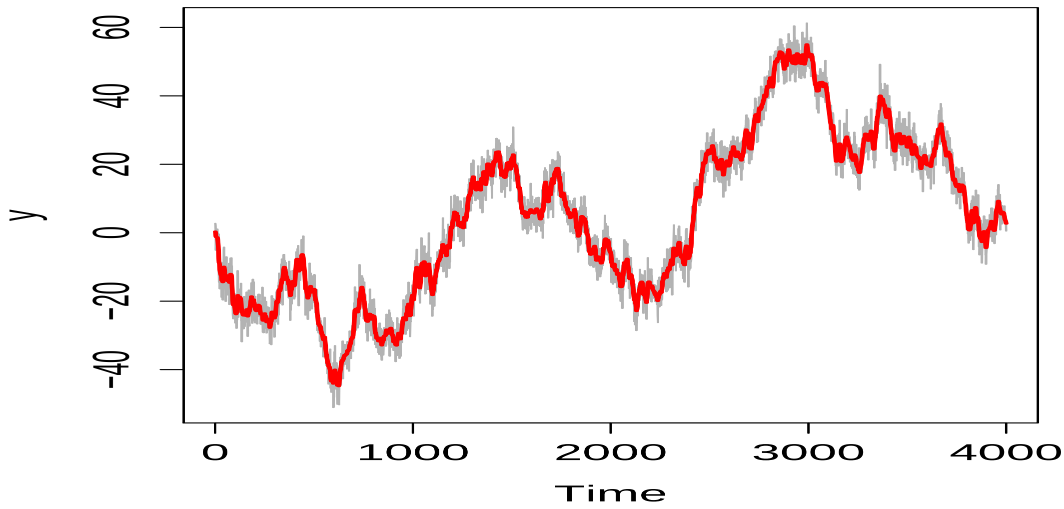


Figure 6.10: Fitted simulated data in Figure 6.1.

Example of fitting the simulated data in Figure 6.2

In this example the simulated data in Figure 6.2 is fitted with the `RW()` function and the fitted hyperparameters are compared with the true hyperparameters.

```
set.seed(23322)
y2 <- mrwAll.sim(N=4000, mu=0, sige=10, sigb=1, order=1, plot=TRUE)
m2 <- RW(y2,plot=T)
```

The fitted mean of model `m2` is identical to the simulated true mean in Figure 6.2, as shown in Figure 6.11, and the fitted hyperparameters of model `m2` agrees with the true hyperparameters, as shown in Table 6.2.

Table 6.2: The true hyperparameters and fitted hyperparameters of model `m2`.

Data	Random walk local level	σ_e	σ_b
y2	true hyperparameters	10	1
	fitted hyperparameters	9.9523	1.022

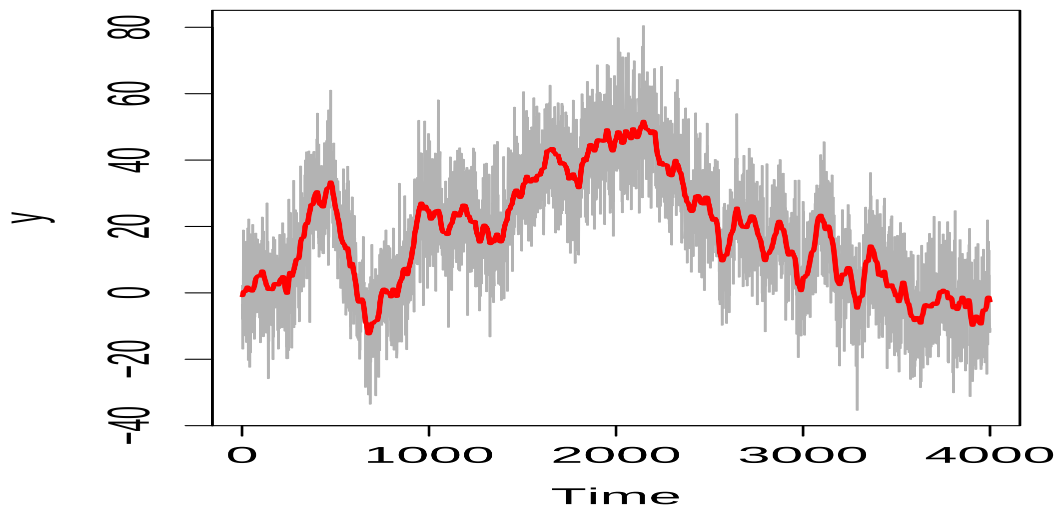


Figure 6.11: Fitted simulated data in Figure 6.2.

6.3.2 Autoregressive local level

The function to fit the autoregressive local level model in R is

`AR(y, weights=rep(1,length(y)), order=3, sig2e=1, sig2b=1, phi1=0.5, phi2=0.5, phi3=0.5, delta=0.1, plot=FALSE)`

The model fitted is given in Section 6.1.5.

`AR()` fits an autoregressive local level model and returns both the fitted values for ar local level model and estimates the hyperparameters of the fitted model. the `delta` is an optional penalty with the same purpose as in `RW()`.

6.3.3 Random walk local level and trend

The function to fit the random walk local level and trend model in R is

`rw.tr(y, weights=rep(1,length(y)), sig2e=1, sig2b=2, sig2d=1, plot=FALSE)`

The model fitted is given in Section 6.1.4.

`rw.tr()` returns both the fitted values for random walk local level with trend and estimates the hyperparameters of the model.

6.3.4 Random walk local level with random coefficient of an explanatory variable

The function to fit the random walk local level with a random coefficient of an explanatory variable model in R is

`rw.exp(y, x, weights=rep(1,length(y)), sig2e=1, sig2b=1, sig2v=1, plot=FALSE)`

The model fitted is given by

$$
\begin{aligned}
y_t &= \gamma_t + \beta_t x_t + e_t \qquad (6.1)\\
\gamma_t &= \gamma_{t-j} + b_t\\
\beta_t &= \beta_{t-1} + v_t
\end{aligned}
$$

where $e_t \sim NO(0, \sigma_e^2)$, $b_t \sim NO(0, \sigma_b^2)$, and $v_t \sim NO(0, \sigma_v^2)$, where γ_t is the random walk local level and x_t is the explanatory variable.

This function needs two time series inputs, `y` for the observations, and `x` for the explanatory variable values. In `rw.exp()` function the coefficient β_t of the explanatory variable changes over time. It models the effect of the explanatory variable on the time series observations. The function returns the fitted values for the local level model and the fitted values of the coefficient of the explanatory variable, and estimates the hyperparameters of the model. The argument `sig2v` is the initial value of the variance for the random error in the coefficient β_t , i.e. σ_v^2.

6.4 Seasonality fitting functions

6.4.1 Seasonality

The function to fit the seasonal effect model in R is

```
seas(y, weights=rep(1,length(y)), sig2e=1, sig2w=1, frequency=4,
plot=FALSE)
```

The model fitted is given in Section 6.2.1.

This fits the seasonal effect model. The argument `frequency` represents the frequency of the observations, where the time series is daily, monthly or quarterly observations.

6.4.2 Random walk local level and seasonal

The function to fit the random walk local level and seasonal model in R is

`rw.seas(y, weights=rep(1,length(y)), sig2e=1, sig2b=1, sig2w=1, frequency=4, plot=FALSE)`.

The model fitted is given in Section 6.2.2.

This fits the random walk local level with seasonal model. The argument `frequency` represents the frequency of the observations, where the time series is for example daily, monthly or quarterly observations. If the data has a quarterly seasonal effect then we set `frequency=4`, for daily `frequency=7` and for monthly `frequency=12`.

6.4.3 Random walk local level with trend and seasonal

The function to fit the random walk local level with trend and seasonal model in R is

`rw.tr.seas(y, weights=rep(1,length(y)), sig2e=1, sig2b=1, sig2d =1, sig2w=1, frequency=4, plot=FALSE)`

The model fitted is given by

$$
\begin{aligned}
y_t &= \gamma_t + s_t + e_t \\
\gamma_t &= \gamma_{t-1} + \psi_t + b_t \\
\psi_t &= \psi_{t-1} + d_t \\
s_t &= -\sum_{m=1}^{M-1} s_{t-m} + w_t
\end{aligned}
$$

where $e_t \sim NO(0, \sigma_e^2)$, $b_t \sim NO(0, \sigma_b^2)$, $d_t \sim NO(0, \sigma_d^2)$, and $w_t \sim NO(0, \sigma_w^2)$ where

γ_t is the random walk local level, ψ_t is the stochastic trend, and s_t is the seasonal effect.

This gives estimation and fitting of the random walk local level with trend and seasonality. It returns and plots (if `plot=TRUE`) the fitted values for the local level, for stochastic trend and for seasonal effects.

6.4.4 Autoregressive local level and seasonal

The function to fit the autoregressive walk local level and seasonal model in R is

```
ar.seas(y, weights=rep(1,length(y)), order=3, sig2e=1, sig2b=1,
sig2w=.5, phi1=0.5, phi2=0.5, phi3=0.5, frequency=4, plot=FALSE)
```

The model fitted is given in Section 6.2.3

This gives estimation and fitting of autoregressive local level with seasonality. The current function `ar.seas()` only allows order up to and including three.

6.4.5 Autoregressive local level with trend and seasonal

The function to fit the autoregressive local level with trend and seasonal model in R is

```
ar.tr.seas(y, weights=rep(1,length(y)), order=1, sig2e=1, sig2b=1,
sig2d=1, sig2w=1, phi=0.5, plot=TRUE)
```

The model fitted is given by

$$\begin{aligned} y_t &= \gamma_t + s_t + e_t \\ \gamma_t &= \sum_{j=1}^{J} \phi_j \gamma_{t-j} + \psi_t + b_t \\ \psi_t &= \sum_{l=1}^{L} \rho_l \psi_{t-l} + d_t \\ s_t &= -\sum_{m=1}^{M-1} s_{t-m} + w_t \end{aligned} \tag{6.2}$$

This fits an ar local level with trend and seasonality model and returfs the fitted values for local level, trend and seasonality and estimates the hyperparameters of each component. The current function `ar.seas()` only allows order up to and including three.

Note:

The fitted mean $\hat{\mu}_t$ (with or without trend and seasonality) given by the fitting functions are exactly identical to the simulated mean μ_t (with or without trend and seasonality) given by the simulation functions, for this reason the author did not plot the fitted mean because is very identical to the simulated mean given in the simulation examples.

Chapter 7

GEST process and simulation

7.1 Introduction

This Chapter introduces the theory of a new stochastic process called the Generalized Structural (GEST) stochastic process, provides new simulated examples of the GEST process in R, fitting the non-Gaussian examples with the GEST model, and derives two theorems for the properties of the GEST process.

Current Gaussian structural time series models provide dynamic linear models for the mean of the conditional normal distribution, or provide a separate model for stochastic volatility. Non-Gaussian structural time series models provide a generalized dynamic linear model for the mean of the conditional exponential family distributions, or a stochastic volatility model.

This chapter introduces a distributional stochastic process[1] for Gaussian and

[1] A *Stochastic process* involves a random variable, e.g. Y_t, which is time (or space) varying. The main properties for distinction between processes are

- the *state space* as the set of all possible observed values. This *space* can be continuous or discrete.
- the *index* t can be continuous or discrete.
- the *nature of dependence* of the random variables, Y_t

non-Gaussian continuous and discrete (count) time series data, including seasonal time series data. The generalized structural time series (GEST) process extends the traditional Gaussian structural time series models by explicitly modelling the conditional (i.e. time varying) location, scale, skewness and kurtosis parameters jointly. The method of estimation of the hyperparameters is demonstrated in the next chapter.

The GEST process is a stochastic process which describes a time varying location (e.g. mean), time varying scale (e.g standard deviation), and time varying shape (e.g skewness and kurtosis parameters) as a random walk or autoregressive process including seasonality.

The parameters of the conditional distribution are treated as signals, for example, if the conditional distribution of the process is Gaussian, it means that the data has two unobserved signals, a location signal and a scale signal, whereas for the skew Student t distribution, there are four unobserved signals, a location signal, a scale signal and two shape (skewness and kurtosis) signals.

This chapter defines the GEST stochastic process, gives its properties and simulates Gaussian and non-Gaussian structural time series models. The next chapter introduces the GEST model for the signals of the assumed conditional distribution of the process. The algorithm for fitting the GEST model, which is described in detail in the next chapter, is based on the RS algorithm for fitting generalized additive models for location, scale and shape (Rigby and Stasinopoulos 2005).

7.2 The GEST process

The GEST process assumes that the random variable Y_t is derived from a probability density function $f_{Y_t}(y_t|\boldsymbol{\theta}_t)$ conditional on $\boldsymbol{\theta}_t$ where ${\boldsymbol{\theta}_t}^\top = (\theta_{1,t}, \ldots, \theta_{K,t})$ is a vector of distribution parameters for $f_{Y_t}()$.

Hence $Y_t|\boldsymbol{\theta}_t \sim \mathcal{D}(\boldsymbol{\theta}_t)$, for conditional distribution $\mathcal{D}$, where each $\theta_{k,t}$ is generated by a random process given by

$$g_k(\theta_{k,t}) = \beta_{k,0} + \gamma_{k,t} \tag{7.1}$$

for $t = 1, 2, \ldots, T$, where

$$\gamma_{k,t} = \sum_{j=1}^{J_k} \phi_{k,j}\gamma_{k,t-j} + b_{k,t} \tag{7.2}$$

for $t = J+1, J+2, \ldots, T$, where, for $k = 1, 2, \ldots, K$, function $g_k()$ is a specified link function, $\gamma_{k,t}$ for $t = 1, 2, \ldots, T$ is an individual structural time series random process and $b_{k,t}$ are random errors, independent from each other mutually and serially, and normally distributed with expected values equal to zero and variance $\sigma^2_{b_k}$. Thus $\mathbf{b_k} \sim N_{n-J_k}(0, \sigma^2_{b_k}\mathbf{I}_{n-J_k})$, where $\mathbf{b_k}^\top = (b_{k,J_k+1}, \ldots, b_{k,T})$ for $k = 1, 2, \ldots, K$.

There are several important points to be made here about a GEST process.

- The probability distribution $f_{Y_t}(y_t|\boldsymbol{\theta}_t)$ can be a continuous or discrete distribution.

- For most practical applications, K, the number of parameters $\boldsymbol{\theta}_t$ in the distribution is less than or equal to four. These four parameters are denoted as $\boldsymbol{\theta}_t = (\mu_t, \sigma_t, \nu_t, \tau_t)$ where μ_t is a time-varying location parameter, σ_t is a time-varying scale parameter and ν_t and τ_t are time-varying shape parameters, which may be related to the time-varying skewness and time-varying kurtosis of the distribution respectively.

- The link function $g_k()$ is used to ensure that the individual parameter is defined on a permissible range. For example, a log link for sigma, i.e. $g_2(\sigma_t) =$

$\log(\sigma_t) = \gamma_{2,t}$, will ensure that $\sigma_t = \exp(\gamma_{2,t})$ is always positive.

- The $\phi_{k,j}$ in equation (7.2) are autoregressive parameters for the individual predictors $\gamma_{k,t}$ for $k = 1, 2, 3, 4$. Note that specific fixed values for $\phi_{k,j}$ for $j = 1, 2, \ldots, J_k$ replaces autoregressive terms with random walk terms for $\gamma_{k,t}$. For example setting $J_k = 1$ and $\phi_{k,1} = 1$ gives a random walk of order 1, while setting $J_k = 2$, $\phi_{k,1} = 2$ and $\phi_{k,2} = -1$ gives a random walk of order 2, for $k = 1, 2, 3, 4$.

- The process can be extended to include a seasonal effect (with M seasons)

$$g_k(\theta_{k,t}) = \beta_{k,0} + \gamma_{k,t} + s_{k,t}$$

where $\gamma_{k,t}$ is given by (7.2) and

$$s_{k,t} = -\sum_{m=1}^{M-1} s_{k,t-m} + w_t$$

- Note that the generation of the GEST process requires four sets of values:

  (i) the constant parameters $\beta_{k,0}$ for $k = 1, 2, \ldots, K$.

  (ii) the AR parameters $\phi_{k,j}$ for $j = 1, 2, \ldots, J_k$ and $k = 1, 2, \ldots, K$,

  (iii) the standard deviations σ_{b_k} of the white noises since $b_{k,t} \sim N\left(0, \sigma_{b_k}^2\right)$ for $k = 1, 2, \ldots, K$,

  (iv) the initial starting values for the distribution parameters.

The GEST process is very flexible and can take familiar patterns of real data situation. The GEST process can be non-stationary and potentially explosive by nature. This is not in general bad, since many physical, economic and financial phenomena variables are themselves explosive. However, some statistical properties are difficult to establish unless additional assumptions about the nature of the GEST process are made.

7.3 Properties of the GEST process

Two theorems are introduced here to show that under certain circumstances the GEST process is stationary with well defined marginal mean and variance. In particular note Theorem 1 assumes: i) identity link function for μ_t and ii) log link function for σ_t. Theorem 2 assumes: i) log link function for μ_t and : ii) log link function for σ_t:

7.3.1 Theorem 1

Theorem 1: *Let μ_t and $c\sigma_t$ (where c is a known constant) be, respectively, the conditional mean and standard deviation (assumed to exist) of the distribution $Y_t|\mu_t, \sigma_t, \nu_t, \tau_t \sim \mathcal{D}(\mu_t, \sigma_t, \nu_t, \tau_t)$ where $\mu_t = \beta_{1,0} + \gamma_{1,t}$ and $\log \sigma_t = \beta_{2,0} + \gamma_{2,t}$ and where*

$$\gamma_{k,t} = \sum_{j=1}^{J_k} \phi_{k,j} \gamma_{k,t-j} + b_{k,t}$$

for $k = 1, 2$, where $b_{1,t}$ and $b_{2,t}$ are mutually and serially independently normally distributed with mean 0 and variances $\sigma^2_{b_1}$ and $\sigma^2_{b_2}$ respectively and where

$$\begin{aligned} \gamma_{1,t} &= \Phi_1(B)^{-1} b_{1,t} = \psi_1(B) b_{1,t} \\ \gamma_{2,t} &= \Phi_2(B)^{-1} b_{2,t} = \psi_2(B) b_{2,t}, \end{aligned}$$

assuming $\Phi_1(B)$ and $\Phi_2(B)$ are invertible, then the GEST process has a stationary mean and variance given by

$$\begin{aligned} E[Y_t] &= \beta_{1,0} \\ V[Y_t] &= S_1\sigma_{b_1}^2 + c^2\exp\left(2\beta_{2,0} + 2S_2\sigma_{b_2}^2\right) \end{aligned}$$

respectively, where $S_k = 1 + \sum_{j=1}^{\infty}\psi_{k,j}^2$ for $k = 1,2$ and where $\psi_k(B) = \Phi_k(B)^{-1} = 1 + \psi_{k,1}B + \psi_{k,2}B^2 + \ldots$ and provided $\Phi_k(B)$ is invertible, where $\Phi_k(B) = 1 - \phi_{k,1}B - \phi_{k,2}B^2 - \ldots - \phi_{k,J_k}B^J$ and B is the backshift time operator, $By_t = y_{t-1}$.

Appendix C1 gives the proof for Theorem 1. Note that Theorem 1 is not affected by the form of the model for ν_t and τ_t. Also Theorem 1 applies to any distribution $\mathcal{D}$ in which μ_t and $c\sigma_t$ are respectively the mean and standard deviation of $\mathcal{D}$. In particular Theorem 1 applies to the normal, $NO(\mu, \sigma)$, skew Student t, $SST(\mu, \sigma, \nu, \tau)$, Power Exponential, $PE(\mu, \sigma, \nu)$, t-family parameterized so σ is the standard deviation, $TF2(\mu, \sigma, \nu)$, and Johnson's Su, $JSU(\mu, \sigma, \nu, \tau)$, distributions, where $c = 1$. It also applies to the logistic, $LO(\mu, \sigma)$, Gumbel, $GU(\mu, \sigma)$, and Reverse Gumbel, $RG(\mu, \sigma)$, where $c \neq 1$ (see Stasinopoulos *et al.*, 2008, for the parametrization of the probability density functions of the distributions).

7.3.2 Theorem 2

Theorem 2: *Let the distribution of $Y_t|\mu_t, \sigma_t, \nu_t, \tau_t \sim \mathcal{D}(\mu_t, \sigma_t, \nu_t, \tau_t)$ have a mean μ_t and variance $v(\mu_t, \sigma_t)$ where $\log\mu_t = \beta_{1,0} + \gamma_{1,t}$, and $\log\sigma_t = \beta_{2,0} + \gamma_{2,t}$ and where*

$$\begin{aligned} \gamma_{1,t} &= \Phi_1(B)^{-1}b_{1,t} = \psi_1(B)b_{1,t} \\ \gamma_{2,t} &= \Phi_2(B)^{-1}b_{2,t} = \psi_2(B)b_{2,t}, \end{aligned}$$

as defined in Theorem 1, assuming $\Phi_1(B)$ and $\Phi_2(B)$ are invertible, then the following give marginal means and variances of the related process:

a) $E[Y_t] = E[\mu_t] = \exp\left(\beta_{1,0} + \frac{1}{2}S_1\sigma_{b_1}^2\right)$

b) $V[Y_t] = V[\mu_t] + E[v(\mu_t, \sigma_t)]$

c) $V[\mu_t] = \exp(2\beta_{1,0})\left[\exp\left(2S_1\sigma_{b_1}^2\right) - \exp\left(S_1\sigma_{b_1}^2\right)\right]$

d) $E[\mu_t^r] = \exp\left(r\beta_{1,0} + \frac{1}{2}r^2S_1\sigma_{b_1}^2\right)$

e) $E[\sigma_t^r] = \exp\left(r\beta_{2,0} + \frac{1}{2}r^2S_2\sigma_{b_2}^2\right)$

for $r > 0$.

Appendix C2 gives the proof for Theorem 2 together with a corollary for Theorem 2 providing the marginal variance of Y_t for four conditional distributions for Y_t, the negative binomial type I and type II, $NBI(\mu,\sigma)$, $NBII(\mu,\sigma)$, the gamma, $GA(\mu,\sigma)$, and inverse Gaussian, $IG(\mu,\sigma)$, distributions (see Stasinopoulos *et al.*, 2008, for the parametrization of the probability (density) functions of the distributions).

7.4 Simulation of the GEST process

This section provides simulations of the GEST process in R, for Gaussian and non-Gaussian continuous and discrete (count) time series data, with conditional distributions for up to four parameters. Any of the 80 distributions in the `gamlss` package in R, (Stasinopoulos and Rigby, 2007), can be used to model the conditional distribution of the response variable. Here, for illustration, the normal, Poisson, negative binomial type 1, Student t and skew Student t conditional distributions are considered. The fitting of the GEST process, including estimation of the hyperparameters, is presented in the next chapter.

7.4.1 GEST process with normal distribution

For the conditional normal distribution, the simulations of the GEST process are a generalization of the simulations presented in the previous chapter, because the GEST process provides dynamic linear models for *both* the mean and the standard deviation jointly. The advantage of the GEST process is the ability of generating different scenarios of a normal stochastic process with both dynamic mean and dynamic standard deviation, as a random walk or autoregressive process, including seasonality.

Gaussian random walk local level model

The Gaussian random walk local level model is defined as:

$$\begin{aligned} y_t &= \gamma_t + e_t \\ \gamma_t &= \gamma_{t-1} + b_t \end{aligned}$$

where e_t and b_t are two independent Gaussian white noise, where $e_t \sim NO(0, \sigma_e^2)$ and $b_t \sim NO(0, \sigma_b^2)$. This model can be written in a distributional form as:

$$\begin{aligned} Y_t|\mu_t, \sigma &\sim NO(\mu_t, \sigma) \\ \mu_t &= \gamma_t \\ \gamma_t &= \gamma_{t-1} + b_t \end{aligned}$$

Note that, in Gaussian local level models the variance is assumed constant, whereas in the GEST process the variance can be modelled explicitly by the log of standard deviation.

Let the parametric conditional distribution $\mathcal{D}$ of the response variable Y_t be the Gaussian distribution and assume an identity link for the mean (μ_t) and a log link for the standard deviation (σ_t).

GEST process for Y_t with random walk order 1 local level for μ_t and a constant for $\log(\sigma_t)$

The GEST process for Y_t with random walk order 1 local level for the mean (μ_t) and a constant for the log standard deviation ($\log(\sigma_t)$) is defined as:

$$\begin{aligned} Y_t|\mu_t, \sigma_t &\sim NO(\mu_t, \sigma_t) \\ \mu_t &= \beta_{1,0} + \gamma_{1,t} \\ \log(\sigma_t) &= \beta_{2,0} \\ \gamma_{1,t} &= \gamma_{1,t-1} + b_{1,t} \end{aligned} \tag{7.3}$$

for $t = 1, 2, \ldots, T$, where $b_{1,t} \sim NO(0, \sigma_b)$. Note that, in a random walk process the constant $\beta_{1,0}$ is confounded with the initial value of the random walk $\gamma_{1,1}$. In the autoregressive (ar) process the constant $\beta_{1,0}$ is the reversion line around which the ar process moves as a stationary process, whereas the random walk process is not a stationary process, and its variance increases as the time increases.

Also

$$\hat{\sigma}_{Y_t} = \exp(\hat{\beta}_{2,0}) \times \hat{\sigma}_e = \hat{\sigma}_t \times \hat{\sigma}_e,$$

where $\hat{\sigma}_e$ is the fitted standard deviation for the errors in the mean local level model.

Below is the first example of the GEST stochastic process for 1000 random observations, generated by assuming that the conditional distribution $f_{Y_t}(y_t|\boldsymbol{\theta}_t)$ of the process is Gaussian $NO(\mu_t, \sigma_t)$. The mean parameter of the distribution of the $NO(\mu_t, \sigma_t)$, for $t = 1, 2, \ldots, T$, is simulated using a random walk order one process.

Note that, in the simulation function `gest.sim()` of the GEST process, `N`: is the number of observations, `mu.init`: is the initial value of the mean process, which is the value for $\beta_{1,0}$ confounded with the initial value of the random walk $\gamma_{1,1}$, `sigma.init`: is the initial value of the sigma process, which is the value for $\sigma_t = \exp(\beta_{2,0})$, `mu.sigb`: is the true value of the standard deviation for the errors in the mean local level, and `sigma.sigb`: is the true value of the standard deviation for the errors in the sigma local level. In this example, $b_{1,t} \sim NO(0, .1)$ and sigma is fixed $\sigma_t = 1$ and $\log(\sigma_t) = \beta_{2,0} = 0$ in model (7.3).

The output from `plot=TRUE` in the above command is given in Figure 7.1, which gives, for $t = 1, 2, \ldots, T$, the simulated series y_t together with: the simulated mean process μ_t and its corresponding predictor process $\eta_{1,t} = \mu_t$ [since μ has the default identity link for a $NO(\mu, \sigma)$ distribution] labelled `mu.eta` in Figure 7.1, the simulated sigma process $\sigma_t = \exp(\beta_{2,0})$, a constant in this example, and its corresponding predictor process $\eta_{2,t} = \log(\sigma_t) = \beta_{2,0}$ [since σ has the default log link for a $NO(\mu, \sigma)$ distribution] labelled `sigma.eta` in Figure 7.1.

The R commands for simulating Figure 7.1 is given in Appendix D.

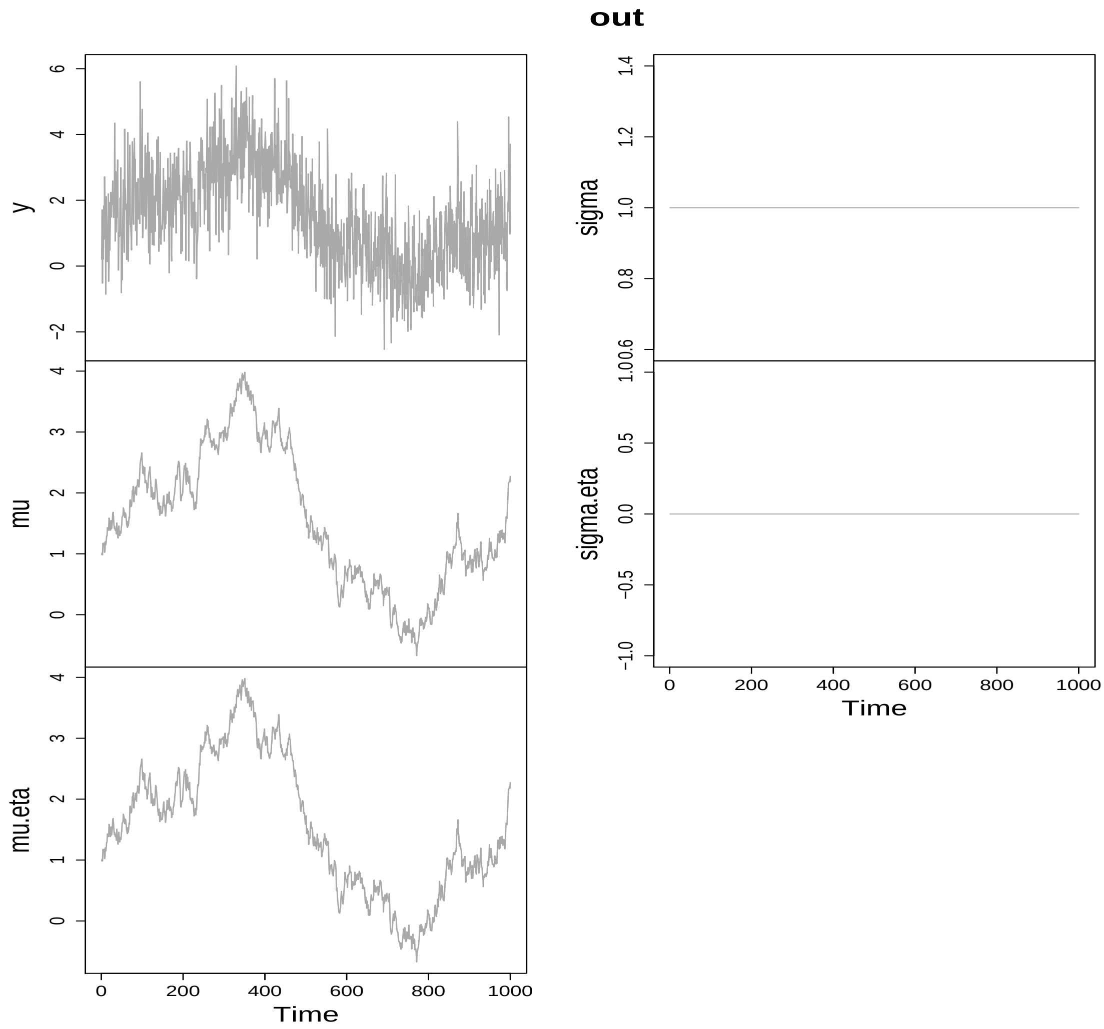


Figure 7.1: A GEST process simulation from a normal distribution with a constant sigma.

GEST process for Y_t with random walk order 1 local level for μ_t and random walk order 1 local level for $\log(\sigma_t)$

The GEST process for Y_t with random walk order 1 local level for the mean (μ_t) and random walk order 1 local level for the log standard deviation ($\log(\sigma_t)$) is defined as:

$$
\begin{aligned}
Y_t|\mu_t, \sigma_t &\sim NO(\mu_t, \sigma_t) \\
\mu_t &= \beta_{1,0} + \gamma_{1,t} \\
\log(\sigma_t) &= \beta_{2,0} + \gamma_{2,t}
\end{aligned}
$$

where

$$
\begin{aligned}
\gamma_{1,t} &= \gamma_{1,t-1} + b_{1,t} \\
\gamma_{2,t} &= \gamma_{2,t-1} + b_{2,t}.
\end{aligned}
$$

Below is the second example of a GEST stochastic process by assuming that the $f_{Y_t}(y_t|\boldsymbol{\theta}_t)$ of the process is Gaussian $NO(\mu_t, \sigma_t)$ and both μ_t and $\log(\sigma_t)$ follow a random walk order 1 process. We simulate each of the distribution parameters of the $NO(\mu_t, \sigma_t)$ for $t = 1, 2, \ldots, n$, using a random walk order one process, where $b_{1,t} \sim N(0, 0.1)$ and $b_{2,t} \sim N(0, 0.05)$.

The R commands for simulating Figure 7.2 is given in Appendix D, and the resulting output from `plot=T` is given in Figure 7.2.

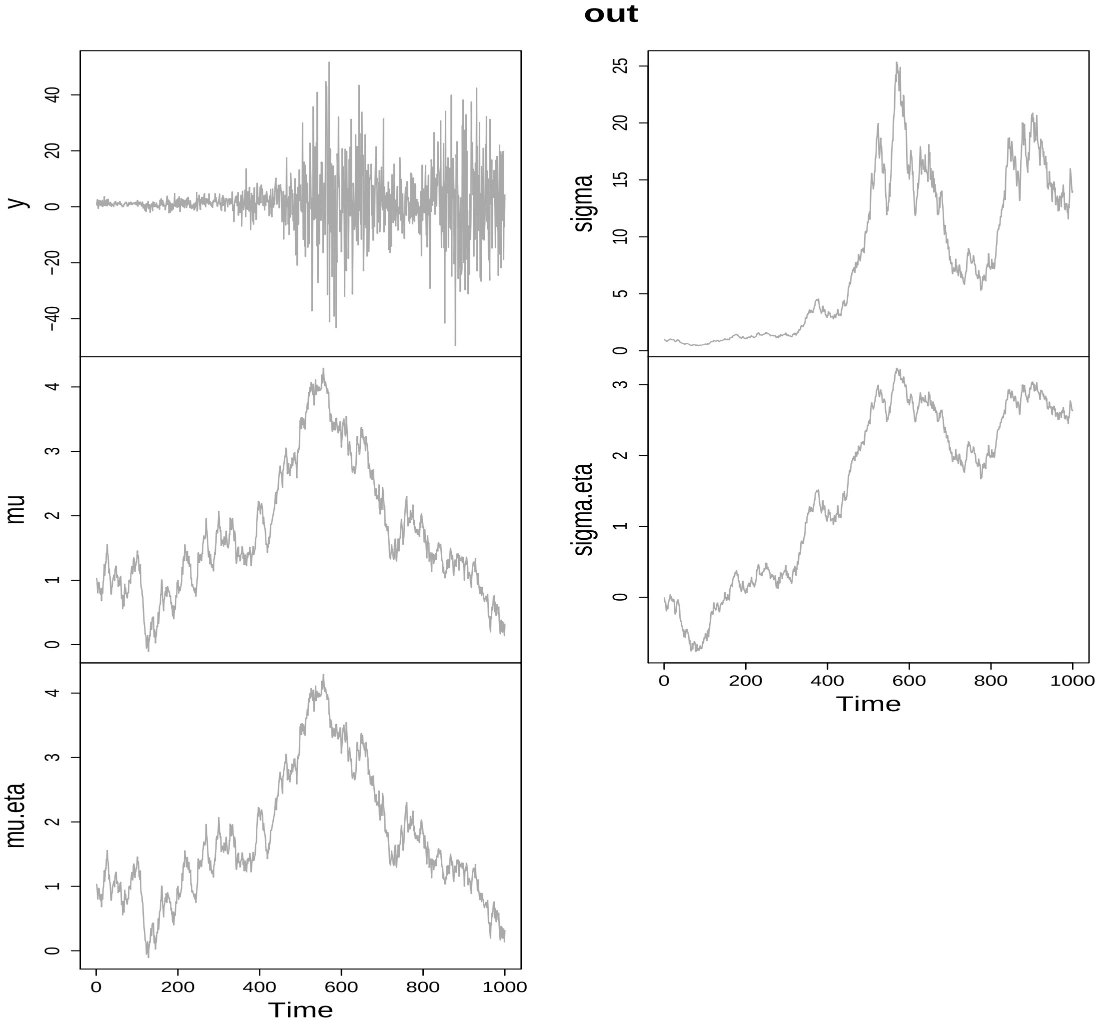


Figure 7.2: A GEST process simulation from a normal distribution with stochastic (random walks order 1) mu and sigma.

GEST process for Y_t with an autoregressive order 1 local level for μ_t and random walk order 1 local level for $\log(\sigma_t)$

The GEST process for Y_t with an autoregressive order 1 local level for the mean (μ_t) and random walk order 1 local level for the log standard deviation ($\log(\sigma_t)$) is defined as:

$$\begin{aligned} Y_t|\mu_t, \sigma_t &\sim NO(\mu_t, \sigma_t) \\ \mu_t &= \beta_{1,0} + \gamma_{1,t} \\ \log(\sigma_t) &= \beta_{2,0} + \gamma_{2,t} \end{aligned}$$

where

$$\begin{aligned} \gamma_{1,t} &= \phi\gamma_{1,t-1} + b_{1,t} \\ \gamma_{2,t} &= \gamma_{2,t-1} + b_{2,t}. \end{aligned}$$

Below is the third example of a GEST stochastic process for 1000 observations generated by assuming that the $f_{Y_t}(y_t|\boldsymbol{\theta}_t)$ of the process is Gaussian $NO(\mu_t, \sigma_t)$. The mean (μ_t) of the distribution parameters of the $NO(\mu_t, \sigma_t)$, for $t = 1, 2, \ldots, T$, follows an autoregressive order one process, and the log standard deviation ($\log(\sigma_t)$) follows a random walk order 1 process. The additional input values in the `gest.sim()` are the value for ϕ, the ar(1) parameter, `mu.phi`, and `mu.type`: the type of the process from which the mean is generated. Hence, $b_{1,t} \sim N(0, 0.1)$, $b_{2,t} \sim N(0, 0.05)$, $\phi = 0.5$ and `mu.type="AR"`.

The R commands for simulating Figure 7.3 is given in Appendix D, and the resulting output from `plot=T` is given in Figure 7.3.

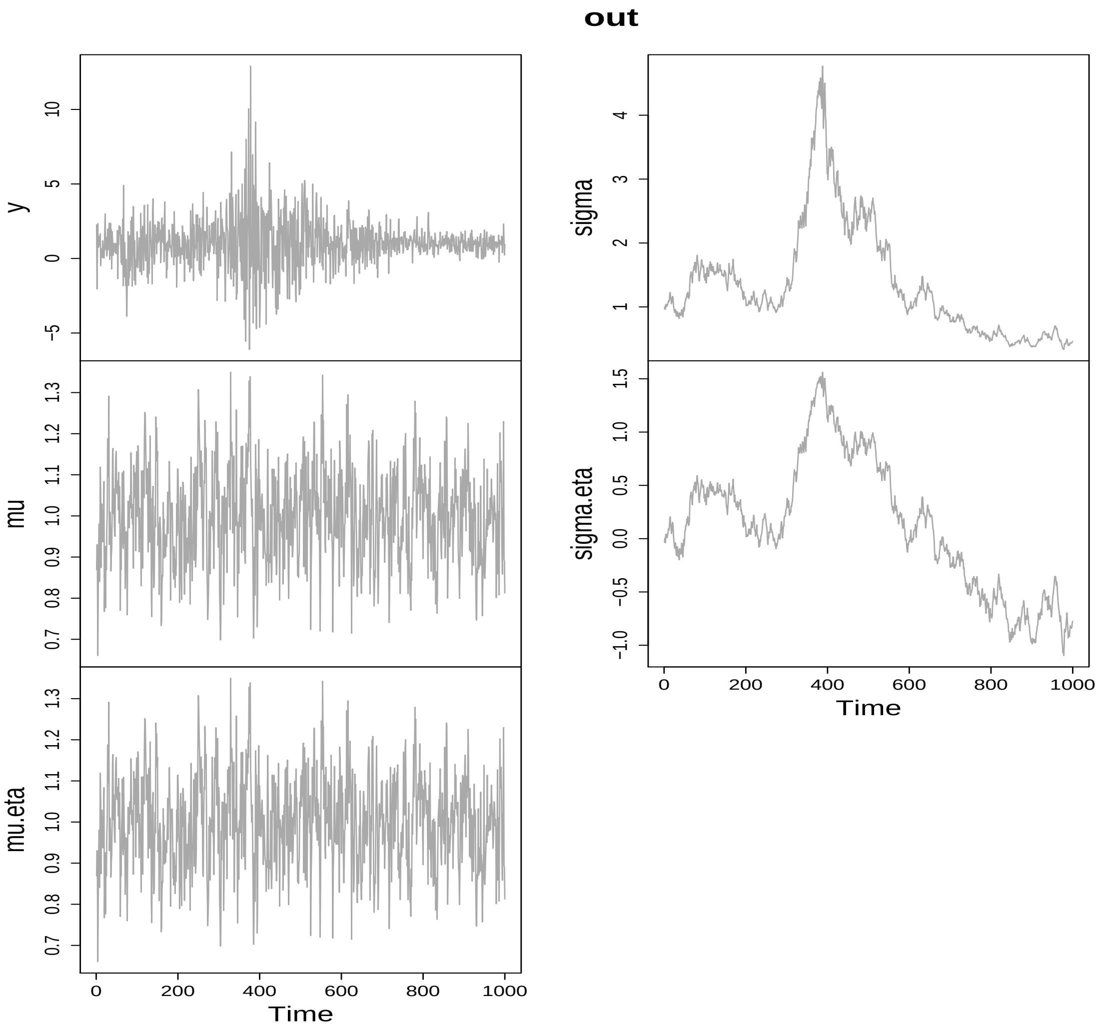


Figure 7.3: A GEST process simulation from a normal distribution with ar(1) for the mean level and rw(1) for log standard deviation.

GEST process for Y_t with a random walk local level and seasonal for μ_t and a constant for $\log(\sigma_t)$

The GEST process for Y_t with a random walk local level and seasonal model for the mean (μ_t) and a constant log standard deviation ($\log(\sigma_t)$) is defined as:

$$\begin{aligned}
Y_t|\mu_t,\sigma_t &\sim NO(\mu_t,\sigma_t) \\
\mu_t &= \beta_{1,0}+\gamma_{1,t}+s_{1,t} \\
\log(\sigma_t) &= \beta_{2,0}
\end{aligned}$$

where

$$\begin{aligned}
\gamma_{1,t} &= \gamma_{1,t-1}+b_{1,t} \\
s_{1,t} &= -\sum_{m=1}^{M-1} s_{1,t-m}+w_t.
\end{aligned}$$

Below is the fourth example of a GEST stochastic process for 240 monthly observations generated by assuming that the $f_{Y_t}(y_t|\boldsymbol{\theta}_t)$ of the process is Gaussian $NO(\mu_t,\sigma)$. The mean is simulated using a random walk order 1 local level and seasonal process, and log standard deviation is constant. The additional input values in the `gest.sim()` are `mu.sigS`: the true value of the standard deviation for the errors in the seasonal process, `frequency`: is the recurring pattern of the data, if the observations are quarterly, then `frequency=4`, daily is 7, and monthly is 12, and `mu.type`: the type of the process from which the mean is generated from. Hence, $b_{1,t}\sim N(0,0.1)$, $w_t\sim N(0,0.01)$, `frequency=12`, and `mu.type="levelSeasonal"`.

The R commands for simulating Figure 7.4 is given in Appendix D, and the resulting output from `plot=T` is given in Figure 7.4.

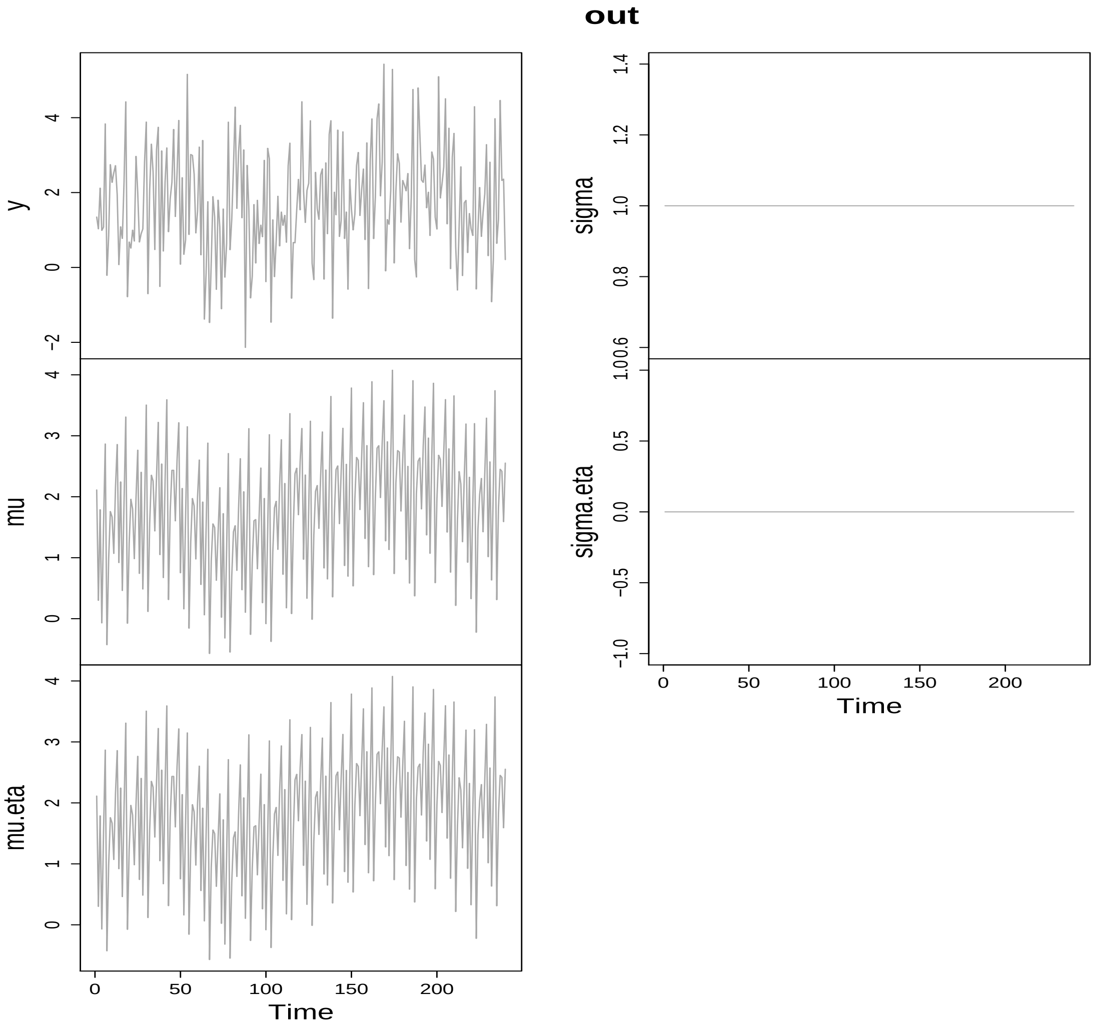


Figure 7.4: A GEST process simulation from a normal distribution with local level and seasonal for the mean and a constant for log standard deviation.

GEST process for Y_t with a random walk local level and seasonal for μ_t and a random walk order 1 local level for $\log(\sigma_t)$

The GEST process for Y_t with a random walk local level and seasonal model for the mean (μ_t) and a random walk order 1 local level for log standard deviation $\log(\sigma_t)$ is defined as:

$$
\begin{aligned}
Y_t|\mu_t, \sigma_t &\sim NO(\mu_t, \sigma_t) \\
\mu_t &= \beta_{1,0} + \gamma_{1,t} + s_{1,t} \\
\log(\sigma_t) &= \beta_{2,0} + \gamma_{2,t}
\end{aligned}
$$

where

$$
\begin{aligned}
\gamma_{1,t} &= \gamma_{1,t-1} + b_{1,t} \\
\gamma_{2,t} &= \gamma_{2,t-1} + b_{2,t} \\
s_{1,t} &= -\sum_{m=1}^{M-1} s_{1,t-m} + w_t
\end{aligned}
$$

Below is the fifth example of a GEST stochastic process for 240 monthly observations generated by assuming that the $f_{Y_t}(y_t|\boldsymbol{\theta}_t)$ of the process is Gaussian $NO(\mu_t, \sigma_t)$. The mean is simulated using a random walk order 1 local level and seasonal process and the log standard deviation is simulated with a random walk order 1 local level, where $b_{1,t} \sim N(0, 0.1)$, $b_{2,t} \sim N(0, 0.06)$, $w_t \sim N(0, 0.01)$, `frequency=12`, and `mu.type="levelSeasonal"`.

The R commands for simulating Figure 7.5 is given in Appendix D, and the resulting output from `plot=T` is given in Figure 7.5.

