Faculty of Social Sciences
University of Helsinki

Essays on mixture autoregressive models with applications to macroeconomics and finance

Savi Virolainen

Doctoral dissertation, to be presented for public discussion with the permission of the Faculty of Social Sciences of the University of Helsinki, in Auditorium 116, Unioninkatu 35, on the 2nd of December, 2022 at 13 o'clock.

Doctoral Programme in Economics,
Faculty of Social Sciences, University of Helsinki

Supervisors:
Professor Markku Lanne, University of Helsinki
Professor Mika Meitz, University of Helsinki
Professor Emeritus Pentti Saikkonen, University of Helsinki

Pre-examiners:
Associate Professor Emanuele Bacchiocchi, University of Bologna
Professor Yoosoon Chang, Indiana University

Opponent:
Professor Luca Fanelli, University of Bologna

ISBN 978-951-51-8734-5 (paperback)
ISBN 978-951-51-8735-2 (PDF)

Unigrafia Oy
Helsinki 2022

Abstract

This dissertation complements a family of mixture autoregressive models based on Gaussian and Student's t distributions by filling the gaps in the previous literature with four self contained essays. This includes univariate models as well as reduced form and structural multivariate models. Empirical applications to macroeconomics and finance demonstrate their usefulness. I have also accompanied this dissertation with open source software, in the form of R packages **uGMAR** and **gmvarkit**, which provide a comprehensive set of tools for estimation and other numerical analysis of the models. The software is distributed through the Comprehensive R Archive Network.

The first essay introduces a new mixture autoregressive model that combines linear Gaussian autoregressions and linear Student's t autoregressions as its mixture components. The model has attractive properties analogous to the Gaussian and Student's t mixture autoregressive models, but it is more flexible as it enables to model series which consist of both conditionally homoskedastic Gaussian regimes and conditionally heteroskedastic Student's t regimes. The usefulness of the model is demonstrated in an empirical application to the monthly U.S. interest rate spread between the 3-month Treasury bill rate and the effective federal funds rate.

The second essay describes the R package **uGMAR**, which provides tools for estimating and analysing the Gaussian mixture autoregressive model, the Student's t mixture autoregressive model, and the Gaussian and Student's t mixture autoregressive model. The model parameters are estimated with the method of maximum likelihood by running multiple rounds of a two-phase estimation procedure in which a genetic algorithm is used to find starting values for a gradient based method. For evaluating the adequacy of the estimated models, **uGMAR** utilizes so-called quantile residuals and provides functions for graphical diagnostics as well as for calculating formal diagnostic tests. **uGMAR** also facilitates simulation from the processes and forecasting future values of the process by a simulation-based Monte Carlo method. I illustrate the use of **uGMAR** with the monthly U.S. interest rate spread between the 10-year and 1-year Treasury rates.

In the third essay, I proceed to multivariate models and introduce a structural Gaussian mixture vector autoregressive model. The shocks are identified by combining simultaneous diagonalization of the reduced form error covariance matrices with constraints on the time-varying impact matrix. This leads to flexible identification conditions, and some the constraints are also testable. In an empirical application to quarterly U.S. data covering the period from 1953Q3 to 2021Q4, my model identifies two regimes: a stable inflation regime and an unstable inflation

regime. The unstable inflation regime is characterized by high or volatile inflation, and it mainly prevails in the 1970's, early 1980's, during the Financial crisis, and in the COVID-19 crisis from 2020Q3 onwards. The stable inflation regime, in turn, is characterized by moderate inflation, and it prevails when the unstable inflation regime does not. While the effects of the monetary policy shock are relatively symmetric in the unstable inflation regime, I find strong asymmetries with respect to the sign and size of the shock as well as to the initial state of the economy in the stable inflation regime. On average, the real effects of the monetary policy shock are somewhat stronger in the stable inflation regime than in the unstable inflation regime.

The last essay introduces a new mixture vector autoregressive model based on Gaussian and Student's t distributions. The model incorporates conditionally homoskedastic linear Gaussian vector autoregressions and conditionally heteroskedastic linear Student's t vector autoregressions as its mixture components. For a pth order model, the mixing weights depend on the full distribution of the preceding p observations. The specific formulation of the mixing weights leads to attractive practical and theoretical properties such as ergodicity and full knowledge of the stationary distribution of $p + 1$ consecutive observations. The empirical application studies asymmetries in the effects of Euro area monetary policy shocks. My model identifies two regimes: a low-growth regime and a high-growth regime. The low-growth regime is characterized by negative (but volatile) output gap, and it mainly prevails after the Financial crisis. The high-growth regime is characterized by positive output gap, and it mainly dominates before the Financial crisis. I find the real effects less enduring for an expansionary than for a contractionary monetary policy shock. On average, the inflationary effects of the monetary policy shock are stronger in the high-growth regime than in the low-growth regime.

Acknowledgements

There are several people I wish to thank for the help with this dissertation project. Above all, I want to thank my supervisors for the substantial help. I am deeply grateful to Professor Markku Lanne for all the valuable advice and useful comments, which have been a tremendous help. I would also like to extend my deepest gratitude to Professor Mika Meitz for all the helpful discussions and comments, and for always taking the time to answer my questions. I cannot begin to express my thanks to Professor Emeritus Pentti Saikkonen, who has helped and supported me countless times with various concerns.

Pentti's guidance in the academic world has had an enormous impact on the path that led me to write a doctoral dissertation on mixture autoregressive models. After supervising my bachelor's thesis on time series analysis, he introduced me to Markku and Mika for a summer job as a research assistant to develop a software package for a mixture autoregressive model. Later, Pentti supervised my master's thesis on a related model, which ended up being introduced also in the first essay of my dissertation. After studying time series analysis in my master's studies and having laid foundation for the first essay, continuing to the doctoral studies under the supervision of Markku, Mika, and Pentti felt like the natural next step for learning more about time series econometrics.

I wish to also thank the pre-examiners of my dissertation, Associate Professor Emanuele Bacchiocchi and Professor Yoosoon Chang, for providing detailed comments. I would also like to thank Professor Luca Fanelli for agreeing to act of as my opponent.

The project would not have been possible without funding. I am grateful for the four-year salaried position in the Doctoral Programme in Economics at the University of Helsinki. I am also grateful to the Academy of Finland for the financial support (Grants 308628 and 347986).

I must thank Docent Leena Kalliovirta, Associate Professor Henri Nyberg, Professor Antti Ripatti, and others, who commented my work in the time series econometrics workshops and helped me to improve the essays. I would also like to thank Associate Professor Daniel Preve for the discussions on numerical optimization, which helped me in the early stages of the development of the software accompanying this dissertation.

I thank my fellow doctoral researchers and other colleagues for the peer support and all the pleasant conversations. Finally, I am thankful to my friends and family for the support.