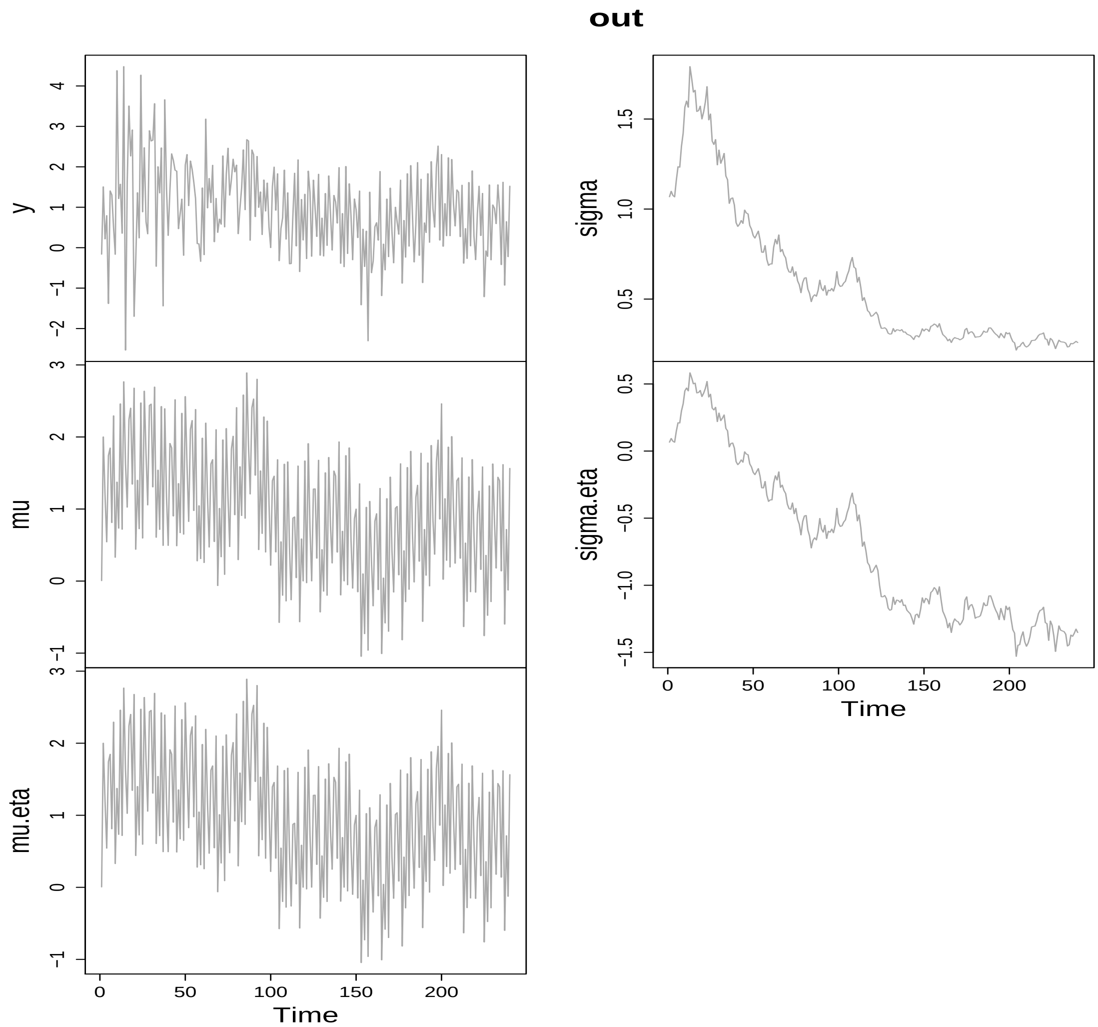


Figure 7.5: A GEST process simulation from a normal distribution with rw(1) local level and seasonal for the mean and rw(1) for log standard deviation.

7.4.2 GEST process with Poisson distribution

The Poisson distribution was defined by the French mathematician Simeon D. Poisson (1781-1840). The sample space of a Poisson distribution is the set of non-negative integers. More specifically, let Y be a random variable, which is Poisson distributed, denoted as $Y \sim PO(\mu)$. The probability density function of the random variable Y is defined as:

$$f(y) = P[Y = y] = e^{-\mu}\ \frac{\mu^y}{y!}\ , \qquad \mu > 0,$$

where $y = 0, 1, 2, ...$

The Poisson distribution plays a similar central role for discrete distributions as the normal distribution for continuous distributions. It is used when events occur randomly in a time interval or in a given space, in the analysis of count data.

From Lindsey (1993), a **Poisson Process**[2] is a stochastic process, where the response is treated as an independent Poisson variable, which is time varying. In the Poisson process the response variable is the count of events per unit of time that occur randomly and independently. The GEST Poisson process is doubly stochastic in discrete time.

[2]The properties that characterize a Poisson process are the following:

- The probability of an event in $(t, t + \delta_t)$ is $w\delta_t + o(\delta_t)$; that is, the probability of an event occurring in a small interval is proportional to its length
- The occurrence of events in $(t, t+\delta_t)$ is independent of what happens before t; the occurrence of events in disjoint intervals are independent
- The probability of more than one event occurring in $(t, t + \delta_t)$ is $o(\delta_t)$; the probability of re-occurrence of the event in a small interval is insignificant.

Note that the time intervals between events in a Poisson process follow the exponential distribution. Assume a random variable N_t counts the number of events in (0,t). This variable follows a Poisson distribution with mean $\mu = E(N_t) = w\,t$, where w is called intensity or rate and in this case is constant. If the rate w is time varying, i.e. $w(t)$, then it is referred as **nonhomogeneous Poisson process**. In addition, if $w(t)$ is stochastic it is called **doubly stochastic** or a **Cox Process**.

Let the parametric conditional distribution $\mathcal{D}$ of the response variable Y_t be the Poisson distribution and assume a log link for the mean (μ_t). The GEST Poisson process for Y_t with random walk local level is defined as:

$$
\begin{aligned}
Y_t|\mu_t &\sim PO(\mu_t) \\
\log(\mu_t) &= \beta_{1,0} + \gamma_{1,t} \\
\gamma_{1,t} &= \gamma_{1,t-1} + b_{1,t}
\end{aligned}
$$

Below is an example of the GEST stochastic process for 1000 random observations, generated by assuming that the conditional distribution $f_{Y_t}(y_t|\boldsymbol{\theta}_t)$ of the process is Poisson $PO(\mu_t)$. The log mean $(\log(\mu_t))$ of the distribution $PO(\mu_t)$ for $t = 1, 2, \ldots, T$ is simulated using a random walk order one process.

Figure 7.6 plots, for $t = 1, 2, \ldots, 1000$, the simulated process y_t, the simulated mean μ_t, and the simulated mean predictor $\eta_{1,t} = \log(\mu_t)$ (labelled mu.eta). The simulated process y_t is also fitted using the GEST model. Figure 7.8 shows the simulated mean μ_t in gray and the fitted μ_t in red.

Figure 7.8 shows the simulated time-varying mean, $\log(\mu_t)$, for the GEST Poisson process (gray line) and the fitted GEST model (red line) for this process.

The R commands for simulating Figure 7.6 is given in Appendix D, the resulting output from `plot=T` is given in Figure 7.6, and the fitted GEST process with the GSET model is given in Figure 7.8.

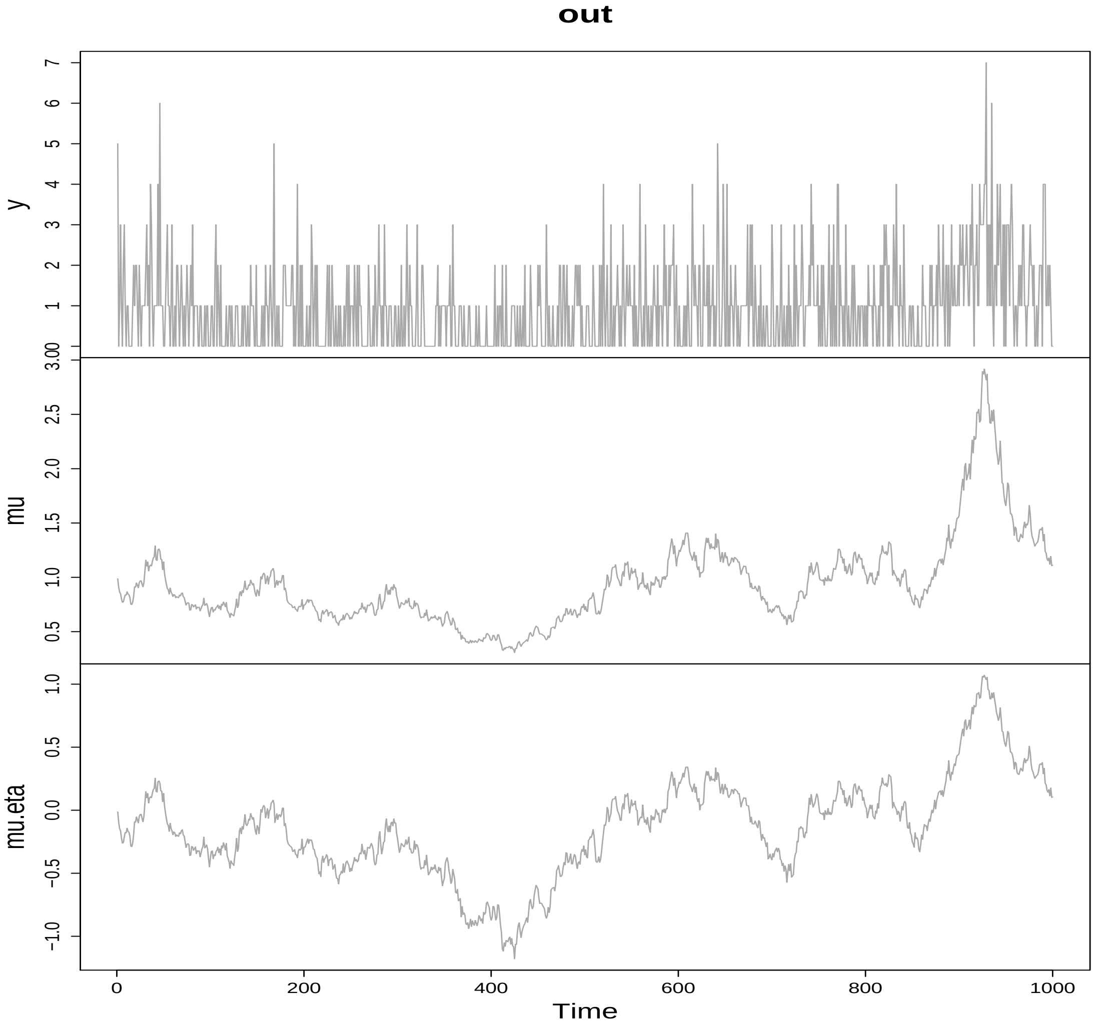


Figure 7.6: A GEST process simulation from a Poisson distribution

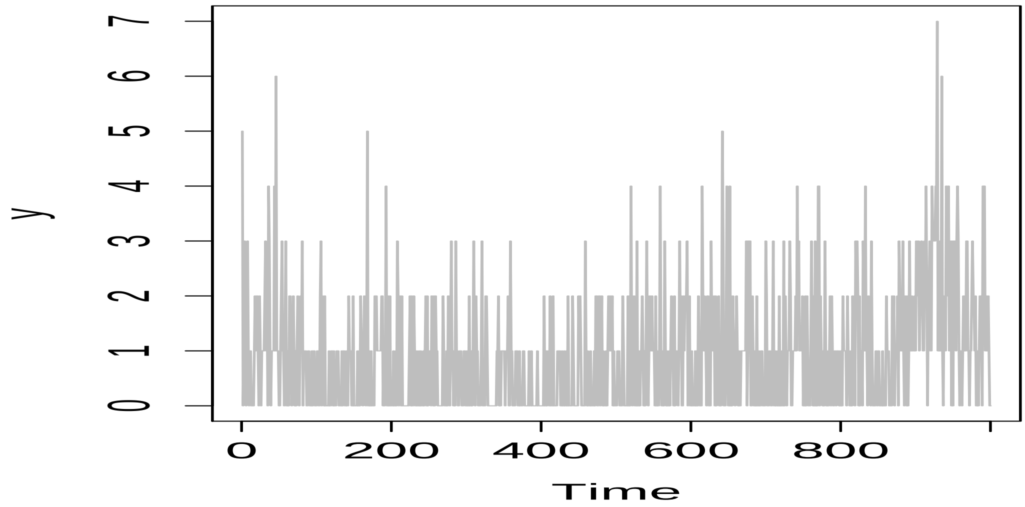


Figure 7.7: the simulated process y_t of the GEST Poisson process

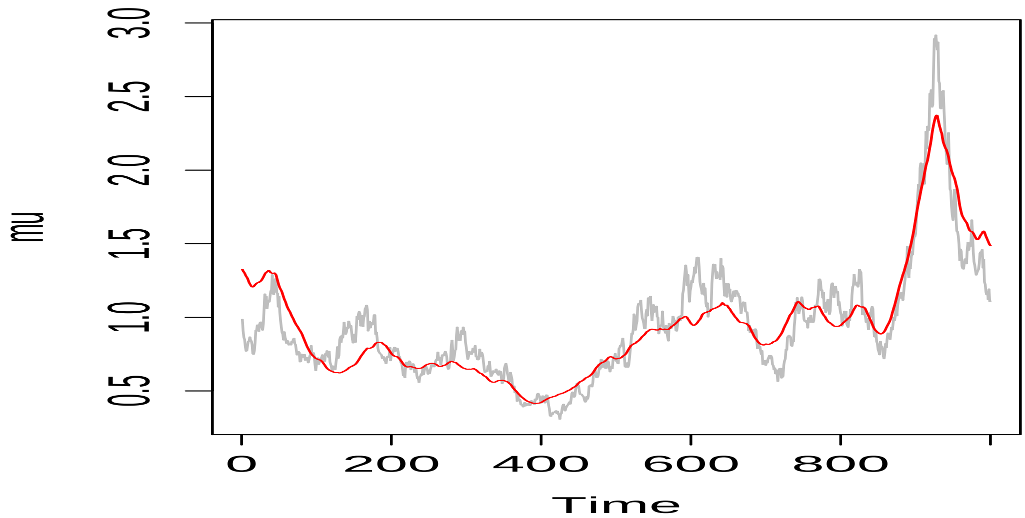


Figure 7.8: The simulated mean mu_t of y_t (in gray) and the fitted GEST model for the μ_t (in red).

7.4.3 GEST process with negative binomial type I distribution

The negative binomial type I distribution (NBI) is a mixed Poisson distribution obtained as the marginal distribution of Y when $Y|\delta \sim PO(\mu\delta)$ and $\delta \sim GA(1, \sigma^{\frac{1}{2}})$, i.e. δ has a gamma distribution with mean 1 and scale parameter $\sigma^{\frac{1}{2}}$ (and hence has dispersion σ). The negative binomial distribution can be highly positively skewed, unlike the Poisson distribution, which is close to symmetric for moderate μ and even closer as μ increases. The extra σ parameter allows the variance to change for a fixed mean, unlike the Poisson distribution for which the variance is fixed equal to the mean. Hence the negative binomial allows modelling of the variance as well as of the mean.

The probability function of the negative binomial distribution type I, denoted here as NBI(μ,σ), is given by

$$p_Y(y|\mu, \sigma) = \frac{\Gamma(y + \frac{1}{\sigma})}{\Gamma(\frac{1}{\sigma})\Gamma(y+1)} \left(\frac{\sigma\mu}{1+\sigma\mu}\right)^y \left(\frac{1}{1+\sigma\mu}\right)^{1/\sigma}$$

for $y = 0, 1, 2, ...$, where $\mu > 0$ and $\sigma > 0$.

The mean of Y is $E(Y) = \mu$ and the variance of Y is $Var(Y) = \mu + \sigma\mu^2$. (From Stasinopoulos *et al.* (2013), p 168 and p 222).

The sigma is the dispersion of the negative binomial type I distribution, if the sigma is equal to zero, the NBI is identical to Poisson, if the sigma is greater than zero, there is an over-dispersion in the data.

Let the parametric conditional distribution $\mathcal{D}$ of the response variable Y_t be the negative binomial type I distribution and assume a log link for the mean, μ_t, and a log link for σ_t.

The GEST process for Y_t with a random walk order 1 local level model for the log mean ($\log(\mu_t)$) and a random walk order 1 local level model for the log sigma ($\log(\sigma_t)$) is defined as:

$$
\begin{aligned}
Y_t|\mu_t, \sigma_t &\sim NBI(\mu_t, \sigma_t) \\
\log(\mu_t) &= \beta_{1,0} + \gamma_{1,t} \\
\log(\sigma_t) &= \beta_{2,0} + \gamma_{2,t}
\end{aligned}
$$

where

$$
\begin{aligned}
\gamma_{1,t} &= \gamma_{1,t-1} + b_{1,t} \\
\gamma_{2,t} &= \gamma_{2,t-1} + b_{2,t}
\end{aligned}
$$

Below is an example of the GEST stochastic process for 1000 random observations, generated by assuming that the conditional distribution $f_{Y_t}(y_t|\boldsymbol{\theta}_t)$ of the process is negative binomial type I, NBI(μ_t, σ_t). The log mean ($\log(\mu_t)$) and the log sigma ($\log(\sigma_t)$) of the NBI(μ_t, σ_t) distribution for $t = 1, 2, \ldots, T$, are generated by a random walk order one process.

Figure 7.9 plots, for $t = 1, 2, \ldots, 1000$, the simulated process y_t, the simulated mean μ_t, the simulated mean predictor $\eta_{1,t} = \log(\mu_t)$ (labelled mu.eta), the simulated standard deviation σ_t, and the simulated standard deviation predictor $\eta_{2,t} = \log(\sigma_t)$ (labelled sigma.eta).

Figure 7.10 shows the the simulated process y_t, the simulated μ_t in gray and the fitted μ_t in red, and the simulated σ_t in gray and the fitted σ_t in red.

The R commands for simulating and fitting Figure 7.9 are given in Appendix D, and the resulting output from `plot=T` is given in Figure 7.9.

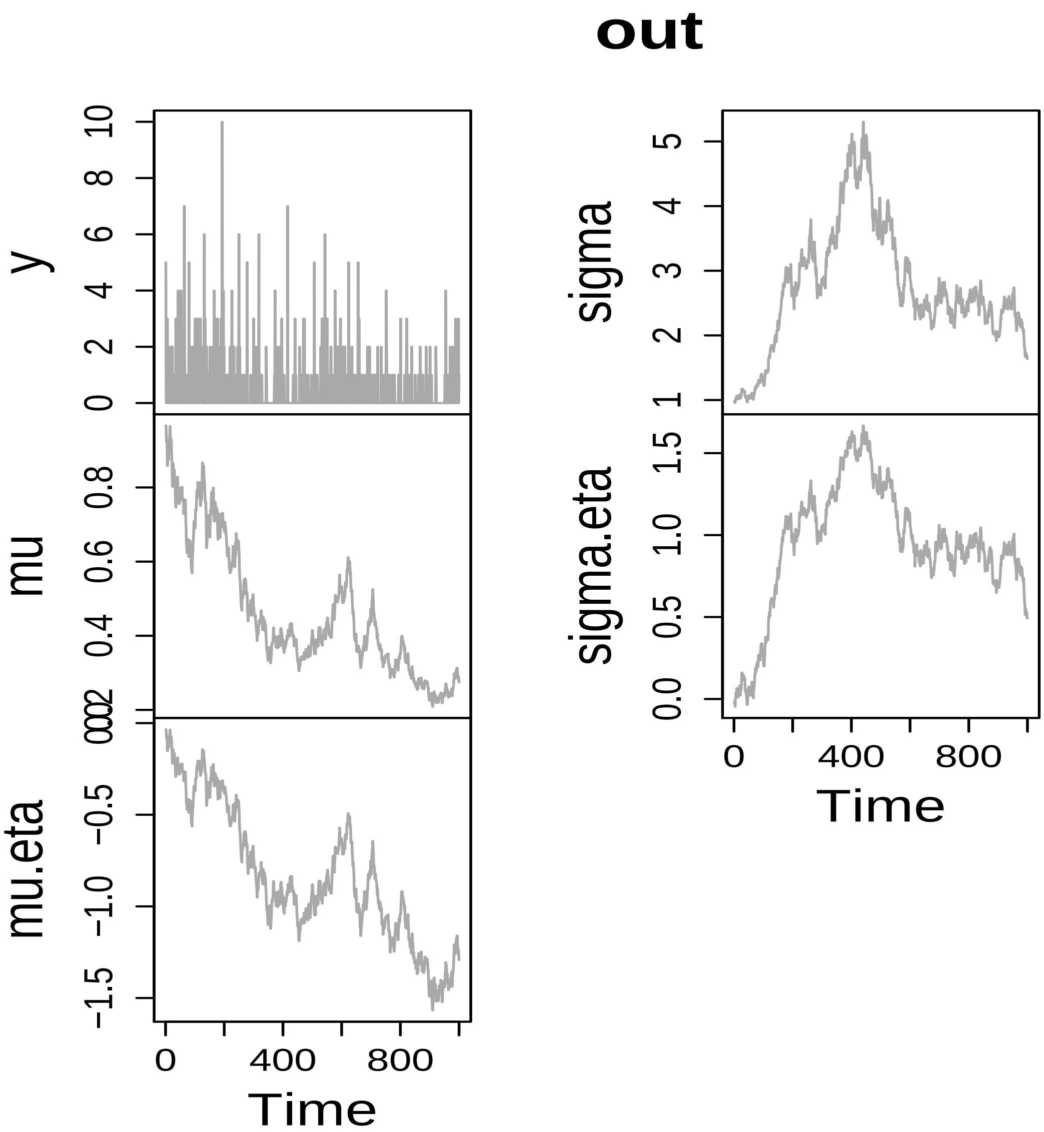


Figure 7.9: A GEST process simulation from a NBI distribution

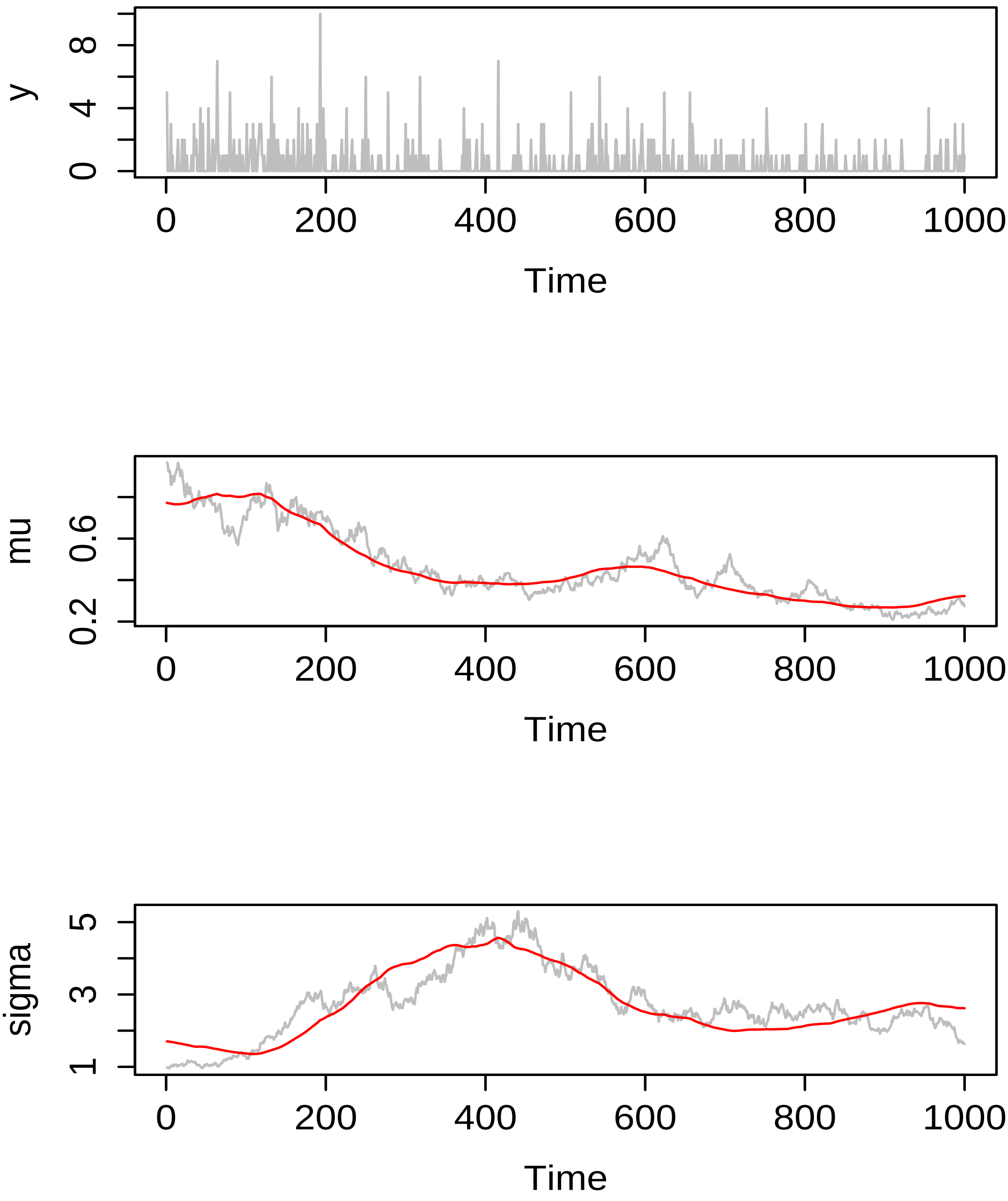


Figure 7.10: The actual simulation (in gray) for μ_t and σ_t for the GEST process and the fitted GEST model (in red) for the μ_t and σ_t.

7.4.4 GEST process with Student t distribution

The Student t family distribution is suitable for modelling leptokurtic data, that is, data with higher kurtosis than the normal distribution. The pdf of the t family distribution, denoted here as TF(μ,σ,ν), is defined by

$$f_Y(y|\mu,\sigma,\nu) = \frac{1}{\sigma B\left(\frac{1}{2},\frac{\nu}{2}\right)\nu^{\frac{1}{2}}}\left[1+\frac{(y-\mu)^2}{\sigma^2\nu}\right]^{-\frac{\nu+1}{2}}$$

for $-\infty < y < \infty$, where $-\infty < \mu < \infty$, $\sigma > 0$ and $\nu > 0$, where $B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$ is the beta function. The mean and variance of Y are given by $E(Y) = \mu$ and $Var(Y) = \sigma^2\nu/(\nu-2)$ when $\nu > 2$. (From Stasinopoulos *et al.* (2013), p 205).

Rigby and Stasinopoulos (2012) reparameterized the Student t family distribution, such that the mean is μ and the standard deviation is σ for degrees of freedom parameter $\nu(> 2)$. The reparameterized Student t family distribution is denoted here as TF2(μ,σ,ν).

Let the parametric conditional distribution $\mathcal{D}$ of the response variable Y_t be the reparameterized Student t family, TF2, distribution and assume an identity link for the mean (μ_t), a log link for the standard deviation (σ_t) and a log shifted link the degrees of freedom (ν_t). The time varying degrees of freedom is the kurtosis signal of the Student t distributional process. If the kurtosis is higher than 30, then the TF process looks the same as a Gaussian process.

The GEST process for Y_t with a random walk order 1 local level model for the mean (μ_t), the log sigma ($\log(\sigma_t)$) and the log shifted degrees of freedom ($\log(\nu_t-2)$) is defined as:

$$\begin{aligned}
Y_t|\mu_t, \sigma_t, \nu_t &\sim TF(\mu_t, \sigma_t, \nu_t)\\
\mu_t &= \beta_{1,0} + \gamma_{1,t}\\
\log(\sigma_t) &= \beta_{2,0} + \gamma_{2,t}\\
\log(\nu_t - 2) &= \beta_{3,0} + \gamma_{3,t}
\end{aligned}$$

where

$$\begin{aligned}
\gamma_{1,t} &= \gamma_{1,t-1} + b_{1,t}\\
\gamma_{2,t} &= \gamma_{2,t-1} + b_{2,t}\\
\gamma_{3,t} &= \gamma_{3,t-1} + b_{3,t}.
\end{aligned}$$

Below is an example of the GEST stochastic process for 1000 random observations, generated by assuming that the conditional distribution $f_{Y_t}(y_t|\boldsymbol{\theta}_t)$ of the process is Student t, TF2(μ_t, σ_t, ν_t). The mean (μ_t), log sigma $(\log(\sigma_t))$ and log shifted degrees of freedom $(\log(\nu_t - 2))$ of the distribution TF2(μ_t, σ_t, ν_t), for $t = 1, 2, \ldots, T$, are simulated with a random walk order one process. Note that the link function $\log(\nu_t - 2)$ is used because, for the TF2 distribution, $\nu > 2$, ensures it has a finite mean μ_t and finite standard deviation σ_t.

The R commands for simulating Figure 7.11 is given in Appendix D, and the resulting output from `plot=T` is given in Figure 7.11.

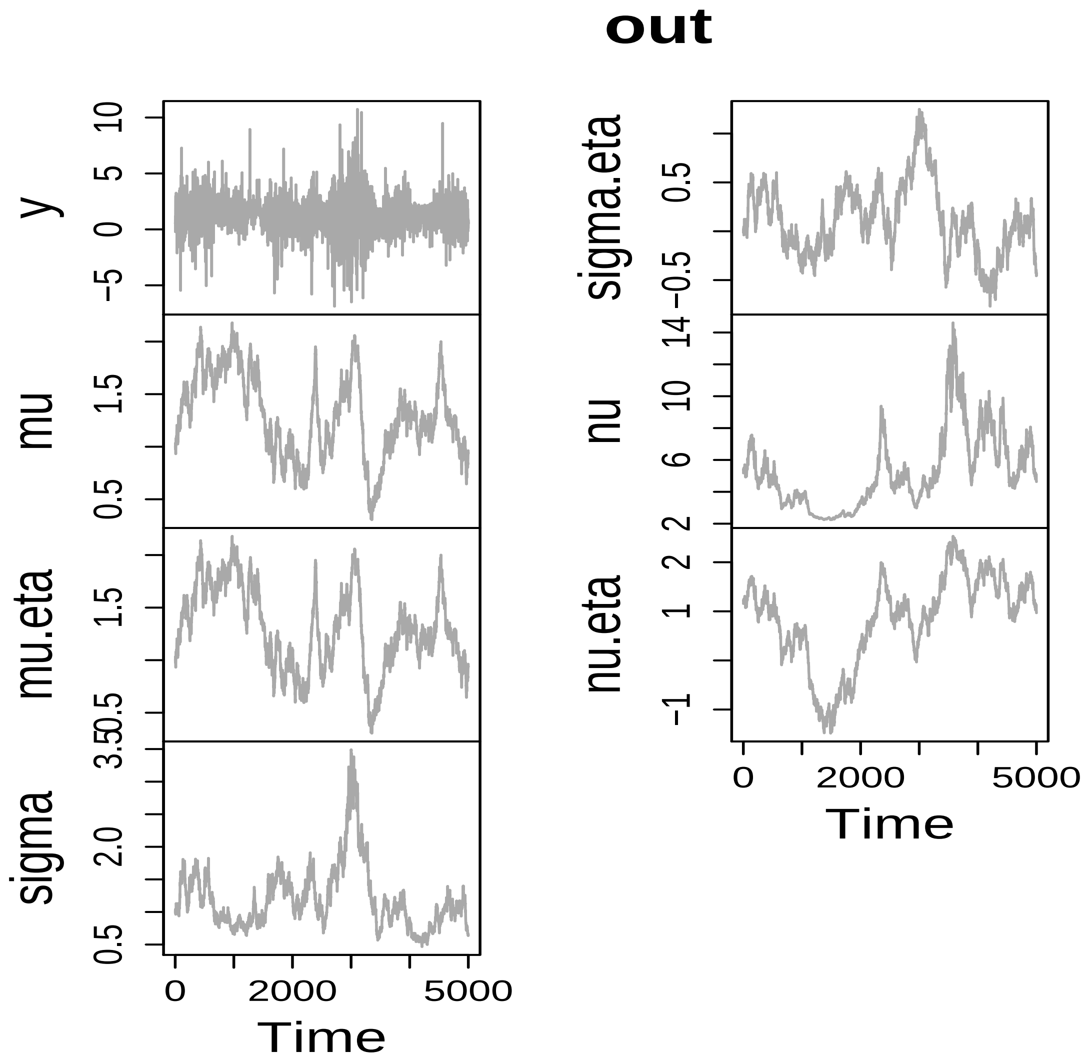


Figure 7.11: A GEST process simulation from a TF2 distribution.

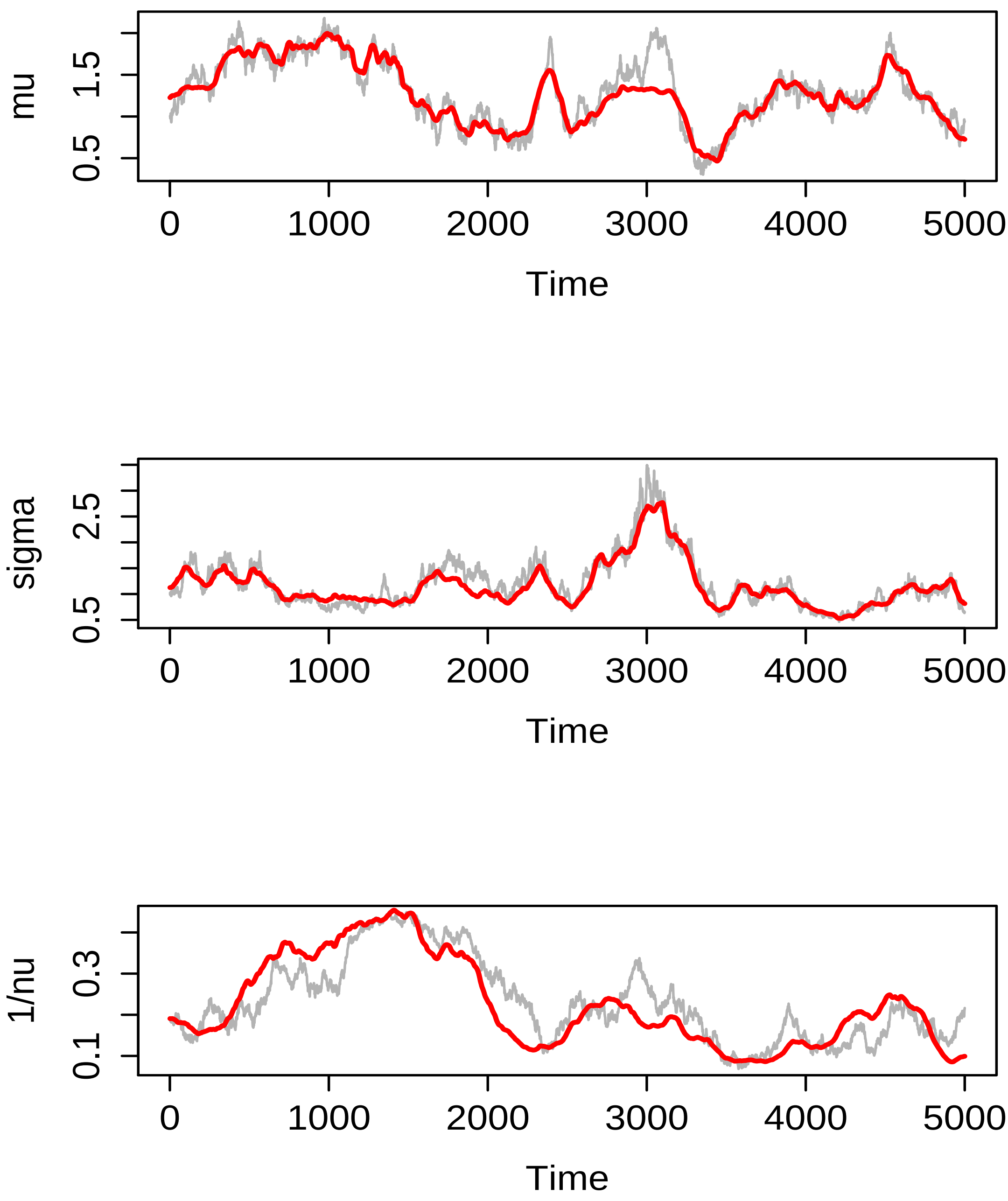


Figure 7.12: The simulated μ_t, σ_t, and $1/\nu_t$ (in gray) of the GEST process and the fitted GEST model (in red) for μ_t, σ_t, $1/\nu_t$ using a TF2 distribution.

7.4.5 GEST process with skew Student t distribution

The skew Student t distribution, SST, is one of the appropriate distribution to model data with positive or negative skewness and high or low kurtosis, where the tails are heavier then the Gaussian distribution. The SST(μ, σ, ν, τ) has mean μ, standard deviation σ, skewness parameter ν, and kurtosis parameter $\tau (> 2)$. (See Appendix B for model detail).

Let the parametric conditional distribution $\mathcal{D}$ of the response variable Y_t be the skew Student t distribution, SST$(\mu_t, \sigma_t, \nu_t, \tau_t)$, and assume an identity link for the mean (μ_t), a log link for the standard deviation $(\log(\sigma_t))$, a log link for the skewness parameter $(\log(\nu_t))$, and a log shifted link for the kurtosis parameter $(\log(\tau_t - 2))$. The GEST process for Y_t with a random walk order 1 local level model for the mean, log sigma, log skewness parameter and log shifted degrees of freedom parameter is defined as:

$$
\begin{aligned}
Y_t|\mu_t, \sigma_t, \nu_t, \tau_t &\sim SST(\mu_t, \sigma_t, \nu_t, \tau_t) \\
\mu_t &= \beta_{1,0} + \gamma_{1,t} \\
\log(\sigma_t) &= \beta_{2,0} + \gamma_{2,t} \\
\log(\nu_t) &= \beta_{3,0} + \gamma_{3,t} \\
\log(\tau_t - 2) &= \beta_{4,0} + \gamma_{4,t}
\end{aligned}
$$

where

$$
\begin{aligned}
\gamma_{1,t} &= \gamma_{1,t-1} + b_{1,t} \\
\gamma_{2,t} &= \gamma_{2,t-1} + b_{2,t} \\
\gamma_{3,t} &= \gamma_{3,t-1} + b_{3,t} \\
\gamma_{4,t} &= \gamma_{4,t-1} + b_{4,t}.
\end{aligned}
$$

Below we generate an example of a GEST stochastic process by assuming that the $f_{Y_t}(y_t|\boldsymbol{\theta}_t)$ of the process is a skew Student t, $\text{SST}(\mu_t, \sigma_t, \nu_t, \tau_t)$. We simulate each of the predictors of the distribution parameters of the $\text{SST}(\mu_t, \sigma_t, \nu_t, \tau_t)$, for $t = 1, 2, \ldots, T$, using a random walk order one process. Note the link function $\log(\tau_t - 2)$ is used because, for the SST distribution, $\tau > 2$, ensures it has a finite mean μ_t and finite standard deviation σ_t. The initial values of the distribution parameters were $\beta_{1,0} = 0$, $\beta_{2,0} = 1$, $\beta_{3,0} = 1$, $\beta_{4,0} = 5$ and the variances of the $b_{k,t}$ innovations were chosen to be $\sigma^2_{b_1} = 0.0001$, $\sigma^2_{b_2} = 0.0009$, $\sigma^2_{b_3} = 0.0004$, and $\sigma^2_{b_4} = 0.0004$.

Figure 7.13 shows the simulated process y_t and the generated time-varying mean μ_t, time-varying standard deviation σ_t, time-varying skewness parameter ν_t, and time-varying kurtosis parameter τ_t. [The reciprocal of τ_t can be plotted for clarity, because very large values of τ_t in the plot make it difficult to see smaller values of τ_t]. Note for the SST distribution $\nu_t < 1$ produces a negatively skewed distribution, while $\nu_t > 1$ produces a positively skewed distribution. The kurtosis increases as $\tau_t > 2$ decreases or $1/\tau_t$ increases.

Figure 7.14 shows the generated (black line) time-varying mean μ_t, time-varying standard deviation σ_t, time-varying skewness parameter ν_t, and time-varying reciprocal of the kurtosis parameter $1/\tau_t$. [The reciprocal of τ_t is plotted for clarity, because very large values of τ_t in the plot make it difficult to see smaller values of τ_t]. It also shows the fitted GEST process (red lines) estimated using the GEST model introduced in chapter 8.

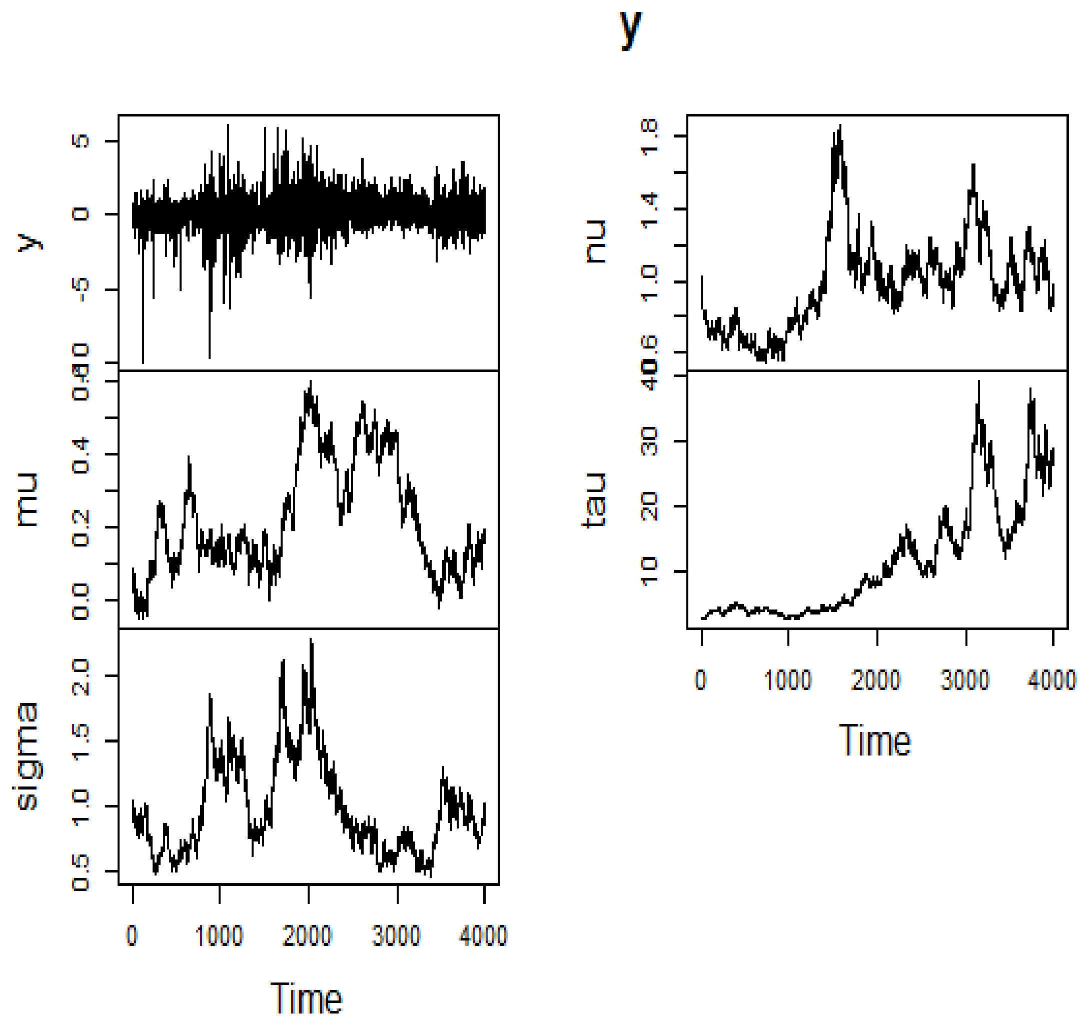
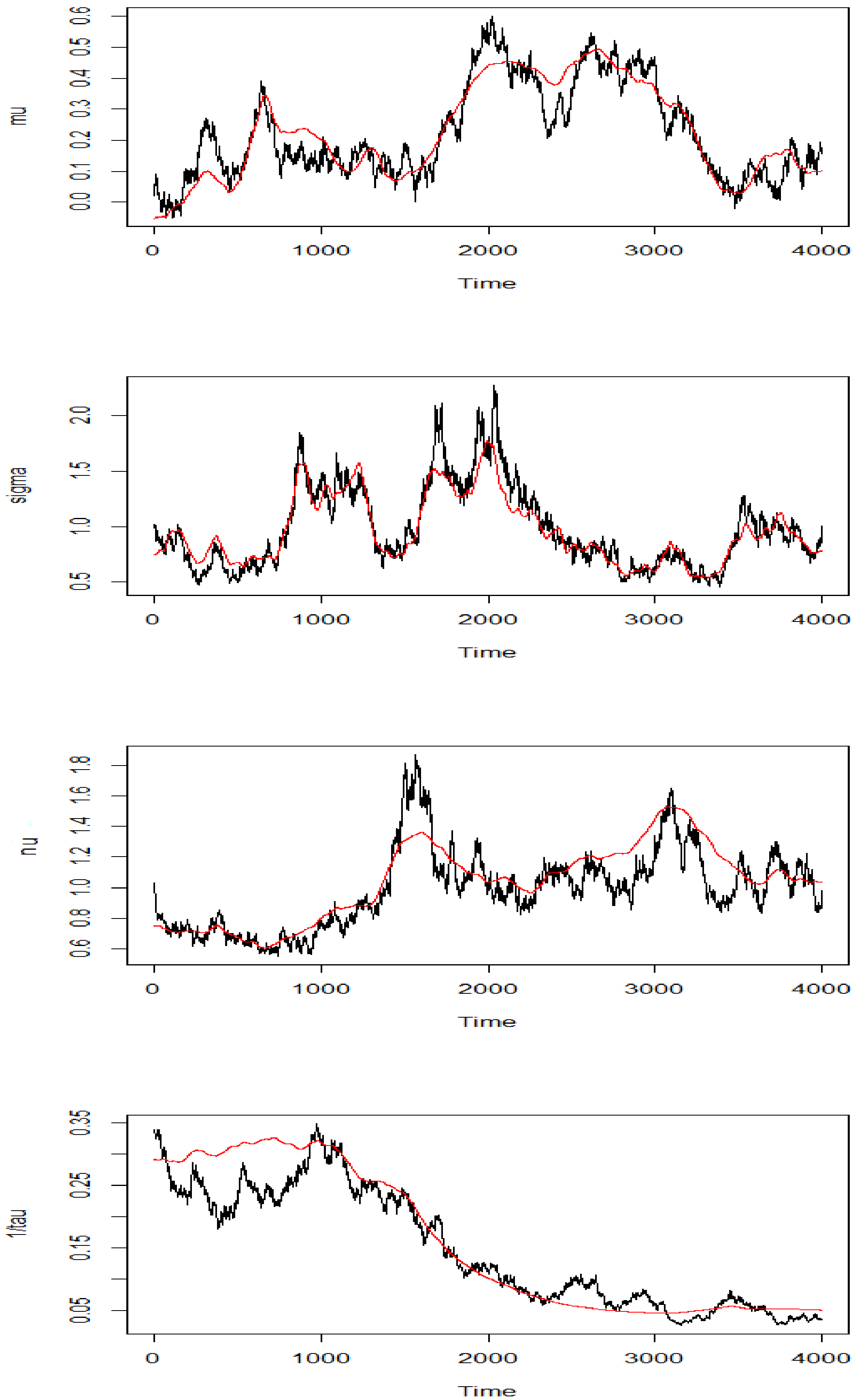


Figure 7.13: A GEST process simulation from a skew Student t-distribution (SST)



s

Figure 7.14: The actual realisations (in black) for μ_t, σ_t, ν_t and $1/\tau_t$ for the GEST process and the fitted GEST model (in red) for μ_t, σ_t, ν_t and $1/\tau_t$.

The `gest.sim()` function

The following function is the simulation function of the GEST process in R. The function `gest.sim()` produces a simulation of Gaussian and non-Gaussian structural time series data. It allows any of the 80 distributions available in the `gamlss` package in R (Stasinopoulos and Rigby, 2007), with up to four distribution parameters, denoted μ, σ, ν and τ. It simulates the predictor for the parameters, μ_t, σ_t, ν_t and τ_t (usually representing location, scale, skewness and kurtosis parameters), as a random walk or autoregressive process, including optionally seasonality. It generates a data for a given conditional distribution, after simulating its parameters.

The general form of the function is

```
gest.sim(N=1000, family=NO, mu.type= c("level", "AR", "levelSeasonal",
"ARSeasonal", "Seasonal"), sigma.type= c("level", "AR", "levelSeasonal",
"ARSeasonal", "Seasonal"), nu.type= c("level", "AR", "levelSeasonal", "AR-
Seasonal", "Seasonal"), tau.type= c("level", "AR", "levelSeasonal", "AR-
Seasonal", "Seasonal"), mu.init=1, sigma.init=1, nu.init=1, tau.init=1,
mu.sigb=.01, sigma.sigb=.01, nu.sigb=.01, tau.sigb=.01, mu.sigS=.01,
sigma.sigS=.01, nu.sigS=.01, tau.sigS=.01, mu.phi=.5, sigma.phi=.5, nu.phi=.5,
tau.phi=.5, mu.order=1, sigma.order=1, nu.order=1, tau.order=1, frequency=12,
mu.Sinit=NULL, sigma.Sinit=NULL, nu.Sinit=NULL, tau.Sinit=NULL, plot=FALSE,
main=NULL)
```

The function `gest.sim()` has multiple options for each simulated parameter: the type of the process for the predictor of each parameter $(\mu_t, \sigma_t, \nu_t, \tau_t)$, if it is a random walk, autoregressive (ar), seasonal, random walk and seasonal, or autoregressive and seasonal process (e.g `mu.type`); the conditional distribution of the observations (`family`); initial values for the parameter predictors (e.g. `mu.init`); the true values of the standard deviations for the innovations of the predictors of the parameters, (e.g. `mu.sigb` for local level for the predictor of the mean, and `mu.sigS` for the seasonal for the predictor of the mean); the true value for the ar process parameter ϕ for the parameter predictors (e.g. `mu.phi`); the order of the random walk or ar process for the parameter predictor (e.g. `mu.order`).

Chapter 8

GEST model and estimation

8.1 Introduction

This Chapter introduces the statistical framework of the GEST model, defines the maximum likelihood estimation methods globally and locally, and gives the GEST algorithm. Also, it introduces a general framework for modelling univariate time series.

The Generalized Structural (GEST) time series model is a univariate parameter-driven model for non-Gaussian time series. It extends the univariate Gaussian structural time series models to a flexible non-Gaussian structural framework, with the potential of modelling variety of phenomena, including continuous or discrete variables with possibly a positive or negative skewness and/or high or low kurtosis.

The GEST model assumes a parametric conditional distribution for the response variable given the past, and allows some or all of the predictors of the distribution parameters to vary stochastically, resulting in a general stochastic model.

The dependent variable, conditional on the past history of the variable, is allowed to come from a parametric distribution with up to four parameters, often representing the location (e.g. mean), scale (e.g. standard deviation), skewness and kurtosis respectively. The predictors for the distribution parameters are modelled jointly and explicitly by a structural model and/or a constant, linear, non-linear, smooth non-parametric, or varying coefficient model to account for the effect of explanatory variables.

The GEST model extends the generalized additive model for location, scale and shape (GAMLSS) model, Rigby and Stasinopoulos (2005), to focus on structural time series modelling. Its applications include modelling time series counts (e.g. discrete counts) using for example a negative binomial conditional distribution, including structural models for the location and/or scale of the distribution, and modelling continuous time series data using for example a skew Student t conditional distribution including structural models for the location, scale, skewness and kurtosis distribution parameters.

Section 8.2 defines the GEST model and presents its characteristics. Section 8.3 provides two methods of estimation of the GEST model's hyperparameters. Section 8.4 explains in detail the local estimation of the hyperparameters. Section 8.5 derives the effective degrees of freedom of the GEST model.

8.2 The GEST model

The GEST model assumes that, conditional of the past, the response variable Y comes from a parametric distribution with probability (density) function $f_Y(y|\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a vector of unknown distribution parameters. The distribution parameter vector $\boldsymbol{\theta}$ is restricted to at most four parameters denoted $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4) = (\mu, \sigma, \nu, \tau)$, where μ is in general a location parameter, σ a scale parameter, and ν and τ are shape parameters (often affecting the skewness and kurtosis respectively).

Each of the distribution parameters (μ, σ, ν, τ) is modelled by a structural time series model and/or linear, non-linear or smooth non-parametric models to account for explanatory variables. Each structural model is a random walk or autoregressive model (not limited to order one), and/or a seasonal effect.

Definition: *Let Y_t be the response variable for $t = 1, 2, \ldots, T$ then the GEST model is defined as:*

$$
\begin{aligned}
Y_t|\mu_t, \sigma_t, \nu_t, \tau_t &\sim \mathcal{D}(\mu_t, \sigma_t, \nu_t, \tau_t) \\
g_1(\mu_t) = \eta_{1,t} &= \mathbf{x}_{1,t}^\top \boldsymbol{\beta}_1 + \gamma_{1,t} \\
g_2(\sigma_t) = \eta_{2,t} &= \mathbf{x}_{2,t}^\top \boldsymbol{\beta}_2 + \gamma_{2,t} \\
g_3(\nu_t) = \eta_{3,t} &= \mathbf{x}_{3,t}^\top \boldsymbol{\beta}_3 + \gamma_{3,t} \\
g_4(\tau_t) = \eta_{4,t} &= \mathbf{x}_{4,t}^\top \boldsymbol{\beta}_4 + \gamma_{4,t}
\end{aligned}
\tag{8.1}
$$

for $t = 1, 2, \ldots, T$, where $\mathcal{D}$ represents the conditional distribution of the response variable, g_k is a known link function (e.g., identity or log link function), $\boldsymbol{\beta}_k$ is a parameter vector of length p_k and the $\mathbf{x_{k,t}}$ are explanatory variable vectors and the $\gamma_{k,t}$ for $k = 1, 2, 3, 4$ are defined as:

$$
\gamma_{k,t} = \sum_{j=1}^{J_k} \phi_{k,j} \gamma_{k,t-j} + b_{k,t}, \tag{8.2}
$$

for $t = J_k + 1, J_k + 2, \ldots, T$, where $b_{k,t}$ are random errors, independent from each other mutually and serially, and normally distributed with expected values equal to zero and variance $\sigma^2_{b_k}$, thus $\mathbf{b_k} \sim N_{n-J_k}(0, \sigma^2_{b_k} \mathbf{I}_{n-J_k})$, where $\mathbf{b_k}^\top = (b_{k,J_k+1}, \ldots, b_{k,T})$ for $k = 1, 2, \ldots, K$.

Characteristic of the GEST model

Regarding the GEST model, it is important to note that:

- The response variable distribution $\mathcal{D}(\mu_t, \sigma_t, \nu_t, \tau_t)$ can be any continuous or discrete distribution.

- Typically the linear term $\mathbf{x}_{k,t}^\top \boldsymbol{\beta}_k$ could include the constant, continuous or categorical explanatory variables and possibly a linear term in time or a fixed seasonal effect, for $k = 1, 2, 3, 4$.

- The explanatory variables can be different for each distribution parameter μ_t, σ_t, ν_t and τ_t.

- To account for non-linearities in the relationship between the parameters of the distribution and the explanatory variables, model (8.1) can be extended to include non-linear and smooth non-parametric models for the distribution parameters μ_t, σ_t, ν_t and τ_t as in Rigby and Stasinopoulos (2005). For example equations in (8.1) can be amended to $\eta_{k,t} = \mathbf{x}_{k,t}^\top \boldsymbol{\beta}_k + \sum_{j=1}^{J_k} s_j(x_j) + \gamma_{k,t}$ where the $s_j()$ are smooth functions e.g. P-splines of Eilers and Marx (1996).

- A distribution parameter model can be extended to include a seasonal effect (with M seasons)

$$g_k(\theta_{k,t}) = \eta_{k,t} = \mathbf{x}_{k,t}^\top \boldsymbol{\beta}_k + \gamma_{k,t} + s_{k,t}$$

  where $\gamma_{k,t}$ is given by (8.2) and

$$s_{k,t} = -\sum_{m=1}^{M-1} s_{k,t-m} + w_t.$$

- The random effects (8.2) $\gamma_{k,t}$ can be extended to include *persistent* explanatory variable effects,

$$\gamma_{k,t} = \sum_{j=1}^{J_k} \phi_{k,j}\gamma_{k,t-j} + v_{k,t}^{\top}\delta_k + b_{k,t} \tag{8.3}$$

  where δ_k is a parameter vector of length q_k and explanatory variable vector $v_{k,t}$ is of length q_k. This term is used in the analysis of the S&P 500 stock index returns, (see Chapter 9), for modelling the leverage effect using asymmetric stochastic volatility (see for example Asai and McAleer, 2005; Omori *et al.*, 2007).

- The GEST model integrates regression-type and time-series-type models for all the distribution parameters (μ_t, σ_t, ν_t and τ_t) of the assumed parametric conditional distribution $\mathcal{D}$ of the response variable, allowing the location, scale, skewness and kurtosis parameters of the conditional distribution $\mathcal{D}$ to change over time. Also the distribution $\mathcal{D}$ can be any parametric (continuous or discrete) distribution and is not necessarily restricted to the assumption of the exponential family distribution.

Below two examples of the GEST model are provided by specifying two different distributions, namely the Gaussian distribution and the skew Student t distribution.

Example 1: Conditional normal distribution

In the GEST model let the parametric conditional distribution $\mathcal{D}$ of the response variable Y_t be the Gaussian distribution and assume an identity link for the mean (μ_t) and a log link for the standard deviation (σ_t). Then the GEST model for Y_t is defined as:

$$
\begin{aligned}
Y_t|\mu_t, \sigma_t &\sim NO(\mu_t, \sigma_t) \\
\mu_t &= \mathbf{x}_{1,t}^\top \boldsymbol{\beta}_1 + \gamma_{1,t} \\
\log(\sigma_t) &= \mathbf{x}_{2,t}^\top \boldsymbol{\beta}_2 + \gamma_{2,t}
\end{aligned}
$$

where the unobserved state, $\gamma_{k,t}$ for $k = 1, 2$, is defined by equation (8.2). For a constant mean model $\mu_t = b_1$, we have the standard stochastic volatility model. Clearly the Gaussian GEST model does not allow explicit modelling of the conditional skewness and kurtosis since the conditional skewness and kurtosis of the Gaussian GEST model above are constants.

Example 2: Conditional skew Student t distribution

To allow explicit modelling of skewness and kurtosis of Y_t we need to assume a more flexible distribution with more than two parameters. The skew Student t distribution, page 269, denoted here as $SST(\mu_t, \sigma_t, \nu_t, \tau_t)$, has four parameters, μ_t is exactly the mean, $\sigma_t > 0$ is exactly the standard deviation, while $\nu_t > 0$ controls the skewness ($\nu_t < 1$ implies negative skewness, while $\nu_t > 1$ implies positive skewness) and $\tau_t > 2$ controls the kurtosis (a lower τ_t implies heavier tails). Hence the SST distribution allows modelling of the mean, standard deviation, skewness and kurtosis. The SST distribution is widely used in financial data analysis but its definition and derivation are more obscure (the reader can consult Appendix B on page 269 for the definition and the derivation of the SST distribution).

It is beneficial to have a skew Student t distribution parametrization such that μ_t is the mean and σ_t is the standard deviation, so that the changes in the mean and standard deviation can be interpreted separately from the changes in the shape of the distribution (resulting from changes in ν_t and τ_t).

Assuming an identity link for the mean (μ_t) and log links for the standard deviation (σ_t), skewness parameter (ν_t), and a log shifted link for the kurtosis parameter (τ_t) of the conditional SST distribution of the response variable Y_t, the GEST model for Y_t is defined as:

$$\begin{aligned} Y_t|\mu_t, \sigma_t, \nu_t, \tau_t &\sim SST(\mu_t, \sigma_t, \nu_t, \tau_t) \\ \mu_t &= \mathbf{x}_{1,t}^\top \boldsymbol{\beta}_1 + \gamma_{1,t} \\ \log(\sigma_t) &= \mathbf{x}_{2,t}^\top \boldsymbol{\beta}_2 + \gamma_{2,t} \\ \log(\nu_t) &= \mathbf{x}_{3,t}^\top \boldsymbol{\beta}_3 + \gamma_{3,t} \\ \log(\tau_t - 2) &= \mathbf{x}_{4,t}^\top \boldsymbol{\beta}_4 + \gamma_{4,t} \end{aligned}$$

where the unobserved state, $\gamma_{k,t}$ for $k = 1, 2, 3, 4$, is defined by equation (8.2).

8.3 Estimation of the GEST model

8.3.1 Introduction

The GEST model, defined by equation (8.1), has four distinct sets of parameters:

(a) $\boldsymbol{\beta}^\top = \left(\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top, \boldsymbol{\beta}_3^\top, \boldsymbol{\beta}_4^\top\right)$ the constants or coefficients of the covariates,

(b) $\boldsymbol{\gamma}^\top = \left(\boldsymbol{\gamma}_1^\top, \boldsymbol{\gamma}_2^\top, \boldsymbol{\gamma}_3^\top, \boldsymbol{\gamma}_4^\top\right)$ the structural terms or the random effect vectors,

(c) $\boldsymbol{\phi}^\top = \left(\boldsymbol{\phi}_1^\top, \boldsymbol{\phi}_2^\top, \boldsymbol{\phi}_3^\top, \boldsymbol{\phi}_4^\top\right)$, the AR coefficients for autoregressive structural terms,

(d) $\boldsymbol{\sigma}_b^\top = (\sigma_{b_1}, \sigma_{b_2}, \sigma_{b_3}, \sigma_{b_4})$, the standard deviations of the normal errors $b_{k,t}$ for $k = 1, 2, 3, 4$, in the structural terms,

where $\boldsymbol{\phi}$ and $\boldsymbol{\sigma}_b$ are referred to as the *hyperparameters* of the GEST model.

GEST algorithm for estimation of $\boldsymbol{\beta}$ and γ given fitted hyperparameters

---

(A) initialise $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \boldsymbol{\theta}_4) = (\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau})$, and set initial $\boldsymbol{\gamma}_k = 0$ for $k = 1, 2, 3, 4$

(B) start the *outer cycle* in order to fit each of the distribution parameter vectors $\boldsymbol{\theta}_k$, for $k = 1, 2, 3, 4$ sequentially until convergence [where $\boldsymbol{\theta}_1 = \boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_T)^\top$, $\boldsymbol{\theta}_2 = \boldsymbol{\sigma}$, $\boldsymbol{\theta}_3 = \boldsymbol{\nu}$, $\boldsymbol{\theta}_4 = \boldsymbol{\tau}$],

(a) start the following *inner cycle* (or "local scoring") for each iteration of the outer cycle in order to fit one of the distribution parameter vectors, $\boldsymbol{\theta}_k$

(i) evaluate the current *iterative response variable* $\mathbf{z}_k$ and current *iterative weights* $\mathbf{W}_k$ (where $\mathbf{z}_k = \boldsymbol{\eta}_k + \mathbf{W}_k^{-1}\mathbf{u}_k$, $\mathbf{W}{=}-\frac{\partial^2 \ell}{\partial \eta \partial \eta^\top}$, or $-E\left[\frac{\partial^2 \ell}{\partial \eta \partial \eta^\top}\right]$ or $\left(\frac{\partial \ell}{\partial \eta}\right)^2$, and $\mathbf{u} = \frac{\partial \ell}{\partial \eta_k}$).

(ii) start the Gauss-Seidel (or "backfitting") algorithm

(I) estimate $\boldsymbol{\beta}_k$ by regressing the current partial residuals $\boldsymbol{\epsilon}_{0k} = \mathbf{z}_k - \boldsymbol{\gamma}_k$ against design matrix $\mathbf{X}_k$ using current weights $\mathbf{W}_k$.

(II) estimate the hyperparameters σ_b^2 and $\boldsymbol{\phi}$ by maximising their local likelihood function Q, and then estimate $\boldsymbol{\gamma}_k$ using the equation $\hat{\gamma}_k = \left[\boldsymbol{\Sigma}^{-1} + \sigma_b^{-2}\mathbf{D}^\top\mathbf{D}\right]^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}$.

(iii) end the Gauss-Seidel algorithm on convergence of $\boldsymbol{\beta}_k$ and $\boldsymbol{\gamma}_k$

(iv) update $\boldsymbol{\theta}_k$ and $\boldsymbol{\eta}_k = g(\boldsymbol{\theta}_k)$.

(b) end the inner cycle on convergence of $\boldsymbol{\theta}_k$.

(C) end the outer cycle when the global deviance $(= -2 * l)$ of the estimated model converges.

It is important to emphasise here that the outer cycle fits a specific distribution parameter vector (e.g. $\boldsymbol{\mu}$), by fixing the other distribution parameter vectors (e.g. $\boldsymbol{\sigma}, \boldsymbol{\nu}$ and $\boldsymbol{\tau}$) to their current maximum values, and the inner cycle uses a "local scoring" or Newton algorithm resulting in an iterative reweighted backfitting. Furthermore, the Gauss-Seidel algorithm in (B)(a)(ii) above is called the "backfitting" algorithm by Hastie and Tibshirani (1990) and Hastie *et al.* (2009).

The joint distribution for all the components of the GEST model is derived by using the conditional probability law:

$$p(A|B) = \frac{p(A \cap B)}{p(B)}$$

where $p(A|B)$ is the conditional probability, $p(A \cap B)$ is the joint probability and $p(B)$ is the marginal probability.

Hence,

$$p(A \cap B) = p(A|B) * p(B). \tag{8.4}$$

The joint distribution for $(\mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\gamma})$ and $(\boldsymbol{\phi}, \boldsymbol{\sigma}_b)$ is given by:

$$\begin{aligned} f(\mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\phi}, \boldsymbol{\sigma}_b) &= f(\mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\gamma}|\boldsymbol{\phi}, \boldsymbol{\sigma}_b) * f(\boldsymbol{\phi}, \boldsymbol{\sigma}_b) \\ &= f(\mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\gamma}|\boldsymbol{\phi}, \boldsymbol{\sigma}_b) * f(\boldsymbol{\phi}) * f(\boldsymbol{\sigma}_b) \end{aligned} \tag{8.5}$$

assuming $\boldsymbol{\phi}$ and $\boldsymbol{\sigma}_b$ have independent priors. Applying the equation (8.4) to (8.5) and assuming that $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ have independent priors (given $\boldsymbol{\phi}$ and $\boldsymbol{\sigma}_b$), implies the following:

$$\begin{aligned}
f(\mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\phi}, \boldsymbol{\sigma}_b) &= f(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\phi}, \boldsymbol{\sigma}_b) * f(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\phi}, \boldsymbol{\sigma}_b) \\
&= f(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\phi}, \boldsymbol{\sigma}_b) * f(\boldsymbol{\beta}, \boldsymbol{\gamma}|\boldsymbol{\phi}, \boldsymbol{\sigma}_b) * f(\boldsymbol{\phi}) * f(\boldsymbol{\sigma}_b) \\
&= f(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\gamma}) * f(\boldsymbol{\beta}|\boldsymbol{\gamma}, \boldsymbol{\phi}, \boldsymbol{\sigma}_b) * f(\boldsymbol{\gamma}|\boldsymbol{\phi}, \boldsymbol{\sigma}_b) * f(\boldsymbol{\phi}) * f(\boldsymbol{\sigma}_b) \\
&= f(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\gamma}) * f(\boldsymbol{\beta}) * f(\boldsymbol{\gamma}|\boldsymbol{\phi}, \boldsymbol{\sigma}_b) * f(\boldsymbol{\phi}) * f(\boldsymbol{\sigma}_b)
\end{aligned}$$

assuming the prior for $\boldsymbol{\beta}$ is independent of $\boldsymbol{\phi}$ and $\boldsymbol{\sigma}_b$ (as well as $\boldsymbol{\gamma}$).

Hence,

$$f(\mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\phi}, \boldsymbol{\sigma}_b) = f(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\gamma}) f(\boldsymbol{\gamma}|\boldsymbol{\phi}, \boldsymbol{\sigma}_b) f(\boldsymbol{\phi}) f(\boldsymbol{\sigma}_b) f(\boldsymbol{\beta}) \tag{8.6}$$

where

$$f(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\gamma}) = \prod_{t=1}^{T} f(y_t|\mu_t, \sigma_t, \nu_t, \tau_t)$$

is the likelihood function, based on the assumed conditional distribution $\mathcal{D}(\mu_t, \sigma_t, \nu_t, \tau_t)$ for Y_t in equation (1.1), $f(\boldsymbol{\gamma}|\boldsymbol{\phi}, \boldsymbol{\sigma}_b)$ is a product of four independent multivariate normal prior distributions for $\boldsymbol{\gamma}_k$ for $k = 1, 2, 3, 4$ (assuming prior independence between the $\boldsymbol{\gamma}_k$ given $\boldsymbol{\phi}$ and $\boldsymbol{\sigma}_b$),

$$f(\boldsymbol{\gamma}|\boldsymbol{\phi}, \boldsymbol{\sigma_b}) = \prod_{k=1}^{4} f(\boldsymbol{\gamma}_k|\boldsymbol{\phi}, \boldsymbol{\sigma_b}).$$

The terms $f(\boldsymbol{\phi}), f(\boldsymbol{\sigma}_b)$ and $f(\boldsymbol{\beta})$ are independent prior distributions for the $\boldsymbol{\phi}$, $\boldsymbol{\sigma}_b$ and $\boldsymbol{\beta}$ parameters respectively.

Assuming a uniform prior for $\boldsymbol{\beta}$ in (8.6) gives

$$f(\mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\phi}, \boldsymbol{\sigma}_b) = f(\mathbf{y}|\boldsymbol{\beta}, \gamma) f(\gamma|\boldsymbol{\phi}, \boldsymbol{\sigma}_b) f(\boldsymbol{\phi}) f(\boldsymbol{\sigma}_b). \tag{8.7}$$

Note that, in a fully Bayesian inference, the posterior distribution of $\boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\phi}$ and $\boldsymbol{\sigma}_b$ can be obtained by using Markov chain Monte Carlo sampling as in Fahrmeir and Tutz (2001).

In addition, assuming a uniform prior for $\boldsymbol{\beta}$, from equation (8.7) we have the posterior distribution of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$,

$$\begin{aligned} f(\boldsymbol{\beta}, \gamma|\boldsymbol{\phi}, \boldsymbol{\sigma}_b, \mathbf{y}) &= \frac{f(\mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\phi}, \boldsymbol{\sigma}_b)}{f(\boldsymbol{\phi}, \boldsymbol{\sigma}_b, \mathbf{y})} \\ &\propto f(\mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\phi}, \boldsymbol{\sigma}_b) \\ &\propto f(\mathbf{y}|\boldsymbol{\beta}, \gamma) f(\gamma|\boldsymbol{\phi}, \boldsymbol{\sigma}_b). \end{aligned} \tag{8.8}$$

Maximizing equation (8.8) gives posterior mode estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, given $\boldsymbol{\phi}, \boldsymbol{\sigma}_b$ and $\mathbf{y}$. By taking the log of equation (8.8), maximizing (8.8) is equivalent to maximizing the *extended (or joint) log likelihood function* (Lee *et. al.* (2006)) for the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, given fixed $\boldsymbol{\phi}$ and $\boldsymbol{\sigma}_b$, defined by:

$$l_e = \log f\left(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\gamma}\right) + \log f\left(\boldsymbol{\gamma}|\boldsymbol{\phi}, \boldsymbol{\sigma}_b\right) \tag{8.9}$$

where

$$\log f\left(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\gamma}\right) = \sum_{t=1}^{T} \log f(y_t|\mu_t, \sigma_t, \nu_t, \tau_t) \tag{8.10}$$

is the log likelihood function and l_e is the log of the extended or joint log likelihood functions.

The maximization of the extended likelihood in (8.9), given fitted hyperparameters, can be achieved by using the GEST algorithm described above on page 151.

The estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ given fitted hyperparameters $(\boldsymbol{\phi}, \boldsymbol{\sigma}_b)$ is based on the *RS algorithm* which maximizes the extended likelihood function (see Rigby and Stasinopoulos (2005) Appendices B2 and C, equations (17) and (19)).

The GEST algorithm provides posterior mode estimates of the sets of parameters of $\boldsymbol{\beta}^\top = \left(\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top, \boldsymbol{\beta}_3^\top, \boldsymbol{\beta}_4^\top\right)$, and $\boldsymbol{\gamma}^\top = \left(\boldsymbol{\gamma}_1^\top, \boldsymbol{\gamma}_2^\top, \boldsymbol{\gamma}_3^\top, \boldsymbol{\gamma}_4^\top\right)$ by maximizing the extended log likelihood for fitted or fixed hyperparameters $\sigma_{b_k}^2$ and $\phi_{k,j}$ for $k = 1, 2, 3, 4$.

8.3.2 Estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ given fixed hyperparameters $\boldsymbol{\phi}$ and $\boldsymbol{\sigma}_b$

The estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ given fixed hyperparameters $(\boldsymbol{\phi}, \boldsymbol{\sigma}_b)$ is based on the *RS algorithm* which maximizes the extended likelihood function (see Rigby and Stasinopoulos (2005) Appendices B2 and C equations (17) and (19)).

The maximization of the extended likelihood in (8.9), given fixed hyperparameters, can be achieved by using the GEST algorithm described above on page 151. In the local estimation procedure for the random effects hyperparameters, step (B)(a)(ii)(II) page 151, in the GEST algorithm for estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ given fitted hyperparameters, is replaced by (B)(a)(ii)(II*).

(B)(a) (ii)(II*) estimate $\boldsymbol{\gamma}_k$ by smoothing the current partial residuals $\boldsymbol{\epsilon}_k = \mathbf{z}_k - \mathbf{X}\boldsymbol{\beta}_k$ over time with weights $\mathbf{W}_k$ [using equation (8.20) with σ_e^2 set to 1].

8.3.3 Global (i.e. external) estimation of hyperparameters $\boldsymbol{\phi}$ and $\boldsymbol{\sigma}_b$

There are two estimation methods of the random effect hyperparameters, a global (i.e. external) and a local (i.e. internal) method. When the random effects hyperparameters are unknown, they can be estimated by maximizing the marginal likelihood, which is obtained by integrating out $\boldsymbol{\gamma}$ from $f(\mathbf{y}, \boldsymbol{\gamma}|\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\sigma}_b)$, or by integrating out both $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ from $f(\mathbf{y}, \boldsymbol{\beta}, \gamma|\boldsymbol{\phi}, \boldsymbol{\sigma}_b)$ for Restricted Maximum Likelihood Estimation, REML.

Both methods considered here (global and local) are in general approximative methods. [However, the Gaussian random effect mean models in chapter 4 provide the exact estimation method of normal random effect hyperparameters for the mean]. For non-Gaussian observations, the integral of both $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ is intractable, and notoriously difficult. Breslow and Clayton (1993) used a Laplace integral approximation to estimate the random effect hyperparameters in generalized linear mixed models (GLMM) models. Lee and Nelder (1996) proposed the extended likelihood rather than the marginal likelihood in hierarchical generalized linear models which allows random effects to be not normally distributed, and estimate the hyperparameters by maximizing the adjusted profile likelihood. Pinheiro and Bates (2000), and Rigby and Stasinopoulos (2005), Section A.2.3. both used a Laplace integral approximation to estimate the random effect hyperparameters.

More information on the Laplace approximation can be found in Tierney and Kadane (1986) and Evans and Swartz (2000, p.62), Severini (2000, Section 2.11).

Restricted maximum likelihood estimation of the hyperparameters $(\boldsymbol{\phi}, \boldsymbol{\sigma}_b)$:

The restricted maximum likelihood estimation of the hyperparameters $(\boldsymbol{\phi}, \boldsymbol{\sigma}_b)$ is defined by:

$$
\begin{aligned}
L(\boldsymbol{\phi}, \boldsymbol{\sigma}_b) &= \int\int f(\mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\gamma}|\boldsymbol{\phi}, \boldsymbol{\sigma}_b) d\boldsymbol{\gamma} d\boldsymbol{\beta} \\
&= \int\int f(\mathbf{y}|\boldsymbol{\beta}, \gamma) f(\gamma|\boldsymbol{\phi}, \boldsymbol{\sigma}_b) d\boldsymbol{\gamma} d\boldsymbol{\beta} \qquad (8.11)
\end{aligned}
$$

where

$$f(\mathbf{y}|\boldsymbol{\beta}, \gamma) = \prod_{t=1}^{T} f(y_t|\boldsymbol{\beta}, \boldsymbol{\gamma})$$

denotes the conditional density function of the responses $\mathbf{y}$ given $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, and $f(\gamma|\boldsymbol{\phi}, \boldsymbol{\sigma}_b)$ is the density function of the random effect $\boldsymbol{\gamma}$ given $\boldsymbol{\phi}$ and $\boldsymbol{\sigma}_b$, where $\boldsymbol{\phi}$ and $\boldsymbol{\sigma}_b$ are hyperparameters, $\boldsymbol{\beta}$ are the regression parameters, and $\boldsymbol{\gamma}$ are the random effects parameters. The likelihood $f(\mathbf{y}|\boldsymbol{\beta}, \gamma) f(\gamma|\boldsymbol{\phi}, \boldsymbol{\sigma}_b)$ is known as the joint or extended likelihood in hierarchical generalized linear models (Lee and Nelder, 1996).

However, this integration is intractable for a non-Gaussian response variable and becomes more difficult as the number of random components increases. The integral approximation of Laplace gives the following approximative marginal log likelihood:

$$l(\boldsymbol{\phi}, \boldsymbol{\sigma}_b) = \log f(\mathbf{y}|\hat{\boldsymbol{\beta}}, \hat{\gamma}) + \log f(\hat{\gamma}|\boldsymbol{\phi}, \boldsymbol{\sigma}_b) - \frac{1}{2} \log \left| \frac{\hat{\mathbf{D}}_{\boldsymbol{\beta}, \boldsymbol{\gamma}}}{2\pi} \right| \qquad (8.12)$$

where $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}})$ are the fitted values of $(\boldsymbol{\beta}, \boldsymbol{\gamma})$ given by maximising the extended likelihood over $(\boldsymbol{\beta}, \boldsymbol{\gamma})$ for given $(\boldsymbol{\phi}, \boldsymbol{\sigma}_b)$, and $\hat{\mathbf{D}}_{\boldsymbol{\beta},\boldsymbol{\gamma}}$ is the second derivative of the extended likelihood with respect to $(\boldsymbol{\beta}, \boldsymbol{\gamma})$ evaluated at $(\boldsymbol{\beta}, \boldsymbol{\gamma}) = (\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}})$.

Note $l(\boldsymbol{\phi}, \boldsymbol{\sigma}_b)$ can be maximized numerically over $(\boldsymbol{\phi}, \boldsymbol{\sigma}_b)$.

For a normal distribution $\hat{\mathbf{D}}_{\boldsymbol{\beta},\boldsymbol{\gamma}}$ is defined as:

$$\hat{\mathbf{D}}_{\boldsymbol{\beta},\boldsymbol{\gamma}} = \begin{pmatrix} X_1^\top W_{11} X_1 & X_1^\top W_{11} & 0 & 0 \\ X_1^\top W_{11} & W_{11} + G_1 & 0 & 0 \\ 0 & 0 & X_2^\top W_{22} X_2 & X_2^\top W_{22} \\ 0 & 0 & X_2^\top W_{22} & W_{22} + G_2 \end{pmatrix}.$$

Maximum likelihood estimation of $\boldsymbol{\beta}$ and hyperparameters $(\boldsymbol{\phi}, \boldsymbol{\sigma}_b)$:

The maximum likelihood estimation of $\boldsymbol{\beta}$ and hyperparameters $(\boldsymbol{\phi}, \boldsymbol{\sigma}_b)$ is defined by:

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\sigma}_b) &= \int f(\mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\gamma} | \boldsymbol{\phi}, \boldsymbol{\sigma}_b) d\boldsymbol{\gamma} \\ &= \int f(\mathbf{y} | \boldsymbol{\beta}, \boldsymbol{\gamma}) f(\boldsymbol{\gamma} | \boldsymbol{\phi}, \boldsymbol{\sigma}_b) d\boldsymbol{\gamma} \end{aligned} \tag{8.13}$$

where

$$f(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\gamma}) = \prod_{t=1}^{T} f(y_t | \boldsymbol{\beta}, \boldsymbol{\gamma})$$

denotes the conditional density function of the responses $\mathbf{y}$ given $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, and $f(\boldsymbol{\gamma}|\boldsymbol{\phi}, \boldsymbol{\sigma}_b)$ is the density function of the random effect $\boldsymbol{\gamma}$ given $\boldsymbol{\phi}$ and $\boldsymbol{\sigma}_b$, where $\boldsymbol{\phi}$ and $\boldsymbol{\sigma}_b$ are hyperparameters, $\boldsymbol{\beta}$ are the regression parameters, and $\boldsymbol{\gamma}$ are the random effects parameters. The likelihood $f(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\gamma}) f(\boldsymbol{\gamma}|\boldsymbol{\phi}, \boldsymbol{\sigma}_b)$, as on page 154 (8.8), is known as the joint or extended likelihood in hierarchical generalized linear models

(Lee and Nelder, 1996).

However, this integration is also intractable for a non-Gaussian response variable and becomes more difficult as the number of random components increases. The integral approximation of Laplace gives the following approximative marginal log likelihood :

$$l(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\sigma}_b) = \log f(\mathbf{y}|\boldsymbol{\beta}, \hat{\gamma}) + \log f(\hat{\gamma}|\boldsymbol{\phi}, \boldsymbol{\sigma}_b) - \frac{1}{2} \log \left| \frac{\hat{\mathbf{D}}_{\boldsymbol{\gamma}}}{2\pi} \right| \tag{8.14}$$

where $\hat{\boldsymbol{\gamma}}$ is the fitted value of $\boldsymbol{\gamma}$ given by maximising the extended likelihood over $\boldsymbol{\gamma}$ for given $(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\sigma}_b)$, and $\hat{\mathbf{D}}_{\boldsymbol{\gamma}}$ is the second derivative of the extended likelihood with respect to $\boldsymbol{\gamma}$ evaluated at $\boldsymbol{\gamma} = \hat{\boldsymbol{\gamma}}$.

Note $l(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\sigma}_b)$ can be maximized numerically over $(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\sigma}_b)$.

For a normal distribution $\hat{\mathbf{D}}_{\boldsymbol{\gamma}}$ is defined as:

$$\hat{\mathbf{D}}_{\boldsymbol{\gamma}} = \mathbf{W} + \mathbf{G},$$

where

$$\mathbf{W} = \begin{pmatrix} W_{11} & 0 \\ 0 & W_{22} \end{pmatrix},$$

and

$$\begin{aligned} W_{11} &= E\left[\frac{\partial^2 \ell}{\partial \eta_1 \partial \eta_1}\right] \\ W_{22} &= E\left[\frac{\partial^2 \ell}{\partial \eta_2 \partial \eta_2}\right] \end{aligned}$$

and

$$\mathbf{G} = \begin{pmatrix} \sigma_{b_1}^{-2}\mathbf{D}_1^\top \mathbf{D}_1 & 0 \\ 0 & \sigma_{b_2}^{-2}\mathbf{D}_2^\top \mathbf{D}_2 \end{pmatrix}.$$

where

$$\mathbf{D_1} = \mathbf{D_2} = \begin{pmatrix} -1 & 1 & & & 0 \\ 0 & -1 & 1 & & \\ & .. & .. & .. & \\ & .. & .. & .. & \\ 0 & & & -1 & 1 \end{pmatrix}$$

For a GEST model with no fixed effects parameters $\boldsymbol{\beta}$,

$$l(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\sigma}_b) = l(\boldsymbol{\phi}, \boldsymbol{\sigma}_b),$$

which can be maximised numerically over $\boldsymbol{\phi}$ and $\boldsymbol{\sigma}_b$.

For example, for a three parameter Student t distribution, $Y_t|\mu_t, \sigma_t, \nu_t \sim TF(\mu_t, \sigma_t, \nu_t)$, the external estimation of the random effects parameters maximizes

$$l(\boldsymbol{\phi}, \boldsymbol{\sigma}_b) \quad = \quad \log f(\mathbf{y}|\boldsymbol{\gamma}) + \sum_{k=1}^{3} \log f(b_k) - \frac{1}{2}\log|\mathbf{W} + \mathbf{G}| + \frac{3}{2}T \log 2\pi$$

where

$$\log f(b_k) = -\frac{1}{2}(T - J)\log(2\pi\sigma_{b_k}^2) - \frac{1}{2}\sigma_{b_k}^{-2}\boldsymbol{\gamma}_k^\top \mathbf{D}_k^\top \mathbf{D}_k \boldsymbol{\gamma}_k$$

and

$$\mathbf{W} = \begin{pmatrix} W_{11} & 0 & 0 \\ 0 & W_{22} & W_{23} \\ 0 & W_{32} & W_{33} \end{pmatrix},$$

$$\mathbf{G} = \begin{pmatrix} \sigma_{b_1}^{-2}\mathbf{D}_1^\top \mathbf{D}_1 & 0 & 0 \\ 0 & \sigma_{b_2}^{-2}\mathbf{D}_2^\top \mathbf{D}_2 & 0 \\ 0 & 0 & \sigma_{b_3}^{-2}\mathbf{D}_3^\top \mathbf{D}_3 \end{pmatrix},$$

where, if $\hat{\mathbf{D}}_{\boldsymbol{\gamma}}$ is replaced by its expected value,

$$\begin{aligned} W_{11} &= E\left[\frac{\partial^2 \ell}{\partial \eta_1 \partial \eta_1}\right] \\ W_{22} &= E\left[\frac{\partial^2 \ell}{\partial \eta_2 \partial \eta_2}\right] \\ W_{33} &= E\left[\frac{\partial^2 \ell}{\partial \eta_3 \partial \eta_3}\right] \\ W_{23} &= E\left[\frac{\partial^2 \ell}{\partial \eta_2 \partial \eta_3}\right] = E\left[\frac{\partial^2 \ell}{\partial \sigma \partial \nu}\right] * \frac{\partial \sigma}{\partial \eta_2} * \frac{\partial \nu}{\partial \eta_3} \\ W_{32} &= E\left[\frac{\partial^2 \ell}{\partial \eta_3 \partial \eta_2}\right] = E\left[\frac{\partial^2 \ell}{\partial \nu \partial \sigma}\right] * \frac{\partial \nu}{\partial \eta_3} * \frac{\partial \sigma}{\partial \eta_2}. \end{aligned}$$

For a four parameter distribution, for example the skew Student t distribution, page 269, $Y_t|\mu_t, \sigma_t, \nu_t, \tau_t \sim SST(\mu_t, \sigma_t, \nu_t, \tau_t)$, the external estimation of the random effects parameters, assuming no fixed effects parameters, maximizes:

$$l(\boldsymbol{\phi}, \boldsymbol{\sigma}_b) = \log f(\mathbf{y}|\boldsymbol{\gamma}) + \sum_{k=1}^{4} \log f(b_k) - \frac{1}{2}\log|\mathbf{W}+\mathbf{G}| + \frac{4}{2}T\log 2\pi$$

where

$$\log f(b_k) = -\frac{1}{2}(T-J)\log(2\pi\sigma^2_{b_k}) - \frac{1}{2}\sigma_{b_k}^{-2}\boldsymbol{\gamma}_k^\top \mathbf{D}_k^\top \mathbf{D}_k \boldsymbol{\gamma}_k$$

and

$$\mathbf{W}+\mathbf{G} = \begin{pmatrix} W_{11}+\sigma_{b_1}^{-2}\mathbf{D}_1^\top\mathbf{D}_1 & W_{12} & W_{13} & W_{14} \\ W_{21} & W_{22}+\sigma_{b_2}^{-2}\mathbf{D}_2^\top\mathbf{D}_2 & W_{23} & W_{24} \\ W_{31} & W_{32} & W_{33}+\sigma_{b_3}^{-2}\mathbf{D}_3^\top\mathbf{D}_3 & W_{34} \\ W_{41} & W_{42} & W_{43} & W_{44}+\sigma_{b_4}^{-2}\mathbf{D}_4^\top\mathbf{D}_4 \end{pmatrix},$$

where, if $\hat{\mathbf{D}}\boldsymbol{\gamma}$ is replaced by its expected value,

$$\begin{aligned}
W_{12} &= E\left[\frac{\partial^2\ell}{\partial\eta_1\partial\eta_2}\right] = E\left[\frac{\partial^2\ell}{\partial\mu\partial\sigma}\right] * \frac{\partial\mu}{\partial\eta_1} * \frac{\partial\sigma}{\partial\eta_2} \\
W_{13} &= E\left[\frac{\partial^2\ell}{\partial\eta_1\partial\eta_3}\right] = E\left[\frac{\partial^2\ell}{\partial\mu\partial\nu}\right] * \frac{\partial\mu}{\partial\eta_1} * \frac{\partial\nu}{\partial\eta_3} \\
W_{23} &= E\left[\frac{\partial^2\ell}{\partial\eta_2\partial\eta_3}\right] = E\left[\frac{\partial^2\ell}{\partial\sigma\partial\nu}\right] * \frac{\partial\sigma}{\partial\eta_2} * \frac{\partial\nu}{\partial\eta_3} \\
W_{24} &= E\left[\frac{\partial^2\ell}{\partial\eta_2\partial\eta_4}\right] = E\left[\frac{\partial^2\ell}{\partial\sigma\partial\tau}\right] * \frac{\partial\sigma}{\partial\eta_2} * \frac{\partial\tau}{\partial\eta_4} \\
W_{34} &= E\left[\frac{\partial^2\ell}{\partial\eta_3\partial\eta_4}\right] = E\left[\frac{\partial^2\ell}{\partial\nu\partial\tau}\right] * \frac{\partial\nu}{\partial\eta_3} * \frac{\partial\tau}{\partial\eta_4}.
\end{aligned}$$

The equation for estimating $\boldsymbol{\gamma}_k$ (within the GEST algorithm for estimating $\boldsymbol{\gamma}$ given fixed hyperparameters $\boldsymbol{\phi}$ and $\boldsymbol{\sigma}_b$, from section 8.3.2)

is given by:

$$\hat{\boldsymbol{\gamma}}_k = \left[\mathbf{W}_{kk} + \sigma_{b_k}^{-2}\mathbf{D}_k^\top\mathbf{D}_k\right]^{-1}\mathbf{W}_{kk}\ \boldsymbol{\epsilon}_k \tag{8.15}$$

(See Rigby and Stasinopoulos (2005) Appendices B2 and C, with a special case of equation (17), where $\mathbf{Z} = \mathbf{I}$ and subscript j is omitted), where $\boldsymbol{\epsilon}_k = \mathbf{z}_k - \mathbf{X}\boldsymbol{\beta}_k$ are the current partial residuals for $k = 1, 2, 3, 4$.

To estimate $(\boldsymbol{\phi}, \boldsymbol{\sigma}_b^2)$ and hence obtain $\hat{\boldsymbol{\gamma}}_k$, the following algorithm is used:

The algorithm for estimating $\boldsymbol{\alpha} = (\boldsymbol{\phi}, \boldsymbol{\sigma}_b)$

1. Select starting values for $\boldsymbol{\alpha} = (\boldsymbol{\phi}, \boldsymbol{\sigma}_b)$.

2. Maximize $l(\boldsymbol{\phi}, \boldsymbol{\sigma}_b)$ over $\boldsymbol{\alpha}$ using a numerical algorithm, where, within the numerical algorithm, $\boldsymbol{\gamma}_k$ for $k = 1, 2, 3, 4$, given $\boldsymbol{\alpha}$, are estimated using the GEST algorithm for fixed hyperparameters from section 8.3.2.

3. Use the maximizing values for $\boldsymbol{\alpha}$ to calculate the maximizing values for $\boldsymbol{\gamma}_k$.

Note that this method is computationally time-consuming. The local method described below is faster and has been found to produce almost identical results in several examples.

8.3.4 Local (i.e. internal) estimation of hyperparameters $\boldsymbol{\phi}$ and $\boldsymbol{\sigma}_b$

The local estimation procedure is based on ideas from Pawitan (2001), section 17.5, page 445-448, for normal linear mixed models estimation method, and on Rigby and Stasinopoulos (2013), as a generalization to several random effects. The local estimation method is described in Pinheiro and Bates (2000), Venables and Ripley (2002), p.297-298, Wood (2006), Section 6.4. The GEST model is based on the Pawitan procedure because the random effects $\boldsymbol{\gamma}_t$ are assumed normal, and uses the local estimation procedure rather than the global method because the local is much faster than the global method and both methods have been found to produce very similar results in several examples. In addition, the local method has been called penalized quasi likelihood (PQL). The local method of the GEST model which is described below uses penalized likelihood.

During the fit of each one of μ, σ, ν, and τ, the corresponding structural parameters [i.e. $\sigma^2_{b_k}$ and $\boldsymbol{\phi}_k$ for $k = 1, 2, 3, 4$ where $\boldsymbol{\phi}_k^\top = (\phi_{k,1}, \phi_{k,2}, \ldots, \phi_{k,J_k})$], are estimated by the internal (i.e. local) marginal maximum likelihood estimation procedure outlined below.

To simplify the notation the subscript k is dropped so θ_t now represents any one of the parameters $(\mu_t, \sigma_t, \nu_t, \tau_t)$:

$$g(\theta_t) = \eta_t = \mathbf{x}_t^\top \boldsymbol{\beta} + \gamma_t \tag{8.16}$$

for $t = 1, 2, \ldots, T$, where γ_t is defined by (8.2) with subscript k omitted, i.e.

$$\gamma_t = \sum_{j=1}^{J} \phi_j \gamma_{t-j} + b_t, \tag{8.17}$$

for $t = J+1, J+2, \ldots, T$.

On the predictor scale (8.16), in the structural model fitting part of the backfitting algorithm [i.e. step (B)(a)(ii)(II) of the GEST fitting algorithm for fitting $\boldsymbol{\gamma}$ page 151] the following local approximate internal model is used:

$$\boldsymbol{\epsilon} = \boldsymbol{\gamma} + \boldsymbol{e}$$

where $\boldsymbol{\epsilon} = \mathbf{z} - \mathbf{X}\boldsymbol{\beta}$ are the current partial residuals, $\boldsymbol{e} \sim N_T(\mathbf{0}, \boldsymbol{\Sigma})$ and $\boldsymbol{\Sigma} = \sigma_e^2 \mathbf{W}^{-1}$, where $\mathbf{z} = \boldsymbol{\eta} + \mathbf{W}^{-1}\mathbf{u}$ is the current pseudo response variable (or iterative response variable), $\mathbf{W}$ is a diagonal matrix of current weights given by one of the following $-\frac{\partial^2 \ell}{\partial \eta \partial \eta^\top}$, $-E\left[\frac{\partial^2 \ell}{\partial \eta \partial \eta^\top}\right]$ or $\left(\frac{\partial \ell}{\partial \eta}\right)^2$, i.e. the observed information, the expected information or the squared score function, depending respectively on whether a Newton-Raphson, Fisher scoring or quasi-Newton algorithm is used, and $\mathbf{u} = \frac{\partial \ell}{\partial \eta}$.

The algorithm, given later in this Section, maximises the local likelihood function Q, given below, directly over the structural model parameters $\boldsymbol{\alpha} = (\sigma_b^2, \boldsymbol{\phi})$, where $\boldsymbol{\phi}^\top = (\phi_1, \phi_2, \ldots, \phi_J)$, using a numerical algorithm.

For $t = 1, 2, \ldots, T$,

$$\epsilon_t = \gamma_t + e_t$$

where γ_t is given in (8.17). Hence,

$$\begin{aligned} e_t &= \epsilon_t - \gamma_t \\ b_t &= \gamma_t - \sum_{j=1}^{J} \phi_j \gamma_{t-j}. \end{aligned}$$

for $t = J+1, J+2, \ldots, T$.

In summary for the local level (autoregressive and random walk) structural model:

$$g(\boldsymbol{\theta}) = \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\gamma}$$

with local approximate internal model given by

$$
\begin{aligned}
\boldsymbol{\epsilon} &= \boldsymbol{\gamma} + \boldsymbol{e} \\
\boldsymbol{\epsilon} &= \mathbf{z} - \mathbf{X}\boldsymbol{\beta} \\
\mathbf{z} &= \boldsymbol{\eta} + \mathbf{W}^{-1}\mathbf{u} \\
\mathbf{D}\boldsymbol{\gamma} &= \mathbf{b} \\
\boldsymbol{\epsilon}|\boldsymbol{\gamma} &\sim N(\boldsymbol{\gamma}, \boldsymbol{\Sigma}) \\
\mathbf{b} &\sim N(0, \sigma_b^2 \mathbf{I_{T-J}})
\end{aligned}
\tag{8.18}
$$

where $\boldsymbol{e} \sim N_T(\mathbf{0}, \boldsymbol{\Sigma})$ and $\boldsymbol{\Sigma} = \sigma_e^2 \mathbf{W}^{-1}$, $\boldsymbol{\Sigma}^{-1} = \sigma_e^{-2}\mathbf{W}$, $\mathbf{u} = \frac{\partial \ell}{\partial \eta}$, $\mathbf{z}$ is the current pseudo response variable and $\boldsymbol{\epsilon}$ are the current partial residuals, $\mathbf{W} = -\frac{\partial^2 \ell}{\partial \eta \partial \eta^\top}$, or $-E\left[\frac{\partial^2 \ell}{\partial \eta \partial \eta^\top}\right]$ or $\left(\frac{\partial \ell}{\partial \eta}\right)^2$. Note that $\boldsymbol{\eta}, \boldsymbol{\gamma}, \mathbf{z}, \mathbf{u}, \boldsymbol{\epsilon}$ and $\boldsymbol{e}$ are vectors of length T, whilst $\mathbf{b} = (b_{J+1}, b_{J+2}, \ldots, b_T)^\top$ is a vector of length $T - J$.

Note that Pawitan (2001) shows a *computational equivalence* between the usual estimation of random effects and their parameters (i.e. integrating out the random effects and maximizing over the fixed and random parameters) and maximizing an objective function Q (in the form of an adjusted profile extended likelihood for the random effects parameters). Given the absence of fixed effects locally, the Q function, maximized over the random effects $\boldsymbol{\gamma}$ given the random effects parameters $\boldsymbol{\alpha} = (\sigma_b^2, \boldsymbol{\phi})$, gives the local likelihood function of $\boldsymbol{\alpha}$. Here locally the random effects are $\boldsymbol{\gamma}$ with parameters $\boldsymbol{\alpha}$ and, generalizing Lee *et al.* (2006), p277-279, the local function Q is given by

$$
Q = \log f(\boldsymbol{\epsilon}|\boldsymbol{\gamma}) + \log f(\boldsymbol{\gamma}) - \frac{1}{2}\log\left|\boldsymbol{\Sigma}^{-1} + \sigma_b^{-2}\mathbf{D}^\top\mathbf{D}\right| + \frac{T}{2}\log 2\pi \tag{8.19}
$$

where T is the number of observations, $\boldsymbol{\epsilon} = \mathbf{z} - \mathbf{X}\boldsymbol{\beta}$ is the vector of current partial residuals, $\mathbf{e} \sim N_T(0, \sigma_e^2 \mathbf{W}^{-1})$, where $\mathbf{e} = (e_1, e_2, \ldots, e_T)^\top$, $\boldsymbol{\Sigma} = \sigma_e^2 \boldsymbol{W}^{-1}$ and $\mathbf{D}$ is defined below. Note that assuming diffuse uniform priors for $(\gamma_1, \ldots, \gamma_J)$ in Q gives

$$f(\boldsymbol{\gamma}) = \prod_{t=J+1}^{T} f(\gamma_t | \boldsymbol{\gamma}_{t-1}) \equiv \prod_{t=J+1}^{T} f(b_t) = f(\mathbf{b})$$

where $\boldsymbol{\gamma}_{t-1} = (\gamma_1, \gamma_2, \ldots, \gamma_{t-1})$, $\mathbf{b} = (b_{J+1}, b_{J+2}, \ldots, b_T)^\top$, $\mathbf{b} \sim N_{T-J}(0, \sigma_b^2 \mathbf{I}_{T-J})$. Maximising Q over $\boldsymbol{\alpha} = (\sigma_b^2, \boldsymbol{\phi})$ gives estimates of σ_b^2 and $\boldsymbol{\phi}$. Then, $\boldsymbol{\gamma}$ is estimated effectively by smoothing the partial residuals using

$$\hat{\boldsymbol{\gamma}} = \left[\boldsymbol{\Sigma}^{-1} + \sigma_b^{-2} \mathbf{D}^\top \mathbf{D}\right]^{-1} \boldsymbol{\Sigma}^{-1} \boldsymbol{\epsilon}. \tag{8.20}$$

Step (B)(a)(ii)(II): the algorithm for estimating $\boldsymbol{\alpha} = (\sigma_e^2, \sigma_b^2, \boldsymbol{\phi})$

1. Select starting values for $\boldsymbol{\alpha} = (\sigma_e^2, \sigma_b^2, \boldsymbol{\phi})$.

2. Maximize Q over $\boldsymbol{\alpha}$ using a numerical algorithm, where $\hat{\boldsymbol{\gamma}}$ given $\boldsymbol{\alpha}$ is estimated by (8.20) before calculating Q in the function evaluating Q.