Helsinki, October 2022
Savi Virolainen

Contents

CONTENTS

Chapter 1

Introduction

1.1 Backround

The analysis of the dynamic consequences of events over time is a major part of empirical macroeconomics and finance. Statistical analysis of time series data dates back to Yule (1927), and a wide range of time series models have been developed since. A major branch of time series literature focuses on autoregressive models that assume the current observation to be a function of the past observations, a random shock, and possibly exogenous variables. Linear autoregressive models, in particular, are popular workhorse models in empirical macroeconomics, and assume the current observation to be a linear function of (often a finite number of) the preceding observations and a random shock.

Linear autoregressive models are relatively simple and very capable of filtering autocorrelation. They are not, however, able to capture all the relevant characteristics of the series when the underlying data generating dynamics vary in time, for instance, depending on the state of the economy. Variation in the dynamics may arise due to wars, crises, business cycle fluctuations, or policy shifts, for example. Various types of time series models capable of capturing such features have been proposed. One of them is the class of mixture autoregressive models introduced by Le, Martin, and Raftery (1996) in which each observation is generated by one of the mixture components (or regimes) that is randomly selected according to the probabilities given by the mixing weights. Different mixture autoregressive models can be created by defining the mixture components and mixing weights in various ways.

Wong and Li (2000), for instance, proposed using Gaussian mixture components with constant mixing weights, allowing for multimodal predictive distributions, while Wong, Chan, and Kam (2009) suggested using more heavy tailed Student's t distributions in financial applications. Wong and Li (2001b) extended the model of Wong and Li (2000) to accommodate autoregressive conditional heteroskedasticity (ARCH) in each regime. Fong, Li, Yau, and Wong (2007), in turn, extended the model of Wong and Li (2000) to the multivariate case, whereas Bentarzi and Djeddou (2014) further extended the multivariate model to incorporate periodically time-varying coefficients. Wong and Li (2001a) introduced a model with Gaussian mixture components and

mixing weights that vary according to endogenous or exogenous variables through a logistic function, and Burgard, Neuenkirch, and Nöckel (2019) proposed a similar type of model in the multivariate setting. Lanne and Saikkonen (2003) defined the mixing weights as functions of a lagged observation and incorporated a generalized ARCH process in each regime. The model of Bec, Rahbek, and Shephard (2008), on the other hand, allows for epochs when the process is seemingly nonstationary, while at the same time the process is stationary.

This dissertation considers mixture (vector) autoregressive models whose mixing weights are constructed so that the greater the relative weighted likelihood of a regime is, the more likely the process is to generate an observation from it. The idea of using mixing weights that are endogenously determined through the relative weighted likelihoods of the regimes originates from Glasbey (2001), who proposed a first order two-regime Gaussian mixture autoregressive (GMAR) model for modelling solar radiation. Kalliovirta, Meitz, and Saikkonen (2015) then extended this model to the general case and studied its theoretical properties extensively, which led to further developments of that type of mixture autoregressive models.

Kalliovirta, Meitz, and Saikkonen (2016) introduced the multivariate counterpart of the GMAR model, the Gaussian mixture vector autoregressive model. Meitz, Preve, and Saikkonen (forthcoming) proposed utilizing a Student's t distribution in the univariate setting and introduced the Student's t mixture autoregressive model with conditionally heteroskedastic regimes. Together, the mixture (vector) autoregressive models with Gaussian or Student's t mixture components and mixing weights determined through the relative weighted likelihoods of the regimes constitute an appealing family of mixture autoregressive models. I complement the rest of this family in this dissertation, including structural versions of the multivariate models.

1.2 Linear autoregressive models

Linear autoregressive models are popular workhorse models of empirical macroeconomics. They are relatively simple and effectively filter autocorrelation, and they are also the component processes of the mixture autoregressive models considered in this dissertation. Suppose y_t $(t = 1, 2, ...)$ is the real valued d-dimensional time series of interest with $d \geq 1$, and denote by $\mathcal{F}_{t-1}$ the σ-algebra generated by the random vectors or scalars $\{y_s, s < t\}$, i.e., $\mathcal{F}_{t-1}$ contains the information about the past of y_t. The benchmark linear autoregressive model of order p assumes that

$$y_t = \phi_0 + \sum_{i=1}^{p} A_i y_{t-1} + \Omega_t^{1/2} \varepsilon_t, \tag{1.2.1}$$

where $\phi_0 \in \mathbb{R}^d$ is an intercept parameter and $A_i \in \mathbb{R}^{d \times d}$ $(i = 1, ..., p)$ are coefficient matrices (or coefficients in the scalar case). The error process ε_t $(t = 1, 2, ...)$ is identically and independently distributed (IID) with zero mean and identity covariance matrix (or unit variance in the scalar case), and ε_t is independent of $\mathcal{F}_{t-1}$. The matrix $\Omega_t^{1/2} \in \mathbb{R}^{d \times d}$, in turn, captures the conditional covariance matrix (or variance) Ω_t of the process y_t (conditionally on the past of y_t), which I assume positive definite. Different linear autoregressions can be created by defining the error

process ε_t and covariance matrix Ω_t in various ways. It is also possible to add, for instance, trend or seasonal components, functions of exogenous variables, or a moving average component to the model.

The family of mixture autoregressive models studied in this dissertation involves two types of linear autoregressions as mixture components: linear Gaussian autoregressions and linear Student's t autoregressions. The linear Gaussian autoregressions assume that the errors ε_t follow a d-dimensional standard normal distribution and that the covariance matrix Ω_t is a positive definite constant. The linear Student's t autoregressions assume that the errors ε_t follow a d-dimensional Student's t distribution with mean zero, identity covariance matrix, and $\nu + dp$ degrees of freedom. The covariance matrix Ω_t is a positive definite constant $(d \times d)$ matrix multiplied by a time-varying scalar that depends on the preceding p observations through their quadratic form and the autoregression coefficients in A_i, $i = 1, ..., p$ (the exact definition is not presented here for brevity). We assume that the Gaussian and Student's t autoregressions both satisfy the stability condition $\det(I_d - \sum_{i=1}^{p} A_i z^i) \neq 0$ for $|z| \leq 1$, where the $(d \times d)$ identity matrix I_d is a scalar for $d = 1$. This leads to stationarity of the linear autoregressions, and as is shown in the dissertation, also to the stationarity of the introduced mixture autoregressions.

The Student's t autoregression is able to capture conditional heteroskedasticity and fatter tailed distributions than the Gaussian autoregression, which is assumed conditionally homoskedastic. However, as the conditional covariance matrix of the Student's t autoregression depends on the preceding p observations through the same parameters as the conditional mean, it does not generally filter autocorrelation as well as the Gaussian autoregression. Both linear autoregressions, nevertheless, assume that the dynamics of the process stay constant over time. Since the dynamics of the economy may vary in time due to wars, crises, business cycle fluctuations, or policy shifts, for example, it is useful to consider mixture models that allow for such variation.