3. Use the maximizing values for $\boldsymbol{\alpha}$ to calculate the maximizing values for $\boldsymbol{\gamma}$.

In step 2, Q is given by

$$\begin{aligned}
Q &= \log f(\boldsymbol{\epsilon}|\boldsymbol{\gamma}) + \log f(\boldsymbol{\gamma}) - \frac{1}{2} \log \left|\boldsymbol{\Sigma}^{-1} + \sigma_b^{-2} \mathbf{D}^\top \mathbf{D}\right| + \frac{T}{2} \log 2\pi \\
\log f(\boldsymbol{\epsilon}|\boldsymbol{\gamma}) &= -\frac{1}{2} \log \left|2\pi \mathbf{W}^{-1}\right| - \frac{1}{2} \sigma_e^{-2} (\boldsymbol{\epsilon} - \boldsymbol{\gamma})^\top \mathbf{W} (\boldsymbol{\epsilon} - \boldsymbol{\gamma}) - \frac{1}{2} T \log \sigma_e^2 \\
\log f(\boldsymbol{\gamma}) &= \log f(\mathbf{b}) = -\frac{1}{2}(T - J) \log \left(2\pi\sigma_b^2\right) - \frac{1}{2} \sigma_b^{-2} \boldsymbol{\gamma}^\top \mathbf{D}^\top \mathbf{D} \boldsymbol{\gamma}^\top
\end{aligned}$$

since $\mathbf{b} = \mathbf{D}\boldsymbol{\gamma} \sim \mathbf{N}(\mathbf{0}, \sigma^{\mathbf{2}}_{\mathbf{b}}\mathbf{I}_{\mathbf{T-J}})$, where $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_T)^\top, \boldsymbol{\Sigma}^{-1} = \sigma_e^{-2}\mathbf{W}$ and

$$\mathbf{D} = \begin{pmatrix} -\phi_J & -\phi_{J-1} & \ldots & \ldots & -\phi_1 & 1 & 0 & 0 & \ldots & \ldots & 0 & 0 \\ 0 & -\phi_J & -\phi_{J-1} & \ldots & \ldots & -\phi_1 & 1 & 0 & \ldots & \ldots & 0 & 0 \\ & & & & & & & & & & & \\ & & & & & & & & & & & \\ 0 & 0 & 0 & \ldots & \ldots & 0 & -\phi_J & -\phi_{J-1} & \ldots & \ldots & -\phi_1 & 1 \end{pmatrix},$$

and the maximizing of Q over $\boldsymbol{\gamma}$ given $\boldsymbol{\alpha}$ gives equation (8.20).

Note that $\mathbf{D}$ is a $(T - J)$ x T matrix and $\boldsymbol{\Sigma}$ is T x T.

8.4 Local estimation functions of the hyperparameters

Here the local level structural model is generalized in Section 8.4.1 to a local level with leverage structural model, in Section 8.4.2 to a local level with seasonal structural model, in Section 8.4.3 to a local level with trend structural model, in Section 8.4.4 to a local level with trend and seasonality structural model, and in Section 8.4.5 to a local level with random coefficient of an explanatory variable structural model.

Note that in Section 5.6 of Chapter 5 of this thesis, the Q function is based on the normal data using the response variable $\mathbf{y}$, this chapter uses pseudo response variable $\boldsymbol{\epsilon}$ which is approximated locally using a normal distribution, hence, the local Q function given by equation (8.19), where $\boldsymbol{\epsilon}$ has replaced $\mathbf{y}$

8.4.1 Local level with persistent effect

Here

$$g(\boldsymbol{\theta}) = \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\gamma},$$

where the model for random effect γ_t includes explanatory variables v_t which have a persistent effect (e.g. the leverage effect) on the distribution parameter,

$$\gamma_t = \sum_{j=1}^{J} \phi_j \gamma_{t-j} + v_t^\top \delta + b_t$$

for $t = J+1, J+2, \ldots, T$. Hence,

$$\mathbf{D}_\gamma \boldsymbol{\gamma} = \mathbf{V}\boldsymbol{\delta} + \mathbf{b},$$

where $\mathbf{D}_\gamma \boldsymbol{\gamma}$ replaces $\mathbf{D}\boldsymbol{\gamma} = \mathbf{b}$ in the local approximate internal model given by (8.18), and where $\mathbf{D}_\gamma$ is given below.

Maximizing Q over $(\sigma_e^2, \sigma_b^2, \boldsymbol{\phi})$ gives estimates of $\sigma_e^2, \sigma_b^2, \boldsymbol{\phi}$, where

$$
\begin{aligned}
Q &= \log f(\boldsymbol{\epsilon}|\boldsymbol{\gamma}) + \log f(\boldsymbol{\gamma}) - \frac{1}{2}\log\left|\mathbf{A} + \sigma_b^{-2}\mathbf{D}^\top\mathbf{D}\right| + \frac{T}{2}\log 2\pi \\
\log f(\boldsymbol{\epsilon}|\boldsymbol{\gamma}) &= -\frac{1}{2}\log|2\pi\boldsymbol{\Sigma}| - \frac{1}{2}(\boldsymbol{\epsilon} - \boldsymbol{\gamma})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{\epsilon} - \boldsymbol{\gamma}) \\
\log f(\boldsymbol{\gamma}) &= -\frac{1}{2}\log\left|2\pi\sigma_b^2\right| - \frac{1}{2}\sigma_b^{-2}\left(\boldsymbol{\gamma}^\top, \boldsymbol{\delta}^\top\right)\mathbf{D}^\top\mathbf{D}\left(\boldsymbol{\gamma}^\top, \boldsymbol{\delta}^\top\right)^\top
\end{aligned}
$$

since $\mathbf{b} = \mathbf{D}_\gamma\boldsymbol{\gamma} - \mathbf{V}\boldsymbol{\delta} = \mathbf{D}(\boldsymbol{\gamma}^\top, \boldsymbol{\delta}^\top)^\top \sim \mathbf{N}(\mathbf{0}, \sigma_\mathbf{b}^\mathbf{2}\mathbf{I_{T-J}})$, where $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_T)^\top$, $\mathbf{D} = \begin{pmatrix} \mathbf{D}_\gamma, & -\mathbf{V} \end{pmatrix}$,

$$
\mathbf{D}_\gamma = \begin{pmatrix}
-\phi_J & -\phi_{J-1} & \ldots & \ldots & -\phi_1 & 1 & 0 & 0 & \ldots & \ldots & 0 & 0 \\
0 & -\phi_J & -\phi_{J-1} & \ldots & \ldots & -\phi_1 & 1 & 0 & \ldots & \ldots & 0 & 0 \\
 & & & & & & & & & & & \\
 & & & & & & & & & & & \\
0 & 0 & 0 & \ldots & \ldots & 0 & -\phi_J & -\phi_{J-1} & \ldots & \ldots & -\phi_1 & 1
\end{pmatrix},
$$

$$
\mathbf{A} = \begin{pmatrix} \boldsymbol{\Sigma}^{-1} & 0 \\ 0 & 0 \end{pmatrix}.
$$

Note that $\mathbf{D}_\gamma$ is a $(T - J)$ x T matrix, $\mathbf{D}$ is $(T - J)$ x $(T + q)$ and $\mathbf{A}$ is $(T + q)$ x $(T + q)$.

The equation for estimating $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$ is given by:

$$
\begin{pmatrix} \boldsymbol{\gamma} \\ \boldsymbol{\delta} \end{pmatrix} = \left[\mathbf{A} + \sigma_b^{-2}\mathbf{D}^\top\mathbf{D}\right]^{-1}\mathbf{A}\begin{pmatrix} \boldsymbol{\epsilon} \\ \mathbf{0} \end{pmatrix}, \tag{8.21}
$$

where **0** is a vector of zeros of length q.

Here are some examples of local level without and with a persistent effect:

(i) Stochastic volatility model without a persistent effect:

$$
\begin{aligned}
Y_t|\mu_t, \sigma_t &\sim NO(\mu_t, \sigma_t) \\
\mu_t &= \beta_{1,0} \\
\log(\sigma_t) &= \beta_{2,0} + \gamma_{2,t} \\
\gamma_{2,t} &= \phi_1 \gamma_{2,t-1} + b_{2,t}
\end{aligned}
\tag{8.22}
$$

(ii) Asymmetric stochastic volatility model with instantaneous effect, since the effect of y_{t-1} on σ_t only lasts for one time point:

$$
\begin{aligned}
Y_t|\mu_t, \sigma_t &\sim NO(\mu_t, \sigma_t) \\
\mu_t &= \beta_{1,0} \\
\log(\sigma_t) &= \beta_{2,0} + \gamma_{2,t} + \beta_{2,1} y_{t-1} \\
\gamma_{2,t} &= \phi_1 \gamma_{2,t-1} + b_{2,t}
\end{aligned}
\tag{8.23}
$$

(iii) Asymmetric stochastic volatility model with a persistent effect, since the effect of y_{t-1} on σ_t persists:

$$
\begin{aligned}
Y_t|\mu_t, \sigma_t &\sim NO(\mu_t, \sigma_t) \\
\mu_t &= \beta_{1,0} \\
\log(\sigma_t) &= \beta_{2,0} + \gamma_{2,t} \\
\gamma_{2,t} &= \phi_1 \gamma_{2,t-1} + \delta_1 v_{1,t-1} + \delta_2 v_{2,t-1} + b_{2,t}
\end{aligned} \tag{8.24}
$$

where $v_{1,t-1} = \text{arcsinh}(y_{t-1})$ (if $y_{t-1} < 0$) and $v_{2,t-1} = \text{arcsinh}(y_{t-1})$ (if $y_{t-1} >= 0$) to account for the leverage effect. The use of the transformed $\sinh(y_{t-1})$ rather than just y_{t-1} was found to reduce occasional extreme spikes in the fitted volatility.

8.4.2 Local level with seasonal effect

Here

$$g(\boldsymbol{\theta}) = \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\gamma} + \mathbf{s}.$$

The local approximate internal model is given by

$$
\begin{aligned}
\boldsymbol{\epsilon} &= \boldsymbol{\gamma} + \mathbf{s} + \boldsymbol{e} \\
\boldsymbol{\epsilon}|\boldsymbol{\gamma}, \mathbf{s} &\sim N(\boldsymbol{\gamma} + \mathbf{s}, \boldsymbol{\Sigma}) \\
\gamma_t &= \sum_{j=1}^{J} \phi_j \gamma_{t-j} + b_t \\
s_t &= -\sum_{m=1}^{M-1} s_{t-m} + w_t \\
\mathbf{b} &= \mathbf{D}_\gamma \boldsymbol{\gamma} \sim N(0, \sigma_b^2 \mathbf{I}_{T-J}) \\
\boldsymbol{\omega} &= \mathbf{D}_s \mathbf{s} \sim N(0, \sigma_w^2 \mathbf{I}_{T-M+1}).
\end{aligned}
$$

Hence $\mathbf{D}(\boldsymbol{\gamma}^\top, \mathbf{s}^\top)^\top \sim N(0, \mathbf{M})$, where $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_T)^\top$, $\mathbf{s} = (s_1, s_2, \ldots, s_T)^\top$, $\mathbf{D} =$ matrix diagonal $(\mathbf{D}_\gamma, \mathbf{D}_s)$ and $\mathbf{M} =$ matrix diagonal $(\sigma_b^2 \mathbf{I}_{T-J}, \sigma_w^2 \mathbf{I}_{T-M+1})$.

Hence

$$
\begin{aligned}
f(\boldsymbol{\epsilon}) &= \int f(\boldsymbol{\epsilon}|\boldsymbol{\gamma}, \mathbf{s}) f(\boldsymbol{\gamma}, \mathbf{s}) d\boldsymbol{\gamma} d\mathbf{s} \\
f(\boldsymbol{\epsilon}|\gamma, \mathbf{s}) &= \frac{1}{|2\pi\Sigma|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\boldsymbol{\epsilon} - \boldsymbol{\gamma} - \mathbf{s})^\top \Sigma^{-1} (\boldsymbol{\epsilon} - \boldsymbol{\gamma} - \mathbf{s})\right] \\
f(\boldsymbol{\gamma}, \mathbf{s}) &= \frac{1}{|2\pi\mathbf{M}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\boldsymbol{\gamma}^\top, \mathbf{s}^\top)\mathbf{D}^\top \mathbf{M}^{-1} \mathbf{D}(\boldsymbol{\gamma}^\top, \mathbf{s}^\top)^\top\right].
\end{aligned}
$$

Maximize Q over $(\sigma_e^2, \sigma_b^2, \sigma_w^2, \boldsymbol{\phi})$ gives estimate of $\sigma_e^2, \sigma_b^2, \sigma_w^2, \boldsymbol{\phi}$ where

$$
\begin{aligned}
Q &= \log f(\boldsymbol{\epsilon}|\boldsymbol{\gamma}, \mathbf{s}) + \log f(\boldsymbol{\gamma}, \mathbf{s}) - \frac{1}{2} \log\left|\mathbf{A} + \mathbf{D}^\top \mathbf{M}^{-1} \mathbf{D}\right| + T \log 2\pi \\
\log f(\boldsymbol{\epsilon}|\boldsymbol{\gamma}, \mathbf{s}) &= -\frac{1}{2} \log|2\pi\boldsymbol{\Sigma}| - \frac{1}{2}\left(\boldsymbol{\epsilon} - \boldsymbol{\gamma} - \mathbf{s}\right)^\top \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{\epsilon} - \boldsymbol{\gamma} - \mathbf{s}\right) \\
\log f(\boldsymbol{\gamma}, \mathbf{s}) &= -\frac{1}{2} \log|2\pi\mathbf{M}| - \frac{1}{2}\left(\boldsymbol{\gamma}^\top \mathbf{s}^\top\right) \mathbf{D}^\top \mathbf{M}^{-1} \mathbf{D} \left(\boldsymbol{\gamma}^\top \mathbf{s}^\top\right)^\top
\end{aligned}
$$

where

$$
\mathbf{D} = \begin{pmatrix} \mathbf{D}_\gamma & 0 \\ 0 & \mathbf{D}_s \end{pmatrix},
$$

$$\mathbf{D}_\gamma = \begin{pmatrix} -\phi_J & -\phi_{J-1} & \dots & \dots & -\phi_1 & 1 & 0 & 0 & \dots & \dots & 0 & 0 \\ 0 & -\phi_J & -\phi_{J-1} & \dots & \dots & -\phi_1 & 1 & 0 & \dots & \dots & 0 & 0 \\ & & & & & & & & & & & \\ & & & & & & & & & & & \\ 0 & 0 & 0 & \dots & \dots & 0 & -\phi_J & -\phi_{J-1} & \dots & \dots & -\phi_1 & 1 \end{pmatrix},$$

$$\mathbf{D}_s = \begin{pmatrix} 1 & 1 & \dots & \dots & \dots & \dots & 1 & 0 & 0 & \dots & \dots & 0 & 0 \\ 0 & 1 & 1 & \dots & \dots & \dots & \dots & 1 & 0 & \dots & \dots & 0 & 0 \\ & & & & & & & & & & & & \\ & & & & & & & & & & & & \\ 0 & 0 & 0 & \dots & \dots & 0 & 1 & 1 & \dots & \dots & \dots & \dots & 1 \end{pmatrix},$$

$$\mathbf{A} = \begin{pmatrix} \mathbf{\Sigma}^{-1} & \mathbf{\Sigma}^{-1} \\ \mathbf{\Sigma}^{-1} & \mathbf{\Sigma}^{-1} \end{pmatrix} = \sigma_{\mathbf{e}}^{-\mathbf{2}} \begin{pmatrix} \mathbf{W} & \mathbf{W} \\ \mathbf{W} & \mathbf{W} \end{pmatrix}.$$

Note that $\mathbf{D}_\gamma$ is a $(T-J)$ x T matrix, $\mathbf{D}_s$ is $(T-M+1)$ x T, $\mathbf{\Sigma} = \sigma_e^2\mathbf{W}^{-1}$ is T x T and M is l x l where $l = 2T - J - M + 1$.

The equation for estimating $\boldsymbol{\gamma}$ and $\mathbf{s}$ is given by:

$$\begin{pmatrix} \boldsymbol{\gamma} \\ \mathbf{s} \end{pmatrix} = \left[\mathbf{A} + \mathbf{D}^\top\mathbf{M}^{-1}\mathbf{D}\right]^{-1}\mathbf{A}\begin{pmatrix} \boldsymbol{\epsilon} \\ \mathbf{0} \end{pmatrix} \tag{8.25}$$

where $\mathbf{0}$ is a vector of zeros of length T.

8.4.3 Local level with trend

Here

$$g(\boldsymbol{\theta}) = \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\gamma}.$$

The local approximate internal model is given by

$$\begin{aligned}
\boldsymbol{\epsilon} &= \boldsymbol{\gamma} + \boldsymbol{\psi} + \boldsymbol{e} \\
\boldsymbol{\epsilon}|\boldsymbol{\gamma},\boldsymbol{\psi} &\sim N(\boldsymbol{\gamma} + \boldsymbol{\psi}, \boldsymbol{\Sigma}) \\
\gamma_t &= \sum_{j=1}^{J} \phi_j \gamma_{t-j} + \psi_t + b_t \\
\psi_t &= \sum_{l=1}^{L} \rho_l \psi_{t-l} + d_t \\
\mathbf{b} &= \mathbf{D}_\gamma \boldsymbol{\gamma} - \boldsymbol{\psi} \sim N(0, \sigma_b^2 \mathbf{I}_{T-J}) \\
\mathbf{d} &= \mathbf{D}_\psi \boldsymbol{\psi} \sim N(0, \sigma_d^2 \mathbf{I}_{T-L})
\end{aligned}$$

where $\mathbf{D}_\gamma$ is given in Section 8.4.2 and $\mathbf{D}_\psi$ is similarly defined.

Hence $\mathbf{D}(\boldsymbol{\gamma}^\top, \boldsymbol{\psi}^\top)^\top \sim N(0, \mathbf{M})$, where $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_T)^\top$, $\boldsymbol{\psi} = (\psi_1, \psi_2, \ldots, \psi_T)^\top$, $\mathbf{M}$ = matrix diagonal $(\sigma_b^2 \mathbf{I}_{T-J}, \sigma_d^2 \mathbf{I}_{T-L})$, and $\mathbf{D}$ is given later.

Maximize Q over $(\sigma_e^2, \sigma_b^2, \sigma_d^2, \boldsymbol{\phi}, \boldsymbol{\rho})$ gives estimate of $\sigma_e^2, \sigma_b^2, \sigma_d^2, \boldsymbol{\phi}, \boldsymbol{\rho}$ where Q is given by

$$\begin{aligned}
Q &= \log f(\boldsymbol{\epsilon}|\boldsymbol{\gamma},\boldsymbol{\psi}) + \log f(\boldsymbol{\gamma},\boldsymbol{\psi}) - \frac{1}{2}\log\left|\mathbf{A} + \mathbf{D}^\top \mathbf{M}^{-1}\mathbf{D}\right| + T\log 2\pi \\
\log f(\boldsymbol{\epsilon}|\boldsymbol{\gamma},\boldsymbol{\psi}) &= -\frac{1}{2}\log|2\pi\boldsymbol{\Sigma}| - \frac{1}{2}(\boldsymbol{\epsilon} - \boldsymbol{\gamma})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{\epsilon} - \boldsymbol{\gamma}) \\
\log f(\boldsymbol{\gamma},\boldsymbol{\psi}) &= -\frac{1}{2}\log|2\pi\mathbf{M}| - \frac{1}{2}\left(\boldsymbol{\gamma}^\top \boldsymbol{\psi}^\top\right)\mathbf{D}^\top \mathbf{M}^{-1}\mathbf{D}\left(\boldsymbol{\gamma}^\top \boldsymbol{\psi}^\top\right)^\top
\end{aligned}$$

The equation for estimating $\boldsymbol{\gamma}$ and $\boldsymbol{\psi}$ is given by

$$\begin{pmatrix} \gamma \\ \boldsymbol{\psi} \end{pmatrix} = \left[\mathbf{A} + \mathbf{D}^{\top}\mathbf{M}^{-1}\mathbf{D}\right]^{-1} \mathbf{A} \begin{pmatrix} \boldsymbol{\epsilon} \\ \mathbf{0} \end{pmatrix} \tag{8.26}$$

$$\mathbf{A} = \begin{pmatrix} \Sigma^{-1} & 0 \\ 0 & 0 \end{pmatrix},$$

$$\mathbf{D} = \begin{pmatrix} \mathbf{D}_{\gamma} & \mathbf{D}_{\gamma\psi} \\ 0 & \mathbf{D}_{\psi} \end{pmatrix},$$

where

$$\mathbf{D}_{\gamma\psi} = (-\mathbf{I}_{T-1}\ \mathbf{0})$$

$\mathbf{D}_{\gamma\psi}$ is $(T-J)$ x T with the first $(J-1)$ rows removed from $\mathbf{D}_{\gamma\psi}$

8.4.4 Local level with trend and seasonality

Here

$$g(\boldsymbol{\theta}) = \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\gamma} + \mathbf{s}.$$

The local approximate internal model is given by

$$
\begin{aligned}
\boldsymbol{\epsilon} &= \boldsymbol{\gamma} + \mathbf{s} + \boldsymbol{e} \\
\boldsymbol{\epsilon}|\boldsymbol{\gamma}, \mathbf{s} &\sim N(\boldsymbol{\gamma} + \mathbf{s}, \boldsymbol{\Sigma}) \\
\gamma_t &= \sum_{j=1}^{J} \phi_j \gamma_{t-j} + \psi_t + b_t \\
\psi_t &= \sum_{l=1}^{L} \rho_l \psi_{t-l} + d_t \\
s_t &= -\sum_{m=1}^{M-1} s_{t-m} + w_t \\
\mathbf{b} &= \mathbf{D}_\gamma \boldsymbol{\gamma} - \boldsymbol{\psi} \sim N(0, \sigma_b^2 \mathbf{I}_{T-J}) \\
\mathbf{d} &= \mathbf{D}_\psi \boldsymbol{\psi} \sim N(0, \sigma_d^2 \mathbf{I}_{T-L}) \\
\boldsymbol{\omega} &= \mathbf{D}_s \mathbf{s} \sim N(0, \sigma_w^2 \mathbf{I}_{T-M+1})
\end{aligned}
$$

where $\mathbf{D}_\gamma$ and $\mathbf{D}_s$ are defined in Section 8.4.2 and $\mathbf{D}_\psi$ is defined similarly to $\mathbf{D}_\gamma$.

Hence $\mathbf{D}(\boldsymbol{\gamma}^\top, \boldsymbol{\psi}^\top, \mathbf{s}^\top)^\top \sim N(0, \mathbf{M})$, where $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_T)^\top$, $\boldsymbol{\psi} = (\psi_1, \psi_2, \ldots, \psi_T)^\top$, $\mathbf{s} = (s_1, s_2, \ldots, s_T)^\top$, $\mathbf{M}$ = matrix diagonal $(\sigma_b^2 \mathbf{I}_{T-J}, \sigma_d^2 \mathbf{I}_{T-L}, \sigma_w^2 \mathbf{I}_{T-M+1})$, and $\mathbf{D}$ is given below.

Maximize Q over $(\sigma_e^2, \sigma_b^2, \sigma_d^2, \sigma_w^2, \boldsymbol{\phi}, \boldsymbol{\rho})$ gives estimate of $\sigma_e^2, \sigma_b^2, \sigma_d^2, \sigma_w^2, \boldsymbol{\phi}, \boldsymbol{\rho}$, where Q is given by

$$
\begin{aligned}
Q &= \log f(\boldsymbol{\epsilon}|\boldsymbol{\gamma}, \mathbf{s}) + \log f(\boldsymbol{\gamma}, \boldsymbol{\psi}, \mathbf{s}) - \frac{1}{2} \log \left|\mathbf{A} + \mathbf{D}^\top \mathbf{M}^{-1} \mathbf{D}\right| + \frac{3T}{2} \log 2\pi \\
\log f(\boldsymbol{\epsilon}|\boldsymbol{\gamma}, \mathbf{s}) &= -\frac{1}{2} \log |2\pi \boldsymbol{\Sigma}| - \frac{1}{2} (\boldsymbol{\epsilon} - \boldsymbol{\gamma} - \mathbf{s})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{\epsilon} - \boldsymbol{\gamma} - \mathbf{s}) \\
\log f(\boldsymbol{\gamma}, \boldsymbol{\psi}, \mathbf{s}) &= -\frac{1}{2} \log |2\pi \mathbf{M}| - \frac{1}{2} \left(\boldsymbol{\gamma}^\top \boldsymbol{\psi}^\top \mathbf{s}^\top\right) \mathbf{D}^\top \mathbf{M}^{-1} \mathbf{D} \left(\boldsymbol{\gamma}^\top \boldsymbol{\psi}^\top \mathbf{s}^\top\right)^\top
\end{aligned}
$$

$$\mathbf{D} = \begin{pmatrix} \mathbf{D}_\gamma & \mathbf{D}_{\gamma\psi} & 0 \\ 0 & \mathbf{D}_\psi & 0 \\ 0 & 0 & \mathbf{D}_s \end{pmatrix}$$

where $\mathbf{D}_{\gamma\psi}$ is defined in Section 8.4.3 and

$$\mathbf{A} = \begin{pmatrix} \boldsymbol{\Sigma}^{-1} & 0 & \boldsymbol{\Sigma}^{-1} \\ 0 & 0 & 0 \\ \boldsymbol{\Sigma}^{-1} & 0 & \boldsymbol{\Sigma}^{-1} \end{pmatrix}.$$

The equation for estimating $\boldsymbol{\gamma}, \boldsymbol{\psi}$ and $\mathbf{s}$:

$$\begin{pmatrix} \gamma \\ \psi \\ \mathbf{s} \end{pmatrix} = \left[\mathbf{A} + \mathbf{D}^\top \mathbf{M}^{-1} \mathbf{D}\right]^{-1} \mathbf{A} \begin{pmatrix} \boldsymbol{\epsilon} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix} \tag{8.27}$$

8.4.5 Local level with random coefficient of an explanatory variable

Here

$$g(\boldsymbol{\theta}) = \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\gamma}.$$

The local approximate internal model is given by

$$
\begin{aligned}
\boldsymbol{\epsilon} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\gamma} + \boldsymbol{e} \\
\boldsymbol{\epsilon}|\boldsymbol{\gamma}, \boldsymbol{\beta} &\sim N(\boldsymbol{\gamma} + \mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}) \\
\gamma_t &= \sum_{j=1}^{J} \phi_j \gamma_{t-j} + b_t \\
\beta_t &= \sum_{j=1}^{J'} \phi'_j \beta_{t-j} + v_t \\
\mathbf{b} &= \mathbf{D}_\gamma \boldsymbol{\gamma} \sim N(0, \sigma_b^2 \mathbf{I}_{T-J}) \\
\mathbf{v} &= \mathbf{D}_\beta \boldsymbol{\beta} \sim N(0, \sigma_v^2 \mathbf{I}_{T-J'})
\end{aligned}
$$

$\mathbf{D}(\boldsymbol{\gamma}^\top, \boldsymbol{\beta}^\top)^\top \sim N(0, \mathbf{M})$, where $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_T)^\top$, $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_T)^\top$, $\mathbf{M} =$ matrix diagonal $(\sigma_b^2 \mathbf{I}_{T-J}, \sigma_v^2 \mathbf{I}_{T-J'})$, and $\mathbf{D}$ is defined below.

Maximize Q over $(\sigma_e^2, \sigma_b^2, \sigma_v^2, \boldsymbol{\phi}, \boldsymbol{\phi}')$ gives estimate of $(\sigma_e^2, \sigma_b^2, \sigma_v^2, \boldsymbol{\phi}, \boldsymbol{\phi}')$ where the Q is given by

$$
\begin{aligned}
Q &= \log f(\boldsymbol{\epsilon}|\boldsymbol{\gamma}, \boldsymbol{\beta}) + \log f(\boldsymbol{\gamma}, \boldsymbol{\beta}) - \frac{1}{2} \log \left| \mathbf{A} + \mathbf{D}^\top \mathbf{M}^{-1} \mathbf{D} \right| + T \log 2\pi \\
\log f(\boldsymbol{\epsilon}|\boldsymbol{\gamma}, \boldsymbol{\beta}) &= -\frac{1}{2} \log |2\pi \boldsymbol{\Sigma}| - \frac{1}{2} (\boldsymbol{\epsilon} - \boldsymbol{\gamma} - \mathbf{X}\boldsymbol{\beta})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{\epsilon} - \boldsymbol{\gamma} - \mathbf{X}\boldsymbol{\beta}) \\
\log f(\boldsymbol{\gamma}, \boldsymbol{\beta}) &= -\frac{1}{2} \log |2\pi \mathbf{M}| - \frac{1}{2} \left(\boldsymbol{\gamma}^\top \boldsymbol{\beta}^\top \right) \mathbf{D}^\top \mathbf{M}^{-1} \mathbf{D} \left(\boldsymbol{\gamma}^\top \boldsymbol{\beta}^\top \right)^\top
\end{aligned}
$$

$$
\mathbf{D} = \begin{pmatrix} \mathbf{D}_\gamma & 0 \\ 0 & \mathbf{D}_\beta \end{pmatrix}
$$

$$
\mathbf{A} = \begin{pmatrix} \boldsymbol{\Sigma}^{-1} & \boldsymbol{\Sigma}^{-1}\mathbf{X} \\ \mathbf{X}^\top \boldsymbol{\Sigma}^{-1} & \mathbf{X}^\top \boldsymbol{\Sigma}^{-1} \mathbf{X} \end{pmatrix}
$$

The equation for estimating $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ is given by

$$\begin{pmatrix} \boldsymbol{\gamma} \\ \boldsymbol{\beta} \end{pmatrix} = \left[\mathbf{A} + \mathbf{D}^{\top}\mathbf{M}^{-1}\mathbf{D}\right]^{-1} \mathbf{A} \begin{pmatrix} \boldsymbol{\epsilon} \\ \mathbf{0} \end{pmatrix} \tag{8.28}$$

8.5 Effective degrees of freedom in the GEST

The total effective degrees of freedom of the fitted model, *df*, combines those of the models for $\theta_1, \theta_2, \theta_3$ and θ_4, i.e. μ, σ, ν and τ, given by df_1, df_2, df_3 and df_4 respectively. Hence,

$$df = df_1 + df_2 + df_3 + df_4$$

where

$$df_k = p_k + d_k$$

for $k = 1, 2, 3, 4$, and p_k is the length of $\boldsymbol{\beta}_k$, while d_k, the effective degrees of freedom for the random effects in the model for θ_k.

8.5.1 Effective degrees of freedom for the local level and seasonal structural model

Here the subscript k is omitted and a local level and seasonal model for θ (8.4.2) is assumed.

Let $\mathbf{B} = \left[\boldsymbol{\Sigma}^{-1} + \sigma_b^{-2}\mathbf{D}^{\top}\mathbf{D}\right]^{-1}\boldsymbol{\Sigma}^{-1}$ and let $\hat{\mathbf{B}}, \hat{\boldsymbol{\Sigma}}, \hat{\mathbf{D}}, \hat{\boldsymbol{\gamma}}$ and $\hat{\sigma}_b^{-2}$ be the values of $\mathbf{B}, \boldsymbol{\Sigma}, \mathbf{D}, \boldsymbol{\gamma}$ and σ_b^{-2} on convergence of the GEST fitting procedure (see Section 8.3.4).

On convergence,

$$\hat{\boldsymbol{\gamma}} = \hat{\mathbf{B}}\boldsymbol{\epsilon}.$$

Hence d, the effective degrees of freedom used in the GEST model, is

$$d = tr\left[\hat{\mathbf{B}}\right] = tr\left\{\left[\hat{\boldsymbol{\Sigma}}^{-1} + \hat{\sigma}_b^{-2}\hat{\mathbf{D}}^\top\hat{\mathbf{D}}\right]^{-1}\hat{\boldsymbol{\Sigma}}^{-1}\right\} \tag{8.29}$$

As d is difficult to calculate directly for large T, it can be calculated by setting

$$\partial Q/\partial\sigma_b^2 = 0$$

$$\partial Q/\partial\sigma_w^2 = 0$$

for the local level and seasonal structural model, and using the following result, (see Rigby and Stasinopoulos, (2013), Appendix A):

$$\frac{\partial}{\partial x}\log|x\mathbf{A}+\mathbf{B}| = tr\left[(x\mathbf{A}+\mathbf{B})^{-1}\mathbf{A}\right]$$

where x is a scalar and $\mathbf{A}$ and $\mathbf{B}$ are r x r matrices (provided $|x\mathbf{A}+\mathbf{B}| \neq 0$).

On convergence of maximising the Q:

$$\begin{aligned}
\frac{\partial Q}{\partial\sigma_b^2} &= \frac{1}{2\sigma_b^4}tr\left\{\left[\mathbf{A}+\mathbf{D}^\top\mathbf{M}^{-1}\mathbf{D}\right]^{-1}\begin{pmatrix}\mathbf{D}_\gamma^\top\mathbf{D}_\gamma & 0\\ 0 & 0\end{pmatrix}\right\} + \frac{1}{2\sigma_b^4}\boldsymbol{\gamma}^\top\mathbf{D}_\gamma^\top\mathbf{D}_\gamma\boldsymbol{\gamma} - \frac{(T-J)}{2\sigma_b^2},\\
\frac{\partial Q}{\partial\sigma_w^2} &= \frac{1}{2\sigma_w^4}tr\left\{\left[\mathbf{A}+\mathbf{D}^\top\mathbf{M}^{-1}\mathbf{D}\right]^{-1}\begin{pmatrix}0 & 0\\ 0 & \mathbf{D}_s^\top\mathbf{D}_s\end{pmatrix}\right\} + \frac{1}{2\sigma_w^4}\mathbf{s}^\top\mathbf{D}_s^\top\mathbf{D}_s\mathbf{s} - \frac{(T-M+1)}{2\sigma_w^2},
\end{aligned}$$

$$\begin{aligned}
\frac{\partial Q}{\partial \sigma_b^2} &= 0 \Rightarrow \frac{1}{2\sigma_b^4} tr \left\{ \left[\mathbf{A} + \mathbf{D}^\top \mathbf{M}^{-1}\mathbf{D}\right]^{-1} \begin{pmatrix} \mathbf{D}_\gamma^\top \mathbf{D}_\gamma & 0 \\ 0 & 0 \end{pmatrix} \right\} = -\frac{1}{2\sigma_b^4} \boldsymbol{\gamma}^\top \mathbf{D}_\gamma^\top \mathbf{D}_\gamma \boldsymbol{\gamma} \\
&+ \frac{(T-J)}{2\sigma_b^2}, \\
\frac{\partial Q}{\partial \sigma_w^2} &= 0 \Rightarrow \frac{1}{2\sigma_w^4} tr \left\{ \left[\mathbf{A} + \mathbf{D}^\top \mathbf{M}^{-1}\mathbf{D}\right]^{-1} \begin{pmatrix} 0 & 0 \\ 0 & \mathbf{D}_s^\top \mathbf{D}_s \end{pmatrix} \right\} = -\frac{1}{2\sigma_w^4} \mathbf{s}^\top \mathbf{D}_s^\top \mathbf{D}_s \mathbf{s} \\
&+ \frac{(T-M+1)}{2\sigma_w^2},
\end{aligned}$$

$$\begin{aligned}
tr \left\{ \left[\mathbf{A} + \mathbf{D}^\top \mathbf{M}^{-1}\mathbf{D}\right]^{-1} \begin{pmatrix} \sigma_b^{-2}\mathbf{D}_\gamma^\top \mathbf{D}_\gamma & 0 \\ 0 & 0 \end{pmatrix} \right\} &= -\frac{1}{\sigma_b^2} \boldsymbol{\gamma}^\top \mathbf{D}_\gamma^\top \mathbf{D}_\gamma \boldsymbol{\gamma} + (T-J), \\
tr \left\{ \left[\mathbf{A} + \mathbf{D}^\top \mathbf{M}^{-1}\mathbf{D}\right]^{-1} \begin{pmatrix} 0 & 0 \\ 0 & \sigma_w^{-2}\mathbf{D}_s^\top \mathbf{D}_s \end{pmatrix} \right\} &= -\frac{1}{\sigma_w^2} \mathbf{s}^\top \mathbf{D}_s^\top \mathbf{D}_s \mathbf{s} + (T-M+1).
\end{aligned}$$

Hence, adding the two traces gives:

$$\begin{aligned}
& tr \left\{ \left[\mathbf{A} + \mathbf{D}^\top \mathbf{M}^{-1}\mathbf{D}\right]^{-1} \begin{pmatrix} \sigma_b^{-2}\mathbf{D}_\gamma^\top \mathbf{D}_\gamma & 0 \\ 0 & \sigma_w^{-2}\mathbf{D}_s^\top \mathbf{D}_s \end{pmatrix} \right\} \\
&= 2T - J - M + 1 - \frac{1}{\sigma_b^2} \boldsymbol{\gamma}^\top \mathbf{D}_\gamma^\top \mathbf{D}_\gamma \boldsymbol{\gamma} - \frac{1}{\sigma_w^2} \mathbf{s}^\top \mathbf{D}_s^\top \mathbf{D}_s \mathbf{s}.
\end{aligned}$$

Knowing that:

$$tr \left\{ \left[\mathbf{A} + \mathbf{D}^\top \mathbf{M}^{-1}\mathbf{D}\right]^{-1} \left[\mathbf{A} + \mathbf{D}^\top \mathbf{M}^{-1}\mathbf{D}\right] \right\} = tr\{\mathbf{I}_{2T}\} = 2T$$

Hence,

$$\begin{aligned} d &= tr\left\{\left[\mathbf{A}+\mathbf{D}^{\top}\mathbf{M}^{-1}\mathbf{D}\right]^{-1}\mathbf{A}\right\} \\ &= 2T - tr\left\{\left[\mathbf{A}+\mathbf{D}^{\top}\mathbf{M}^{-1}\mathbf{D}\right]^{-1}\begin{pmatrix} \sigma_b^{-2}\mathbf{D}_\gamma^{\top}\mathbf{D}_\gamma & 0 \\ 0 & \sigma_w^{-2}\mathbf{D}_s^{\top}\mathbf{D}_s \end{pmatrix}\right\} \end{aligned}$$

$$\begin{aligned} d &= 2T - \left\{2T - J - M + 1 - \frac{1}{\sigma_b^2}\boldsymbol{\gamma}^{\top}\mathbf{D}_\gamma^{\top}\mathbf{D}_\gamma\boldsymbol{\gamma} - \frac{1}{\sigma_w^2}\mathbf{s}^{\top}\mathbf{D}_s^{\top}\mathbf{D}_s\mathbf{s}\right\} \\ &= J + M - 1 + \frac{1}{\sigma_b^2}\boldsymbol{\gamma}^{\top}\mathbf{D}_\gamma^{\top}\mathbf{D}_\gamma\boldsymbol{\gamma} + \frac{1}{\sigma_w^2}\mathbf{s}^{\top}\mathbf{D}_s^{\top}\mathbf{D}_s\mathbf{s}. \end{aligned} \tag{8.30}$$

Hence, for each distribution parameter, d is calculated using the values $\hat{\boldsymbol{\gamma}}$, $\hat{\sigma}_b^2$ and $\hat{\sigma}_w^2$ on convergence of the GEST fitting algorithm.

8.5.2 Effective degrees of freedom for the local level structural model

The effective degrees of freedom for local level without seasonal effect is equal to:

$$d = J + \frac{1}{\sigma_b^2}\boldsymbol{\gamma}^{\top}\mathbf{D}_\gamma^{\top}\mathbf{D}_\gamma\boldsymbol{\gamma}.$$

Hence, for each distribution parameter, d is calculated using the values $\hat{\boldsymbol{\gamma}}$, $\hat{\sigma}_b^2$ on convergence of the GEST fitting algorithm.

8.5.3 Effective degrees of freedom for the local level with trend and seasonal structural model

The effective degrees of freedom for the local level with trend and seasonal structural model (8.4.4) is given by setting:

$$\partial Q / \partial \sigma_b^2 = 0$$

$$\partial Q / \partial \sigma_d^2 = 0$$

$$\partial Q / \partial \sigma_w^2 = 0$$

Using the following result, (see Rigby and Stasinopoulos, (2013), Appendix A):

$$\frac{\partial}{\partial x} \log |x\mathbf{A} + \mathbf{B}| = tr\left[(x\mathbf{A} + \mathbf{B})^{-1}\mathbf{A}\right]$$

where x is a scalar and $\mathbf{A}$ and $\mathbf{B}$ are r x r matrices (provided $|x\mathbf{A} + \mathbf{B}| \neq 0$).

On convergence of maximising the Q:

$$\begin{aligned}
\frac{\partial Q}{\partial \sigma_b^2} &= \frac{1}{2\sigma_b^4} tr \left\{ \left[\mathbf{A} + \mathbf{D}^\top \mathbf{M}^{-1}\mathbf{D}\right]^{-1} \begin{pmatrix} \mathbf{D}_\gamma^\top \mathbf{D}_\gamma & \mathbf{D}_\gamma^\top \mathbf{D}_{\gamma\psi} & 0 \\ \mathbf{D}_{\gamma\psi}^\top \mathbf{D}_\gamma & \mathbf{D}_{\gamma\psi}^\top \mathbf{D}_{\gamma\psi} & 0 \\ 0 & 0 & 0 \end{pmatrix} \right\} + \frac{1}{2\sigma_b^4} \boldsymbol{\gamma}^\top \mathbf{D}_\gamma^\top \mathbf{D}_\gamma \boldsymbol{\gamma} \\
&\quad - \frac{(T-J)}{2\sigma_b^2}, \\
\frac{\partial Q}{\partial \sigma_d^2} &= \frac{1}{2\sigma_d^4} tr \left\{ \left[\mathbf{A} + \mathbf{D}^\top \mathbf{M}^{-1}\mathbf{D}\right]^{-1} \begin{pmatrix} 0 & 0 & 0 \\ 0 & \mathbf{D}_\psi^\top \mathbf{D}_\psi & 0 \\ 0 & 0 & 0 \end{pmatrix} \right\} + \frac{1}{2\sigma_d^4} \boldsymbol{\psi}^\top \mathbf{D}_\psi^\top \mathbf{D}_\psi \boldsymbol{\psi} \\
&\quad - \frac{(T-J-L+1)}{2\sigma_d^2}, \\
\frac{\partial Q}{\partial \sigma_w^2} &= \frac{1}{2\sigma_w^4} tr \left\{ \left[\mathbf{A} + \mathbf{D}^\top \mathbf{M}^{-1}\mathbf{D}\right]^{-1} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \mathbf{D}_s^\top \mathbf{D}_s \end{pmatrix} \right\} + \frac{1}{2\sigma_w^4} \mathbf{s}^\top \mathbf{D}_s^\top \mathbf{D}_s \mathbf{s} \\
&\quad - \frac{(T-M+1)}{2\sigma_w^2},
\end{aligned}$$

Setting

$$\frac{\partial Q}{\partial \sigma_b^2} = 0 \Rightarrow$$

$$\frac{1}{2\sigma_b^4} tr \left\{ \left[\mathbf{A} + \mathbf{D}^\top \mathbf{M}^{-1}\mathbf{D}\right]^{-1} \begin{pmatrix} \mathbf{D}_\gamma^\top \mathbf{D}_\gamma & \mathbf{D}_\gamma^\top \mathbf{D}_{\gamma\psi} & 0 \\ \mathbf{D}_{\gamma\psi}^\top \mathbf{D}_\gamma & \mathbf{D}_{\gamma\psi}^\top \mathbf{D}_{\gamma\psi} & 0 \\ 0 & 0 & 0 \end{pmatrix} \right\} = -\frac{1}{2\sigma_b^4} \boldsymbol{\gamma}^\top \mathbf{D}_\gamma^\top \mathbf{D}_\gamma \boldsymbol{\gamma} + \frac{(T-J)}{2\sigma_b^2},$$

$$\frac{\partial Q}{\partial \sigma_d^2} = 0 \Rightarrow$$

$$\frac{1}{2\sigma_d^4} tr \left\{ \left[\mathbf{A}+\mathbf{D}^\top\mathbf{M}^{-1}\mathbf{D}\right]^{-1} \begin{pmatrix} 0 & 0 & 0 \\ 0 & \mathbf{D}_\psi^\top\mathbf{D}_\psi & 0 \\ 0 & 0 & 0 \end{pmatrix} \right\} = -\frac{1}{2\sigma_d^4}\boldsymbol{\psi}^\top\mathbf{D}_\psi^\top\mathbf{D}_\psi\boldsymbol{\psi} + \frac{(T-J-L+1)}{2\sigma_d^2},$$

$$\frac{\partial Q}{\partial \sigma_w^2} = 0 \Rightarrow$$

$$\frac{1}{2\sigma_w^4} tr \left\{ \left[\mathbf{A}+\mathbf{D}^\top\mathbf{M}^{-1}\mathbf{D}\right]^{-1} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \mathbf{D}_s^\top\mathbf{D}_s \end{pmatrix} \right\} = -\frac{1}{2\sigma_w^4}\mathbf{s}^\top\mathbf{D}_s^\top\mathbf{D}_s\mathbf{s} + \frac{(T-M+1)}{2\sigma_w^2}.$$

Hence, adding the three traces gives:

$$\begin{aligned} tr\left\{\left[\mathbf{A}+\mathbf{D}^\top\mathbf{M}^{-1}\mathbf{D}\right]^{-1}\mathbf{D}^\top\mathbf{M}^{-1}\mathbf{D}\right\} &= \frac{1}{\sigma_b^2}(\boldsymbol{\gamma}^\top\boldsymbol{\psi}^\top)\begin{pmatrix} \mathbf{D}_\gamma^\top\mathbf{D}_\gamma & \mathbf{D}_\gamma^\top\mathbf{D}_{\gamma\psi} \\ \mathbf{D}_\gamma^\top\mathbf{D}_\gamma & \mathbf{D}_\gamma^\top\mathbf{D}_{\gamma\psi} \end{pmatrix}\begin{pmatrix} \boldsymbol{\gamma} \\ \boldsymbol{\psi} \end{pmatrix} \\ &+\frac{1}{\sigma_d^2}\boldsymbol{\psi}^\top\mathbf{D}_\psi^\top\mathbf{D}_\psi\boldsymbol{\psi} + \frac{1}{\sigma_w^2}\mathbf{s}^\top\mathbf{D}_s^\top\mathbf{D}_s\mathbf{s} \\ &+(T-J)+(T-J-L+1)+(T-M+1). \end{aligned}$$

Knowing that:

$$tr\left\{\left[\mathbf{A}+\mathbf{D}^\top\mathbf{M}^{-1}\mathbf{D}\right]^{-1}\left[\mathbf{A}+\mathbf{D}^\top\mathbf{M}^{-1}\mathbf{D}\right]\right\} = tr\{\mathbf{I}_{3T}\} = 3T$$

Hence,

$$d = tr\left\{\left[\mathbf{A}+\mathbf{D}^\top\mathbf{M}^{-1}\mathbf{D}\right]^{-1}\mathbf{A}\right\} = 3T - tr\left[\left(\mathbf{A}+\mathbf{D}^\top\mathbf{M}^{-1}\mathbf{D}\right)^{-1}\mathbf{D}^\top\mathbf{M}^{-1}\mathbf{D}\right]$$

Hence the effective degrees of freedom d is given by

$$\begin{aligned} d &= 2J + L + M - 2 \\ &\quad + \frac{1}{\sigma_b^2}(\boldsymbol{\gamma}^\top \boldsymbol{\psi}^\top) \begin{pmatrix} \mathbf{D}_\gamma^\top \mathbf{D}_\gamma & \mathbf{D}_\gamma^\top \mathbf{D}_{\gamma\psi} \\ \mathbf{D}_\gamma^\top \mathbf{D}_\gamma & \mathbf{D}_\gamma^\top \mathbf{D}_{\gamma\psi} \end{pmatrix} \begin{pmatrix} \boldsymbol{\gamma} \\ \boldsymbol{\psi} \end{pmatrix} \\ &\quad + \frac{1}{\sigma_d^2} \boldsymbol{\psi}^\top \mathbf{D}_\psi^\top \mathbf{D}_\psi \boldsymbol{\psi} + \frac{1}{\sigma_w^2} \mathbf{s}^\top \mathbf{D}_s^\top \mathbf{D}_s \mathbf{s}. \end{aligned} \tag{8.31}$$

Note that the final formula for the effective degrees of freedom d, (8.31), has divisors $\left(\frac{1}{\sigma_b^2}, \frac{1}{\sigma_d^2}, \frac{1}{\sigma_w^2}\right)$. The estimates of these variances are very small in Chapter 9, e.g. p220, so the inverse of these variances will be very large, with a very large effective degrees of freedom. However, the fitted model in p220 is the random walk of order 2 local level and deterministic smooth seasonal. In the random walk of order 2 local level the fitted local level is a very smooth curve with no much variability, (see Figure 9.14 on page 229), so the variance is very small and the inverse of the variance is very large, but because the smoothing matrix $\mathbf{D}_\gamma$ is a second order differencing matrix, the $\hat{\boldsymbol{\gamma}}^\top \mathbf{D}_\gamma^\top \mathbf{D}_\gamma \hat{\boldsymbol{\gamma}}$ becomes smaller and it compensates the very large value of the inverse of the variance. Hence, as the random walk have a higher order, the fitted local level will be very smooth with a very small variance, but the higher order differencing smoothing matrix makes the sum of the squared differenced fitted values smaller to compensate the very large value of the inverse of the variance, so when the $\hat{\boldsymbol{\gamma}}^\top \mathbf{D}_\gamma^\top \mathbf{D}_\gamma \hat{\boldsymbol{\gamma}}$ is multiplied by the inverse of the variance, the effective degrees of freedom becomes smaller.