1.3 Mixture autoregressive models

Mixture autoregressive models can be described as collections of linear autoregressive models, each of which is referred to as a mixture component, a component process, or a regime. This dissertation considers mixture autoregressive models in which each observation is generated by a mixture component that is randomly selected according to the probabilities given by the mixing weights. For concreteness, I will next formalize this definition.

Proceeding with the notation introduced in Section 1.2, I consider mixture autoregressive models with autoregressive order p and M mixture components for which we have

$$y_t = \sum_{m=1}^{M} s_{m,t} \left(\phi_{m,0} + \sum_{i=1}^{p} A_{m,i} y_{t-i} + \Omega_{m,t}^{1/2} \varepsilon_{m,t} \right), \tag{1.3.1}$$

where $s_{m,t}$, $m = 1, ..., M$, are unobservable regime variables such that at each t, exactly one of them takes the value one and the others take the value zero according to the ($\mathcal{F}_{t-1}$-measurable)

probabilities given by the mixing weights $\alpha_{m,t} \equiv \mathrm{P}(s_{m,t} = 1|\mathcal{F}_{t-1})$ that satisfy $\sum_{m=1}^{M} \alpha_{m,t} = 1$. The quantities $\phi_{m,0}$, $A_{m,i}$, and $\Omega_{m,t}$ as well as the IID error processes $\varepsilon_{m,t}$, $m = 1, ..., M$, are assumed to satisfy the same properties for each regime m as the linear autoregression (1.2.1) defined in Section 1.2. That is, at each time point t, the process reduces to linear autoregression corresponding to the regime m that is selected, while the regime is selected randomly according to the probabilities given by the mixing weights $\alpha_{m,t}$. Different mixture autoregressive models can be created by defining the mixture components and mixing weights in various ways.

The constant mixing weights proposed by Wong and Li (2000), Wong and Li (2001b), Fong *et al.* (2007), Wong *et al.* (2009), and Bentarzi and Djeddou (2014), to name a few, allow for multimodal predictive distributions but do not let the regime-switching probabilities to vary in time. Logistic mixing weights utilized, for example, by Wong and Li (2001a) and Burgard *et al.* (2019) allow the regime-switching probabilities to depend on the level of endogenous or exogenous variables.[1] The mixing weights proposed by Lanne and Saikkonen (2003), in turn, depend the level of a lagged observation, whereas the ones introduced by Bec *et al.* (2008) depend on the magnitude of lagged observations (irrespective of their signs).

This dissertation considers mixing weights that, for a pth order model, depend on the full distribution of the preceding p observations. This allows the regime-switching probabilities to depend on the level, variability, kurtosis, and temporal as well as contemporaneous dependence of the past observations. Specifically, the mixing weights are constructed so that the greater the relative weighted likelihood of a regime is, the more likely the process is to generate an observation from it. This is an appealing feature for forecasting, and it also facilitates associating statistical characteristics and economic interpretations to the regimes. Moreover, it leads to attractive theoretical properties such as ergodicity and full knowledge of the stationary distribution of $p + 1$ consecutive observations, as is shown in Chapters 2 and 5 (and in Kalliovirta *et al.*, 2015, 2016, Meitz *et al.*, forthcoming).

To exemplify how models identify statistical regimes in the data, Figure 1.1 presents the quarterly percentage change of the real U.S. GDP from 1947Q2 to 2021Q4, where the shaded areas are the NBER based U.S. recessions.[2] I fitted the Gaussian and Student's mixture autoregressive model (Virolainen, forthcoming, and Chapter 2 of this dissertation) to the series with one Gaussian mixture component, one Student's t mixture component, and autoregressive order two. The diagnostic tests of Kalliovirta (2012) show that this model is adequate.[3] Then, I depicted the estimated mixing weights of the Gaussian regime to the figure with blue dashed lines.

As the mixing weights in Figure 1.1 show, the Gaussian regime obtains a large probability often during recessions but also after them. The Gaussian regime does not, therefore, perfectly match the periods of recessions. But this is not a deficiency, however. On the contrary, it

[1] In the model of Wong and Li (2001a), exogenous variables are, in addition to the mixing weights, allowed to enter the linear autoregressions on the right side of (1.3.1).

[2] The series were retrieved from the Federal Reserve Bank of St. Louis database.

[3] The normality test as well as the autocorrelation and conditional heteroskedasticity tests taking into account $1, 2, ..., 20$ lags all passed at all the conventional levels of significance without and with the simulation procedure using a sample of length 10000.

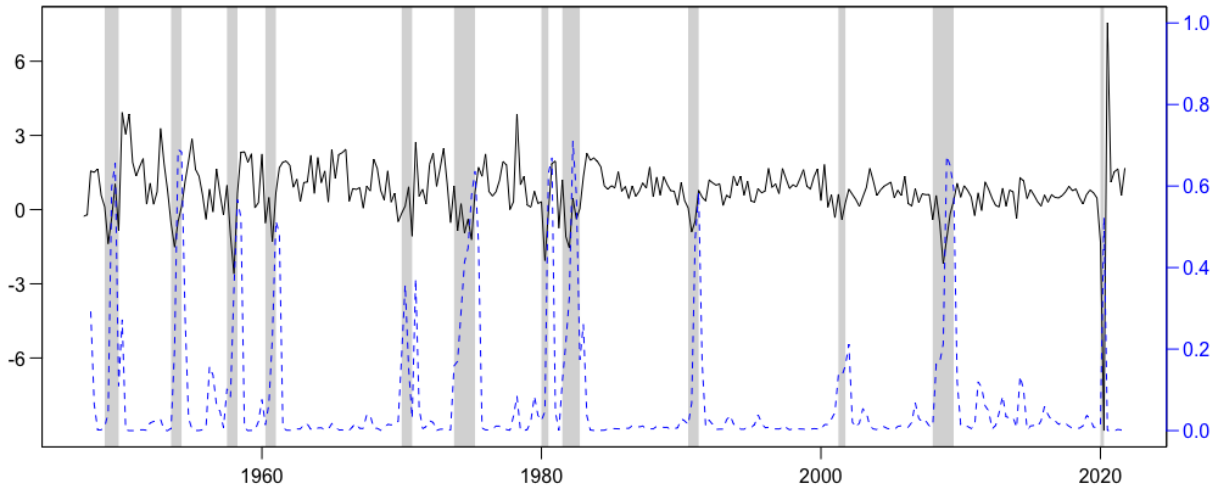

Figure 1.1: The quarterly percentage change of the real U.S. GDP from 1947Q2 to 2021Q4 (black solid line) together with the estimated mixing weights of the Gaussian regime of the fitted Gaussian and Student's t mixture autoregressive model with one Gaussian and one Student's t regime and autoregressive order two (blue dashed line). The shaded areas are the NBER based U.S. recessions.

reveals that the dynamics of the series do not seem to vary particularly between recessions and expansions only but often also in the periods following the recessions (and occasionally at other times too).

1.4 Structural autoregressive models

The reduced form mixture autoregressive models discussed in the previous section are useful for forecasting and studying the statistical relations of the variables. Particularly due to the endogenously determined mixing weights, the models also facilitate identifying statistical regimes from the data, which may also have economic interpretations. However, the interest is often in the causal effects of events such as fiscal or monetary policy actions, for example. To that end, it is useful to study the effects of exogenous shocks, as proposed by Sims (1980).

It is not very useful to study the effects of the reduced form shocks, $u_t \equiv \sum_{m=1}^{M} s_{m,t}\Omega_{m,t}^{1/2}\varepsilon_{m,t}$ in the mixture model (1.3.1) (or $\Omega_t^{1/2}\varepsilon_t$ in the linear model (1.2.1)), because the economy may react simultaneously (i.e., within the time period) to a shock when it arrives. This endogenous response of the economy would then be present in the reduced form shock, and the effects of this reduced form shock would not thereby isolate the causal effects of the (unexpected) movements of the variable of interest (see, e.g., Ramey, 2016).

To elaborate with an example, consider a bivariate system of production and inflation, and suppose that one is interested in the causal effects of unexpected movements of inflation (i.e., the effects of a shock to inflation). If an exogenous shock arrives to production, the supply changes

and prices start adjusting to the new level of supply. Then, the consequent (potential) simultaneous change in inflation enters the reduced form shock of inflation. Therefore, the reduced form shock of inflation contains both, the exogenous shock to inflation and the simultaneous adjustment of inflation to the shock to production. The effects of this reduced form shock would, hence, partially be some of the effects of the endogenous adjustment to the shock to production. Similarly, some of the effects of the shock to inflation (potentially) enter the reduced form shock of production and are not present in the reduced form shock of inflation. The effects of the reduced form shock of inflation would not thereby reveal the causal effects of a shock to inflation.

Chapters 4 and 5 introduce structural versions of the (multivariate) models, which facilitate tracing out the causal effects of the shocks. Structural shocks, e_t, are defined as orthogonal, serially uncorrelated exogenous shocks to the system and do not thereby suffer from the above-described endogeneity problem of the reduced form shocks. They are recovered from the reduced form shock u_t with the transformation

$$e_t = B_t^{-1} u_t \tag{1.4.1}$$

by finding a non-singular $(d \times d)$ impact matrix B_t (which may be time-varying) that orthogonalizes the reduced form shocks, i.e., such that the conditional covariance matrix of e_t (typically conditional on the past of y_t) is a diagonal matrix. A common normalization is to assume that it is an identity matrix, $\text{Cov}(e_t|\mathcal{F}_{t-1}) = I_d$. That is, the objective is to find B_t such that

$$B_t^{-1}\text{Cov}(u_t|\mathcal{F}_{t-1})B_t'^{-1} = I_d. \tag{1.4.2}$$

There are in general multiple solutions to B_t, as there are d^2 variables and $d(d+1)/2$ unique equations in (1.4.2). To recover the structural shocks, the model should, hence, be constrained so that the solution is unique. The identification problem is then to find the constraints that recover the shocks of interest from the set of all possible structural shocks. Various types of solutions to the identification problem have been proposed in the literature. Some of them are discussed in Ramey (2016), for instance.

In Chapter 4, I adopt the solution proposed by Lanne and Lütkepohl (2010) and Lanne, Lütkepohl, and Maciejowsla (2010) and show that the impact matrix is unique up to ordering and signs of its columns for the structural Gaussian mixture vector autoregressive model. Then, I derive a general set of conditions for uniquely identifying any subset of the shocks. Based on these results, Chapter 5 establishes the same identification conditions also for the structural Student's t mixture vector autoregressive model and for the structural Gaussian and Student's t mixture vector autoregressive model. In the empirical applications, I study the effects of U.S. monetary policy shocks in Chapter 4, whereas in Chapter 5, I study the effects of Euro area monetary policy shocks.

1.5 Summary of the essays

This section briefly summarizes the contents of the four self-contained essays of this dissertation. The aim of these essays is to fill gaps in the previous literature by complementing a family of mixture autoregressive models and introducing the software for numerical analysis of these models. The first essay (Chapter 2) generalizes the Gaussian mixture autoregressive (GMAR) model (Kalliovirta *et al.*, 2015) and the Student's t mixture autoregressive (StMAR) model (Meitz *et al.*, forthcoming) to the Gaussian and Student's t mixture autoregressive (G-StMAR) model by combining Gaussian and Student's t mixture components. The second essay (Chapter 3) describes the R package **uGMAR**, which introduces a comprehensive set of tools for the numerical analysis of the GMAR, StMAR, and G-StMAR models. The third and fourth essays are concerned with the multivariate versions of these models. The third essay (Chapter 4) introduces a structural version of the Gaussian mixture vector autoregressive model (Kalliovirta *et al.*, 2016) with time-varying impact matrix and statistically identified shocks. Finally, the fourth essay (Chapter 5) introduces the multivariate counterparts of the StMAR model and the G-StMAR model, including their structural versions. The methods put forth in these chapters are accommodated in the R package **gmvarkit**, which provides a comprehensive set of tools for numerical analysis of the multivariate models. The R package **gmvarkit** is not explicitly described in any of the essays, but it works similarly to **uGMAR**.

1.5.1 A mixture autoregressive model based on Gaussian and Student's *t*-distributions

Mixture autoregressive models are useful for modelling series in which the data generating dynamics vary in time, for instance, due to wars, crises, business cycle fluctuations, or policy shifts. Mixture autoregressive models can be described as collections of (typically linear) autoregressive models, which are called mixture components, components processes, or regimes. At each time point, the process generates an observation from one of its mixture components that is randomly selected according to the probabilities given by the mixing weights.

In the first essay, I introduce a new mixture autoregressive model that is a combination of the Gaussian mixture autoregressive (GMAR) model (Kalliovirta *et al.*, 2015) and the Student's t mixture autoregressive (StMAR) model (Meitz *et al.*, forthcoming). This model, referred to as the G-StMAR model, accommodates conditionally homoskedastic linear Gaussian autoregressions and conditionally heteroskedastic Student's t autoregressions as its mixture components. The G-StMAR model is obtained from Equation (1.3.1) in Section 1.3 with $d = 1$ and assuming that the first $M_1 \leq M$ mixture components are linear Gaussian autoregressions and the rest $M - M_1$ mixture components are linear Student's t autoregressions.

For a pth order model, the mixing weights are determined through the full distribution of the previous p observations, which allows the regime-switching probabilities to depend on the level, variability, temporal dependence, and kurtosis of the past observations. Specifically, the mixing weights are constructed so that the greater the weighted relative likelihood of a regime is, the

more likely the process is to generate an observation from it. This facilitates associating specific characteristics and giving economic interpretations to the regimes. The specific formulation of the mixing weights also leads to attractive theoretical properties, such as ergodicity and full knowledge of the stationary distribution of $p + 1$ consecutive observations.