Chapter 9

Examples in the GEST

9.1 Introduction

This chapter provides practical illustrations of the capabilities of the GEST model fitting procedure and diagnostic facilities to model univariate non-Gaussian time series data. Two examples of continuous data: pound/dollar daily exchange rates and Standard and Poor 500 stock index, and two examples of counts data: van drivers killed in road accidents in the UK and polio incidence in the United States, are modelled and analysed using the GEST model with different distributions and their best model is chosen using Akaike information criteria (AIC) (Akaike, 1983).

The purpose of modeling the pound/dollar daily exchange rates, is to model the stochastic volatility with the GEST model and compare the GEST estimates with previous models; for modelling the Standard and Poor 500 stock index returns, is to model the asymmetric stochastic volatility, the conditional skewness parameter and the conditional kurtosis parameter jointly with the GEST model and compare the GEST estimates with the GARCH and APARCH models; and the purpose for modelling counts data (e.g. van drivers killed in the UK, and the polio incidence in the US), is to model the stochastic seasonality with GEST model, investigate whether the data is overdispersed by using the negative binomial type I conditional distribution, applying different models for the local level, for example random walk order 1, random walk order 2, autoregressive order 1, including explanatory variables.

Modelling time series of counts with classical Gaussian models is inappropriate and it is necessary to consider non-Gaussian time series models. Univariate time series of counts are discrete counts with distinct and non-negative integer values. Analysing discrete data is one of the rapidly developing areas in time series modelling, Zeger (1988) modelled counts time series with a Poisson distribution, Davis *et al.* (2000) modelled the counts by the Poisson distribution given a latent process and Davis *et al.* (2009) used a negative binomial distribution for the same data.

9.2 Pound sterling and US dollar exchange rate

The data in this example are the pound sterling and US dollar daily exchange rates from 01-10-1981 to 28-06-1985. Harvey *et al.* (1994), Shephard and Pitt (1997), Kim *et al.* (1998) and Durbin and Koopman (2000) fitted a stochastic volatility model to pound/dollar exchange rates' returns with a conditional normal distribution, to model the volatility clustering effect of the returns. Their stochastic volatility model is defined as:

$$\begin{aligned} y_t &= \sigma_t \epsilon_t = \sigma \epsilon_t \exp\left(\frac{h_t}{2}\right) \\ h_t &= \phi h_{t-1} + \eta_t \end{aligned} \tag{9.1}$$

where $\epsilon_t \sim N(0,1)$, and $\eta_t \sim N(0, \sigma_\eta^2)$.

The efficiency of Harvey *et al.* (1994) estimator was improved by Shephard and

Pitt (1997) by using MCMC method, and this was again improved in speed by Kim *et al.* (1998). Durbin and Koopman (2000) developed an importance-sampling method for fitting the stochastic volatility model.

Using the GEST model, let Y_t be the pound/dollar exchange rates' returns, and consider the conditional normal distribution, $NO(y_t|\mu_t,\sigma_t)$, of the response variable Y_t.

9.2.1 Conditional normal distribution

$$\begin{aligned} Y_t|\mu_t,\sigma_t &\sim NO(\mu_t,\sigma_t) \\ \mu_t &= \beta_{1,0} \\ \log(\sigma_t) &= \beta_{2,0}+\gamma_{2,t} \\ \gamma_{2,t} &= \phi\gamma_{2,t-1}+b_{2,t} \end{aligned} \tag{9.2}$$

where $b_{2,t} \sim N(0,\sigma_b^2)$.

Table 9.1: Model comparison of the estimated parameters

Stochastic Volatility	σ_η^2	ϕ
Harvey *et al.*	0.0069	0.9912
Durbin and Koopman, classical approach	0.01165	0.9866
Durbin and Koopman, Bayesian approach	0.007425	0.9731
GEST model	0.007182	0.9744

The R commands for fitting the GEST model are given in Appendix D

From Table (9.1), the GEST estimation of the hyperparameters are very similar to the Bayesian approach of Durbin and Koopman (2000), indicating an accurate fit of the GEST model for the stochastic volatility of the pound sterling and US dollar exchange rates' returns. Note that the GEST models the stochastic volatility with $\log(\sigma_t)$ and not with $\log(\sigma_t^2)$.

Figure (9.1) shows the time series of the returns of the pound/dollar daily exchange rates from 01-10-1981 to 28-06-1985, and Figure (9.2) shows the fitted stochastic volatility of the GEST model. Clearly the returns exhibit volatility clustering effect (large changes in returns tend to cluster together, resulting in persistence of the amplitudes of return changes; or large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes; Mandelbrot, 1963). The GEST models the volatility clustering effect of the pound/dollar returns with a stochastic volatility model for $\log(\sigma_t)$ as an autoregressive order 1 process. As shown in Figure (9.2), $\hat{\sigma}_t$ increases when the volatility clustering effect is high, and decreases when the clustering is low. Figure (9.3) shows the QQ plot of the residuals of the fitted stochastic volatility with the GEST using the normal distribution, they appear satisfactory.

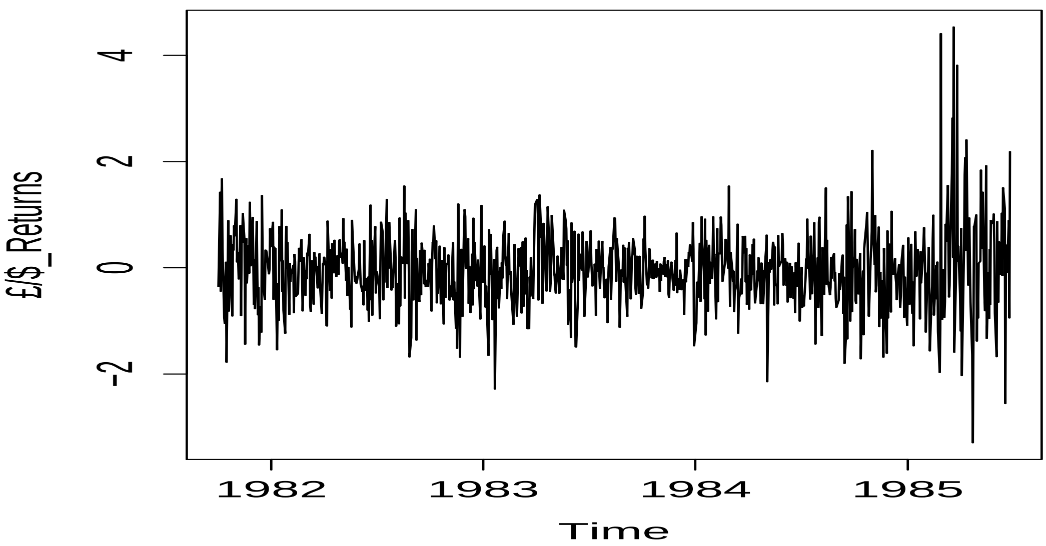


Figure 9.1: The returns for pound/dollar daily exchange rates from 01-10-1981 to 28-06-1985

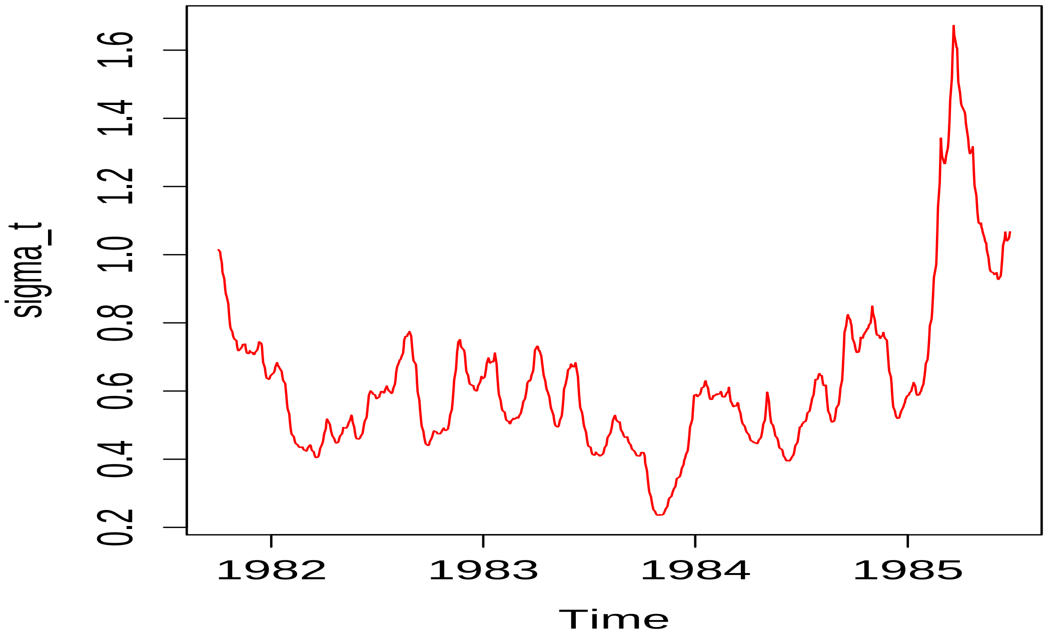


Figure 9.2: The fitted stochastic volatility with the GEST model for the pound/dollar daily returns.

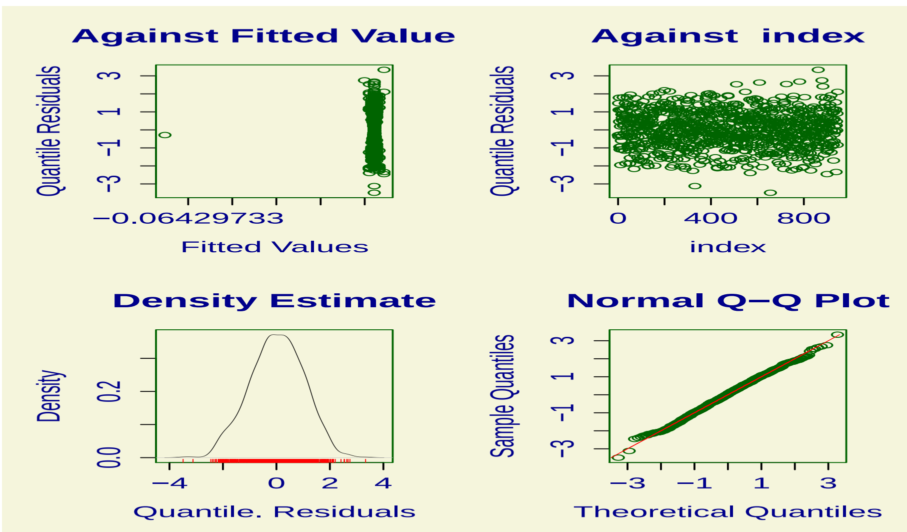


Figure 9.3: The QQ plot of the residuals of the fitted stochastic volatility with the GEST model using a normal distribution to pound/dollar daily returns.

9.3 Standard and Poor 500 stock index

9.3.1 Introduction

In this example, the GEST model is illustrated by an application to financial daily returns of the S&P 500 stock index. The data, taken from the Yahoo.finance website, are daily closing prices of the S&P 500 stock index from 02/01/1980 to 31/12/2012, i.e. 8324 daily observations. Here the flexibility of the GEST model is demonstrated in modelling the returns of the S&P 500 index based upon time-varying estimates of distribution parameters μ_t, σ_t, ν_t and τ_t, representing location, scale, skewness and kurtosis parameters of the conditional distribution.

In this section 9.3.1, the author compares different GEST models and selects the best model using the Akaike information criterion (AIC). Then, in sections 9.3.4 and 9.3.5, the chosen GEST model is compared with the APARCH model using AIC and normalized probability integral transform (normalized PIT) residuals to assess the adequacy of each fitted model. Finally, in section 9.3.6, the chosen GEST model is extended for modelling the conditional mean of S&P 500 stock index returns as a random walk order 2 process, in addition to stochastic volatility, skewness and kurtosis. The skew Student-t (SST) distribution, which is a skew heavy tailed distribution, is used in the GEST model for the conditional distribution of the S&P 500 daily returns.

9.3.2 Conditional skew Student t distribution

Let P_t be the price at time t and $y_t = ln(P_t/P_{t-1}) * 100$ be the return of the S&P 500 over the period 02/01/1980 to 31/12/2012. The conditional probability density function $f(y_t|\mu_t, \sigma_t, \nu_t, \tau_t)$ of the S&P 500 index returns y_t is modelled by a skew t-distribution, SST described in Appendix B, using the GEST to model the stochastic volatility, stochastic skewness and stochastic kurtosis parameters of the returns using an autoregressive model for $\log(\sigma_t)$, and a random walk for $\log(\nu_t)$ and $\log(\tau_t - 2)$. The model is given by

$$
\begin{aligned}
Y_t|\mu_t, \sigma_t, \nu_t, \tau_t &\sim SST(\mu_t, \sigma_t, \nu_t, \tau_t) \\
\mu_t &= \beta_{1,0} \\
\log(\sigma_t) &= \beta_{2,0} + \gamma_{2,t} \\
\log(\nu_t) &= \beta_{3,0} + \gamma_{3,t} \\
\log(\tau_t - 2) &= \beta_{4,0} + \gamma_{4,t}
\end{aligned}
\tag{9.3}
$$

where

$$
\begin{aligned}
\gamma_{2,t} &= \phi_1 \gamma_{2,t-1} + b_{2,t} \\
\gamma_{3,t} &= \gamma_{3,t-1} + b_{3,t} \\
\gamma_{4,t} &= \gamma_{4,t-1} + b_{4,t}.
\end{aligned}
$$

This is model `m1` in Table 9.2. Note that $\beta_{2,0}$ is the reversion line for $\log(\sigma_t)$ around which the autoregressive process $\gamma_{2,t}$ varies.

9.3.3 Leverage effect in volatility model

Leverage effects enable $\log(\sigma_t)$ to respond asymmetrically to positive and negative values of y_t, the return of the S&P 500, and are typically incorporated into the GEST model by including two variables $v_{1,t-1}$ and $v_{2,t-1}$ in the random effects $\gamma_{2,t}$, where $v_{1,t-1} = \text{arcsinh}(y_{t-1})$ if $y_{t-1} < 0$ and $v_{2,t-1} = \text{arcsinh}(y_{t-1})$ if $y_{t-1} >= 0$.

The use of the transformed $\text{arcsinh}(y_{t-1})$ rather than just y_{t-1} in the GEST model was found to reduce occasional extreme spikes in the fitted volatility. A very similar technique was used by Nelson (1991) in the exponential GARCH (EGARCH) model to capture leverage effects, and by Glosten, Jagannathan and Runkle (1993) in the GJR-GARCH model, where Glosten, Jagannathan and Runkle (1993) included a variable in which the squared observations are multiplied by an indicator that takes a value of unity when an observation is negative and zero otherwise; Asai and McAleer (2005) used Nelson's (1991) technique for modelling the asymmetric stochastic volatility model.

In model (9.3), asymmetric stochastic volatility terms were included in the model for $\log(\sigma_t)$ to account for leverage effect giving

$$\gamma_{2,t} = \phi_1 \gamma_{2,t-1} + \delta_1 v_{1,t-1} + \delta_2 v_{2,t-1} + b_{2,t} \tag{9.4}$$

where $v_{1,t-1} = \text{arcsinh}(y_{t-1})$ (if $y_{t-1} < 0$) and $v_{2,t-1} = \text{arcsinh}(y_{t-1})$ (if $y_{t-1} >= 0$).

This is model `m2` in Table 9.2. Submodels of model `m2` where ν and/or τ is constant over time are also given in Table 9.2. Effectively, Table 9.2 compares between the submodels of the GEST model `m2` to check whether we need a random walk (rw) model for the skewness parameter and/or a random walk model for the kurtosis parameter or just a constant for one or both parameters. Therefore, we fit five submodels to the S&P 500 data and summarise the Akaike information criterion (AIC). Note that the "ar with lev." model for σ_t is given by equation (9.4).

The model selected with minimum AIC is `m2` giving fitted model:

Table 9.2: Submodels of model `m2`

model	μ_t	σ_t	ν_t	τ_t	df	AIC
`m1`	const	ar	rw	rw	349.4901	22005.87
`m2`	const	ar with lev.	rw	rw	311.0655	**21862.74**
`m3`	const	ar with lev.	rw	const	334.4506	21871.31
`m4`	const	ar with lev.	const	rw	299.5491	21874.33
`m5`	const	ar with lev.	const	const	330.4522	21879.61

$$
\begin{aligned}
Y_t|\mu_t,\sigma_t,\nu_t,\tau_t &\sim SST(\mu_t,\sigma_t,\nu_t,\tau_t) \\
\mu_t &= 0.0266 \\
\log(\sigma_t) &= 0.1362+\gamma_{2,t} \\
\log(\nu_t) &= -0.0693+\gamma_{3,t} \\
\log(\tau_t-2) &= 2.451+\gamma_{4,t}
\end{aligned} \tag{9.5}
$$

where

$$
\begin{aligned}
\gamma_{2,t} &= 0.9855\gamma_{2,t-1} - 0.0658 v_{1,t-1} - 0.0712 v_{2,t-1} + b_{2,t} \\
\gamma_{3,t} &= \gamma_{3,t-1} + b_{3,t} \\
\gamma_{4,t} &= \gamma_{4,t-1} + b_{4,t}
\end{aligned}
$$

The fitted values for the variances are $\hat{\sigma}^2_{b_2} = 0.00336, \hat{\sigma}^2_{b_3} = 3.2459e^{-05}$ and $\sigma^2_{b_4} = 0.00199$

Figure (9.4) shows the time series of the returns of the S&P 500 stock index over the period 02/01/1980 to 31/12/2012, and the fitted stochastic volatility of the GEST model, for the return of the S&P 500. It is clear that the returns exhibit the volatility clustering effect (large changes in returns tend to cluster together, resulting in persistence of the amplitudes of return changes). When the returns are volatile the fitted stochastic volatility ($\hat{\sigma}_t$) increases, as is clear in the period when the financial markets had a crisis of a credit crunch in 2008. When the returns are not volatile the fitted stochastic volatility decreases.

Figure (9.5) shows the conditional skewness (ν_t) and the inverse of the conditional kurtosis ($1/\tau_t$) of the GEST model for the return of the S&P 500. These two parameters model the shape of the conditional distribution of the returns. When ($\nu_t > 1$) the distribution of returns is positively skewed, and when ($\nu_t < 1$) the distribution of returns is negatively skewed. Also the returns shows some spikes, which occurs occasionally without the clustering effect. These are large positive or negative returns without persistence, indicating high kurtosis rather than high volatility. These spikes are measured by the conditional kurtosis in the GEST model. A large value of ($1/\tau_t$) indicates a high kurtosis, and small value of ($1/\tau_t$) indicates a low kurtosis.

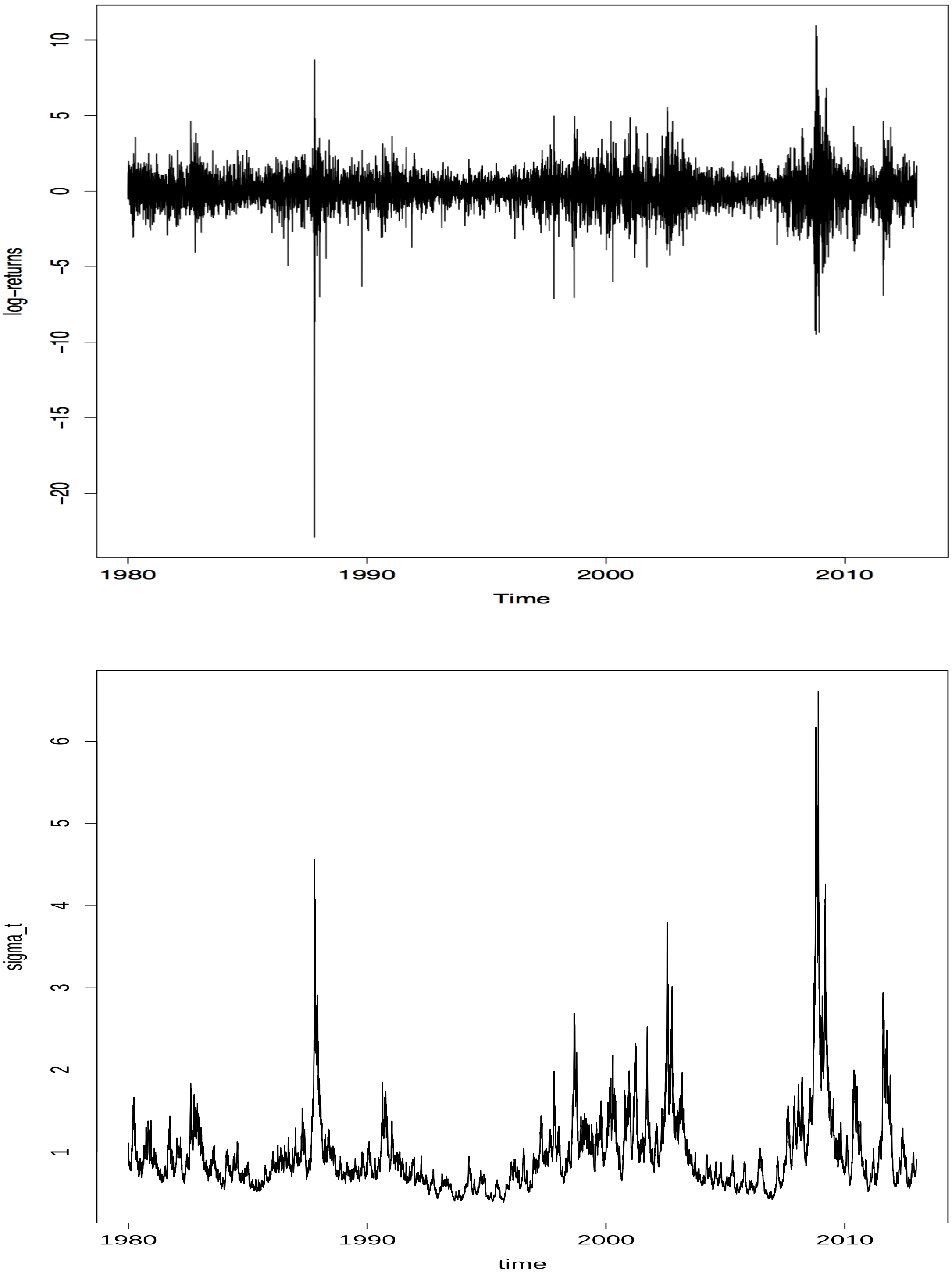


Figure 9.4: Returns y_t and fitted σ_t for model m2.

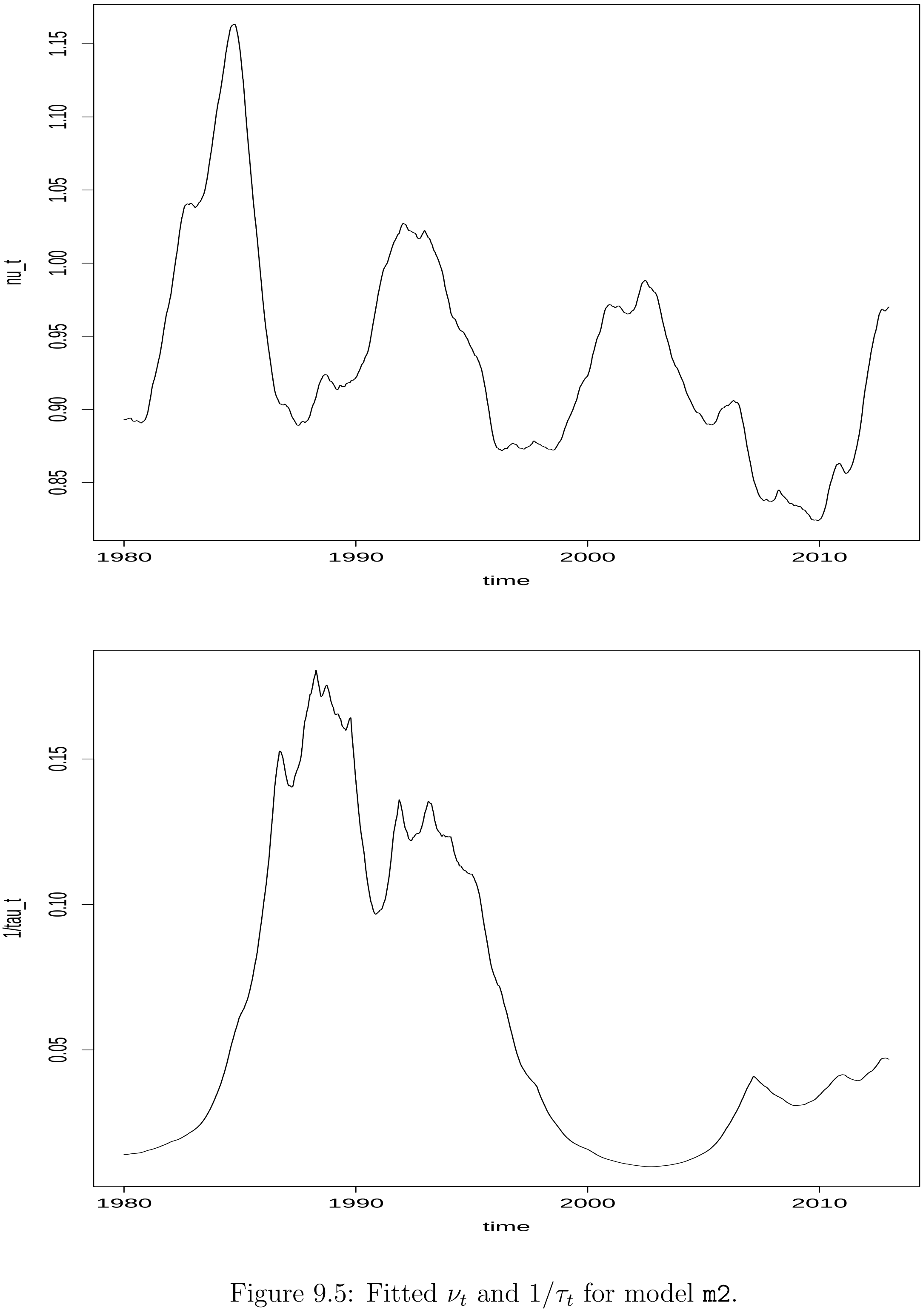


Figure 9.5: Fitted ν_t and $1/\tau_t$ for model `m2`.

9.3.4 Comparing GEST model `m2` and APARCH(1,1) model

The chosen GEST model `m2` is now compared with alternative models. Asymmetric power ARCH (APARCH) models are used to model volatility of the S&P 500 stock index returns with the SST distribution in order to see how well they capture the asymmetry and the fat tails of the asset returns compared with the GEST model. To measure the goodness of fit, we use the global deviance (equals to - twice the maximum log-likelihood). The Akaike information criterion (AIC) was used to choose the best fitted model. Using the `fGarch` package available in `R` (Wurtz *et al.*, 2006) we fit the GARCH(1,1) model, introduced by Bollerslev (1986), the GJR-GARCH(1,1) model, introduced by Glosten, Jagannathan and Runkle (1993) to allow for leverage effect, and the APARCH(1,1) model, introduced by Ding, Granger and Engle (1993), which adds the flexibility of a varying exponent. The GEST model `m2` has the lowest AIC followed by the APARCH(1,1) model.

Table 9.3: Model comparison between the GEST, GARCH, GJR-GARCH, and APARCH

Information Criteria	GEST	GARCH	GJR-GARCH	APARCH
Global Deviance	21240.61	22294.64	22615.48	22141.96
AIC	**21862.74**	22306.65	22625.48	**22157.96**

9.3.5 Residual analysis for GEST model `m2` and APARCH(1,1) model

In this section we compare the residual analysis for the GEST model `m2` with that of the APARCH(1,1) model.

Interpreting the shape of the worm plots

The residuals used here are called the normalized probability integral transform (normalized PIT) residuals, (Rosenblatt, 1952; Mitchell and Wallis, 2011) or normalized quantile residuals (Dunn and Smyth, 1996) and are defined by

$$\hat{r}_t = \Phi^{-1}(\hat{u}_t)$$

where

$$\hat{u}_t = F_{Y_t}(y_t|\hat{\mu}_t, \hat{\sigma}_t, \hat{\nu}_t, \hat{\tau}_t)$$

where $\hat{u}_t$ are the PIT residuals, F_Y is the cumulative distribution function of the conditional distribution of Y_t and Φ^{-1} is the inverse cumulative distribution function of a standard normal $N(0, 1)$ variable. The reason for using these residuals is that the true residuals r_t have a standard normal distribution if the model is correct. Hence the residuals $\hat{r}_t$ can be compared with a normal distribution using detrended QQ plots.

Figures 9.6 and 9.7 display detail diagnostic plots for the residuals using a worm plot developed by Van Buuren and Fredriks (2001). In this plot the range of time is split into six intervals with equal numbers of days. The six time ranges are listed and displayed in horizontal steps in the chart above the worm plot in Figures 9.6 and 9.7. A detrended normal QQ plot of the residuals in each interval is then displayed, for the lowest time range in the bottom left hand plot, in rows to the highest time range in the top right hand plot in both figures.

The worm plot allows detection of inadequacies in the model fit within specific time ranges. From Figure 9.6, the detrended QQ plots for the fitted APARCH model show inadequate fits to the data within most of the six time ranges, indicating differences between the model and the empirical, mean, variance, skewness and kurtosis of the residuals, within the corresponding time range to the individual QQ plot.

Table 9.4 below gives the interpretation of worm plots. Column 1 gives the shape of the worm plot, column 3 shows how the residuals differ from a standard normal distribution, while column 4 shows the corresponding inadequacy of the assumed model in explaining the response variable. For example, a worm plot with an U-shape indicates residuals with positive skewness, implying that the skewness of the model's response variable is too low and a higher (i.e. more positively or less negatively) skew distribution is required. Column 2 of Table 9.4 is described in section 9.3.5

Applying the interpretation of Table 9.4 to the APARCH model in Figure 9.6, indicates that in worm plot 1 the residuals are positively skewed (the model skewness is too low), in worm plot 3 the variance of residuals is too low (the model variance is too high), in worm plot 4 and 6 the residuals are negatively skewed (the model skewness is too high), in worm plot 5 residuals are platy-kurtotic (model kurtosis is too high). Clearly the constant (conditional) skewness and the constant (conditional) kurtosis in the APARCH model is inadequate. From Figure 9.7, the detrended QQ plots for the GEST model show adequate fits to the data within most of the six time ranges, indicating a reasonable fit to the data within time ranges. Note that the single QQ plot for all residuals from each of the APARCH and GEST models (not shown here) appear relatively satisfactory but splitting the QQ plot into the six time periods shows inadequacies, particularly in the APARCH model.

Table 9.4: The different shapes for the worm plots (first column) and corresponding guide range of Z statistics (second column), interpreted with respect to the normalized PIT residuals (third column) and the model response variable (fourth column).

Shape of worm plot	**Z stats.**	**Normalized PIT resid.**	**Response var.**
level: below the origin	Z1 < -2	mean too small	mean too large
level: above the origin	Z1 > 2	mean too large	mean too small
line: negative slope	Z2 < -2	variance too small	variance too large
line: positive slope	Z2 > 2	variance too large	variance too small
U-shape	Z3 < -2	positive skewness	skewness too high
inverted U-shape	Z3 > 2	negative skewness	skewness too low
S-shape with left bent up	Z4 < -2	platykurtosis (i.e. tails too heavy)	kurtosis too high
S-shape with left bent down	Z4 > 2	leptokurtosis (ie. tails too light)	kurtosis too low

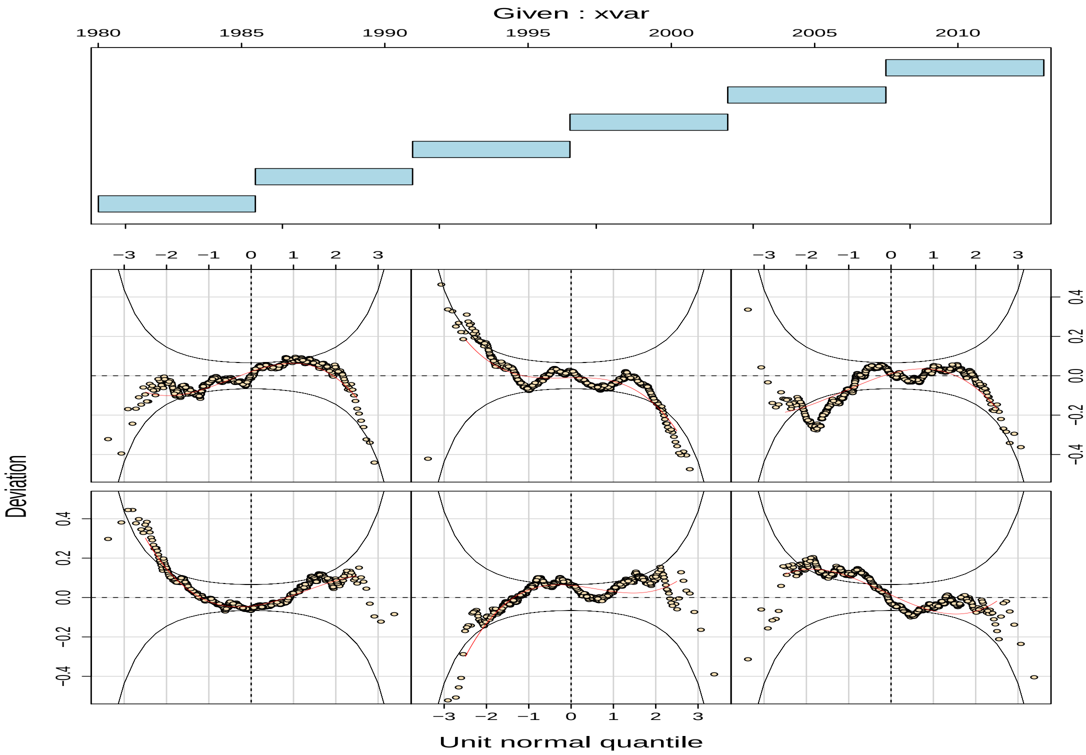


Figure 9.6: Worm plot of the APARCH model

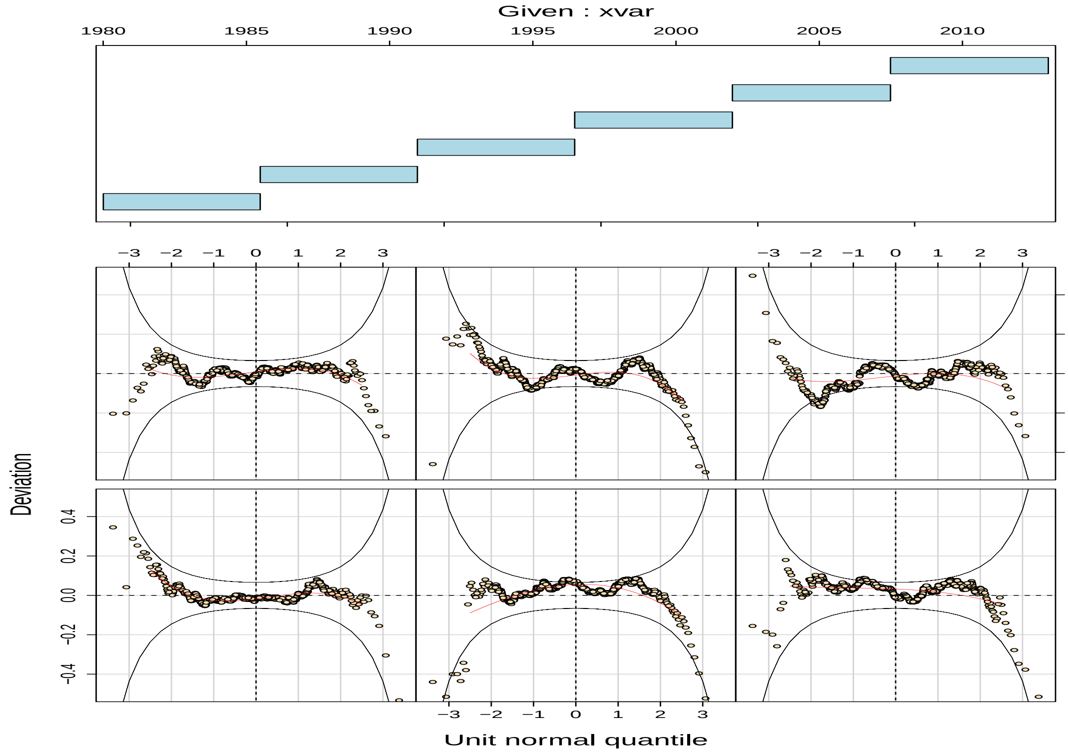


Figure 9.7: Worm plot of the GEST model

Interpreting the Z statistics

Model diagnosis is further investigated by calculating Z statistics to test the normality of the residuals within time groups (Royston and Wright, 2000).

Let G be the number of time groups and let $\{r_{gi}, i = 1, 2, .., n_i\}$ be the residuals in time group g, with mean $\bar{r}_g$ and standard deviation s_g, for $g = 1, 2, .., G$. The following statistics $\mathsf{Z}_{g1}, \mathsf{Z}_{g2}, \mathsf{Z}_{g3}, \mathsf{Z}_{g4}$ are calculated from the residuals in group g to test whether the residuals in group g have population mean 0, variance 1, skewness 0 and kurtosis 3, where $\mathsf{Z}_{g1} = n_g^{1/2}\bar{r}_g$, $\mathsf{Z}_{g2} = \left\{s_g^{2/3} - [1 - 2/(9n_g - 9)]\right\} / \{2/(9n_g - 9)\}^{1/2}$ and Z_{g3} and Z_{g4} are test statistics for skewness and kurtosis given by D'Agostino *et al.* (1990), in their equations (13) and (19) respectively. Provided the number of groups G is sufficiently large then the Z_{gj} values should have approximately standard normal distributions under the null hypothesis that the true residuals are standard normally distributed. We suggest as a rough guide values of $|\mathsf{Z}_{gj}|$ greater than 2 be considered as indicative of significant inadequacies in the model. Note that significant positive (or negative) values $\mathsf{Z}_{gj} > 2$ (or $\mathsf{Z}_{gj} < -2$) for $j = 1, 2, 3$ or 4 indicate respectively that the residuals in time group g have a higher (or lower) mean, variance, skewness or kurtosis than the assumed standard normal distribution. See Table 9.4 for the interpretation.

Table 9.5 gives the values of Z_{gj} obtained from the APARCH fitted model. The significant negative values of Z_{g2} are Z_{32} and Z_{52} indicating that the residual variance is too low (or equivalently that the fitted APARCH model variance or volatility is too high) within the corresponding time group, i.e. interval of time t for the group. The significant negative values of Z_{g3} are Z_{23} and Z_{63} indicating that the residual skewness is too low (or equivalently the model skewness is too high) while the significant positive value of Z_{g3} is Z_{13} indicating that the residual skewness is too high (or equivalently the model skewness is too low). The significant negative value of Z_{g4} is Z_{54} indicating that the residual kurtosis is too low (or equivalently the model kurtosis is too high) while the significant positive values of Z_{g4} are Z_{24} and Z_{34}, indicating that the residual kurtosis is too high (or equivalently the model kurtosis is too low). Clearly a constant skewness and constant kurtosis in APARCH model is inadequate.

Table 9.6 gives the values of Z_{gj} obtained from the GEST fitted `m2` model. There is only one significant value Z_{23}, indicating the residual skewness is too low (or equivalently the model skewness is too high) for the corresponding time period of the group of observations.

In conclusion, the residual analysis shows that the GEST fitted model does fit the data better than the APARCH model.

Table 9.5: Z statistics of APARCH

group, g	time, t	Z1	Z2	Z3	Z4
1	0.5 to 1387.5	-0.14	-0.13	**3.44**	-1.44
2	1387.5 to 2775.5	1.27	1.51	**-2.75**	**3.14**
3	2775.5 to 4162.5	0.69	**-3.38**	0.53	**2.06**
4	4162.5 to 5549.5	0.08	1.94	-1.48	-1.85
5	5549.5 to 6937.5	-0.58	**-2.16**	-0.53	**-2.49**
6	6937.5 to 8324.5	-0.69	1.83	**-2.58**	-1.08

Table 9.6: Z statistics of GEST

group, g	time, t	Z1	Z2	Z3	Z4
1	0.5 to 1387.5	-0.28	-0.13	0.84	-1.71
2	1387.5 to 2775.5	1.14	0.23	**-2.16**	0.34
3	2775.5 to 4162.5	0.99	-0.66	-0.39	-0.05
4	4162.5 to 5549.5	0.20	0.38	-0.37	-1.11
5	5549.5 to 6937.5	-0.30	-0.37	-0.10	-1.82
6	6937.5 to 8324.5	-0.62	0.48	-0.49	-0.60

9.3.6 Extended model for the S&P 500

Here the GEST model `m2` for the S&P 500 data is extended to include a random walk order 2 models for μ_t, ν_t and τ_t and autoregressive ar(1) with leverage model for σ_t, i.e.

$$\begin{aligned}
Y_t|\mu_t, \sigma_t, \nu_t, \tau_t &\sim SST(\mu_t, \sigma_t, \nu_t, \tau_t) \\
\mu_t &= \beta_{1,0} + \gamma_{1,t} \\
\log(\sigma_t) &= \beta_{2,0} + \gamma_{2,t} \\
\log(\nu_t) &= \beta_{3,0} + \gamma_{3,t} \\
\log(\tau_t - 2) &= \beta_{4,0} + \gamma_{4,t}
\end{aligned} \tag{9.6}$$

where

$$
\begin{aligned}
\gamma_{1,t} &= 2\gamma_{1,t-1} - \gamma_{1,t-2} + b_{1,t} \\
\gamma_{2,t} &= \phi_1 \gamma_{2,t-1} + \delta_1 v_{1,t-1} + \delta_2 v_{2,t-1} + b_{2,t} \\
\gamma_{3,t} &= 2\gamma_{3,t-1} - \gamma_{3,t-2} + b_{3,t} \\
\gamma_{4,t} &= 2\gamma_{4,t-1} - \gamma_{4,t-2} + b_{4,t}
\end{aligned}
$$

This provides a fitted model for the return mean μ_t, given in Figure 9.8, which corresponds to economic cycles of rising S&P 500 or "booms" (with positive mean returns i.e. $\mu_t > 0$), and declining S&P 500 or "busts" (with negative mean returns i.e. $\mu_t < 0$). Figure 9.9 gives the fitted σ_t, ν_t and $1/\tau_t$ for model 9.7. They are similar to Figures (9.4) and (9.4), except ν_t and τ_t are smoother. The reason for considering GEST model 9.7 was that it provides an interesting model for the return mean μ_t and also has a reduced $AIC = 21815.54$ compared with $AIC = 21862.74$ for GEST model `m2`.

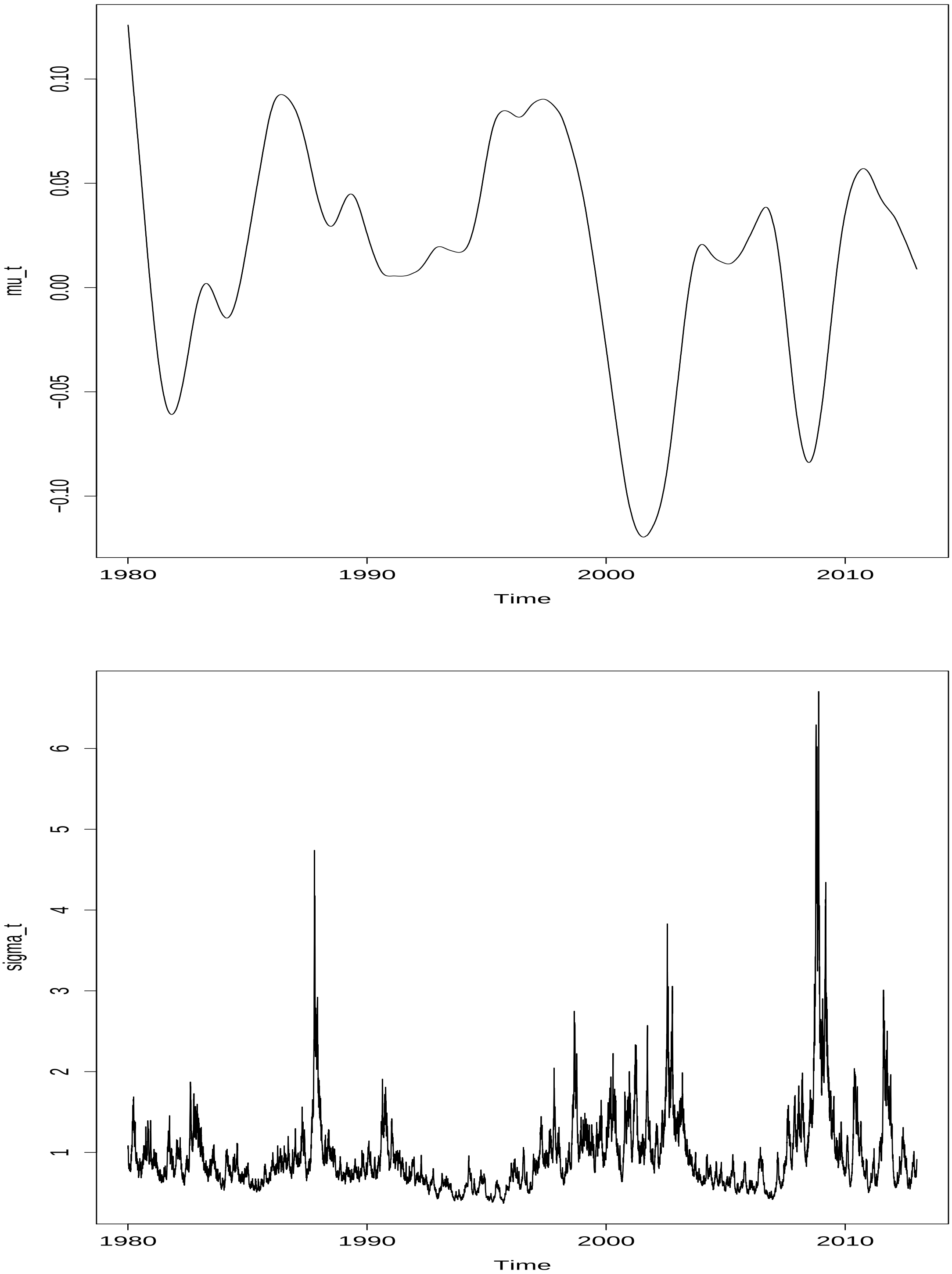


Figure 9.8: Fitted μ_t and σ_t for extended GEST model of equation (15).

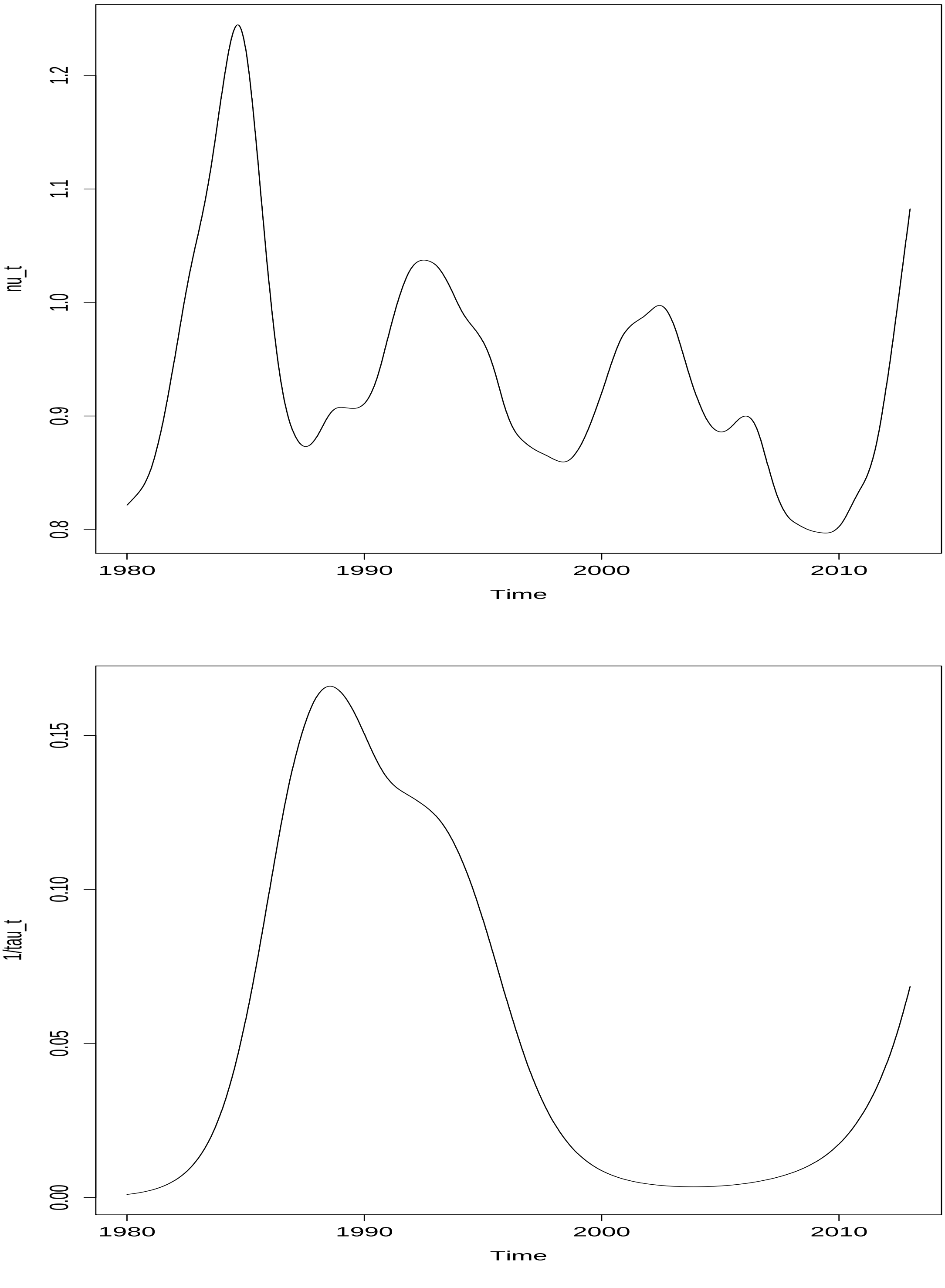


Figure 9.9: Fitted ν_t and $1/\tau_t$ for extended GEST model of equation (15).