The formulation of the mixing weights leads to attractive properties, but their complex dependence on the preceding observations through the autoregressive parameters makes the estimation of the model parameters challenging in practice. Considering estimation by the method of maximum likelihood, the definition of the mixing weights particularly induces a large number of modes to the surface of the log-likelihood function, and large areas to the parameter space, where it is flat in multiple directions. Following Dorsey and Mayer (1995) and Meitz, Preve, and Saikkonen (2018), Meitz *et al.* (forthcoming), I propose using a two-phase estimation procedure, where a genetic algorithm is used to find starting values for a gradient based variable metric algorithm (Nash, 1990, algorithm 21, implemented by R Core Team (2022)). I also describe the modified genetic algorithm that is implemented in the accompanying R package **uGMAR** (Virolainen, 2018b).

It turns out that the G-StMAR model is a limiting case of a StMAR model with the t-distributions of some regimes tending to normal distributions as the degrees of freedom parameters tend to infinity. Based in this observation, I propose a model selection procedure in which one should first find a suitable StMAR model. Then, if the StMAR model contains large degrees of freedom parameter estimates, one should switch the corresponding regimes to the Gaussian type by estimating the appropriate G-StMAR model. As opposed to the limiting StMAR model, the advantage of the G-StMAR model is that it removes the redundant degrees of freedom parameters from the model, and it is free from numerical problems induced by weak identification of very large degrees of freedom parameters.

The usefulness of the G-StMAR model is demonstrated in an empirical application to the monthly U.S. interest rate spread between the 3-month Treasury bill (TB) rate and the effective federal funds (FF) rate. The G-StMAR model identifies three regimes for the spread, with a switch from a Student's t regime to a Gaussian regime arising from a switch in the economic regime, namely, to a regime where the zero lower bound limits the movements of the interest rates. The two Student's t regimes accommodate eras of low mean and high variability and high mean and moderate variability. The former Student's t regime dominates often when the market possibly anticipates decreases in the FF rate or has increased preference for safety, whereas the latter one mostly prevails when the Fed is arguably not expected to significantly decrease the FF rate target.

My findings are consistent with Sarno and Thornton (2003), who found that the FF rate seems to adjust to the TB rate, supporting the hypothesis that the market anticipates movements of the FF rate, moving the TB rate, and hence the spread, in advance. As opposed to modelling the series with a StMAR model containing an overly large degrees of freedom parameter estimate, switching to the more parsimonious G-StMAR model allows to numerically compute approximate standard errors for the estimates, and moreover, to perform Kalliovirta's (2012) quantile residual tests, which turned out to be useful in model selection.

1.5.2 uGMAR: a family of mixture autoregressive models in R

Mixture autoregressive models are useful for analysing time series that exhibit nonlinear, regime-switching features. The Gaussian mixture autoregressive (GMAR) model (Kalliovirta *et al.*, 2015), the Student's t mixture autoregressive (StMAR) model (Meitz *et al.*, forthcoming), and the Gaussian and Student's t mixture autoregressive (G-StMAR) model (Virolainen, forthcoming) constitute a (univariate) family of such models that I refer to as the GSMAR models. A GSMAR process generates each observation from one of its mixture components, which are either conditionally homoskedastic linear Gaussian autoregressions or conditionally heteroskedastic linear Student's t autoregressions. The mixture component that generates each observation is randomly selected according to the probabilities determined by the mixing weights that, for a pth order model, depend on the full distribution of the previous p observations. Consequently, the regime-switching probabilities may depend on the level, variability, kurtosis, and temporal dependence of the past observations. The specific formulation of the mixing weights also leads to attractive theoretical properties such as ergodicity and full knowledge of the stationary distribution of $p + 1$ consecutive observations.

The second essay describes the R package **uGMAR** providing a comprehensive set of easy-to-use tools for GSMAR modelling, including unconstrained and constrained maximum likelihood estimation of the model parameters, quantile residual based model diagnostics, simulation from the processes, and forecasting. The emphasis is on estimation, as it can be rather tricky. In particular, due to the endogenously determined mixing weights, the log-likelihood function has a large number of modes, and in large areas of the parameter space, the log-likelihood function is flat in multiple directions. The global maximum point of the log-likelihood function is also frequently located very near the boundary of the parameter space. It turns out, however, that such near-the-boundary estimates often maximize the log-likelihood function for rather a technical reason, and it might be more appropriate to prefer an alternative estimate based on the largest local maximum point that is clearly in the interior of the parameter space.

The model parameters are estimated by running multiple rounds of a two-phase estimation procedure in which a modified genetic algorithm is used to find starting values for a gradient based variable metric algorithm. Because of the multimodality of the log-likelihood function, some of the estimation rounds may end up in different local maximum points, thereby enabling the researcher to build models not only based on the global maximum point but also on the local ones. The estimated models can be conveniently examined with the summary and plot methods. For evaluating their adequacy, **uGMAR** utilizes quantile residual diagnostics in the framework presented in Kalliovirta (2012), including graphical diagnostics as well as Kalliovirta's (2012) diagnostic tests that take into account uncertainty about the true parameter value. Following Kalliovirta *et al.* (2015) and Meitz *et al.* (forthcoming), forecasting is based on a Monte Carlo simulation method.

I illustrate the use of **uGMAR** with the monthly U.S. interest rate spread between the 10-year and 1-year Treasury rates. Illustrative examples are given for unconstrained and constrained estimation, examining the properties of the estimates and the estimated model, testing hypotheses regarding the parameters, evaluating the adequacy of the model with quantile residual diagnos-

tics, creating a GSMAR model object with specific parameter values, simulating observations from a GSMAR process, and forecasting future values of a GSMAR process. In addition to illustrating the use of **uGMAR**, I discuss the problem of model selection.

1.5.3 Structural Gaussian mixture vector autoregressive model with application to the asymmetric effects of monetary policy shocks

Tracing out the effects of an economic shock is a major task in econometrics. A popular approach is to consider a set of key variables and utilize a structural vector autoregressive (SVAR) or structural vector error correction (SVEC) model for the purpose. They have well established theoretical grounds (see Kilian and Lütkepohl, 2017, and the references therein) and are accommodated by many popular statistical software packages. Linear SVAR and SVEC models are not, however, suitable for modelling series in which the underlying data generating dynamics are nonlinear or the shocks have different effects in different states of the economy. Models capable of capturing such features include mixture models, such as the mixture vector autoregressive model (Fong *et al.*, 2007), the mixture periodic vector autoregressive model (Bentarzi and Djeddou, 2014), the Gaussian mixture vector autoregressive (GMVAR) model (Kalliovirta *et al.*, 2016), and the logit mixture vector autoregressive model (Burgard *et al.*, 2019).

In the third essay, I introduce a structural version of the GMVAR model. In the structural GMVAR (SGMVAR) model of autoregressive order p, the regime-switching dynamics are endogenously determined by the full distribution of the previous p observations. At each time point, the greater the relative weighted likelihood of a regime is, the more likely the process is to generate an observation from it, which facilitates giving economic interpretations to the regimes. The specific formulation of the mixing weights also leads to attractive theoretical properties, such as ergodicity and fully known stationary distribution of $p + 1$ consecutive observations.

The effects of structural shocks depend on the initial values of the included variables and they are also allowed to vary according to the sign and size of the shock due to possible regime-switches. Consequently, the (generalized) impulse response functions reflect the prevailing macroeconomic conditions that are transmitted to the regime-switching probabilities through the level, variability, and temporal as well as contemporaneous dependence of the past observations. The impact matrix of the SGMVAR model is time-varying and constructed so that it captures the conditional heteroskedasticity of the reduced form error, thereby enabling standardization of the conditional variance of each structural shock to a constant. The initial effects of a constant-sized structural shock are, hence, amplified according to the conditional variance of the reduced form error, also reflecting the prevailing state of the economy.

I make use of the conditional heteroskedasticity of the reduced form error and show that the SGMVAR model generally identifies the structural shocks up to ordering and sign, but does not reveal which column of the impact matrix is related to which shock. Specifically, I obtain a local solution to the identification problem discussed in Section 1.4 by constraining the model so that entries of the impact matrix are not allowed to vary relative to the other entries in the same column but are allowed vary in magnitude and relative to the other columns. Since the

impact matrix is also subject to estimation error, I make use of the solution proposed by Lanne and Lütkepohl (2010) and Lanne *et al.* (2010) and derive general conditions for formal, global identification of any subset of the shocks. This leads to flexible identification conditions, and some of the constraints are also testable. For impulse response analysis, I utilize the generalized impulse response function (Koop, Pesaran, and Potter, 1996) and propose a Monte Carlo algorithm for its estimation by making use of the known stationary distribution of the SGMVAR process.

In an empirical application, I study asymmetries in the expected effects of monetary policy shocks in the U.S. using quarterly data covering the period from 1954Q3 to 2021Q4. The data includes four variables: cyclical component of the logarithmized real GDP (separated from the trend with the one-sided Hodrick-Prescott filter), the log-difference of GDP implicit price deflator, the log-difference of producer price index, and an interest rate variable. My policy variable is the interest rate variable, which is the effective federal funds rate from 1954Q3 to 2008Q2 and the Wu and Xia (2016) shadow rate from 2008Q3 to 2021Q4. I identify the monetary policy shock by assuming that it moves output and producer price inflation to the opposite direction from the interest rate variable at impact, and that it does not move inflation at impact. The Wald test accepts the zero constraint with the p-value 0.92. Two of the other shocks are assumed to move inflation at impact and one of them is assumed to move output and producer price inflation to the opposite directions.

My SGMVAR model identifies two regimes: a stable inflation regime and an unstable inflation regime. The unstable inflation regime is characterized by high or volatile inflation, and it mainly prevails in the 1970's, early 1980's, during the Financial crisis, and in the COVID-19 crisis from 2020Q3 onwards. The stable inflation regime, in turn, is characterized by moderate inflation, and it prevails when the stable inflation regime does not. I find the effects of the monetary policy shock relatively symmetric in the unstable inflation regime, as it rarely causes a switch to the stable inflation regime. A contractionary (expansionary) monetary policy shock appears to first increase (decrease) inflation after which the inflation significantly decreases (increases) for several years. The strong contraction (expansion) in the cyclical component of the GDP lasts for roughly three years and is followed by a small short-term expansion (contraction) before the response decays to zero.

In the stable inflation regime, the (generalized) impulse responses are strongly asymmetric with respect to the sign and size of the monetary policy shock as well as to the initial state of the economy. A contractionary shock causes, on average, roughly a three-year hump-shaped contraction of the GDP, but it also seems to increase inflation by driving the economy towards the unstable inflation regime. A small expansionary shock does not move prices much on average, but a large expansionary shock often drives the economy towards the unstable inflation regime and propagates high and persistent inflation. High inflation is followed by a significant monetary policy tightening and persistent contraction of the GDP after the initial expansion. On average, the real effects of the monetary policy shock are found somewhat stronger in the stable inflation regime than in the unstable inflation regime.

1.5.4 Gaussian and Student's *t* mixture vector autoregressive model with application to the asymmetric effects of monetary policy shocks in the Euro area

Mixture autoregressive models are useful for modelling series in which the data generating dynamics vary in time. The GMAR model (Kalliovirta *et al.*, 2015), the StMAR model (Meitz *et al.*, forthcoming), and the G-StMAR model (Virolainen, forthcoming, also Chapter 2 of this dissertation) constitute an appealing family of univariate mixture autoregressive models based on Gaussian and Student's t distributions. Kalliovirta *et al.* (2016) introduced a multivariate version of the GMAR model, while Chapter 4 of this dissertation proposes a structural version of this multivariate model. This essay complements the rest of this family by introducing the multivariate counterparts of the StMAR model and the G-StMAR model, including their structural versions. The accompanying R package **gmvarkit** (Virolainen, 2018a) provides a comprehensive set of tools for maximum likelihood estimation and other numerical analysis of the introduced models.

I introduce a new mixture vector autoregressive model that accommodates conditionally homoskedastic linear Gaussian vector autoregressions (VAR) and conditionally heteroskedastic linear Student's t VARs as its mixture components. The model, referred to as the G-StMVAR model, is obtained from definition (1.3.1) in Section 1.3 with $d \geq 2$ and by assuming that the first $M_1 \leq M$ mixture components are Gaussian VARs and the rest $M - M_1$ mixture components are Student's t VARs. If there are only Gaussian regimes, the GMVAR model of Kalliovirta *et al.* (2016) is obtained. If there are only Student's t regimes, the multivariate counterpart of the StMAR model of Meitz *et al.* (forthcoming) is obtained, which I refer to as the StMVAR model. In the presence of Gaussian and Student's t regimes, the multivariate counterpart of the G-StMAR model of Virolainen (forthcoming) is obtained.