9.4 Van drivers killed in UK

In time series analysis of road traffic safety, it is often required to assess the effect of road safety measures on the development in traffic safety over time. The data in this example are the monthly number of light goods vehicle drivers killed in road accidents from 1969 to 1984 in the UK. The data is available in R as `vandrivers` within the package `sspir` (Dethlefsen and Lundbye-Christensen, 2006). It consists of 192 observations of counts, where, on January 31st, 1983, a seat belt legislation law was introduced in the UK.

Durbin and Koopman (2000) modeled this data to measure the effect of the seat belt legislation on the number of deaths of van drivers in road traffic accident in the UK. They used a conditional Poisson distribution with a trend, seasonality and intervention in the structural mean model. The intervention parameter is a dummy variable or an indicator variable which takes zero values before the legislation period and a unity afterward, to measure the effect of the seat belt law. In particular, their model consist of a random walk local level and seasonality with an intervention term, `seatbelt`.

Let Y_t be the monthly number of light goods van drivers killed in road accidents in the UK from January 1969 to December 1984, and consider; the conditional Poisson distribution $PO(Y_t|\mu_t)$ of the response variable Y_t in the GEST model, and the conditional negative binomial type I distribution $NBI(Y_t|\mu_t, \sigma_t)$ of the response variable Y_t in the GEST model, with and without the intervention variable, to measure the effect of the seat belt legislation on road safety. In addition, with deterministic seasonality and stochastic seasonality to check whether there were any changes in the seasonality pattern in the data.

9.4.1 Conditional Poisson distribution

Here we consider conditional Poisson distribution models with each of a random walk order 1 and a random walk order 2 local levels, each of which with (i) no seasonal effect, (ii) a deterministic seasonal effect factor, (iii) a deterministic smooth seasonal effect, (iv) a stochastic seasonal effect, each of which (I) without an intervention variable, (II) with an intervention variable. The results are summarized later in Tables 1.1. to 1.4.

RW(1) local level model, `m1`

$$\begin{aligned} Y_t|\mu_t &\sim PO(\mu_t) \\ \log(\mu_t) &= \beta_{1,0} + \gamma_{1,t} \\ \gamma_{1,t} &= \gamma_{1,t-1} + b_{1,t} \end{aligned}$$

where $b_{1,t} \sim NO(0, \sigma_b^2)$.

RW(1) local level and deterministic seasonal effect factor model, `m2`

$$\begin{aligned} Y_t|\mu_t &\sim PO(\mu_t) \\ \log(\mu_t) &= \beta_{1,0} + \xi_t + \gamma_{1,t} \\ \gamma_{1,t} &= \gamma_{1,t-1} + b_{1,t} \end{aligned}$$

where $b_{1,t} \sim NO(0, \sigma_b^2)$ and where ξ is a factor with 12 levels, for fixed seasonality, i.e. $\xi^\top = (1, 2, \ldots, 12, 1, 2, \ldots, 12, 1, 2, \ldots, 12)$.

RW(1) local level and deterministic smooth seasonal effect model, `m15`

$$\begin{aligned} Y_t|\mu_t &\sim PO(\mu_t) \\ \log(\mu_t) &= \beta_{1,0} + \beta_{1,1}x_{1,t} + \beta_{1,2}x_{2,t} + \beta_{1,3}x_{3,t} + \beta_{1,4}x_{4,t} + \gamma_{1,t} \\ \gamma_{1,t} &= \gamma_{1,t-1} + b_{1,t} \end{aligned}$$

where $b_{1,t} \sim NO(0, \sigma_b^2)$.

RW(1) local level and stochastic seasonal model, `m3`

$$\begin{aligned} Y_t|\mu_t &\sim PO(\mu_t) \\ \log(\mu_t) &= \beta_{1,0} + \gamma_{1,t} + s_{1,t} \end{aligned}$$

where

$$\begin{aligned} \gamma_{1,t} &= \gamma_{1,t-1} + b_{1,t} \\ s_{1,t} &= -\sum_{m=1}^{M-1} s_{1,t-m} + w_t \end{aligned}$$

where $b_{1,t} \sim NO(0, \sigma_b^2)$ and $w_{1,t} \sim NO(0, \sigma_w^2)$.

RW(1) local level with intervention variable model, `m4`

$$\begin{aligned} Y_t|\mu_t &\sim PO(\mu_t) \\ \log(\mu_t) &= \beta_{1,0} + \beta_{1,1}\zeta_t + \gamma_{1,t} \\ \gamma_{1,t} &= \gamma_{1,t-1} + b_{1,t} \end{aligned}$$

where $b_{1,t} \sim NO(0, \sigma_b^2)$ and where ζ is an intervention variable with $\zeta_t = 0$ for $t < t_I$ and $\zeta_t = 1$ for $t \geqslant t_I$ where, $t_I = 170$, is the point corresponding to the month February 1983, following the intervention on January 31st 1983.

RW(1) local level with intervention variable and deterministic seasonal effect factor model, `m5`

$$
\begin{aligned}
Y_t|\mu_t &\sim PO(\mu_t) \\
\log(\mu_t) &= \beta_{1,0} + \beta_{1,1}\zeta_t + \xi_t + \gamma_{1,t} \\
\gamma_{1,t} &= \gamma_{1,t-1} + b_{1,t}
\end{aligned}
$$

where $b_{1,t} \sim NO(0, \sigma_b^2)$.

RW(1) local level with intervention variable and deterministic smooth seasonal effect model, `m16`

$$
\begin{aligned}
Y_t|\mu_t &\sim PO(\mu_t) \\
\log(\mu_t) &= \beta_{1,0} + \beta_{1,1}\zeta_t + \beta_{1,2}x_{1,t} + \beta_{1,3}x_{2,t} + \beta_{1,4}x_{3,t} + \beta_{1,5}x_{4,t} + \gamma_{1,t} \\
\gamma_{1,t} &= \gamma_{1,t-1} + b_{1,t}
\end{aligned}
$$

where $b_{1,t} \sim NO(0, \sigma_b^2)$.

RW(1) local level with intervention variable and stochastic seasonal model, `m6`

$$
\begin{aligned}
Y_t|\mu_t &\sim PO(\mu_t) \\
\log(\mu_t) &= \beta_{1,0} + \beta_{1,1}\zeta + \gamma_{1,t} + s_{1,t}
\end{aligned}
$$

where

$$
\begin{aligned}
\gamma_{1,t} &= \gamma_{1,t-1} + b_{1,t} \\
s_{1,t} &= -\sum_{m=1}^{M-1} s_{1,t-m} + w_t
\end{aligned}
$$

where $b_{1,t} \sim NO(0, \sigma_b^2)$ and $w_{1,t} \sim NO(0, \sigma_w^2)$.

RW(2) local level model, `m20`

$$
\begin{aligned}
Y_t|\mu_t &\sim PO(\mu_t) \\
\log(\mu_t) &= \beta_{1,0} + \gamma_{1,t} \\
\gamma_{1,t} &= 2\gamma_{1,t-1} - \gamma_{1,t-2} + b_{1,t}
\end{aligned}
$$

where $b_{1,t} \sim NO(0, \sigma_b^2)$.

RW(2) local level and deterministic seasonal effect factor model, `m21`

$$
\begin{aligned}
Y_t|\mu_t &\sim PO(\mu_t) \\
\log(\mu_t) &= \beta_{1,0} + \xi_t + \gamma_{1,t} \\
\gamma_{1,t} &= 2\gamma_{1,t-1} - \gamma_{1,t-2} + b_{1,t}
\end{aligned}
$$

where $b_{1,t} \sim NO(0, \sigma_b^2)$ and where ξ is a factor with 12 levels, for fixed seasonality, i.e. $\xi^\top = (1, 2, \ldots, 12, 1, 2, \ldots, 12, 1, 2, \ldots, 12)$.

RW(2) local level and deterministic smooth seasonal effect model, `m22`

$$
\begin{aligned}
Y_t|\mu_t &\sim PO(\mu_t) \\
\log(\mu_t) &= \beta_{1,0} + \beta_{1,1}x_{1,t} + \beta_{1,2}x_{2,t} + \beta_{1,3}x_{3,t} + \beta_{1,4}x_{4,t} + \gamma_{1,t} \\
\gamma_{1,t} &= 2\gamma_{1,t-1} - \gamma_{1,t-2} + b_{1,t}
\end{aligned}
$$

where $b_{1,t} \sim NO(0, \sigma_b^2)$.

RW(2) local level and stochastic seasonal model, `m23`

$$
\begin{aligned}
Y_t|\mu_t &\sim PO(\mu_t) \\
\log(\mu_t) &= \beta_{1,0} + \gamma_{1,t} + s_{1,t}
\end{aligned}
$$

where

$$
\begin{aligned}
\gamma_{1,t} &= 2\gamma_{1,t-1} - \gamma_{1,t-2} + b_{1,t} \\
s_{1,t} &= -\sum_{m=1}^{M-1} s_{1,t-m} + w_t
\end{aligned}
$$

where $b_{1,t} \sim NO(0, \sigma_b^2)$ and $w_{1,t} \sim NO(0, \sigma_w^2)$.

RW(2) local level with intervention variable model, `m24`

$$\begin{aligned} Y_t|\mu_t &\sim PO(\mu_t) \\ \log(\mu_t) &= \beta_{1,0} + \beta_{1,1}\zeta_t + \gamma_{1,t} \\ \gamma_{1,t} &= 2\gamma_{1,t-1} - \gamma_{1,t-2} + b_{1,t} \end{aligned}$$

where $b_{1,t} \sim NO(0, \sigma_b^2)$ and where ζ is an intervention variable with $\zeta_t = 0$ for $t < t_I$ and $\zeta_t = 1$ for $t \geqslant t_I$ where, $t_I = 170$, is the point corresponding to the month February 1983, following the intervention on January 31st 1983.

RW(2) local level with intervention variable and deterministic seasonal effect factor model, `m25`

$$\begin{aligned} Y_t|\mu_t &\sim PO(\mu_t) \\ \log(\mu_t) &= \beta_{1,0} + \beta_{1,1}\zeta_t + \xi_t + \gamma_{1,t} \\ \gamma_{1,t} &= 2\gamma_{1,t-1} - \gamma_{1,t-2} + b_{1,t} \end{aligned}$$

where $b_{1,t} \sim NO(0, \sigma_b^2)$.

RW(2) local level with intervention variable and deterministic smooth seasonal effect model, `m26`

$$\begin{aligned}
Y_t|\mu_t &\sim PO(\mu_t) \\
\log(\mu_t) &= \beta_{1,0} + \beta_{1,1}\zeta_t + \beta_{1,2}x_{1,t} + \beta_{1,3}x_{2,t} + \beta_{1,4}x_{3,t} + \beta_{1,5}x_{4,t} + \gamma_{1,t} \\
\gamma_{1,t} &= 2\gamma_{1,t-1} - \gamma_{1,t-2} + b_{1,t}
\end{aligned}$$

where $b_{1,t} \sim NO(0, \sigma_b^2)$.

RW(2) local level with intervention variable and stochastic seasonal model, `m27`

$$\begin{aligned}
Y_t|\mu_t &\sim PO(\mu_t) \\
\log(\mu_t) &= \beta_{1,0} + \beta_{1,1}\zeta_t + \gamma_{1,t} + s_{1,t}
\end{aligned}$$

where

$$\begin{aligned}
\gamma_{1,t} &= 2\gamma_{1,t-1} - \gamma_{1,t-2} + b_{1,t} \\
s_{1,t} &= -\sum_{m=1}^{M-1} s_{1,t-m} + w_t
\end{aligned}$$

where $b_{1,t} \sim NO(0, \sigma_b^2)$ and $w_{1,t} \sim NO(0, \sigma_w^2)$.

Table 9.7: RW(1) local level models without `seatbelt` using the conditional Poisson distribution

model	$\hat{\sigma}_b^2$	$\hat{\sigma}_w^2$	x_t	ξ_t	df	cond. AIC
`m1`	0.0009535645	-	-	-	10.61	965.53
`m2`	0.0009204082	-	-	y	22.24	958.12
`m15`	0.0009076958	-	y	-	15.00	**951.35**
`m3`	0.0009120539	1.259696e-07	-	-	21.77	957.64

The R commands for fitting these models are given in Appendix D.

Table 9.8: RW(1) local level models with `seatbelt` using the conditional Poisson distribution

model	$\hat{\sigma}_b^2$	$\hat{\sigma}_w^2$	x_t	ξ_t	$\hat{\beta}_{1,1}$	df	cond. AIC
`m4`	0.0005873255	-	-	-	-0.3156	9.67	965.02
`m5`	0.0005943976	-	-	y	-0.2662	21.32	958.20
`m16`	0.0005771665	-	y	-	-0.27792	14.08	**951.27**
`m6`	0.0005843044	8.039543e-09	-	-	-0.2697	20.94	957.84

The R commands for fitting these models are given in Appendix D.

where y=yes in Tables 9.7, 9.8, 9.9, and 9.10.

Table 9.9: RW(2) local level models without `seatbelt` using the conditional Poisson distribution

model	$\hat{\sigma}_b^2$	$\hat{\sigma}_w^2$	x_t	ξ_t	df	cond. AIC
m20	1.979532e-07	-	-	-	4.48	963.79
m21	1.923868e-07	-	-	y	15.57	954.26
m22	1.88806e-07	-	y	-	8.53	**947.58**
m23	8.15026e-08	6.076289e-08	-	-	16.04	956.11

The R commands for fitting these models are given in Appendix D.

Table 9.10: RW(2) local level models with `seatbelt` using the conditional Poisson distribution

model	$\hat{\sigma}_b^2$	$\hat{\sigma}_w^2$	x_t	ξ_t	$\hat{\beta}_{1,1}$	df	cond. AIC
m24	6.798457e-08	-	-	-	-0.2162	4.89	964.10
m25	7.800994e-08	-	-	y	-0.17751	16.04	955.17
m26	6.995681e-08	-	y	-	-0.18796	8.97	**948.36**
m27	2.061154e-09	7.11798e-09	-	-	-0.2437	15.12	955.89

The R commands for fitting these models are given in Appendix D.

According to Tables 9.7, 9.8, 9.9, and 9.10 the best model with a smallest conditional Akaike information criteria, (cond. AIC)[1], is `m22`, RW(2) local level and deterministic smooth seasonal effect. The deterministic smooth seasonal effect is modelled by sine and cosine pairs with both annual and semiannual cycles. The seasonality in this data is constant over time, and it is better to be modeled by smooth sine and cosine pairs.

The model which Durbin and Koopman (2000) fitted to the van drivers was a structural mean model with a random walk local level and stochastic seasonal for the conditional Poisson distribution. Their parameter estimates for the random walk local level and the seasonal disturbances were $\hat{\sigma}_\eta = 0.0245$ and $\hat{\sigma}_w = 0$ respectively, with a conclusion that the seasonal effect is constant over time. Their parameter

[1]The conditional AIC = -2 maximum conditional log-likelihood + 2 effective degrees of freedom, and the global deviance = -2 maximum conditional log-likelihood, both are conditional on the fitted random effect $\hat{\gamma}_t$.

estimate for the seat belt intervention variable was $\lambda = -0.280$, which corresponds to a reduction in the number of deaths of 24%. Durbin and Koopman's (2000) model is equivalent to model `m6` from Table (9.8). The estimated hyperparameters for `m6` are $\sigma_b = 0.02417239$ and $\sigma_w = 0.0000896635$ for the random walk local level and the seasonal disturbances respectively, which implies a fixed seasonal effect in the data, and $\beta_{1,1} = -0.2697$ which corresponds to a reduction in the number of deaths of 24%.

However, model `m22` is better then `m6`, has a smaller AIC, implies that the GEST model `m22`, RW(2) local level and deterministic smooth seasonal effect model, improves the RW(1) local level with intervention variable and stochastic seasonal model of Durbin and Koopman (2000).

For analysing the statistical significance of the seat belt on the reduction of the number of deaths of van drivers in the UK , model `m26`, RW(2) local level with intervention variable and deterministic smooth seasonal effect, is used.

The model `m26` models the random effect with a random walk order 2 local level, the fixed effect of the seasonality and the fixed effect of the seat belt. Testing whether the fixed effect of the seat belt is significant, we refit model `m26`, and call it model `m261`, with fixed hyperparameters for the random walk order 2 local level, estimated from model `m26`, and without the seat belt variable, and test whether the difference in the global deviances of `m26` and `m261` is bigger than the value of the Chi square statistic with one degree of freedom at 5% significance level, ($\chi^2_{1,0.05} = 3.841$). If the fixed effect of seat belt is statistically significant, implies that the seat belt variable has dropped the deviance significantly, and is needed in the model.

Note that, model `m261` is equivalent to model `m22` but with fixed values for hyperparameters taken from the fitted hyperparameters of model `m26`. Also note that the effect of the seat belt in model `m26` is smaller, $\beta_{1,1} = -0.18796$, corresponding to an

17.14% reduction in the mean number of deaths [since $\exp(-0.18796) = 0.8286479$].

Conditional test for the seat belt intervention variable in `m26`

Testing whether the fixed effect of the seat belt legislation law had a significant impact on the reduction of the number of deaths of van drivers in the UK, is obtained approximately using the conditional test of $\beta_{1,1}$ to be significantly negative in model `m26` is necessary. The null hypothesis is $(H_0 : \beta_{1,1} = 0)$ against $(H_1 : \beta_{1,1} \neq 0)$ with 5% significance level.

If H_0 is accepted, then $\exp(\hat{\beta}_{1,1})$ is equal to one, implies that there was no reduction in the mean number of deaths of van drivers. However, if H_0 is rejected and H_1 is accepted, then $\exp(\hat{\beta}_{1,1}) = \exp(-0.18796) = 0.8286479$, implies that the mean level has dropped by 0.1713521, corresponding to 17.13% reduction in the mean number of deaths of van drivers.

Using a $\chi^2_{1,0.05}$ test with one degree of freedom at 5% significance level:

$H_0 : \beta_{1,1} = 0$ against $H_1 : \beta_{1,1} \neq 0$.

Table 9.11: Conditional test for the seat belt intervention variable in `m26`

model	$\hat{\sigma}_e^2$	$\hat{\sigma}_b^2$	$\hat{\beta}_{1,1}$	df	cond. AIC	cond. deviance
`m26`	0.8399006	6.995681e-08	-0.18796	8.97	948.36	930.41
`m261`	0.8399006	6.995681e-08	-	9.17	949.87	931.53

The R commands for fitting these models are given in Appendix D.

From Table 9.11, the difference in deviances is equal to $1.12 < 3.841$. Hence, H_0 is accepted and H_1 is rejected. This indicates $\beta_{1,1}$ is not statistically significant and should not be included in the model, in which case model `m22` is the best model, and the seat belt legislation law does not have a statistical significant effect on the reduction of the mean number of death of van drivers.

The 95% profile confidence interval for $\beta_{1,1}$ is equal to (-0.4379, 0.1261) as shown in Figure 9.10. The 95% profile confidence interval is obtained by fixing the hyper-parameters in model `m261` and offsetting $\beta_{1,1}\zeta_t$, where $\beta_{1,1}$ takes a sequence of values (e.g. from min=-0.6 to max=0.5), and ζ_t is the seat belt intervention variable. The 95% profile confidence interval is obtained by using the function `prof.term()` available in R within the package `gamlss` (Stasinopoulos *et al.*, 2008). This function plots the profile deviance for model `m261` from the minimum value of $\beta_{1,1}$ to the maximum value of $\beta_{1,1}$, gives the 95% profile confidence interval for $\beta_{1,1}$ and the estimate of $\beta_{1,1}$ at the minimum global deviance, as shown in Figure 9.10.

The lower bound of the 95% profile confidence interval for $\beta_{1,1}$ is given by:

$\beta_{1,1} = -0.4379$, the $\exp(-0.4379) = 0.6454$ which corresponds to a 35.46% minimum reduction in the mean number of deaths of van drivers killed in road accidents in the UK.

The upper bound of the 95% profile confidence interval for $\beta_{1,1}$ is given by:

$\beta_{1,1} = 0.1261$, the $\exp(0.1261) = 1.1344$ which corresponds to a 13.44% maximum increase in the mean number of deaths of van drivers killed in road accidents in the UK.

The estimate of $\beta_{1,1}$ is given by:

$\hat{\beta}_{1,1} = -0.18796$, the $\exp(-0.18796) = 0.8286479$ which corresponds to a 17.14% reduction in the mean number of deaths of van drivers killed in road accidents in the UK, predicted by the model `m26`, but not statistically significant because the 95% profile confidence interval for $\beta_{1,1}$ includes the zero, as shown in Figure 9.10. The R commands are given in Appendix D.

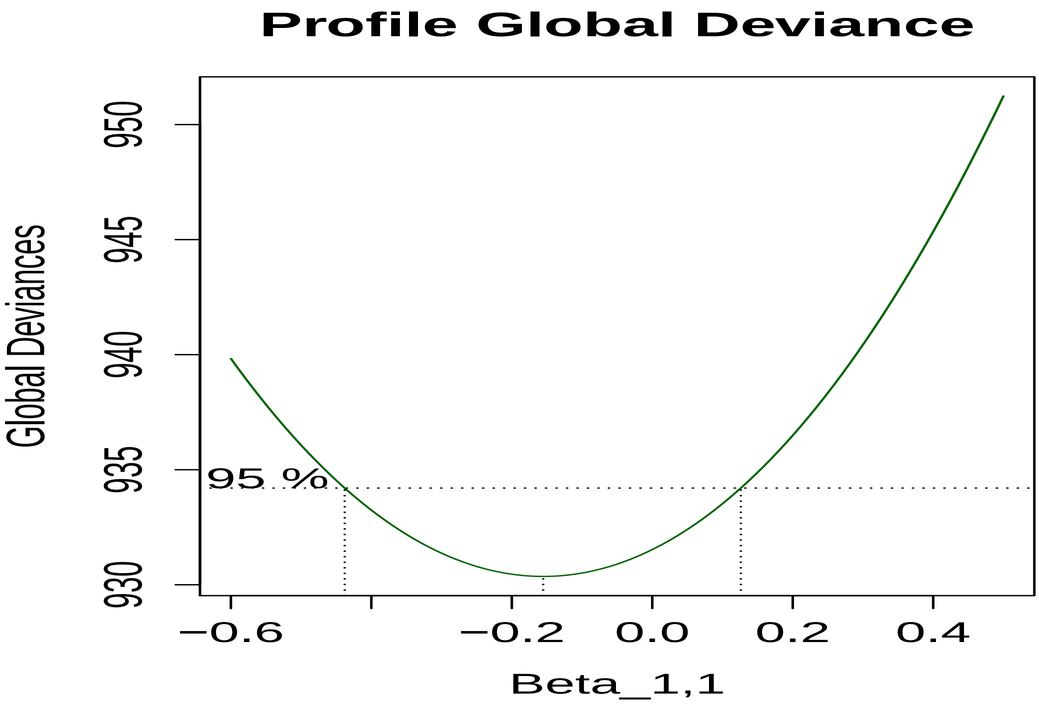


Figure 9.10: The 95% profile confidence interval for $\beta_{1,1}$

95% confidence interval for the hyperparameter σ_b^2 in `m22`

To calculate the 95% two-sided confidence interval for the hyperparameter σ_b^2, the following transformation is needed for the GEST model `m22`. The GEST estimates the $\log(\hat{\sigma}_b^2)$ and calculates its standard error by inverting the Hessian matrix obtained from the fitted model `m22`. This standard error is for the predictor for the hyperparameter.

Hence, the 95% confidence level for $\log(\hat{\sigma}_b^2)$ is given by

$\log(\hat{\sigma}_b^2)$`-1.96(se(`$\log(\hat{\sigma}_b^2)$`)) = -15.4825459 - 1.96(1.2879334) = -18.0069`

$\log(\hat{\sigma}_b^2)$`+1.96(se(`$\log(\hat{\sigma}_b^2)$`)) = -15.4825459 + 1.96(1.2879334) = -12.9582`

Now

$\hat{\sigma}_b^2$ `= exp(-15.4825459)=1.9e-07` and

`exp(-18.0069)= 1.5e-08` and

`exp(-12.9582)= 2.4e-06`.

Hence

`[1.5e-08, 2.4e-06] is the 95% confidence level for` σ_b^2.

9.4.2 Conditional negative binomial type I distribution

Here we consider conditional negative binomial type I, (NBI), distribution models with each of a random walk order 1 and a random walk order 2 local levels, each of which with (i) no seasonal effect, (ii) a deterministic seasonal effect factor, (iii) a deterministic smooth seasonal effect, (iv) a stochastic seasonal effect, each of which (I) without an intervention variable, (II) with an intervention variable. The $\log(\sigma_t) = \beta_{2,0}$, a constant for all NBI models.

For example RW(1) local level model, `b1`, is defined as

$$
\begin{aligned}
Y_t|\mu_t, \sigma_t &\sim NBI(\mu_t, \sigma_t) \\
\log(\mu_t) &= \beta_{1,0} + \gamma_{1,t} \\
\log(\sigma_t) &= \beta_{2,0} \\
\gamma_{1,t} &= \gamma_{1,t-1} + b_{1,t}
\end{aligned}
$$

where $b_{1,t} \sim NO(0, \sigma_b^2)$.

The results are summarized later in Tables 9.12, 9.13, 9.14, and 9.15.

where y=yes in Tables 9.12, 9.13, 9.14, and 9.15.

Table 9.12: RW(1) local level models without `seatbelt` using the conditional NBI distribution

model	$\hat{\sigma}_b^2$	$\hat{\sigma}_w^2$	x_t	ξ_t	$\hat{\beta}_{2,0}$	df	cond. AIC
`b1`	0.0009535645	-	-	-	-36.09	11.61	967.53
`b2`	0.0008421035	-	-	y	-36.09	21.81	959.20
`b15`	0.0009076958	-	y	-	-36.09	16.00	**953.35**
`b3`	0.0009120418	2.821023e-08	-	-	-36.09	22.77	959.64

The R commands for fitting these models are given in Appendix D.

Table 9.13: RW(1) local level models with `seatbelt` using the conditional NBI distribution

model	$\hat{\sigma}_b^2$	$\hat{\sigma}_w^2$	x_t	ξ_t	$\hat{\beta}_{1,1}$	$\hat{\beta}_{2,0}$	df	cond. AIC
`b4`	0.0005873425	-	-	-	-0.3156	-36.09	10.67	967.02
`b5`	0.0005944848	-	-	y	-0.2661	-36.09	22.32	960.20
`b16`	0.0005771708	-	y	-	-0.27791	-36.09	15.08	**953.27**
`b6`	0.0005844664	3.603998e-08	-	-	-0.2696	-36.09	21.94	959.84

The R commands for fitting these models are given in Appendix D.

Table 9.14: RW(2) local level models without `seatbelt` using the conditional NBI distribution

model	$\hat{\sigma}_b^2$	$\hat{\sigma}_w^2$	x_t	ξ_t	$\hat{\beta}_{2,0}$	df	cond. AIC
`b17`	1.979661e-07	-	-	-	-36.09	5.48	965.79
`b18`	1.923878e-07	-	-	y	-36.09	16.57	956.26
`b19`	1.888009e-07	-	y	-	-36.09	9.53	**949.58**
`b20`	8.153892e-08	1.843182e-07	-	-	-36.09	17.04	958.11

The R commands for fitting these models are given in Appendix D.

Table 9.15: RW(2) local level models with `seatbelt` using the conditional NBI distribution

model	$\hat{\sigma}_b^2$	$\hat{\sigma}_w^2$	x_t	ξ_t	$\hat{\beta}_{1,1}$	$\hat{\beta}_{2,0}$	df	cond. AIC
`b21`	6.798399e-08	-	-	-	-0.2162	-36.09	5.89	966.10
`b22`	7.804827e-08	-	-	y	-0.17741	-36.09	17.04	957.17
`b23`	6.996993e-08	-	y	-	-0.18794	-36.09	9.97	**950.36**
`b24`	2.061154e-09	2.768264e-07	-	-	-0.2437	-36.09	16.12	957.91

The R commands for fitting these models are given in Appendix D.

From the above Table 9.12, 9.13, 9.14, and 9.15, `b19` is the best model according to Akaike information criteria, but in comparison with model `m22`, the AIC has not improved, which indicates model `m22` is still the best model. Hence, the Poisson distribution is the best conditional distribution for this data. The conditional test for $\beta_{1,1}$ for model `b19` is not needed, since `m22` has a better AIC using the Poison distribution.

The time series of van drivers killed in road accidents in the UK from January 1969 to December 1984, and the fitted RW(2) local level of model `m22` without seasonality, are plotted together in Figure 9.14. We also plot the data and the fitted RW(1) local level with seat belt intervention variable of model `m16` in Figure 9.11, the fitted RW(2) local level with seat belt intervention variable of model `m26` in Figure 9.12, and the fitted RW(1) local level of model `m15` in Figure 9.13.

Figure 9.15 illustrates the time series of the van drivers killed in the UK and the decomposition of fitted model `m22` into two components, RW(2) local level and deterministic seasonal. The R commands for plotting Figure 9.14 are given in Appendix D.

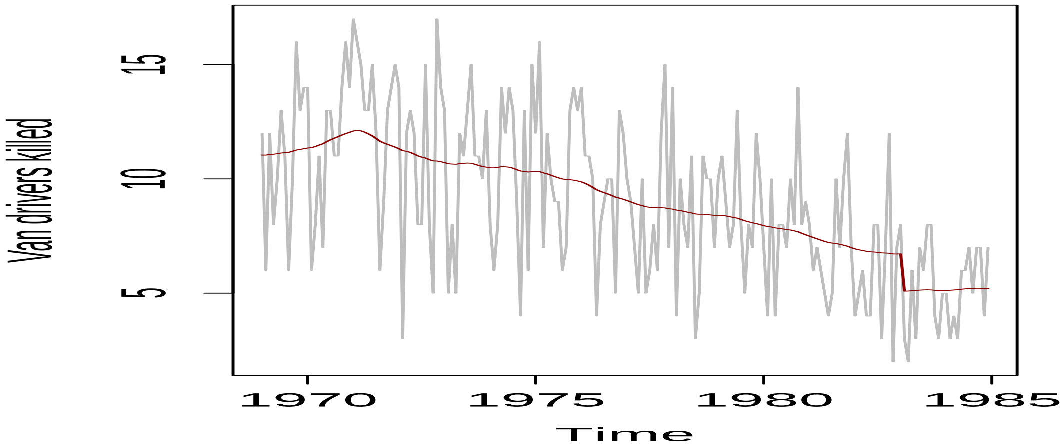


Figure 9.11: Monthly number of light goods van drivers killed in road accidents in the UK from January 1969 to December 1984 in gray, and the fitted RW(1) local level with intervention variable in red.

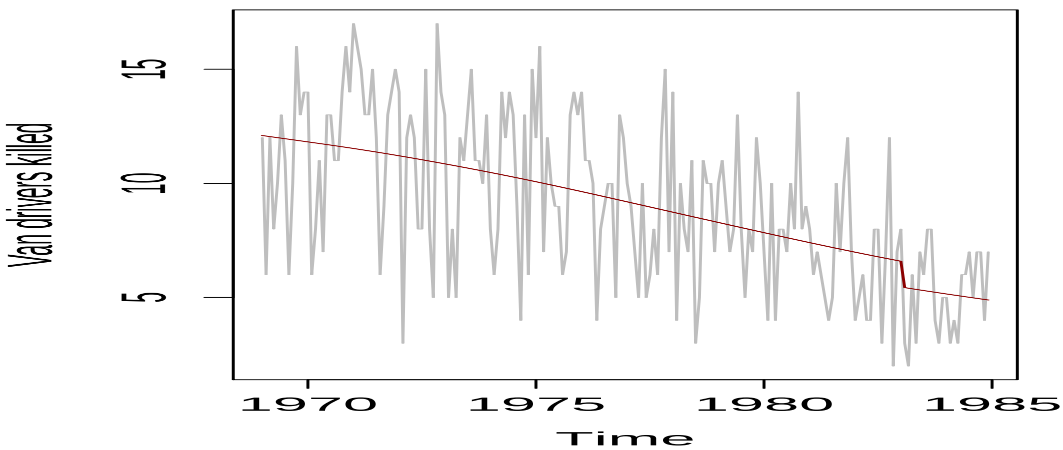


Figure 9.12: Monthly number of light goods van drivers killed in road accidents in the UK from January 1969 to December 1984 in gray, and the fitted RW(2) local level with intervention variable in red.

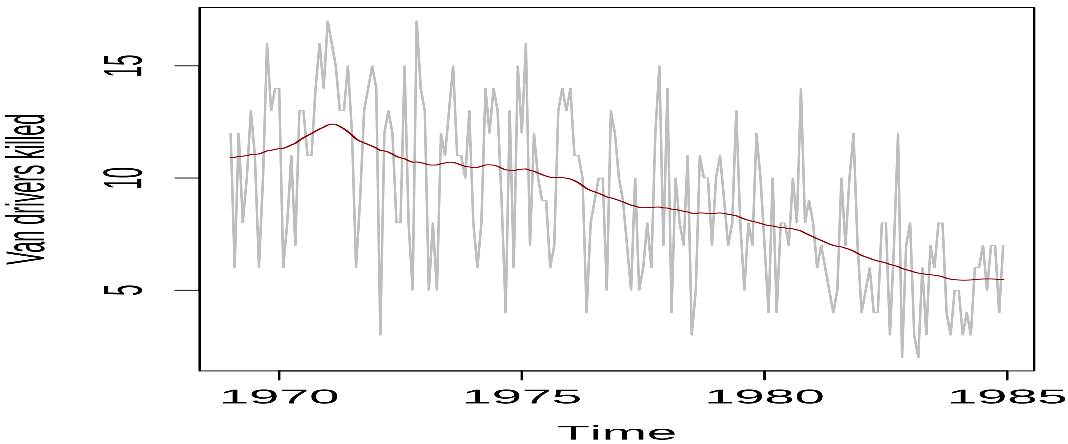


Figure 9.13: Monthly number of light goods van drivers killed in road accidents in the UK from January 1969 to December 1984 in gray, and the fitted RW(1) local level in red.

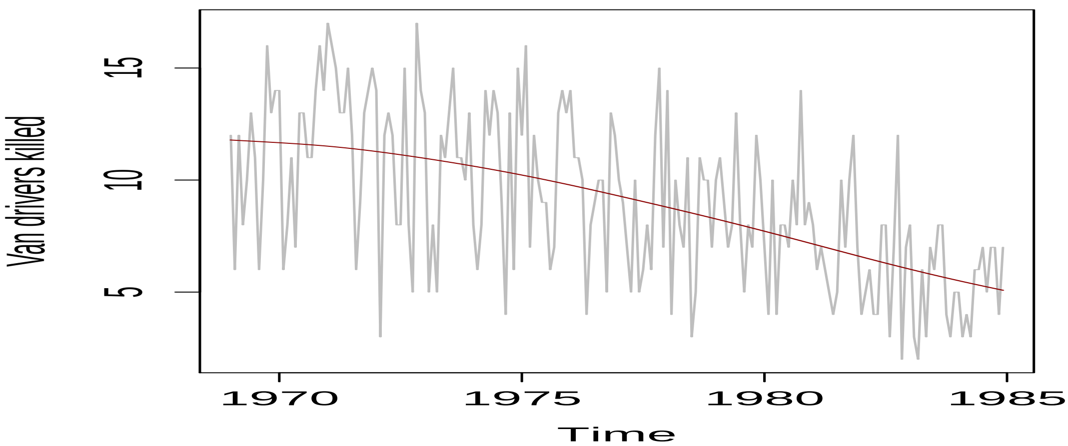


Figure 9.14: Monthly number of light goods van drivers killed in road accidents in the UK from January 1969 to December 1984 in gray, and the fitted RW(2) local level in red.

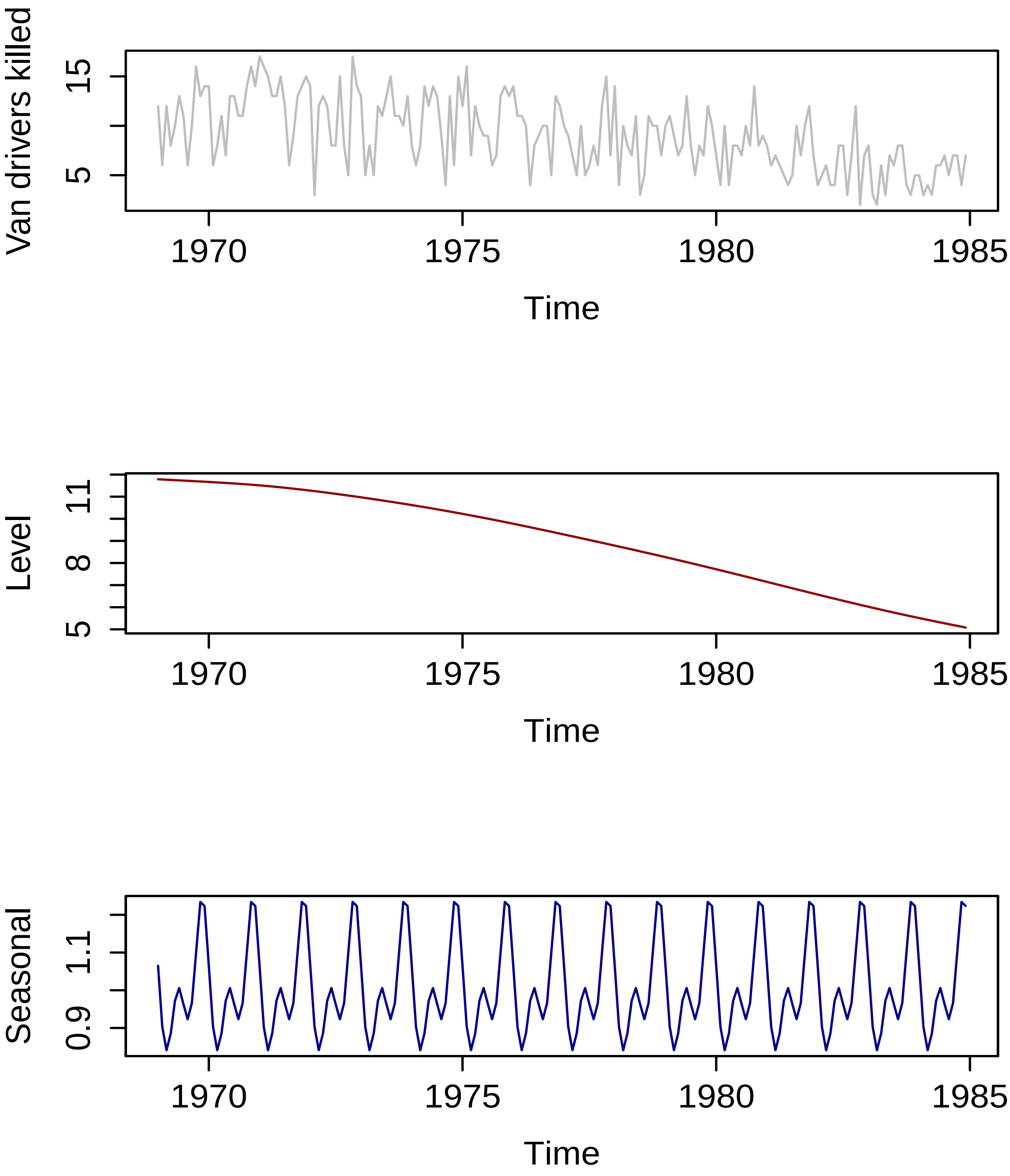


Figure 9.15: The fitted RW(2) local level and smooth seasonal of model `m22` for monthly number of light goods van drivers killed in road accidents in the UK from January 1969 to December 1984.

9.5 Polio incidence in the United States

The data of this example is a time series of the monthly number of cases of poliomyelitis reported in the U.S. Centres for Disease Control from January 1970 to December 1983. The data is available in R as `polio` within the package `gamlss` (Stasinopoulos *et al.*, 2008).

The polio data, consist of 168 observations, were originally modelled by Zeger (1988) who used a parameter-driven approach with the Poisson conditional distribution, in which a first order autoregressive model was used for the latent process. Using an alternative observation-driven approach, Li (1994) compared a second order moving average log linear Poisson process to a second order Markov autoregressive model of Zeger and Qaqish (1988). Fokianos (2000) improved the fit by applying a truncated Poisson model. Benjamin *et al.* (2003) fitted a negative binomial and Poisson GARMA models to the data, reporting a significantly better fit for the negative binomial model with deterministic seasonal, and Davis and Wu (2009) fitted the negative binomial distribution to the polio data.

All the previous models used a deterministic seasonal model. They modelled seasonality with a sine and a cosine for 6 months and 12 months, which is a smooth deterministic cyclical model for seasonality, and no one has considered a stochastic seasonality model for the polio data. In addition, Benjamin *et al.* (2003) and Davis and Wu (2009) reported that the negative binomial distribution fits better than the Poisson distribution for the polio data.

Hence, the main interest of the author for modelling the polio data is investigating whether the data is overdispersed by fitting the negative binomial conditional distribution to the polio data, and investigating whether the seasonal effect is stochastic or deterministic. For these reasons the GEST model has the flexility to model the stochastic seasonality with a structural time series model and check whether the seasonality is changing over time, and has the flexibility to use different discrete conditional distributions to model to the overdispersion in the data.

Let Y_t be the monthly number of cases of poliomyelitis reported in the U.S. from January 1970 to December 1983, and consider; the conditional Poisson distribution $PO(Y_t|\mu_t)$ of the response variable Y_t in the GEST model, and the conditional negative binomial type I distribution $NBI(Y_t|\mu_t, \sigma_t)$ of the response variable Y_t in the GEST model, with and without trend in time, and with deterministic seasonality and stochastic seasonality to check whether there were any changes in the seasonal effect in the polio data.

9.5.1 Conditional Poisson distribution

Here we consider conditional Poisson distribution models with each of a random walk order 1 and autoregressive order 1 local levels, each of which with (i) no seasonal effect, (ii) a deterministic smooth seasonal effect, (iii) a deterministic seasonal effect factor, (iv) a stochastic seasonal effect, each of which (I) without a fixed linear trend in time, (II) with a fixed linear trend in time. The results are summarized later in Tables 9.16, 9.17, 9.18, and 9.19.

RW(1) local level model, `p1`

$$\begin{aligned} Y_t|\mu_t &\sim PO(\mu_t) \\ \log(\mu_t) &= \beta_{1,0} + \gamma_{1,t} \\ \gamma_{1,t} &= \gamma_{1,t-1} + b_{1,t} \end{aligned}$$

where $b_{1,t} \sim NO(0, \sigma_b^2)$.

RW(1) local level and deterministic smooth seasonal effect model, `p2`

$$
\begin{aligned}
Y_t|\mu_t &\sim PO(\mu_t) \\
\log(\mu_t) &= \beta_{1,0} + \beta_{1,1}x_{1,t} + \beta_{1,2}x_{2,t} + \beta_{1,3}x_{3,t} + \beta_{1,4}x_{4,t} + \gamma_{1,t} \\
\gamma_{1,t} &= \gamma_{1,t-1} + b_{1,t}
\end{aligned}
$$

where $b_{1,t} \sim NO(0, \sigma_b^2)$.

RW(1) local level and deterministic seasonal effect factor model, `p3`

$$
\begin{aligned}
Y_t|\mu_t &\sim PO(\mu_t) \\
\log(\mu_t) &= \beta_{1,0} + \xi_t + \gamma_{1,t} \\
\gamma_{1,t} &= \gamma_{1,t-1} + b_{1,t}
\end{aligned}
$$

where $b_{1,t} \sim NO(0, \sigma_b^2)$ and where ξ is a factor with 12 levels, for fixed seasonality, $\xi^\top = (1, 2, \ldots, 12, 1, 2, \ldots, 12, 1, 2, \ldots, 12)$.

RW(1) local level and stochastic seasonal model, `p4`

$$
\begin{aligned}
Y_t|\mu_t &\sim PO(\mu_t) \\
\log(\mu_t) &= \beta_{1,0} + \gamma_{1,t} + s_{1,t}
\end{aligned}
$$

where

$$
\begin{aligned}
\gamma_{1,t} &= \gamma_{1,t-1} + b_{1,t} \\
s_{1,t} &= -\sum_{m=1}^{M-1} s_{1,t-m} + w_t
\end{aligned}
$$

where $b_{1,t} \sim NO(0, \sigma_b^2)$ and $w_{1,t} \sim NO(0, \sigma_w^2)$.

RW(1) local level with fixed linear trend model, `p5`

$$
\begin{aligned}
Y_t|\mu_t &\sim PO(\mu_t) \\
\log(\mu_t) &= \beta_{1,0} + \beta_{1,1}t + \gamma_{1,t} \\
\gamma_{1,t} &= \gamma_{1,t-1} + b_{1,t}
\end{aligned}
$$

where $b_{1,t} \sim NO(0, \sigma_b^2)$.

RW(1) local level with fixed linear trend and deterministic smooth seasonal effect model, `p6`

$$
\begin{aligned}
Y_t|\mu_t &\sim PO(\mu_t) \\
\log(\mu_t) &= \beta_{1,0} + \beta_{1,1}t + \beta_{1,2}x_{1,t} + \beta_{1,3}x_{2,t} + \beta_{1,4}x_{3,t} + \beta_{1,5}x_{4,t} + \gamma_{1,t} \\
\gamma_{1,t} &= \gamma_{1,t-1} + b_{1,t}
\end{aligned}
$$

where $b_{1,t} \sim NO(0, \sigma_b^2)$.

RW(1) local level with fixed linear trend and deterministic seasonal effect factor model, `p7`

$$
\begin{aligned}
Y_t|\mu_t &\sim PO(\mu_t) \\
\log(\mu_t) &= \beta_{1,0} + \beta_{1,1}t + \xi_t + \gamma_{1,t} \\
\gamma_{1,t} &= \gamma_{1,t-1} + b_{1,t}
\end{aligned}
$$

where $b_{1,t} \sim NO(0, \sigma_b^2)$ and where ξ is a factor with 12 levels, for fixed seasonality, $\xi^\top = (1, 2, \ldots, 12, 1, 2, \ldots, 12, 1, 2, \ldots, 12)$.

RW(1) local level with fixed linear trend and stochastic seasonal model, `p8`

$$
\begin{aligned}
Y_t|\mu_t &\sim PO(\mu_t) \\
\log(\mu_t) &= \beta_{1,0} + \beta_{1,1}t + \gamma_{1,t} + s_{1,t}
\end{aligned}
$$

where

$$
\begin{aligned}
\gamma_{1,t} &= \gamma_{1,t-1} + b_{1,t} \\
s_{1,t} &= -\sum_{m=1}^{M-1} s_{1,t-m} + w_t
\end{aligned}
$$

where $b_{1,t} \sim NO(0, \sigma_b^2)$ and $w_{1,t} \sim NO(0, \sigma_w^2)$.

AR(1) local level model, `g1`

$$
\begin{aligned}
Y_t|\mu_t &\sim PO(\mu_t) \\
\log(\mu_t) &= \beta_{1,0} + \gamma_{1,t} \\
\gamma_{1,t} &= \phi_1\gamma_{1,t-1} + b_{1,t}
\end{aligned}
$$

where $b_{1,t} \sim NO(0, \sigma_b^2)$.

AR(1) local level and deterministic smooth seasonal effect model, `g2`

$$\begin{aligned} Y_t|\mu_t &\sim PO(\mu_t) \\ \log(\mu_t) &= \beta_{1,0} + \beta_{1,1}x_{1,t} + \beta_{1,2}x_{2,t} + \beta_{1,3}x_{3,t} + \beta_{1,4}x_{4,t} + \gamma_{1,t} \\ \gamma_{1,t} &= \phi_1\gamma_{1,t-1} + b_{1,t} \end{aligned}$$

where $b_{1,t} \sim NO(0, \sigma_b^2)$.

AR(1) local level and deterministic seasonal effect factor model, `g3`

$$\begin{aligned} Y_t|\mu_t &\sim PO(\mu_t) \\ \log(\mu_t) &= \beta_{1,0} + \xi_t + \gamma_{1,t} \\ \gamma_{1,t} &= \phi_1\gamma_{1,t-1} + b_{1,t} \end{aligned}$$

where $b_{1,t} \sim NO(0, \sigma_b^2)$ and where ξ is a factor with 12 levels, for fixed seasonality, $\xi^\top = (1, 2, \ldots, 12, 1, 2, \ldots, 12, 1, 2, \ldots, 12)$.