Both types of mixture components have the same form for the conditional mean, a linear function of the preceding p observations, $\phi_{m,0} + \sum_{i=1}^{p} A_{m,i} y_{t-i}$ for the regime m in Equation (1.3.1). But the conditional covariance matrices $\Omega_{m,t}$ in (1.3.1) are different. The linear Gaussian VARs have constant conditional covariance matrices, i.e., $\Omega_{m,t} = \Omega_m$ for $m = 1, ..., M_1$. The conditional covariance matrices of the linear Student's t VARs, in turn, consist of a constant covariance matrix that is multiplied by a time-varying scalar that depends on the quadratic form of the previous p observations. That is, $\Omega_{m,t} = \omega_{m,t}\Omega_m$ for $m = M_1 + 1, ..., M$, where $\omega_{m,t}$ is a time-varying scalar. In this sense, the conditional covariance matrix is of ARCH type. But since it is just a time-varying scalar multiplying the constant covariance matrix, it is not as general as the conventional multivariate ARCH process that allows the entries of the conditional covariance matrix to vary relative to each other (e.g., Lütkepohl, 2005, Section 16.3). The specific formulation of the conditional covariance matrix is, nonetheless, convenient for establishing stationary properties similar to the linear Gaussian VARs. My specification of the conditional covariance matrix is also parsimonious, as it only depends on the degrees of freedom and the autoregressive parameters (in addition to the parameters in the constant covariance matrix). This is particularly advantegeous in the context of mixture VARs, as the large number

of parameters may often be problem even without an ARCH component.

For a pth order model, the mixing weights depend on the full distribution of the previous p observations, and they are constructed so that the greater the relative weighted likelihood of a regime is, the more likely the process is to generate an observation from it. Similarly to the other models in the family, this facilitates associating statistical characteristics and giving economic interpretations to the regimes. The specific formulation of the mixing weights also leads to attractive practical and theoretical properties such as ergodicity and full knowledge of the stationary distribution of $p + 1$ consecutive observations. By making use of stationarity and ergodicity, I show that the maximum likelihood estimator of a stationary G-StMVAR model is strongly consistent, and therefore, it has the conventional limiting distribution under the conventional high level conditions. In contrast to the GMVAR model, my model is able to capture excess kurtosis and conditional heteroskedasticity within the regimes.

It turns out that the G-StMVAR model is a limiting case of the StMVAR model with the degrees of freedom parameters of some of the regimes tending to infinity. The GMVAR model is obtained if the degrees of freedom parameters of all the regimes tend to infinity. Hence, if a StMVAR model is fitted to a series generated by a process in which some of the regimes are linear Gaussian VARs, the degrees of freedom parameters of these regimes are (asymptotically) expected to get large estimates. In empirical applications, the numbers of Gaussian and Student's t regimes can, therefore, be selected by first finding a suitable StMVAR model. Then, if some of the regimes obtain a large degrees of freedom parameter estimate, they should be accommodated by switching to the appropriate G-StMVAR model. As opposed to a StMVAR model with very large degrees of freedom parameter estimates, the G-StMVAR model avoids the numerical problems caused by their weak identification. These results are analogous to the univariate models discussed in Chapter 2.

In addition to the reduced form model, I propose a structural version of the G-StMVAR model that generalizes the SGMVAR model introduced in Chapter 4 to accommodate conditionally heteroskedastic Student's t regimes. The SG-StMVAR model incorporates a time-varying impact matrix that varies according to the conditional variance of the reduced form error. As a consequence of a single (time-varying) impact matrix, identification of the shocks requires that the error term covariance matrices are simultaneously diagonalized in all regimes. Together with a constant normalization of the structural error's conditional covariance matrix, this condition generally leads to uniquely identified shocks up to ordering and sign. Hence, as long as one is willing to assume a single (time-varying) impact matrix, its columns characterize the estimated impact effects of the shocks, but it is not revealed which column is related to which shock. Because the impact matrix is also subject to estimation error, further constraints may be needed for labelling the shocks. The identification conditions are the same as for the SGMVAR model, however.

The empirical application in the fourth essay studies asymmetries in the expected effects of the monetary policy shock in the Euro area and considers a monthly data covering the period from January 1999 to December 2021. The data includes four variables: cyclical component of the logarithmized industrial production index (separated from the trend with the linear projection

filter proposed by Hamilton, 2018, and referred to as the output gap hereafter), the log-difference of the harmonized consumer price index, the log-difference of Brent crude oil price, and an interest rate variable. My policy variable is the interest rate variable, which is the Euro overnight index average from January 1999 to October 2008 and the Wu and Xia (2016) shadow rate from November 2008 to December 2021.

I fit a two-regime StMVAR model to the series, but because none of the degrees of freedom parameter estimates is large, I do not consider incorporating Gaussian regimes. The StMVAR model identifies two regimes: a low-growth regime and a high-growth regime. The low-growth regime is characterized by negative (but volatile) output gap, and it mainly prevails after the collapse of Lehman Brothers in the Financial crisis but obtains large mixing weights also during and before the early 2000's recession. The high-growth regime is characterized by positive output gap and it mainly dominates before the Financial crisis. I identify the monetary policy shock by assuming that it moves the output gap and the interest rate variable to the opposite directions at impact, and that it does not move inflation nor oil price inflation at impact. The Wald test accepts the zero constraints individually and jointly at all the conventional levels of significance. Two of the other three shocks are assumed to move inflation at impact and one of them is assumed to move oil price inflation at impact.

I find strong asymmetries with respect to the initial state of the economy and the sign of the shock, but asymmetries with respect to the size of the shock are weak. The real effects are less enduring for an expansionary shock than for a contractionary shock. Particularly in the high-growth regime, a contractionary shock persistently drives the economy towards the low-growth regime, which translates to a very persistent decrease in the output gap. The inflationary effects of the monetary policy shock are stronger in the high-growth regime than in the low-growth regime, and in the latter the price level does not move much on average.

Bibliography

Bec F., Rahbek A., Shephard N. (2008). "The ACR Model: A Multivariate Dynamic Mixture Autoregression." *Oxford Bulletin of Economics and Statistics*, **70**(5), 583–618.

Bentarzi M., Djeddou L. (2014). "On Mixture Periodic Vector Autoregressive Models." *Communications in Statistics - Simulation and Computation*, **43**(10), 2325–2352.

Burgard J., Neuenkirch M., Nöckel M. (2019). "State-Dependent Transmission of Monetary Policy in the Euro Area." *Journal of Money, Credit and Banking*, **51**(7), 2053–2070.

Dorsey R., Mayer W. (1995). "Genetic algorithms for estimation problems with multiple optima, nondifferentiability, and other irregular features." *Journal of Business and Economic Statistics*, **13**(1), 53–66.

Fong P., Li W., Yau C., Wong C. (2007). "On a mixture vector autoregressive model." *The Canadian Journal of Statistics*, **35**(1), 135–150.

Glasbey C. (2001). "Non-linear autoregressive time series with multivariate Gaussian mixtures as marginal distributions." *Journal of Royal Statistical Society: Series C*, **50**(2), 143–154.

Hamilton J. D. (2018). "WHY YOU SHOULD NEVER USE THE HODRICK-PRESCOTT FILTER." *The Review of Economics and Statistics*, **100**(5), 831–843.

Kalliovirta L. (2012). "Misspecification tests based on quantile residuals." *The Econometrics Journal*, **15**(2), 358–393.