AR(1) local level and stochastic seasonal model, `g4`

$$\begin{aligned} Y_t|\mu_t &\sim PO(\mu_t) \\ \log(\mu_t) &= \beta_{1,0} + \gamma_{1,t} + s_{1,t} \end{aligned}$$

where

$$\begin{aligned} \gamma_{1,t} &= \phi_1\gamma_{1,t-1} + b_{1,t} \\ s_{1,t} &= -\sum_{m=1}^{M-1} s_{1,t-m} + w_t \end{aligned}$$

where $b_{1,t} \sim NO(0, \sigma_b^2)$ and $w_{1,t} \sim NO(0, \sigma_w^2)$.

AR(1) local level with fixed linear trend model, g5

$$\begin{aligned} Y_t|\mu_t &\sim PO(\mu_t) \\ \log(\mu_t) &= \beta_{1,0} + \beta_{1,1}t + \gamma_{1,t} \\ \gamma_{1,t} &= \phi_1\gamma_{1,t-1} + b_{1,t} \end{aligned}$$

where $b_{1,t} \sim NO(0, \sigma_b^2)$.

AR(1) local level with fixed linear trend and deterministic smooth seasonal effect model, g6

$$\begin{aligned} Y_t|\mu_t &\sim PO(\mu_t) \\ \log(\mu_t) &= \beta_{1,0} + \beta_{1,1}t + \beta_{1,2}x_{1,t} + \beta_{1,3}x_{2,t} + \beta_{1,4}x_{3,t} + \beta_{1,5}x_{4,t} + \gamma_{1,t} \\ \gamma_{1,t} &= \phi_1\gamma_{1,t-1} + b_{1,t} \end{aligned}$$

where $b_{1,t} \sim NO(0, \sigma_b^2)$.

AR(1) local level with fixed linear trend and deterministic seasonal effect factor model, g7

$$\begin{aligned} Y_t|\mu_t &\sim PO(\mu_t) \\ \log(\mu_t) &= \beta_{1,0} + \beta_{1,1}t + \xi_t + \gamma_{1,t} \\ \gamma_{1,t} &= \phi_1\gamma_{1,t-1} + b_{1,t} \end{aligned}$$

where $b_{1,t} \sim NO(0, \sigma_b^2)$ and where ξ is a factor with 12 levels, for fixed seasonality, $\xi^\top = (1, 2, \ldots, 12, 1, 2, \ldots, 12, 1, 2, \ldots, 12)$.

AR(1) local level with fixed linear trend and stochastic seasonal model, g8

$$
\begin{aligned}
Y_t|\mu_t &\sim PO(\mu_t) \\
\log(\mu_t) &= \beta_{1,0} + \beta_{1,1} t + \gamma_{1,t} + s_{1,t}
\end{aligned}
$$

where

$$
\begin{aligned}
\gamma_{1,t} &= \phi_1 \gamma_{1,t-1} + b_{1,t} \\
s_{1,t} &= -\sum_{m=1}^{M-1} s_{1,t-m} + w_t
\end{aligned}
$$

where $b_{1,t} \sim NO(0, \sigma_b^2)$ and $w_{1,t} \sim NO(0, \sigma_w^2)$.

Table 9.16: RW(1) local level models without `trend` using the conditional Poisson distribution

model	$\hat{\sigma}_b^2$	$\hat{\sigma}_w^2$	x_t	ξ_t	df	cond. AIC
`p1`	0.2188918	-	-	-	43.96	495.52
`p2`	0.07116778	-	y	-	28.74	494.94
`p3`	0.08117827	-	-	y	38.33	496.73
`p4`	0.07903896	0.00829891	-	-	41.95	**493.12**

The R commands for fitting these models are given in Appendix D.

Table 9.17: RW(1) local level models with `trend` using the conditional Poisson distribution

model	$\hat{\sigma}_b^2$	$\hat{\sigma}_w^2$	x_t	ξ_t	$\hat{\beta}_{1,1}$	df	cond. AIC
`p5`	0.219989	-	-	-	-0.003753	45.08	497.52
`p6`	0.0711074	-	y	-	-0.003821	29.72	497.02
`p7`	0.08114721	-	-	y	-0.004001	39.31	498.79
`p8`	0.07892945	0.008317149	-	-	-0.004254	42.93	**495.19**

The R commands for fitting these models are given in Appendix D.

where y=yes in Tables 9.16, 9.17, 9.18, and 9.19.

Table 9.18: AR(1) local level models without `trend` using the conditional Poisson distribution

model	$\hat{\sigma}_b^2$	$\hat{\sigma}_w^2$	x_t	ξ_t	$\hat{\phi}_1$	df	cond. AIC
g1	0.7503972	-	-	-	0.401733	95.53	515.45
g2	0.6581206	-	y	-	0.345565	94.90	516.27
g3	0.6019258	-	-	y	0.4111022	99.10	522.23
g4	0.6818061	5.878952e-06	-	-	0.372089	95.63	**512.40**

The R commands for fitting these models are given in Appendix D.

Table 9.19: AR(1) local level models with `trend` using the conditional Poisson distribution

model	$\hat{\sigma}_b^2$	$\hat{\sigma}_w^2$	x_t	ξ_t	$\hat{\beta}_{1,1}$	$\hat{\phi}_1$	df	cond. AIC
g5	0.7693205	-	-	-	-0.004031	0.3532904	98.07	518.05
g6	0.6956179	-	y	-	-0.004521	0.2528894	99.09	519.62
g7	0.6510392	-	-	y	-0.004751	0.3135243	104.41	526.97
g8	0.7302721	3.132148e-07	-	-	-0.004772	0.2788073	100.15	**516.09**

The R commands for fitting these models are given in Appendix D.

The first two Tables, 9.16 and 9.17, summarize the random walk local level models with and without a fixed linear trend, and with deterministic seasonal and stochastic seasonal, and Tables 9.18, and 9.19 summarize the autoregressive local level models with and without a fixed linear trend, and with deterministic seasonal and stochastic seasonal. The random walk local level models have a better conditional AIC compared with conditional AIC of the autoregressive local level models for the polio data, in which case, the best model with the smallest AIC is model `p4`, RW(1) local level and stochastic seasonal model, given in Figure 9.17.

In addition, the local level and stochastic seasonal models have a smallest AIC than the local level and deterministic seasonal models in all the tables, implying that the seasonality of the polio is changing over time and needs a stochastic seasonal model rather than a deterministic seasonal model.

Figure 9.16 shows the time series of the polio data, from January 1970 to December 1983 , and the fitted of the RW(1) local level of model `p4` without plotting seasonality.

Figure 9.17 shows the time series of the polio data and the decomposition of fitted model `p4` into two components, RW(1) local level and stochastic seasonal. It is clear that the fitted seasonality is changing over time and is not fixed, hence, the seasonality of the polio is stochastic rather than deterministic.

95% confidence interval for the hyperparameters σ_b^2 and σ_w^2 in `p4`

To calculate the 95% two-sided confidence interval for the hyperparameters σ_b^2 and σ_w^2, the following transformations are needed for the GEST model `p4`. The GEST estimates the $\log(\hat{\sigma}_b^2)$ and $\log(\hat{\sigma}_w^2)$ and calculates their standard errors by inverting the Hessian matrix obtained from the fitted model `p4`, these standard errors are for the predictors for the hyperparameters.

Hence, the 95% confidence level for $\log(\sigma_b^2)$ is given by

$\log(\hat{\sigma}_b^2)$`-1.96(se(`$\log(\hat{\sigma}_b^2)$`)) = -2.53781433 - 1.96(0.4557181)= -3.431022`

$\log(\hat{\sigma}_b^2)$`+1.96(se(`$\log(\hat{\sigma}_b^2)$`)) = -2.53781433 + 1.96(0.4557181) = -1.644607`

Now

$\hat{\sigma}_b^2$ `=` exp`(-2.53781433)= 0.07903896` and

exp`(-3.431022)= 0.03235386` and

exp`(-1.644607)= 0.1930884`.

Hence,

`[0.1930884, 0.03235386] is 95% confidence level for` σ_b^2.

The 95% confidence level for $\log(\sigma_w^2)$ is given by

$\log(\hat{\sigma}_w^2)$`-1.96(se(`$\log(\hat{\sigma}_w^2)$`)) = -4.79163116 - 1.96(1.2339636)= -7.2102`

$\log(\hat{\sigma}_w^2)$`+1.96(se(`$\log(\hat{\sigma}_w^2)$`)) = -4.79163116 + 1.96(1.2339636) = -2.373063`

Now

$\hat{\sigma}_w^2$ `=` exp`(-4.79163116)= 0.008298909` and

exp`(-7.2102)= 0.0007390093` and

exp`(-2.373063)= 0.09319483`.

Hence,

`[0.0007390093, 0.09319483] is 95% confidence level for` σ_w^2.

9.5.2 Conditional negative binomial type I distribution

Here we consider conditional negative binomial type I, (NBI), distribution models with each of a random walk order 1 and autoregressive order 1 local levels, each of which with (i) no seasonal effect, (ii) a deterministic smooth seasonal effect, (iii) a deterministic seasonal effect factor, (iv) a stochastic seasonal effect, each of which (I) without a fixed linear trend in time, (II) with a fixed linear trend in time. The $\log(\sigma_t) = \beta_{2,0}$, a constant for all NBI models. The results are summarized later in

Tables 9.20, 9.21, 9.22, and 9.23.

Table 9.20: RW(1) local level models without `trend` using the conditional negative binomial distribution

model	$\hat{\sigma}_b^2$	$\hat{\sigma}_w^2$	ξ_t	x_t	$\hat{\beta}_{2,0}$	df	cond. AIC
`n1`	0.2188917	-	-	-	-36.05	44.96	497.52
`n2`	0.04895041	-	-	y	-2.108	25.07	497.93
`n3`	0.06100436	-	y	-	-2.721	35.27	500.07
`n4`	0.07546877	0.007867358	-	-	-36.05	41.87	**495.36**

The R commands for fitting these models are given in Appendix D.

Table 9.21: RW(1) local level models with `trend` using the conditional negative binomial distribution

model	$\hat{\sigma}_b^2$	$\hat{\sigma}_w^2$	ξ_t	x_t	$\hat{\beta}_{1,1}$	$\hat{\beta}_{2,0}$	df	cond. AIC
`n5`	0.2193205	-	-	-	-0.001671	-36.05	46.01	499.52
`n6`	0.04881961	-	-	y	-0.000558	-2.105	26.04	499.95
`n7`	0.06056373	-	y	-	-0.003242	-2.704	36.17	502.15
`n8`	0.07886942	0.008314509	-	-	-0.003267	-36.05	43.92	**497.18**

The R commands for fitting these models are given in Appendix D.

Table 9.22: AR(1) local level models without `trend` using the conditional negative binomial distribution

model	$\hat{\sigma}_b^2$	$\hat{\sigma}_w^2$	ξ_t	x_t	$\hat{\beta}_{2,0}$	$\hat{\phi}_1$	df	cond. AIC
`e1`	0.7504561	-	-	-	-36.06	0.4017	96.53	517.43
`e2`	0.6580777	-	-	y	-36.05	0.3455	95.90	518.27
`e3`	0.601919	-	y	-	-36.05	0.4111	100.10	524.23
`e4`	0.6821032	1.429029e-06	-	-	-36.05	0.3719	96.64	**514.40**

The R commands for fitting these models are given in Appendix D.

Table 9.23: AR(1) local level models with `trend` using the conditional negative binomial distribution

model	$\hat{\sigma}_b^2$	$\hat{\sigma}_w^2$	ξ_t	x_t	$\hat{\beta}_{1,1}$	$\beta_{2,0}$	$\hat{\phi}_1$	df	cond. AIC
`e5`	0.7693097	-	-	-	-0.00403	-36.06	0.353	99.06	520.02
`e6`	0.6956196	-	-	y	-0.004521	-36.05	0.259	100.09	521.62
`e7`	0.6510741	-	y	-	-0.004751	-36.05	0.313	105.40	528.96
`e8`	0.7301156	1.6325e-05	-	-	0.004772	-36.05	0.279	101.15	**518.09**

The R commands for fitting these models are given in Appendix D.

From Tables 9.20, 9.21, 9.22, and 9.23, n4 is the best model according to Akaike information criteria, but in comparison with model p4, the AIC has not improved with the conditional negative binomial distribution, which indicates that model p4 is still the best model. Hence, the Poisson distribution is the best conditional distribution for the polio data. Note that the estimate of $\log(\hat{\sigma}_t) = \hat{\beta}_{2,0} = -36.05$, hence, $\hat{\sigma}_t = 0$, implies that the variance of the NBI is equal to the variance of Poisson, which indicates that there is no dispersion in the polio data.

Benjamin *et al.* (2003) fitted negative binomial and Poisson GARMA(0,2) models with deterministic seasonal and with and without trend, and their best chosen model was a negative binomial GARMA(0,2) model without trend with the AIC = 504.9. The GEST model has a smaller AIC than the GARMA(0,2) model and disagrees with the GARMA(0,2) in regards to the best fitted distribution for the polio. The R commands for plotting Figure 9.16 and Figure 9.17 are given in Appendix D.

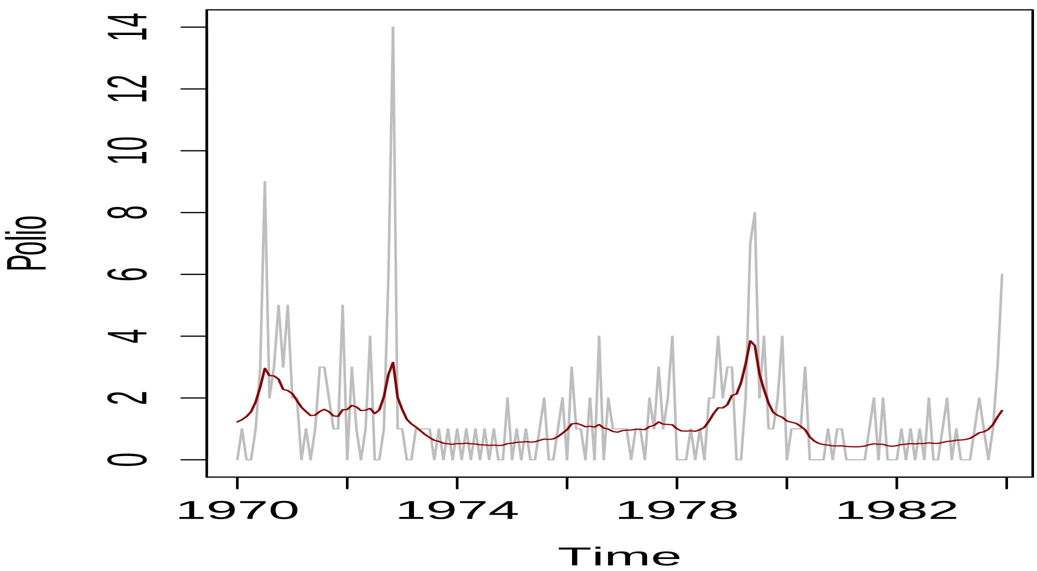


Figure 9.16: Monthly number of polio cases in the U.S. from 1970 to 1983 in gray and the fitted local level of model p4 in red.

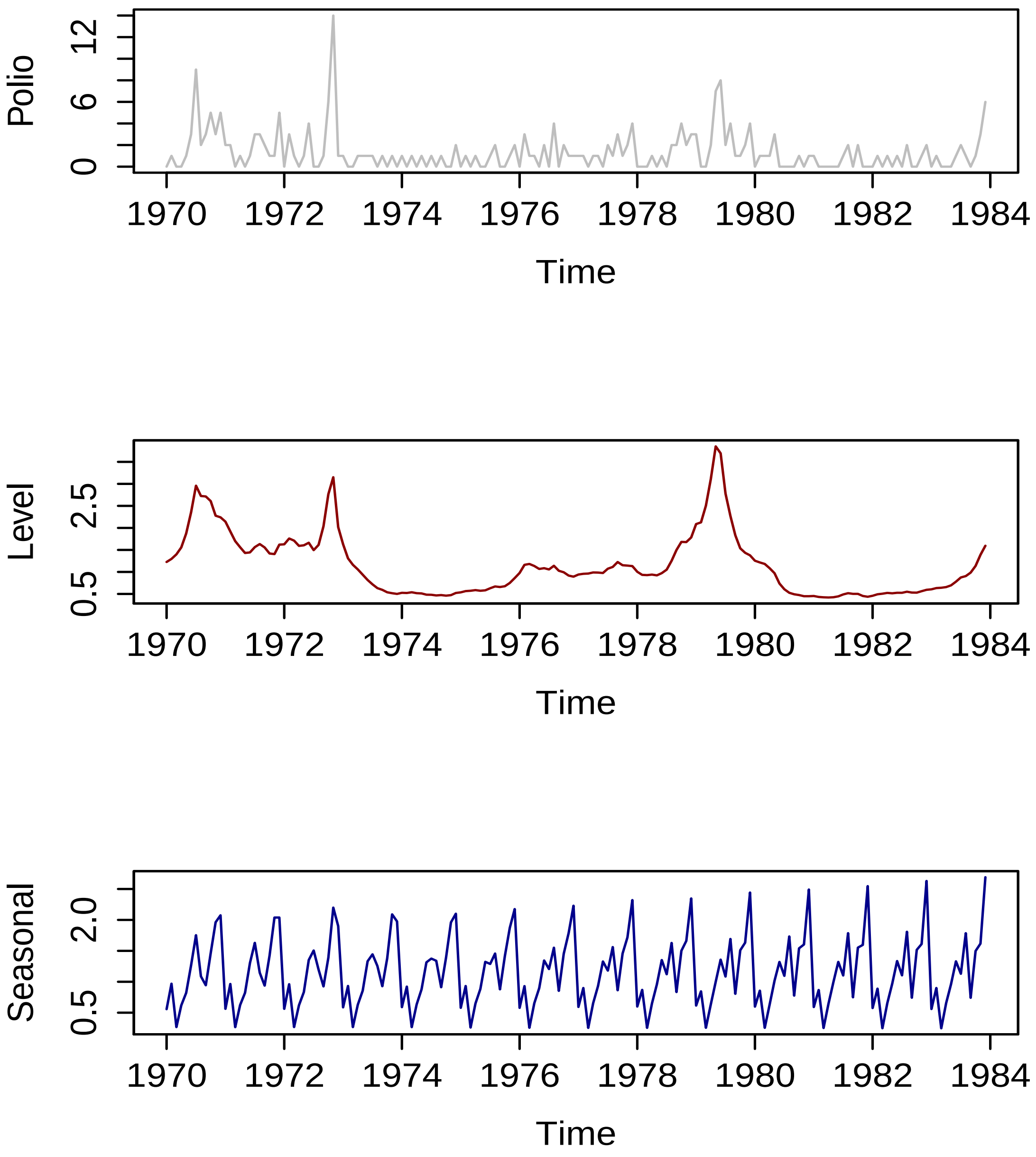


Figure 9.17: Monthly number of polio cases in the U.S. from 1970 to 1983 in gray and the fitted local level in red and the fitted stochastic seasonality in blue of model p4.

Chapter 10

Conclusion and future developments

The thesis presents a new approach for modelling univariate time series, namely the generalized structural time series (GEST) model. The proposed GEST model primarily addresses the difficulty in modelling time-varying skewness and time-varying kurtosis (beyond mean and dispersion time series models) to better describe the non-Gaussian movements in a time series. Proofs of some of the properties of the GEST process are given in Appendix C.

This chapter outlines the main contributions of the thesis in time series analysis and proposes some directions for future developments of the GEST model.

10.1 Originality of the GEST process

The thesis introduces a novel and general stochastic process, namely the GEST process, for Gaussian and non-Gaussian, continuous and discrete, seasonal and non-seasonal time series data.

The GEST process extends the traditional Gaussian processes to non-Gaussian processes and is implemented with 80 conditional distributions, in either stationary and non-stationary situations, by modelling time varying-mean, time-varying variance, time-varying skewness and time-varying kurtosis. Stationarity properties of the GEST process under specific conditions are explored in two theorems in chapter 7.

10.2 Originality of the GEST model

In addition, this thesis makes a number of original contributions to the area of structural time series modelling by developing a new class of univariate time series models, the Generalized Structural (GEST) time series model which extends the traditional univariate Gaussian structural models to non-Gaussian situation.

The GEST model allows modelling of any or all the parameters of the conditional distribution, of Y_t given the past, using structural terms and explanatory terms (e.g. linear and/or non-linear parametric terms and/or smoothing terms in explanatory variables).

For example, the GEST model extends the current Gaussian structural framework of modeling the conditional mean and conditional variance to include two more parameters for modelling conditional skewness and conditional kurtosis in a non-Gaussian structural framework. In this extended structural model, all the parameters are modeled jointly and explicitly via an autoregressive structural term and a seasonal structural term, together with explanatory terms.

The GEST model is a general regression model with a stochastic nature based on the GEST process. A method of estimating the parameters of the GEST model is described and explained in Chapter 8. The S&P 500 returns are used as one example to show that the GEST model outperforms the traditional GARCH type models as a better representation of the past observations. More applications of the GEST model to counts time series, e.g. van drivers killed in the UK and Polio incidence in the US are illustrated.

10.3 Important applications of the GEST model

There are several important points to make here:

- The GEST model allows the use of a flexible parametric distribution $\mathcal{D}(\mu_t, \sigma_t, \nu_t, \tau_t)$ for the dependent variable, including highly skew and/or kurtotic distributions such as the generalized beta type 2 (including the special case of the generalized Pareto) of McDonald and Xu (1995), power exponential of Nelson (1991), Johnson's SU of Johnson et al. (1994), Gumbel of Crowder et al. (1991), Box-Cox Cole-Green of Cole and Green (1992), Sinh-arcsinh of Jones and Pewsey (2009) and skewed t-family distributions.

- The use of a flexible parametric distribution allows: a) the fitting of the GEST model using the penalised likelihood estimation algorithm discussed in section chapter 8, and b) the use of a variety of diagnostic tools (from both the econometric or standard statistical literature) for model checking and selection (see, for example, chapter 9).

- The GEST model expands the systematic part of time series models to allow the stochastic modelling of any or all of the distribution parameters $(\mu_t, \sigma_t, \nu_t, \tau_t)$ as structural terms and (if necessary) linear, non-linear and smooth functions of independent variables (see chapter 8).

- The structural terms for each distribution parameter of the conditional distribution can be a random walk or autoregressive (of any order) and can include seasonal and/or leverage effects.

- The GEST model can model seasonal count data, with a random walk or autoregressive and seasonal structural terms, using a Poisson, negative binomial, Delaporte, or Sichel conditional distribution. If the data is counts without seasonal pattern, then the GEST with a random walk or autoregressive term can be used.

- Note, the conditional distribution of the response variable given the past can be *any* continuous or discrete distribution. In fact any of the 80 or so distributions implemented in the R package `gamlss.dist`, Stasinopoulos *et al.* (2008), can be used. This provides a very flexible way of analysing physical phenomena time series containing rare events, e.g. flooding, or spells of high counts as in epidemiological data or economic time series data like inflation data.

- Time-varying mean, variance, skewness and kurtosis are of interest in themselves and provide information on various aspects of a time series. The GEST model provides a useful framework based upon time-varying estimates of the distribution parameters $(\mu_t, \sigma_t, \nu_t, \tau_t)$.

Chapter 9 demonstrates the flexibility of the GEST model to model the stochastic seasonal effect of the Polio incidents in the U.S.; to have different orders for the random walk local level in the van drivers killed in the UK; and to capture the time-varying skewness and kurtosis of the returns of the S&P 500 stock index. Furthermore, a variety of diagnostic tools have also been used to compare the adequacy of the GEST model with the APARCH model for the returns of the S&P 500 stock index.

10.4 Limitations and future developments

Clearly a large variety of models can be fitted within the GEST framework, but further development is needed to establish the model capability for forecasting.

Further work includes:

- Assessing forecasting with GEST using a recursive forecasting algorithm, and

- a simulation study to explore the properties of hyperparameters estimates and reliability of the standard errors.

  These two further works on the GEST have been considered by the author, but due to time constraint of this research and the time consumption of both algorithms, simulation study and forecasting, have been proposed for future developments of the GEST.

- There is a possibility that the GEST model is over-fitting the parameters of the conditional distribution when modelling the time series. Furthermore, the results are all based on in-sample fit, and the over-fitting could be due to the flexibility of the GEST model and the complexity of its structure in fitting time series, which is good for explaining the past, but not necessarily a good model for forecasting the future. However, in empirical volatility studies, the more challenging and interesting aspect of a new model or method is its ability to forecast volatility more accurately. Thus, one of the important future development of the GEST model is forecasting the stochastic volatility and comparing the forecast of the GEST model with other models. In addition, forecasting the shape parameters (skewness and kurtosis parameters) of the conditional distribution and comparing the in-sample fit with out-sample fit are both important for the development of the GEST model.

- The Q function uses a multivariate normal distribution with weights, alternatively using the Kalman filter with weights can be another way of estimation of the hyperparameters for non-Gaussian time series observations within the GEST model.

- Fitting the parameter τ_t of the conditional distribution using the GEST model can be difficult when the sample size of the observations is small. A possible solution for capturing movements of the parameter τ_t is to increase the sample size so the tail of the conditional distribution is changing over time, or to use a sophisticated optimization tool by coding the algorithm of the GEST model in JAVA.

- Finally a possible development is to speed up the global estimation method and use it instead of the local estimation method, using C programming language.

Appendix A

Derivations of Chapter 4

A.1 Derivations of Section 4.2

Let

$$f(y) = \int f(y|\gamma)f(\gamma)d\gamma,$$

where

$$f(y|\gamma) \quad = \quad \frac{1}{|2\pi\Sigma|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(y - X\beta - Z\gamma)^{\top}\Sigma^{-1}(y - X\beta - Z\gamma)\right\},$$

and

$$f(\gamma)^{1} \quad = \quad \frac{1}{|2\pi D|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\gamma^{\top} D^{-1}\gamma)\right\},$$

hence

[1] This is the prior density for γ.

$$
\begin{aligned}
f(y) &= \int \frac{1}{|2\pi\Sigma|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(y - X\beta - Z\gamma)^\top \Sigma^{-1} (y - X\beta - Z\gamma)\right\} \\
&\quad \times \frac{1}{|2\pi D|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\gamma^\top D^{-1}\gamma)\right\} d\gamma
\end{aligned}
$$

$$
\begin{aligned}
&= |2\pi\Sigma|^{-\frac{1}{2}} |2\pi D|^{-\frac{1}{2}} \int \exp\{-\frac{1}{2}\left[(y - X\beta - Z\gamma)^\top \Sigma^{-1}(y - X\beta - Z\gamma) + (\gamma^\top D^{-1}\gamma)\right]\} d\gamma \\
&= |2\pi\Sigma|^{-\frac{1}{2}} |2\pi D|^{-\frac{1}{2}} \int \exp\{-\frac{1}{2}[\gamma^\top (Z^\top \Sigma^{-1} Z + D^{-1})\gamma + (y - X\beta)^\top \Sigma^{-1} (y - X\beta) \\
&\quad -2(y - X\beta)^\top \Sigma^{-1} Z\gamma]\} d\gamma.
\end{aligned}
$$

Writing

$$
\hat{\gamma} = (Z^\top \Sigma^{-1} Z + D^{-1})^{-1} Z^\top \Sigma^{-1} (y - X\beta)
$$

$$
\begin{aligned}
f(y) &= |2\pi\Sigma|^{-\frac{1}{2}} |2\pi D|^{-\frac{1}{2}} \int \exp\{-\frac{1}{2}[(\gamma - \hat{\gamma})^\top (Z^\top \Sigma^{-1} Z + D^{-1})(\gamma - \hat{\gamma}) \\
&\quad +(y - X\beta)^\top \Sigma^{-1}(y - X\beta) - \hat{\gamma}^\top (Z^\top \Sigma^{-1} Z + D^{-1})\hat{\gamma}]\} d\gamma \\
&= |2\pi\Sigma|^{-\frac{1}{2}} |2\pi D|^{-\frac{1}{2}} |2\pi(Z^\top \Sigma^{-1} Z + D^{-1})|^{-\frac{1}{2}} \exp\{-\frac{1}{2}[(y - X\beta)^\top \Sigma^{-1}(y - X\beta) \\
&\quad -(y - X\beta)^\top \Sigma^{-1} Z (Z^\top \Sigma^{-1} Z + D^{-1})^{-1} Z^\top \Sigma^{-1}(y - X\beta)]\} \\
&= |2\pi\Sigma|^{-\frac{1}{2}} |2\pi D|^{-\frac{1}{2}} |2\pi(Z^\top \Sigma^{-1} Z + D^{-1})|^{-\frac{1}{2}} \exp\{-\frac{1}{2}(y - X\beta)^\top \\
&\quad [\Sigma^{-1} - \Sigma^{-1} Z (Z^\top \Sigma^{-1} Z + D^{-1})^{-1} Z^\top \Sigma^{-1}](y - X\beta)\} \\
&= |2\pi\Sigma|^{-\frac{1}{2}} |2\pi D|^{-\frac{1}{2}} |2\pi(Z^\top \Sigma^{-1} Z + D^{-1})|^{-\frac{1}{2}} \exp\{-\frac{1}{2}(y - X\beta)^\top V^{-1}(y - X\beta)\}
\end{aligned}
$$

where

$$V^{-1} = \Sigma^{-1} - \Sigma^{-1}Z(Z^\top\Sigma^{-1}Z + D^{-1})^{-1}Z^\top\Sigma^{-1},$$

and

$$\int \exp\left\{-\frac{1}{2}(\gamma-\hat{\gamma})^\top(Z^\top\Sigma^{-1}Z + D^{-1})(\gamma-\hat{\gamma})\right\} d\gamma = |2\pi(Z^\top\Sigma^{-1}Z + D^{-1})|^{-\frac{1}{2}},$$

since if

$$\begin{aligned} \gamma &\sim N\left(\hat{\gamma}, [Z^\top\Sigma^{-1}Z + D^{-1}]^{-1}\right), \\ f(\gamma) &= |2\pi(Z^\top\Sigma^{-1}Z + D^{-1})|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}(\gamma-\hat{\gamma})^\top(Z^\top\Sigma^{-1}Z + D^{-1})(\gamma-\hat{\gamma})\right\}, \end{aligned}$$

and

$$\int f(\gamma)d\gamma = 1.$$

Pawitan (2001), page 446 shows that

$$|V| = |\Sigma||D||Z^\top\Sigma^{-1}Z + D^{-1}|.$$

Hence

$$f(y) = |2\pi V|^{-\frac{1}{2}} \exp\{-\frac{1}{2}(y - X\beta)^\top V^{-1}(y - X\beta)\}. \tag{A.1}$$

An alternative method for deriving the marginal distribution of y is to use the characteristic function:

$$\begin{aligned} \text{if } y &\sim N(X\beta + Z\gamma, \Sigma), \\ \text{and } \gamma &\sim N(0, D), \\ \text{then } E_y\left[\exp(iT^\top y)\right] &= E_\gamma\left[E_{y/\gamma}[\exp(iT^\top y)|\gamma]\right], \\ \text{where } E_{y/\gamma}\left[\exp(iT^\top y)|\gamma\right] &= \exp\left(iT^\top[X\beta + Z\gamma] - \frac{1}{2}T^\top\Sigma T\right), \end{aligned}$$

so

$$\begin{aligned} E_\gamma\left[E_{y/\gamma}[\exp(iT^\top y)|\gamma]\right] &= E_\gamma\left[\exp\left(iT^\top[X\beta + Z\gamma] - \frac{1}{2}T^\top\Sigma T\right)\right] \\ &= \exp\left(iT^\top X\beta - \frac{1}{2}T^\top\Sigma T\right) E_\gamma\left[\exp(iT^\top Z\gamma)\right] \\ &= \exp\left(iT^\top X\beta - \frac{1}{2}T^\top\Sigma T\right) \exp\left(iT^\top Z0 - \frac{1}{2}T^\top ZDZ^\top T\right) \\ &= \exp\left(iT^\top X\beta - \frac{1}{2}T^\top(\Sigma + ZDZ)^\top T\right). \end{aligned}$$

Hence

$$\begin{aligned} E_y\left[\exp(iT^\top y)\right] &= \exp\left(iT^\top X\beta - \frac{1}{2}T^\top(\Sigma + ZDZ)^\top T\right), \\ \text{so } y &\sim N\left(X\beta, \Sigma + ZDZ^\top\right), \\ \text{where } \Sigma + ZDZ^\top &= V, \\ \text{then } y &\sim N\left(X\beta, V\right). \end{aligned}$$

A.2 Derivations of Section 4.3

From 4.2 the marginal distribution of y is normal with mean $\mathbf{X}\beta$ and variance V such that:

$$\begin{aligned} E(y) &= E_\gamma[E_y(y|\gamma)] = E_\gamma[X\beta + Z\gamma] = X\beta \\ V(y) &= E_\gamma[V_y(y|\gamma)] + V_\gamma[E_y(y|\gamma)] \\ &= E_\gamma[\Sigma] + V_\gamma[X\beta + Z\gamma] \\ &= \Sigma + ZDZ^\top = V \end{aligned} \tag{A.2}$$

The marginal log-likelihood of the fixed parameters (β, θ) is

$$\log L(\beta, \theta) = -\frac{1}{2}\log|V| - \frac{1}{2}(y - X\beta)^\top V^{-1}(y - X\beta) \tag{A.3}$$

For fixed θ, taking the derivatives of the marginal log-likelihood $\log L(\beta, \theta)$ with respect to β, we find the estimate of β as the solution of

$$\begin{aligned} \left(X^\top V^{-1}X\right)\beta &= X^\top V^{-1}y \\ \hat{\beta} &= \left(X^\top V^{-1}X\right)^{-1} X^\top V^{-1}y \end{aligned} \tag{A.4}$$

The standard errors for $\hat{\beta}$ can be calculated from the observed Fisher information of β:

$$I(\hat{\beta}) = X^\top V^{-1} X \tag{A.5}$$

A.3 Derivations of Section 4.4

The conditional distribution of y given γ is multivariate normal with mean and variance given in (4.2) and the random effects γ is multivariate normal with mean 0 and variance D. The log-likelihood of the random effects γ is given by

$$\begin{aligned}\log L(\beta, \theta, \gamma) &= -\frac{1}{2}\log|\Sigma| - \frac{1}{2}(y - X\beta - Z\gamma)^\top \Sigma^{-1}(y - X\beta - Z\gamma) \\ &\quad -\frac{1}{2}\log|D| - \frac{1}{2}\gamma^\top D^{-1}\gamma\end{aligned} \tag{A.6}$$

Given the fixed parameters (β, θ), taking the derivative of the log-likelihood with respect to γ

$$\frac{\partial \log L}{\partial \gamma} = Z^\top \Sigma^{-1}(y - X\beta - Z\gamma) - D^{-1}\gamma \tag{A.7}$$

setting (A.7) to zero we obtain the estimate of the random effects γ.

The estimate $\hat{\gamma}$ is the solution of

$$\begin{aligned}(Z^\top \Sigma^{-1} Z + D^{-1})\gamma &= Z^\top \Sigma^{-1}(y - X\beta) \\ \hat{\gamma} &= (Z^\top \Sigma^{-1} Z + D^{-1})^{-1} Z^\top \Sigma^{-1}(y - X\beta)\end{aligned} \tag{A.8}$$

This estimate is known as the best linear unbiased predictor (BLUP).

The second derivative matrix of the log-likelihood with respect to γ is

$$\frac{\partial^2 \log L}{\partial\gamma\partial\gamma^\top} = -Z^\top\Sigma^{-1}Z - D^{-1},$$

so the observed Fisher information matrix is equal to

$$I(\hat{\gamma}) = Z^\top\Sigma^{-1}Z + D^{-1}. \tag{A.9}$$

Assuming the fixed effects are known, the standard errors for $\hat{\gamma}$ (also interpreted as the prediction error for a random parameter) can be computed as the square root of the diagonal elements of $I(\hat{\gamma})^{-1}$, Pawitan (2001), page 442.

Note that the estimates of β in (A.4) and γ in (A.8) are the joint maximizer of $\log L(\beta, \theta, \gamma)$ at fixed θ. Specifically, the derivative of $\log L(\beta, \theta, \gamma)$ with respect to β is

$$\frac{\partial \log L}{\partial\beta} = X^\top\Sigma^{-1}(y - X\beta - Z\gamma) \tag{A.10}$$

Combining (A.10) with (A.7) and setting them to zero, we have

$$\begin{pmatrix} X^\top\Sigma^{-1}X & X^\top\Sigma^{-1}Z \\ Z^\top\Sigma^{-1}X & Z^\top\Sigma^{-1}Z + D^{-1} \end{pmatrix} \begin{pmatrix} \beta \\ \gamma \end{pmatrix} = \begin{pmatrix} X^\top\Sigma^{-1}y \\ Z^\top\Sigma^{-1}y \end{pmatrix}. \tag{A.11}$$

The estimates obtained from solving this simultaneous equation are exactly those of equation (A.4) and (A.8), Pawitan (2001), page 444-445.

A.4 Derivations of Section 4.5

The variance of y is equal to $V = \Sigma + ZDZ^\top$. The following identities are from Pawitan (2001), page 445-446.

$$\begin{aligned}
V &= \Sigma + ZDZ^\top \\
V^{-1} &= \Sigma^{-1} - \Sigma^{-1}Z(Z^\top\Sigma^{-1}Z + D^{-1})^{-1}Z^\top\Sigma^{-1} \\
|V| &= |\Sigma||D||Z^\top\Sigma^{-1}Z + D^{-1}| \\
\hat{\gamma} &= DZ^\top V^{-1}(y - X\hat{\beta}) \\
&= (Z^\top\Sigma^{-1}Z + D^{-1})^{-1}Z^\top\Sigma^{-1}(y - X\beta) \\
V^{-1}(y - X\hat{\beta}) &= \Sigma^{-1}(y - X\hat{\beta} - Z\hat{\gamma}) \\
(y - X\hat{\beta})^\top V^{-1}(y - X\hat{\beta}) &= (y - X\hat{\beta} - Z\hat{\gamma})^\top\Sigma^{-1}(y - X\hat{\beta} - Z\hat{\gamma}) + \hat{\gamma}D^{-1}\hat{\gamma}
\end{aligned}$$

The determinant of the variance, $|V| = |\Sigma||D||Z^\top\Sigma^{-1}Z + D^{-1}|$, is derived from the following partitioned matrix result.

Let

$$A = \begin{pmatrix} \Sigma & Z \\ Z^\top & -D^{-1} \end{pmatrix},$$

then the determinant of A is equal to:

$$\begin{aligned}
|A| &= |-D^{-1}||\Sigma + ZDZ^\top| \\
&= |\Sigma||-Z^\top\Sigma^{-1}Z - D^{-1}|.
\end{aligned}$$

The following three steps give a summary of the marginal log-likelihood for θ:

1 By integrating out γ from $f(y|\beta,\gamma)f(\gamma)$ and replacing β by $\hat{\beta} = \hat{\beta}(\theta)$, the profile of the marginal log-likelihood for θ is given by

$$pl(\theta) = -\frac{1}{2}\log|V| - \frac{1}{2}(y - X\hat{\beta})^\top V^{-1}(y - X\hat{\beta}).$$

2 Applying further identity for $|V|$ and an identity for $(y - X\hat{\beta})^\top V^{-1}(y - X\hat{\beta})$ given by Pawitan (2001), page 445-446, to $pl(\theta)$, we can rewrite $pl(\theta)$ for computational purpose as:

$$\begin{aligned} pl(\theta) &= -\frac{1}{2}\log|\Sigma| - \frac{1}{2}(y - X\hat{\beta} - Z\hat{\gamma})^\top \Sigma^{-1}(y - X\hat{\beta} - Z\hat{\gamma}) \\ &\quad -\frac{1}{2}\log|D| - \frac{1}{2}\hat{\gamma}^\top D^{-1}\hat{\gamma} - \frac{1}{2}\log|Z^\top \Sigma^{-1} Z + D^{-1}| \end{aligned} \qquad \text{(A.12)}$$

3 By integrating out both β and γ from $f(y|\beta,\gamma)f(\gamma)$, the modified profile likelihood for θ is given by

$$pl_m(\theta) = -\frac{1}{2}\log|V| - \frac{1}{2}\log|X^\top V^{-1} X| - \frac{1}{2}(y - X\hat{\beta})^\top V^{-1}(y - X\hat{\beta})$$

where $-\frac{1}{2}\log|X^\top V^{-1} X|$ is the extra REML term.

4 Removing hats from $\hat{\beta}$ and $\hat{\gamma}$ in the profile of the marginal log-likelihood for θ (A.12) will be called a Q function. Maximising the Q function over β, γ and θ is also maximising $pl(\theta)$.

$$pl(\theta) \equiv Q(\beta, \theta, \gamma)$$

Pawitan (2001), page 446.

A.5 Derivations of Section 4.6

Assume the variance matrices are of the form

$$\Sigma = \sigma_e^2 W$$
$$D = \sigma_\gamma^2 M,$$

where W and M are known matrices of rank N and q respectively, W and M are assumed identity matrices in the estimation, and where the determinant of Σ and the determinant of D are equal to:

$$|\Sigma| = \sigma_e^{2N}|W|$$
$$|D| = \sigma_\gamma^{2q}|M|.$$

An alternative method of estimation is the *alternating* method:

(a) Given θ, calculate $\hat{\beta}_\theta$ and $\hat{\gamma}_\theta$ using:

$$\begin{aligned} \hat{\beta}_\theta &= (X^\top V^{-1}X)^{-1}X^\top V^{-1}y \\ \hat{\gamma}_\theta &= (Z^\top \Sigma^{-1}Z + D^{-1})^{-1}Z^\top \Sigma^{-1}(y - X\beta) \end{aligned}$$

(b) Given $\hat{\beta}_\theta$ and $\hat{\gamma}_\theta$, update σ_e^2 and σ_γ^2 using

$$\begin{aligned} \sigma_e^2 &= \frac{e^\top W^{-1} e}{N-d} \\ \sigma_\gamma^2 &= \frac{\hat{\gamma}_\theta^\top M^{-1} \hat{\gamma}_\theta}{d} \end{aligned}$$

where

$$e = y - X\hat{\beta}_\theta - Z\hat{\gamma}_\theta,$$

$$N = \text{the number of observations,}$$

and

$$d = \textbf{trace}\{(Z^\top W^{-1} Z + \lambda M^{-1})^{-1} Z^\top W^{-1} Z\},$$

where

$$\lambda = \sigma_e^2/\sigma_\gamma^2.$$

The justification for this is given below.

Let

$$\begin{aligned} Q(\theta, \hat{\beta}_\theta, \hat{\gamma}_\theta) &= -\frac{N}{2}\log\sigma_e^2 - \frac{1}{2\sigma_e^2} e^\top W^{-1} e \\ &\quad -\frac{q}{2}\log\sigma_\gamma^2 - \frac{1}{2\sigma_\gamma^2}\gamma^\top M^{-1}\gamma \\ &\quad -\frac{1}{2}\log|\sigma_e^{-2} Z^\top W^{-1} Z + \sigma_\gamma^{-2} M^{-1}| \end{aligned}$$

where $e = y - X\hat{\beta}_\theta - Z\hat{\gamma}_\theta$ is the error vector, and the constant term is not included in the Q, but is equal to $\frac{N}{2}\log(2\pi)$.

The derivatives of the Q with respect to σ_e^2 and σ_γ^2 are given by:

$$\begin{aligned}
\frac{\partial Q}{\partial \sigma_e^2} &= -\frac{N}{2\sigma_e^2} + \frac{1}{2\sigma_e^4} e^\top W^{-1} e \\
&\quad + \frac{1}{2\sigma_e^4}\mathbf{trace}\{(\sigma_e^{-2} Z^\top W^{-1} Z + \sigma_\gamma^{-2} M^{-1})^{-1} Z^\top W^{-1} Z\} \\
\frac{\partial Q}{\partial \sigma_\gamma^2} &= -\frac{q}{2\sigma_\gamma^2} + \frac{1}{2\sigma_\gamma^4} \gamma^\top M^{-1} \gamma \\
&\quad + \frac{1}{2\sigma_\gamma^4}\mathbf{trace}\{(\sigma_e^{-2} Z^\top W^{-1} Z + \sigma_\gamma^{-2} M^{-1})^{-1} M^{-1}\}
\end{aligned}$$

Pawitan (2001), page 448.

Setting both derivatives of the Q with respect to σ_e^2 and σ_γ^2 to 0, we obtain the following equations for the estimates of σ_e^2 and σ_γ^2:

$$\begin{aligned}
\hat{\sigma}_e^2 &= \frac{1}{N}[e^\top W^{-1} e + \mathbf{trace}\{(\sigma_e^{-2} Z^\top W^{-1} Z + \sigma_\gamma^{-2} M^{-1})^{-1} Z^\top W^{-1} Z\}] \\
\hat{\sigma}_\gamma^2 &= \frac{1}{q}[\gamma^\top M^{-1} \gamma + \mathbf{trace}\{(\sigma_e^{-2} Z^\top W^{-1} Z + \sigma_\gamma^{-2} M^{-1})^{-1} M^{-1}\}.
\end{aligned}$$

Note that, the derivatives of the determinant is based on the following result, (see Rigby and Stasinopoulos, (2013), Appendix A):

$$\frac{\partial}{\partial x} \log |x\mathbf{A} + \mathbf{B}| = \mathbf{trace}\left[(x\mathbf{A} + \mathbf{B})^{-1}\mathbf{A}\right]$$

where x is a scalar and $\mathbf{A}$ and $\mathbf{B}$ are r x r matrices (provided $|x\mathbf{A} + \mathbf{B}| \neq 0$).

Hence,

$$\begin{aligned}
\frac{\partial}{\partial \sigma_e^2} \log |\sigma_e^{-2} Z^\top W^{-1} Z + \sigma_\gamma^{-2} M^{-1}| &= -\frac{1}{\sigma_e^4}\mathbf{trace}\left[(\sigma_e^{-2} Z^\top W^{-1} Z + \sigma_\gamma^{-2} M^{-1})^{-1} Z^\top W^{-1} Z\right] \\
\frac{\partial}{\partial \sigma_\gamma^2} \log |\sigma_e^{-2} Z^\top W^{-1} Z + \sigma_\gamma^{-2} M^{-1}| &= -\frac{1}{\sigma_\gamma^4}\mathbf{trace}\left[(\sigma_e^{-2} Z^\top W^{-1} Z + \sigma_\gamma^{-2} M^{-1})^{-1} M^{-1}\right].
\end{aligned}$$

Alternatively, σ_e^2 and σ_γ^2 can be estimated using the following equations:

$$\begin{aligned}
&\mathbf{trace}\{(\hat{\sigma}_e^{-2} Z^\top W^{-1} Z + \hat{\sigma}_\gamma^{-2} M^{-1})^{-1} Z^\top W^{-1} Z\} \\
&= \mathbf{trace}\{\hat{\sigma}_e^2 (Z^\top W^{-1} Z + \hat{\lambda} M^{-1})^{-1} Z^\top W^{-1} Z\} \\
&= \hat{\sigma}_e^2 d
\end{aligned}$$

where $\hat{\lambda} = \hat{\sigma}_e^2 / \hat{\sigma}_\gamma^2$, and

$$d = \mathbf{trace}\{(Z^\top W^{-1} Z + \hat{\lambda} M^{-1})^{-1} Z^\top W^{-1} Z\}$$

.

Hence,

$$\hat{\sigma}_e^2 = \frac{1}{N}\left[e^\top W^{-1} e + \hat{\sigma}_e^2 d\right]$$

from page 264.

$$\begin{aligned}
N\hat{\sigma}_e^2 &= \left[e^\top W^{-1} e + \hat{\sigma}_e^2 d\right] \\
\hat{\sigma}_e^2 &= \frac{e^\top W^{-1} e}{N - d}.
\end{aligned}$$

Also,

$$\mathbf{trace}\{(\hat{\sigma}_e^{-2} Z^\top W^{-1} Z + \hat{\sigma}_\gamma^{-2} M^{-1})^{-1} (\hat{\sigma}_e^{-2} Z^\top W^{-1} Z + \hat{\sigma}_\gamma^{-2} M^{-1})\} = I_q = q,$$

$$\begin{aligned}&\mathbf{trace}\{\hat{\sigma}_e^{-2}(\hat{\sigma}_e^{-2}Z^\top W^{-1}Z+\hat{\sigma}_\gamma^{-2}M^{-1})^{-1}Z^\top W^{-1}Z\}\\ +&\mathbf{trace}\{\hat{\sigma}_\gamma^{-2}(\hat{\sigma}_e^{-2}Z^\top W^{-1}Z+\hat{\sigma}_\gamma^{-2}M^{-1})^{-1}M^{-1}\}=q,\end{aligned}$$

$$\mathbf{trace}\{\hat{\sigma}_\gamma^{-2}(\hat{\sigma}_e^{-2}Z^\top W^{-1}Z+\hat{\sigma}_\gamma^{-2}M^{-1})^{-1}M^{-1}\}=q-d.$$

Hence,

$$\begin{aligned}\hat{\sigma}_\gamma^2 &= \frac{1}{q}\left[\gamma^\top M^{-1}\gamma+(q-d)\hat{\sigma}_\gamma^2\right]\\ q\hat{\sigma}_\gamma^2 &= \left[\gamma^\top M^{-1}\gamma+(q-d)\hat{\sigma}_\gamma^2\right]\\ \hat{\sigma}_\gamma^2 &= \frac{\gamma^\top M^{-1}\gamma}{d}\end{aligned}$$

where d is the model degrees of freedom.

A.6 Derivations of Section 4.7

The marginal distribution of y is Gaussian with following mean and variance:

$$\begin{aligned}E(y) &= X\beta\\ V(y) &= V=\Sigma+Z_1D_1Z_1^\top+Z_2D_2Z_2^\top\end{aligned}$$

Given $\theta=(\sigma_e^2,\sigma_{\gamma_1}^2,\sigma_{\gamma_2}^2)$, the estimates of β,γ_1,γ_2 are the solution of

$$\begin{pmatrix} X^\top\Sigma^{-1}X & X^\top\Sigma^{-1}Z_1 & X^\top\Sigma^{-1}Z_2 \\ Z_1^\top\Sigma^{-1}X & Z_1^\top\Sigma^{-1}Z_1 + D_1^{-1} & Z_1^\top\Sigma^{-1}Z_2 \\ Z_2^\top\Sigma^{-1}X & Z_2^\top\Sigma^{-1}Z_1 & Z_2^\top\Sigma^{-1}Z_2 + D_2^{-1} \end{pmatrix} \begin{pmatrix} \beta \\ \gamma_1 \\ \gamma_2 \end{pmatrix} = \begin{pmatrix} X^\top\Sigma^{-1}y \\ Z_1^\top\Sigma^{-1}y \\ Z_2^\top\Sigma^{-1}y \end{pmatrix}.$$

The profile log-likelihood of θ is

$$\log L(\theta) = -\frac{1}{2}\log|V| - \frac{1}{2}(y - X\hat{\beta})^\top V^{-1}(y - X\hat{\beta}) \tag{A.13}$$

and the Q function is equal to:

$$\begin{aligned} Q = & -\frac{1}{2}\log|\Sigma| - \frac{1}{2}e^\top\Sigma^{-1}e \\ & -\frac{1}{2}\log|D_1| - \frac{1}{2}\gamma_1{}^\top D_1^{-1}\gamma_1 - \frac{1}{2}\log|Z_1^\top\Sigma^{-1}Z_1 + D_1^{-1}| \\ & -\frac{1}{2}\log|D_2| - \frac{1}{2}\gamma_2{}^\top D_2^{-1}\gamma_2 - \frac{1}{2}\log|Z_2^\top\Sigma^{-1}Z_2 + D_2^{-1}| \end{aligned} \tag{A.14}$$

where $e = y - X\beta - Z_1\gamma_1 - Z_2\gamma_2$, using the assumption that γ_1 and γ_2 are independent, and Z_1 and Z_2 are orthogonal in the sense that $Z_1^\top\Sigma^{-1}Z_2 = 0$.

Let $\Sigma = \sigma_e^2 W, D_1 = \sigma_{\gamma_1}^2 M_1, D_2 = \sigma_{\gamma_2}^2 M_2$, where W, M_1, M_2 are knows matrices of rank N, q_1, q_2 respectively.

By taking the derivatives of Q with respect to all the parameters and setting them to zero we obtain the equations for the variances as follows:

$$
\begin{aligned}
\sigma_e^2 &= \frac{1}{N}[e^\top W^{-1}e + \mathbf{trace}\{(\sigma_e^{-2}Z_1^\top W^{-1}Z_1 + \sigma_{\gamma_1}^{-2}M_1^{-1})^{-1}Z_1^\top W^{-1}Z_1\} \\
&\quad +\mathbf{trace}\{(\sigma_e^{-2}Z_2^\top W^{-1}Z_2 + \sigma_{\gamma_2}^{-2}M_2^{-1})^{-1}Z_2^\top W^{-1}Z_2\}] \\
\sigma_{\gamma_1}^2 &= \frac{1}{q_1}[\gamma_1^\top M_1^{-1}\gamma_1 + \mathbf{trace}\{(\sigma_e^{-2}Z_1^\top W^{-1}Z_1 + \sigma_{\gamma_1}^{-2}M_1^{-1})^{-1}M_1^{-1}\}] \\
\sigma_{\gamma_2}^2 &= \frac{1}{q_2}[\gamma_2^\top M_2^{-1}\gamma_2 + \mathbf{trace}\{(\sigma_e^{-2}Z_2^\top W^{-1}Z_2 + \sigma_{\gamma_2}^{-2}M_2^{-1})^{-1}M_2^{-1}\}]
\end{aligned}
$$

Alternatively $\sigma_e^2, \sigma_{\gamma_1}^2, \sigma_{\gamma_2}^2$ can be evaluated as

$$
\begin{aligned}
\sigma_e^2 &= \frac{e^\top W^{-1}e}{N - d_1 - d_2} \\
\sigma_{\gamma_1}^2 &= \frac{\gamma_1^\top M_1^{-1}\gamma_1}{d_1} \\
\sigma_{\gamma_2}^2 &= \frac{\gamma_2^\top M_2^{-1}\gamma_2}{d_2}
\end{aligned}
$$

where d_1 and d_2 are the degrees of freedom such that:

$$d_i = \mathbf{trace}\{(Z_i^\top W^{-1}Z_i + \lambda_i M_i^{-1})^{-1}Z_i^\top W^{-1}Z_i\}$$

for $i = 1, 2$ and $\lambda_i = \sigma_e^2/\sigma_{\gamma_i}^2$, Pawitan (2001), page 454.

Appendix B

Skew Student t distribution

A skewed Student t distribution was used to allow for skewness and kurtosis in the conditional distribution of financial returns initially by Hansen (1994) and subsequently by Fernandez and Steel (1998) using an alternative parametrization. Fernandez and Steel (1998) consider a shifted and scaled t distribution with τ degrees of freedom, i.e. $\mu_0 + \sigma_0 T$ where $T \sim t_\tau$, denoted here by $TF(\mu_0, \sigma_0, \tau)$, and splice together at μ_0 two differently scaled distributions, $Y_1 \sim TF(\mu_0, \sigma_0/\nu, \tau)$ below μ_0 and $Y_2 \sim TF(\mu_0, \sigma_0\nu, \tau)$ above μ_0, The resulting distribution is denoted here by $Y \sim ST3(\mu_0, \sigma_0, \nu, \tau)$. Wurtz *et al.* (2006) reparameterized the skew t distribution of Fernandez and Steel (1998) so that in the new parametrization μ is the mean and σ is the standard deviation, denoted here by $Y \sim SST(\mu, \sigma, \nu, \tau)$, where

$$\begin{aligned}
f_Y(y|\mu,\sigma,\nu,\tau) &= \frac{2}{(1+\nu^2)}\left\{f_{Y_1}(y)I(y<\mu_0)+\nu^2 f_{Y_2}(y)I(y\geq\mu_0)\right\} \\
&= \frac{c}{\sigma_0}\left\{1+\frac{(y-\mu_0)^2}{\sigma_0^2\tau}\left[\nu^2 I(y<\mu_0)+\frac{1}{\nu^2}I(y\geq\mu_0)\right]\right\}^{-(\tau+1)/2}
\end{aligned}$$

for $-\infty < y < \infty$, $-\infty < \mu < \infty, \sigma > 0, \nu > 0$, and $\tau > 2$, where $c = 2\nu / \left[(1+\nu^2)B(\frac{1}{2},\frac{\tau}{2})\tau^{1/2}\right]$,

$$\mu_0 = \mu - \sigma m/s,$$

and

$$\sigma_0 = \sigma/s$$

and

$$m = 2\tau^{1/2}(\nu^2 - 1) / \left[(\tau - 1)\nu B(\frac{1}{2},\frac{\tau}{2})\right]$$

and

$$s^2 = \left\{\tau\left(\nu^3 + \nu^{-3}\right) / \left[(\tau-2)(\nu+\nu^{-1})\right]\right\} - m^2.$$

Hence $Y \sim SST(\mu,\sigma,\nu,\tau) = ST3(\mu_0,\sigma_0,\nu,\tau)$ has mean μ and variance σ^2 since $E(Y) = \mu_0 + \sigma_0 E(Z_0) = (\mu - \sigma m/s) + (\sigma/s)m = \mu$ and $V(Y) = \sigma_0^2 V(Z_0) = (\sigma^2/s^2)s^2 = \sigma^2$, where $Z_0 = (Y - \mu_0)/\sigma_0 \sim ST3(0,1,\nu,\tau)$ and where $E(Z_0) = m$ and $V(Z_0) = s^2$ provided $\tau > 2$ from Fernandez and Steel (1998) p360. Note that $Z = (Y-\mu)/\sigma \sim SST(0,1,\nu,\tau)$ has mean 0 and variance 1.

Appendix C

Proof of the Theorems

C.1 Theorem 1 Proof

$Y_t|\mu_t, \sigma_t, \nu_t, \tau_t \sim \mathcal{D}(\mu_t, \sigma_t, \nu_t, \tau_t)$ where $\mu_t = \beta_{1,0} + \gamma_{1,t}$, and $\log \sigma_t = \beta_{2,0} + \gamma_{2,t}$. Applying the law of iterated expectations,

a)

$$\begin{aligned} E[Y_t] &= E\left[E\left(Y_t|\mu_t, \sigma_t, \nu_t, \tau_t\right)\right] = E[\mu_t], \\ \text{so } E[\mu_t] &= \beta_{1,0} + E\left(\gamma_{1,t}\right). \\ \text{If } \Phi_1(B)\gamma_{1,t} &= b_{1,t}, \\ \text{then } \gamma_{1,t} &= \Phi_1(B)^{-1} b_{1,t} = \psi_1(B) b_{1,t}, \end{aligned}$$

where $\Phi_1(B)$ is assumed to be invertible. Then

$$\begin{aligned} E[\gamma_{1,t}] &= \psi_1(B)E(b_{1,t}) = 0, \text{ as } b_{1,t} \sim N(0, \sigma_b^2), \\ \text{so } E[\mu_t] &= \beta_{1,0}. \\ \text{Hence } E[Y_t] &= \beta_{1,0} \end{aligned}$$

b)

$$
\begin{aligned}
V[Y_t] &= V\left[E\left(Y_t|\mu_t, \sigma_t, \nu_t, \tau_t\right)\right] + E\left[V\left(Y_t|\mu_t, \sigma_t, \nu_t, \tau_t\right)\right], \\
\text{so } V[Y_t] &= V[\mu_t] + c^2 E[\sigma_t^2], \\
\text{but V}[\mu_\text{t}] &= V[\gamma_{1,t}] = V[\psi_1(B)b_{1,t}] = S_1\sigma_{b_1}^2, \\
\text{and E}[\sigma_\text{t}^2] &= E\left[\exp\left(2\beta_{2,0} + 2\gamma_{2,t}\right)\right] = \exp\left(2\beta_{2,0}\right) E\left[\exp\left(2\gamma_{2,t}\right)\right], \\
\text{where E}\left[\exp\left(2\gamma_{2,\text{t}}\right)\right] &= E\left[\exp\left(2\psi_2(B)b_{2,t}\right)\right] = \prod_{j=0}^{\infty} E\left[\exp\left(2\psi_{2,j}b_{2,t-j}\right)\right], \\
\text{as } \gamma_{2,t} &= \psi_2(B)b_{2,t}, \text{and assuming independence of } b_{k,t}.
\end{aligned}
$$

Applying the following lemma:

if $\boldsymbol{b} \sim N(0, \sigma_b^2)$ then $E\left[\exp(r\psi b)\right] = \exp\left(\frac{1}{2}r^2\psi^2\sigma_b^2\right)$.

$$
\begin{aligned}
\text{so } E\left[\exp\left(2\gamma_{2,t}\right)\right] &= \prod_{j=0}^{\infty} \exp\left(2\psi_{2,j}^2\sigma_{b_2}^2\right) = \exp\left(2S_2\sigma_{b_2}^2\right), \\
\text{where S}_\text{k} &= 1 + \sum_{j=1}^{\infty} \psi_{k,j}^2, \\
\text{and } \psi_{2,0} &= 1. \\
\text{Hence V}[\text{Y}_\text{t}] &= S_1\sigma_{b_1}^2 + c^2 \exp\left(2\beta_{2,0} + 2S_2\sigma_{b_2}^2\right).
\end{aligned}
$$

C.2 Theorem 2 Proof

a) $E[Y_t] = E\left[E\left(Y_t|\mu_t, \sigma_t, \nu_t, \tau_t\right)\right] = E[\mu_t] = \exp\left(\beta_{1,0} + \frac{1}{2}S_1\sigma_{b_1}^2\right)$, from d) below.

b) $V[Y_t] = V\left[E\left(Y_t|\mu_t, \sigma_t, \nu_t, \tau_t\right)\right] + E\left[V\left(Y_t|\mu_t, \sigma_t, \nu_t, \tau_t\right)\right] = V[\mu_t] + E\left[v(\mu_t, \sigma_t)\right]$.

c) $V[\mu_t] = E[\mu_t^2] - \{E[\mu_t]\}^2 = \exp(2\beta_{1,0})\left[\exp(2S_1\sigma_{b_1}^2) - \exp(S_1\sigma_{b_1}^2)\right]$, from d) below.

d) $E[\mu_t^r] = E\left[\exp(r\beta_{1,0} + r\gamma_{1,t})\right] = \exp(r\beta_{1,0})E\left[\exp(r\psi_1(B)b_{1,t})\right]$.

$$E\left[\exp(r\psi_1(B)b_{1,t})\right] = \prod_{j=1}^{\infty} E\left[\exp(r\psi_{1,j}b_{1,t-j}\right] = \prod_{j=1}^{\infty} \exp\left(\frac{1}{2}r^2\psi_{1,j}^2\sigma_{b_1}^2\right)$$

$$= \exp\left(\frac{1}{2}r^2 S_1\sigma_{b_1}^2\right),$$

since if $\boldsymbol{b} \sim N(0, \sigma_b^2)$, $E\left[\exp(r\psi b)\right] = \int_{-\infty}^{\infty} \exp(r\psi b)\frac{1}{\sqrt{2\pi}\sigma_b}\exp\left[-\frac{b^2}{2\sigma_b^2}\right]db$
$= \exp\left(\frac{1}{2}r^2\psi^2\sigma_b^2\right)\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma_b}\exp\left[-\frac{1}{2\sigma_b^2}(b - r\psi\sigma_b^2)^2\right]db = \exp\left(\frac{1}{2}r^2\psi^2\sigma_b^2\right)$.

e) As for d).

Corollary 1 Theorem 2 applies to the following distributions, the negative binomial type 1 and 2, NBI and $NBII$, respectively, the gamma, GA, and inverse Gaussian, IG, distributions. The results below for the marginal mean $E[Y_t]$ and variance $V[Y_t]$ of Y_t use the results of Theorem 2.

1. $Y_t|\mu_t, \sigma_t \sim NBI(\mu_t, \sigma_t)$ with mean μ_t and variance $\mu_t + \sigma_t\mu_t^2$, then $E[Y_t] = E[\mu_t]$ and $V[Y_t] = V[\mu_t] + E[\mu_t] + E[\sigma_t]E[\mu_t^2]$. Note $\sigma_t = 0$ gives $Y_t|\mu_t \sim PO(\mu_t)$.

2. $Y_t|\mu_t, \sigma_t \sim NBII(\mu_t, \sigma_t)$ with mean μ_t and variance $\mu_t + \sigma_t\mu_t$, then $E[Y_t] = E[\mu_t]$ and $V[Y_t] = V[\mu_t] + E[\mu_t] + E[\sigma_t]E[\mu_t]$

3. $Y_t|\mu_t, \sigma_t \sim GA(\mu_t, \sigma_t)$ with mean μ_t and variance $\sigma_t^2\mu_t^2$, then $E[Y_t] = E[\mu_t]$ and $V[Y_t] = V[\mu_t] + E[\sigma_t^2]E[\mu_t^2]$. Note $\sigma_t = 1$ gives $Y_t|\mu_t \sim EXP(\mu_t)$.

4. $Y_t|\mu_t, \sigma_t \sim IG(\mu_t, \sigma_t)$ with mean μ_t and variance $\sigma_t^2\mu_t^3$, then $E[Y_t] = E[\mu_t]$ and $V[Y_t] = V[\mu_t] + E[\sigma_t^2]E[\mu_t^3]$.

Appendix D

R commands

The simulation and fitting functions are programmed in R. They are available by the author and will be in a public domain in a package called `gest` in R statistical software.

D.1 R commands for chapter 5

The data used here can be downloaded in a zip file from this web site:

http://www.ssfpack.com/CKbook.html.

The R fitting commands and output in Table 5.1:

```
NorwayFinland <- read.delim("<PATH>/NorwayFinland.txt", header=F)
Nor  <- (NorwayFinland$V2)
lNor <- log(Nor)
tNor <- (NorwayFinland$V1)
lnorw <- zoo(lNor,order.by=tNor)
m1 <- RW(lnorw,plot=T)
c(m1$sig2e, m1$sig2b, m1$value.of.Q)
```

The R fitting commands and output in Table 5.2:

```
UK.KSI <- read.delim("<PATH>/UKdriversKSI.txt", header=F)
KSI <- ts(UK.KSI, start = c(1969, 1), freq = 12)
lksi <- log(KSI)
m2 <- RW(lksi,plot=T)
c(m2$sig2e, m2$sig2b, m2$value.of.Q)
```

The R fitting commands and output in Table 5.3:

```
NorwayFinland <- read.delim("<PATH>/NorwayFinland.txt", header=F)
Fin<- (NorwayFinland$V3)
lfin<- log(Fin)
m3 <- rw.tr(lfin,plot=T)
c(m3$sig2e, m3$sig2b, m3$sig2d, m3$value.of.Q)
```

The R fitting commands and output in Table 5.4:

```
UK.KSI <- read.delim("<PATH>/UKdriversKSI.txt", header=F)
KSI <- ts(UK.KSI, start = c(1969, 1), freq = 12)
lksi <- log(KSI)
m4 <- rw.tr(lksi,plot=T)
c(m4$sig2e, m4$sig2b, m4$sig2d, m4$value.of.Q)
```

The R fitting commands and output in Table 5.5:

```
UK.inflation<- read.delim("<PATH>/UKinflation.txt", header=F)
UKinf <- ts(UK.inflation, start = c(1950, 1), freq = 4)
m5 <- rw.seas(UKinf,plot=T)
c(m5$sig2e, m5$sig2b, m5$sig2w, m5$value.of.Q)
```

The R fitting commands and output in Table 5.6:

```
UK.KSI <- read.delim("<PATH>/UKdriversKSI.txt", header=F)
KSI <- ts(UK.KSI, start = c(1969, 1), freq = 12)
lksi <- log(KSI)
m6 <- rw.seas(lksi,plot=T, frequency=12)
c(m6$sig2e, m6$sig2b, m6$sig2w, m6$value.of.Q)
```

The R fitting commands and output in Table 5.7:

```
library(FinTS)
data(q.jnj)
m7 <- ar.seas(as.ts(q.jnj),plot=T)
c(m7$sig2e, m7$sig2b, m7$sig2w, m7$phi1)
m8 <- ar.seas(as.ts(log(q.jnj)),plot=T)
c(m8$sig2e, m8$sig2b, m8$sig2w, m8$phi1)
```

The R fitting commands and output in Table 5.8:

```
lpp <- read.delim("<PATH>/logUKpetrolprice.txt", header=F)
UK.KSI <- read.delim("<PATH>/UKdriversKSI.txt", header=F)
KSI <- ts(UK.KSI, start = c(1969, 1), freq = 12)
lksi <- log(KSI)
m9 <- rw.exp(lksi, lpp, plot=T)
c(m9$sig2e, m9$sig2b, m9$sig2v)
c(m9$beta[1], m9$value.of.Q)
```

D.2 R commands for chapter 7

Simulation of Figure 7.1

```
set.seed(1002)

y<-gest.sim(N=1000, mu.init=1, sigma.init=1, mu.sigb=.1, sigma.sigb=0,
plot=T, family=NO)
```

Simulation of Figure 7.2

```
set.seed(1223)

y <-gest.sim(N=1000, mu.init=1, sigma.init=1, mu.sigb=.1, sigma.sigb=.05,
plot=T, family=NO)
```

Simulation of Figure 7.3

```
set.seed(1244)

y<-gest.sim(N=1000,  mu.init=1, sigma.init=1, mu.sigb=.1, sigma.sigb=.05,
mu.type="AR", mu.phi=.5, plot=T, family=NO)
```

Simulation of Figure 7.4

```
set.seed(1240)

y<-gest.sim(N=240, mu.sigb=.1, mu.sigS=.01, sigma.sigb=0,
mu.type="levelSeasonal", frequency=12, plot=T, family=NO)
```

Simulation of Figure 7.5

```
set.seed(1246)

y<-gest.sim(240, mu.sigb =.1, mu.sigS=.01, sigma.sigb=.06,
 mu.type="levelSeasonal", frequency=12, plot=T, family=NO)
```

Simulation of Figure 7.6 and fitting of Figure 7.8

```
set.seed(1224)

y<-gest.sim(N=1000, mu.sigb =.05, plot=T, family=PO)

Y <- y[,1]

m1 <- gamlss(Y~rw(Y), family=PO)
```

Simulation of Figure 7.9 and fitting of Figure 7.10

```
set.seed(12029)
y <-gest.sim(1000, mu.sigb =.03, sigma.sigb=.03, plot=T, family=NBI)
Y <- y[,1]
m1 <- gamlss(Y~rw(Y),sigma.fo=~rw(Y), family=NBI)
op <- par(mfrow=c(3,1))
plot(Y, col="gray", ylab="y")
plot(y[,2], col="gray", ylab="mu")
lines(fitted(m1, "mu"), col="red")
plot(y[,4], col="gray", ylab="sigma")
lines(fitted(m1, "sigma"), col="red")
```

Simulation of Figure 7.11 and fitting of Figure 7.12

```
set.seed(19984)
y<-gest.sim(5000, mu.sigb=.03, sigma.sigb=.03, nu.sigb=.04,
            nu.init=5, plot=T, family=TF2)
Y <- y[,1]
m1 <- gamlss(Y~rw(Y),sigma.fo=~rw(Y), nu.fo=~rw(Y),family=TF2)
op <- par(mfrow=c(4,1))
plot(Y, col="gray", ylab="y")
op <- par(mfrow=c(3,1))
plot(y[,2], col="gray", ylab="mu")
lines(fitted(m1, "mu"), col="red")
plot(y[,4], col="gray", ylab="sigma")
lines(fitted(m1, "sigma"), col="red")
plot(1/y[,6], col="gray", ylab="1/nu")
lines(1/(fitted(m1, "nu")), col="red")
```

D.3 R commands for chapter 9

D.3.1 Pound sterling and US dollar exchange rate

```
sv<-gamlss(Returns~1, sigma.fo=~ar(Returns), family=NO, data=da)
```

The data used in this example is ”sv.dat”, and can be downloaded in a zip file from:

http://www.ssfpack.com/DKbook.html.

D.3.2 Standard and Poor 500 stock index

```
da <- read.csv("<PATH>/SP-80-dec2012.csv")
y <- as.ts(da$Returns)
D <- as.Date(da[,1], "%d/%m/%Y")
Da <- blag(y,lags=1,from.lag=0, omit.na = F, value=0)
# using the asinh transformation
.y._Neg <- asinh(ifelse(Da[,2]<0,Da[,2],0))
.y._Pos <- asinh(ifelse(Da[,2]>=0,Da[,2],0))
#-------------------------------
m1 <- gamlss(y~1, sigma.fo=~ar(y), nu.fo=~rw(y), tau.fo=~rw(y),
family=SST, n.cyc=40)
#-------------------------------
m2 <- gamlss(y~1, sigma.fo=~ar(y, .y._Neg, .y._Pos, include.leverage=T),
nu.fo=~rw(y), tau.fo=~rw(y), family=SST, n.cyc=40)
#-------------------------------
m3 <- gamlss(y~1, sigma.fo=~ar(y, .y._Neg, .y._Pos, include.leverage=T),
nu.fo=~rw(y), tau.fo=~1, family=SST, n.cyc = 40)
#------------------------------
```

```
m4 <- gamlss(y~1, sigma.fo=~ar(y, .y._Neg, .y._Pos, include.leverage=T),
nu.fo=~1, tau.fo=~rw(y), family=SST, n.cyc=40)
#-------------------------------
m5 <- gamlss(y~1, sigma.fo=~ar(y, .y._Neg, .y._Pos, include.leverage=T),
nu.fo= ~1, tau.fo=~1, family = SST, n.cyc = 40)
#-------------------------------
m6 <- gamlss(y~rw(y,order=2), sigma.fo=~ar(y, .y._Neg, .y._Pos,
include.leverage=T), nu.fo=~rw(y,order=2), tau.fo=~rw(y,order=2),
family=SST, n.cyc = 40)
```

D.3.3 Van drivers killed in UK

```
library(gamlss)
library(sspir)
data(vandrivers)
tvd<-1:192
x1 <-cos(2*pi*tvd/12)
x2 <-sin(2*pi*tvd/12)
x3 <-cos(2*pi*tvd/6)
x4 <-sin(2*pi*tvd/6)
mvd<-as.factor(cycle(vandrivers$y))
```

Using the conditional Poisson distribution (PO)

```
m1 <-gamlss(y~rw(y),family=PO,data=vandrivers)
m2 <-gamlss(y~rw(y)+mvd,family=PO,data=vandrivers)
m15<-gamlss(y~rw(y)+x1+x2+x3+x4,family=PO,data=vandrivers)
m3 <-gamlss(y~srw(y,frequency=12),family=PO,data=vandrivers)
```

```
m4 <-gamlss(y~rw(y)+seatbelt,family=PO,data=vandrivers)
m5 <-gamlss(y~rw(y)+seatbelt+mvd,family=PO,data=vandrivers)
m16<-gamlss(y~rw(y)+seatbelt+x1+x2+x3+x4,family=PO,data=vandrivers)
m6 <-gamlss(y~srw(y,frequency=12)+seatbelt,family=PO,data=vandrivers)
m20<-gamlss(y~rw(y,order=2),family=PO, data=vandrivers)
m21<-gamlss(y~rw(y,order=2)+mvd,family=PO,data=vandrivers)
m22<-gamlss(y~rw(y,order=2)+x1+x2+x3+x4,family=PO, data=vandrivers)
m23<-gamlss(y~srw(y,order=2,frequency=12),family=PO,data=vandrivers)
m24<-gamlss(y~rw(y,order=2)+seatbelt,family=PO,data=vandrivers)
m25<-gamlss(y~rw(y,order=2)+seatbelt+mvd,family=PO,data=vandrivers)
m26<-gamlss(y~rw(y,order=2)+seatbelt+x1+x2+x3+x4,family=PO,data=vandrivers)
m27<-gamlss(y~srw(y,order=2,frequency=12)+seatbelt,family=PO,data=vandrivers
```

Using the conditional negative binomial distribution (NBI)

```
b1 <-gamlss(y~rw(y),family=NBI,data=vandrivers)
b2 <-gamlss(y~rw(y)+mvd,family=NBI,data=vandrivers)
b15<-gamlss(y~rw(y)+ x1+x2+x3+x4,family=NBI,data=vandrivers)
b3 <-gamlss(y~srw(y,frequency=12),family=NBI,data=vandrivers)
b4 <-gamlss(y~rw(y)+seatbelt,family=NBI,data=vandrivers)
b5 <-gamlss(y~rw(y)+seatbelt+mvd,family=NBI,data=vandrivers)
b16<-gamlss(y~rw(y)+seatbelt+x1+x2+x3+x4,family=NBI,data=vandrivers)
b6 <-gamlss(y~srw(y,frequency=12)+seatbelt,family=NBI,data=vandrivers)
b17<-gamlss(y~rw(y,order=2),family=NBI,data=vandrivers)
b18<-gamlss(y~rw(y,order=2)+mvd,family=NBI,data=vandrivers)
b19<-gamlss(y~rw(y,order=2)+x1+x2+x3+x4,family=NBI,data=vandrivers)
b20<-gamlss(y~srw(y,order=2,frequency=12),family=NBI,data=vandrivers)
```

```
b21<-gamlss(y~rw(y,order=2)+seatbelt,family=NBI,data=vandrivers)
b22<-gamlss(y~rw(y,order=2)+seatbelt+mvd,family=NBI,data=vandrivers)
b23<-gamlss(y~rw(y,order=2)+seatbelt+x1+x2+x3+x4,family=NBI,data=vandrivers)
b24<-gamlss(y~srw(y,order=2,frequency=12)+seatbelt,family=NBI,data=vandriver
```

Conditional test for $\beta_{1,1}$ in model m26

```
m26<-gamlss(y~rw(y,order=2)+seatbelt+x1+x2+x3+x4,family=PO,data=vandrivers)
m261<-gamlss(y~rw(y,order=2,sig2e.fix=TRUE,sig2b.fix=TRUE,
sig2e=0.8399006,sig2b=6.99568e-08)+x1+x2+x3+x4,family=PO,data=vandrivers)
```

95% profile confidence interval for $\beta_{1,1}$ in model m26

```
mod<-quote(gamlss(y~offset(this*seatbelt)+ rw(y,order=2,sig2e.fix=TRUE,
sig2b.fix=TRUE,sig2e=0.8399006,sig2b=6.99568e-08)+x1+x2+x3+x4,
family=PO, data=vandrivers))
prof.term(mod, min=-0.6, max=0.5, xlab="Beta_1,1")
```

Plotting Figure 9.12

```
library(zoo)
ti <- time(vandrivers$y)
z1 <- -0.18796*(vandrivers$seatbelt)
z2 <- exp(2.21495+(fitted(m26$mu.coefSmo[[1]]))+z1)
z3 <- zoo(z2,order.by=ti)
plot(vandrivers$y, col="gray", ylab="Van drivers killed")
lines(z3, col="darkred")
```

Plotting Figure 9.14

```
library(zoo)
ti <- time(vandrivers$y)
z2 <- exp(2.19736+(fitted(m22$mu.coefSmo[[1]])))
z3 <- zoo(z2,order.by=ti)
plot(vandrivers$y, col="gray", ylab="Van drivers killed")
lines(z3, col="darkred")
```

Plotting Figure 9.15

```
svd1 <- (0.09786 *x1)+ (-0.06923 *x2)+ (0.10381 *x3) +(-0.04517*x4)
S1 <- exp(zoo(svd1,order.by=ti))
op <- par(mfrow=c(3,1))
plot(vandrivers$y, col="gray", ylab="Van drivers killed")
plot(z3, col="darkred", ylab="Level", xlab="Time")
plot(S1, col="darkblue", ylab="Seasonal", xlab="Time")
```

D.3.4 Polio incidence in the United States

```
library(gamlss)
tp <-1:168
x1 <-cos(2*pi*tp/12)
x2 <-sin(2*pi*tp/12)
x3 <-cos(2*pi*tp/6)
x4 <-sin(2*pi*tp/6)
mp <-as.factor(cycle(polio))
```

Using the conditional Poisson distribution (PO)

```
p1 <-gamlss(polio~rw(polio), family=PO, data=polio)

p2 <-gamlss(polio~rw(polio) + x1+x2+x3+x4, family=PO, data=polio)

p3 <-gamlss(polio~rw(polio)+ mp, family=PO, data=polio)

p4 <-gamlss(polio~srw(polio,frequency=12), family=PO, data=polio)

p5 <-gamlss(polio~rw(polio) + tp, family=PO, data=polio)

p6 <-gamlss(polio~rw(polio) + tp + x1+x2+x3+x4, family=PO, data=polio)

p7 <-gamlss(polio~rw(polio) + tp + mp, family=PO, data=polio)

p8 <-gamlss(polio~srw(polio,frequency=12) + tp, family=PO, data=polio)

g1 <-gamlss(polio~ar(polio), family=PO, data=polio)

g2 <-gamlss(polio~ar(polio) + x1+x2+x3+x4, family=PO, data=polio)

g3 <-gamlss(polio~ar(polio) + mp, family=PO, data=polio)

g4 <-gamlss(polio~sar(polio,frequency=12), family=PO, data=polio)

g5 <-gamlss(polio~ar(polio) + tp, family=PO, data=polio)

g6 <-gamlss(polio~ar(polio) + tp + x1+x2+x3+x4, family=PO, data=polio)

g7 <-gamlss(polio~ar(polio) + tp + mp, family=PO, data=polio)

g8 <-gamlss(polio~sar(polio,frequency=12) + tp, family=PO, data=polio)
```

Using the conditional negative binomial distribution (NBI)

```
n1 <-gamlss(polio~rw(polio), family=NBI, data=polio)

n2 <-gamlss(polio~rw(polio)+ x1+x2+x3+x4, family=NBI, data=polio)

n3 <-gamlss(polio~rw(polio)+ mp, family=NBI, data=polio)

n4 <-gamlss(polio~srw(polio,frequency=12), family=NBI, data=polio)

n5 <-gamlss(polio~rw(polio)+ tp, family=NBI, data=polio)

n6 <-gamlss(polio~rw(polio)+ tp + x1+x2+x3+x4, family=NBI, data=polio)

n7 <-gamlss(polio~rw(polio)+ tp + mp, family=NBI, data=polio)
```

```
n8 <-gamlss(polio~srw(polio,frequency=12) + tp, family=NBI, data=polio)

e1 <-gamlss(polio~ar(polio), family=NBI, data=polio)

e2 <-gamlss(polio~ar(polio)+ x1+x2+x3+x4, family=NBI, data=polio)

e3 <-gamlss(polio~ar(polio)+ mp, family=NBI, data=polio)

e4 <-gamlss(polio~sar(polio,frequency=12), family=NBI, data=polio)

e5 <-gamlss(polio~ar(polio)+ tp, family=NBI, data=polio)

e6 <-gamlss(polio~ar(polio)+ tp + x1+x2+x3+x4, family=NBI, data=polio)

e7 <-gamlss(polio~ar(polio)+ tp + mp, family=NBI, data=polio)

e8 <-gamlss(polio~sar(polio,frequency=12) + tp, family=NBI, data=polio)
```

Plotting Figure 9.16

```
library(zoo)

ti <- time(polio)

Z2 <- exp(0.3535 +(fitted(p4$mu.coefSmo[[1]])[,2]))

Z3 <- zoo(Z2,order.by=ti)

plot(polio, col="gray", ylab="Polio")

lines(Z3, col="darkred")
```

Plotting Figure 9.17

```
library(zoo)

sp <- (fitted(p4$mu.coefSmo[[1]])[,3])

S1 <- exp(zoo(sp,order.by=ti))

op <- par(mfrow=c(3,1))

plot(polio, col="gray", ylab="Polio")

plot(Z3, col="darkred", ylab="Level", xlab="Time")

plot(S1, col="darkblue", ylab="Seasonal", xlab="Time")
```

Bibliography

Akaike, H. (1983). Information measures and model selection. *Bulletin of the International Statistical Institute*, **50**, 277-290.

Anderson, B.D.O. and Moore, J.B. (1979). *Optimal Filtering*. Englewood Cliffs: Prentice Hall.

Asai, M. and McAleer, M. (2005). Dynamic asymmetric leverage in stochastic volatility models. *Econometric Reviews*, **24**, 317-332.

Baillie, R.T., Bollerslev, T., and Mikkelsen, H.O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, **74**, 3-30.

Benjamin, M., Rigby, R.A. and Stasinopoulos, D.M. (2003). Generalized autoregressive moving average models. *J. Am. Statist. Ass.*, **98**, 214-223.

Bollerslev, T. (1986). Generalised autoregressive conditional heteroskedasticity. *Journal of Econometrics*, **31**, 307-327.

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics*, **69**, 542-546

Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (1994). *Time Series Analysis: Forecasting and Control.* Prentice Hall.

Breslow, N.E. and Clayton, D.G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, **88**, 9-25.

Briet, O.J.T., Amerasinghe, P.H. and Vounatsou, P. (2013). Generalized seasonal autoregressive integrated moving average models for count data with application to malaria time series with low case numbers. PLoS ONE 8(6): e65761. doi:10.1371/journal.pone.0065761.

Brockwell, P.J. and Davis, R.A. (1996). *Time Series: Theory and Methods.* Springer-Verlag.

Broda, S.A, Haas, M., Krause, J., Paolella, M.S and Steude, S.C. (2013). Stable mixture GARCH models. *Journal of Econometrics*, **172**, 292-306.

Brooks, C., Burke, S.P., Heravi, S. and Persand, G. (2005). Autoregressive conditional kurtosis. *J. Fin. Econometrics*, **3**, 399-421.

Chib, S., Nardari, F. and Shephard, N. (2002). Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics*, *108*, 281-316.

Choy, S.T.B., Wan, W.Y. and Chan, C.M. (2008). Bayesian student-t stochastic volatility models via scale mixtures. *Advances in Econometrics*, **23**, 595-618. Special issue on Bayesian Econometrics Methods.

Cole, T.J. and Green, P.J. (1992). Smoothing reference centile curves: the lms method and penalized likelihood. *Statistics in Medicine*, **11**, 1305-1319.

Commandeur, J.J.F. and Koopman, S.J. (2007). *An Introduction to State Space Time Series Analysis.* Oxford University Press.

Cox, D.R. (1981). Statistical analysis of time series: some recent developments. *Scandinavian Journal of Statistics*, **8**, 93-115.

Crowder, M.J., Kimber, A.C., Smith R.L. and Sweeting, T.J. (1991). *Statistical Analysis of Reliability Data.* Chapman and Hall, London.

D'Agostino, R.B., Balanger, A. and D'Agostino, R.B. (1990). A suggestion for using powerful and informative tests of normality. *American Statistician*, **44**, 316-321.

Davidson, R. (2012). Statistical inference in the presence of heavy tails. *The Econometrics Journal*, **15**, C31-C53.

De Rossi, G. and A. Harvey (2009). Quantiles, expectiles and splines. *Journal of econometrics*, **152**, 179-185.

Dethlefsen, C., Lundbye-Christensen, S. and Luther Christensen, A. (2009). sspir: State Space Models in R. R package version 0.2.8, URL http://CRAN.R-project.org/package=sspir.

Diebold, F.X., Gunther, T.A. and Tay, T.S. (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review*, **39**, 863-883.

Ding, Z., Granger, C.W.J. and Engle, R.F. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, **1**, 83-106.

Dunn, P.K. and Smyth, G.K. (1996). Randomised quantile residuals. *Journal of Computational and Graphical Statistics*, **5**, 236-244.

Durbin, J. and Koopman, S.J. (2000). Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian Perspectives. *Journal of the Royal Statistical Society: Series B*, **62**, 3-56.

Durbin, J. and Koopman, S.J. (2012). *Time Series Analysis by State Space Methods.* Oxford University Press.

Eilers, P. H. C. and Marx, B.D. (1996). Flexible smoothing with B-splines and penalties (with comments and rejoinder). *Statistical Science*, **11**, 89-121.

Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, **50**, 987-1008.

Evans, M. and Swartz, T. (2000). *Approximating Integrals via Monte Carlo and Deterministic Methods.* Oxford: Oxford University Press.

Fahrmeir, L. (1992). Conditional mode estimation by extended Kalman filtering for multivariate dynamic generalised linear models. *J. Am. Statist. Ass.*, **87**, 501-509.

Fahrmeir, L. and Tutz, G. (2001). *Multivariate Statistical Modelling based on Generalized Linear Models.* 2nd edn. New York: Springer.

Fama, E.F. (1965). The behavior of stock market prices. *Journal of Business*, **38**, 34-105.

Fernandez, C. and Steel, M.F.J., (1998). On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, **93**, 359-371.

Fokianos, K. (2001). Truncated poisson regression for time series of counts. *Scandinavian Journal of Statistics*, **28**, 645-659.

Forsberg, L. and Bollerslev, T., (2002). Bridging the gap between the distribution of realized (ECU) volatility and ARCH modelling (of the EURO): the GARCH-NIG model. *J. Appl. Econ.*, **17**, 535-548.

Ghalanos, A. (2013). rugarch: Univariate GARCH models. R package version 1.2-7.

Giraitis, L., Leipus, R., Robinson, P. and Surgailis, D. (2004). LARCH, leverage and long memory. *Journal of Financial Econometrics*, **2**, 177-210.

Glosten, L., Jagannathan, R. and Runkle, D. (1993). On the relation between expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, **48**, 1779-1801.

Granger, C.W.J. (2005). The past and future of empirical finance: some personal comments. *Journal of Econometrics*, **129**, 35-40.

Hansen, B.E. (1994). Autoregressive conditional density estimation. *International Economic Review*, **35**, 705-730.

Harvey, A.C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Cambridge.

Harvey, A.C. and Durbin, J. (1986). The effects of seat belt legislation on British road casualties: a case study in structural time series modelling (with discussion). *J. R. Statist. Soc.* A, **149**, 187-227.

Harvey, A.C., Ruiz, E. and Shephard, N. (1994). Multivariate stochastic variance models. *Review of Economic Studies*, **61**, 247-264.

Harvey, A.C. and Shephard, N. (1996). The estimation of an asymmetric stochastic volatility model for asset returns. *Journal of Business and Economic Statistics*, **14**, 429-434.

Harvey, C.R. and Siddique, A. (1999). Autoregressive conditional skewness. *Journal of Financial and Quantitative Analysis*, **34**, 465-487.

Hastie, T. J. and Tibshirani, R.J. (1990). *Generalized Additive Models.* London: Chapman and Hall.

Hastie, T. J., Tibshirani, R.J and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* 2nd ed, New York, N.Y. : Springer.

Jacquier, E., Polson, N.G. and Rossi, P.E. (2004). Stochastic volatility models: univariate and multivariate extensions. *Journal of Econometrics*, **122**, 185-212.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions*, 2nd ed. Wiley, New York.

Jondeau, E. and Rockinger, M. (2003). Conditional volatility, skewness, and kurtosis: existence, persistence, and comovements. *Journal of Economic Dynamics and Control*, **27**, 1699-1737.

Jondeau, E. and Rockinger, M. (2009). The impact of shocks on higher moments. *Journal of Financial Econometrics*, **7**, 77-105.

Jones, M.C. and Faddy, M.J. (2003). A skew extension of the t-distribution, with application. *Journal of the Royal Statistical Society: Series B*, **65**, 159-174.

Jones, M.C. and Pewsey, A. (2009). Sinh-arcsinh distributions. *Biometrika*, **96**, 761-780.

Kalman, R.E. (1960). A new approach to linear prediction and filtering problems. *J. Basic Engineering Transactions ASME* D, **82**, 35-45.

Kalman, R.E. and Bucy, R. (1961). New results in filtering and prediction theory, *J. Basic Engineering Transactions ASME* D, **83**, 95-108.

Kim, S., Shephard, N. and Chib, S. (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies*, **65**, 361-393.

Kitagawa, G. (1987). Non-Gaussian state-space modelling of nonstationary time series (with discussion). *J. Am. Statist. Ass.*, **82**, 1032-1063.

Kitagawa, G. (1989). Non-Gaussian seasonal adjustment. *Comput. Math. Applic.*, **18**, 503-514.

Kitagawa, G. (1990). The two-filter formula for smoothing and an implementation of the Gaussian-sum smoother. *Technical Report.* Institute of Statistical Mathematics, Tokyo.

Kitagawa, G. and W. Gersch, (1996). *Smoothness Priors Analysis of Time Series.* Springer- Verlag.

Lanne, M. and Pentti, S. (2007). Modeling Conditional Skewness in Stock Returns. *The European Journal of Finance*, **13**, 691-704.

Lee, Y. and Nelder, J.A. (1996). Hierarchical generalized linear models (with discussion). *Journal of the Royal Statistical Society: Series B*, **58**, 619-678.

Lee, Y., Nelder, J.A and Pawitan, Y. (2006). *Generalized Linear Models With Random Effects: Unified Analysis Via H-Likelihood.* Chapman & Hall/CRC

Li, W.K. (1994). Time series models based on generalized linear models: some further results. *Biometrika*, **50**, 506-511.

Mandelbrot, B. (1963). New methods in statistical economics. *Journal of Political Economy*, **71**, 421-440.

Matsumoto, M. and Nishimura, T. (1998). Mersenne Twister: A 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation*, **8**, 3âĂŞ30.

McDonald, J.B. and Xu, Y. J. (1995). A generalisation of the beta distribution with applications. *Journal of Econometrics*, **66**, 133-152.

Mitchell, J. and Wallis, K.F. (2011). Evaluating density forecasts: forecast combinations, model mixtures, calibration and sharpness. *Journal of Applied Econometrics*, **26**, 1023-1040.

Mittnik, S., Paolella, M. S. and Rachev, S.T. (2002). Stationarity of stable power-GARCH processes. *Journal of Econometrics*, **106**, 97-107.

Nagahara, Y. (2003). Non-Gaussian filter and smoother based on the Pearson distribution system. *Journal of Time Series Analysis*, **24**, 721-738.

Nagahara, Y. and Kitagawa, G. (1999). A non-Gaussian stochastic volatility model. *Journal of Computational Finance*, **2**, 33-47.

Nakajima, J. and Omori, Y. (2009). Leverage, heavy-tails and correlated jumps in stochastic volatility models. *Computational Statistics and Data Analysis*, **53**, 2535-2553.

Nakajima, J. and Omori, Y. (2012). Stochastic volatility model with leverage and asymmetrically heavy-tailed error using GH skew Student's t-distribution. *Computational Statistics and Data Analysis*, **56**, 3690-3704.

Nakajima, J., Kunihama, T., Omori, Y. and Fruhwirth-Schnatter, S. (2012). Generalized extreme value distribution with time-dependence using the AR and MA models in state space form. *Computational Statistics and Data Analysis*, **56**, 3241-3259.

Nelder, J.A. (2000). Discussion of Durbin and Koopman (2000). *Journal of the Royal Statistical Society: Series B*, **62**, 38.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: a new approach. *Econometrica*, **59**, 347-370.

Omori, Y., Chib, S., Shephard, N., and Nakajima, J. (2007). Stochastic volatility with leverage: fast likelihood inference. *Journal of Econometrics*, **140**, 425-449.

Pawitan, Y. (2001). *In All Likelihood: Statistical Modelling and Inference Using Likelihood.* Oxford University Press.

Petris, G., Petrone, S. and Campagnoli, P. (2009). *Dynamic Linear Models with R.* Springer-Verlag, New York.

Pinheiro, J.C. and Bates, D.M. (2000). *Mixed-Effects Models in S and S-PLUS.* New York: Springer-Verlag.

Polasek, W. and Pai, J. (1998). Autoregressive moving average models with t and hyperbolic innovations. WWZ-Discussion Paper, 9816.

Rigby, R.A. and Stasinopoulos, D.M. (2005). Generalized additive models for location, scale and shape (with discussion). *Journal of the Royal Statistical Society: Series C*, **54**, 507-554.

Rigby, R.A. and Stasinopoulos, D.M. (2013). Automatic smoothing parameter selection in GAMLSS with an application to centile estimation. *Statistical Methods in Medical Research*, **1**, 1-15.

Rockinger, M. and Jondeau, E. (2002). Entropy densities with an application to autoregressive conditional skewness and kurtosis. *Journal of Econometrics*, **106**, 119-142.

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *The Annals of Mathematical Statistics*, **23**, 470-472.

Royston, P. and Wright, E.M. (2000). Goodness-of-fit statistics for age-specific reference intervals. *Statistics in Medicine*, **19**, 2943-2962.

Rue, H., Martino, S. and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B*, **71**, 319-392.

Shephard, N. (1994). Partial non-Gaussian state space. *Biometrika*, **81**, 115-131.

Shephard, N. (2005). *Stochastic Volatility: Selected Readings.* Oxford University Press, Oxford.

Shephard, N. and Pitt, M. K. (1997). Likelihood analysis of non-Gaussian measurement time series. *Biomnetrika*, **84**, 653-667.

Shumway, R.H. and Stoffer, D.S. (2011). *Time Series Analysis and Its Applications With R Examples.* Third Ed., Springer.

Stasinopoulos, D. M., Rigby, R.A and Akantziliotou, C. (2008). *Instructions on how to use the GAMLSS package in R*, Second Edition, STORM Research Centre, London Metropolitan University, London.

Taylor, S.J. (1994). Modelling stochastic volatility. *Mathematical Finance*, **4**, 183-204.

Tierney, L. and Kadane, J.B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, **81**, 82-86.

Tsiotas, G. (2012). On generalised asymmetric stochastic volatility models. *Computational Statistics and Data Analysis*, **56**, 151-172.

Venables, W.N. and Ripley, B.D. (2002). *Modern Applied Statistics with S.* Forth ed. New York: Springer-Verlag.

Wang, J.J.J, Chan, J.S.K. and Choy, S.T.B (2011). Stochastic volatility models with leverage and heavy-tailed distributions: A Bayesian approach using scale mixtures. *Computational Statistics and Data Analysis*, **55**, 852-862.

West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models.* second ed. Springer, New York.

West, M., Harrison, P. J. and Migon, H. S. (1985). Dynamic generalized linear models and Bayesian forecasting (with discussion). *J. Am. Statist. Ass.*, **80**, 73-97.

Wiener, N. (1949). *Extrapolation, Interpolation and Smoothing of Stationary Time Series with Engineering Applications.* Wiley.

Wilhelmsson, A. (2009). Value at Risk with time varying variance, skewness and kurtosis-the NIG-ACD model. *Econometrics Journal*, **12**, 82-104.

Wood, S.N. (2006). *Generalized Additive Models: An Introduction with R.* Chapman & Hall/CRC.

Wurtz, D., Chalabi, Y. and Luksan, L. (2006). Parameter estimation of ARMA models with GARCH/APARCH errors an R and SPlus software implementation. *Journal of Statistical Software*. Draft.

Yu, J. (2005). On leverage in a stochastic volatility model. *Journal of Econometrics*, **127**, 165-178.

Zeger, S.L. (1988). A regression model for time series of counts. *Biometrika*, **75**, 621-629.

Zeger, S.L. and Qaqish, B. (1988). Markov regression models for time series: a quasi-likelihood approach. *Biometrics*, **44**, 1019-1031.