Kalliovirta L., Meitz M., Saikkonen P. (2015). "A Gaussian Mixture Autoregressive Model for Univariate Time Series." *Journal of Time Series Analysis*, **36**(2), 247–266.

Kalliovirta L., Meitz M., Saikkonen P. (2016). "Gaussian mixture vector autoregression." *Journal of Econometrics*, **192**(2), 465–498.

Kilian L., Lütkepohl H. (2017). *Structural Vector Autoregressive Analysis*. 1st edition. Cambridge University Press, Cambridge.

Koop G., Pesaran M., Potter S. (1996). "Impulse response analysis in nonlinear multivariate models." *Journal of Econometrics*, **74**(1), 119–147.

Lanne M., Lütkepohl H. (2010). "Structural Vector Autoregressions With Nonnormal Residuals." *Journal of Business & Economic Statistics*, **28**(1), 159–168.

Lanne M., Lütkepohl H., Maciejowsla K. (2010). "Structural vector autoregressions with Markov switching." *Journal of Economic Dynamics and Control*, **34**(2), 121–131.

Lanne M., Saikkonen P. (2003). "Modeling the U.S. Short-Term Interest Rate by Mixture Autoregressive Processes." *Journal of Financial Econometrics*, **1**(1), 96–125.

Le N., Martin R., Raftery A. (1996). "Modeling Flat Stretches, Bursts, and Outliers in Time Series Using Mixture Transition Distribution Models." *Journal of the American Statistical Association*, **91**(436), 1504–1515.

Lütkepohl H. (2005). *New Introduction to Multiple Time Series Analysis*. 1st edition. Springer, Berlin.

Meitz M., Preve D., Saikkonen P. (2018). ***StMAR Toolbox**: A MATLAB Toolbox for Student's t Mixture Autoregressive Models*.

Meitz M., Preve D., Saikkonen P. (forthcoming). "A mixture autoregressive model based on Student's t-distribution." *Communications in Statistics - Theory and Methods*.

Nash J. (1990). *Compact Numerical Methods for Computers. Linear Algebra and Function Minimization*. 2nd edition. Adam Hilger, Bristol and New York.

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ramey V. A. (2016). "Macroeconomic Shocks and Their Propagation." In JB Taylor, H Uhlig (eds.), *Handbook of Macroeconomics*, volume 2, chapter 2. Elsevier Science B.V.

Sarno L., Thornton D. (2003). "The dynamic relationship between the federal funds rate and the Treasury bill rate: An empirical investigation." *Journal of Banking & Finance*, **27**(6), 1079–1110.

Sims A. (1980). "Macroeconomics and Reality." *Econometrica*, **48**(1), 1–48.

Virolainen S. (2018a). ***gmvarkit**: Estimate Gaussian and Student's t Mixture Vector Autoregressive Models*. R package version 2.0.3 available at CRAN: `https://CRAN.R-project.org/package=gmvarkit`.

Virolainen S. (2018b). ***uGMAR**: Estimate Univariate Gaussian and Student's t Mixture Autoregressive Models*. R package version 3.4.2 available at CRAN: `https://CRAN.R-project.org/package=uGMAR`.

BIBLIOGRAPHY

Virolainen S. (forthcoming). "A mixture autoregressive model based on Gaussian and Student's t-distributions." *Studies in Nonlinear Dynamics & Econometrics*.

Wong C., Chan W., Kam P. (2009). "A Student's t-mixture autoregressive model with applications to heavy-tailed financial data." *Biometrika*, **96**(3), 751–760.

Wong C., Li W. (2000). "On a mixture autoregressive model." *Journal of the Royal Statistical Society*, **62**(1), 95–115.

Wong C., Li W. (2001a). "On a logistic mixture autoregressive model." *Biometrika*, **88**(3), 833–846.

Wong C., Li W. (2001b). "On a Mixture Autoregressive Conditional Heteroskedastic Model." *Journal of the American Statistical Association*, **96**(455), 982–995.

Wu J., Xia F. (2016). "Measuring the Macroeconomic Impact of Monetary Policy at the Zero Lower Bound." *Journal of Money, Credit and Banking*, **48**(2-3), 253–291.

Yule G. U. (1927). "On a Method of Investigating Periodicities in Disturbed Series, with Special Reference to Wolfer's Sunspot Numbers." *Philosophical Transactions of the Royal Society London. Series A*, **226**, 267–298.

BIBLIOGRAPHY

Chapter 2

A mixture autoregressive model based on Gaussian and Student's *t*-distributions[1]

2.1 Introduction

Recently, Kalliovirta, Meitz, and Saikkonen (2015) introduced a mixture autoregressive model based on Gaussian distribution with very attractive features. The Gaussian mixture autoregressive (GMAR) model has linear Gaussian autoregressions as its component models and mixing weights that, for a pth order model, depend on the full distribution of the p past observations. The specific formulation of the mixing weights leads to ergodicity and full knowledge of the stationary distribution of $p + 1$ consecutive observations. Moreover, it allows regime switches to depend on the level, variability, and temporal dependence of the past observations.

Meitz, Preve, and Saikkonen (forthcoming) proposed a mixture autoregressive model closely related to the GMAR model but based on Student's t-distribution. The Student's t mixture autoregressive (StMAR) model has linear Student's t autoregressions as its component models and mixing weights constructed analogously to the GMAR model, leading to similar theoretical and practical properties. The linear Student's t autoregressions have the same form for the conditional mean as the linear Gaussian autoregressions (a linear function of the past observations) but different conditional variance. In particular, the conditional variances of the linear Student's t autoregressions depend on quadratic forms of past observations, whereas in the Gaussian case the conditional variances of the component models are constants. Utilization of the t-distribution does hence not only allow the StMAR model to account for larger kurtosis than the GMAR model but also stronger forms of conditional heteroskedasticity.

In this essay, I propose a generalization of the GMAR and StMAR models. The G-StMAR model accommodates both linear Gaussian autoregressions and linear Student's t autoregressions as its component models, and its mixing weights are constructed analogously to the GMAR and StMAR models, leading to similar attractive features. It thus enables to model series which

[1] This essay has been accepted to the journal Studies in Nonlinear Dynamics & Econometrics (Virolainen, forthcoming).

consist of regimes with time varying conditional variance and excess kurtosis as well as regimes with constant conditional variance and zero excess kurtosis. It turns out that the G-StMAR model is a limiting case of a StMAR model with the t-distributions of some regimes tending to normal distributions as the degrees of freedom parameters tend to infinity. As opposed to the limiting StMAR model, the advantage of the G-StMAR model is that it removes the redundant degrees of freedom parameters from the model and is free from numerical problems induced by weak identification of very large degrees of freedom parameters.

I demonstrate the usefulness of the G-StMAR model in an empirical application to the monthly U.S. interest rate spread between the 3-month Treasury bill (TB) rate and the effective federal funds (FF) rate. My G-StMAR model identifies three regimes for the spread, with a GMAR type regime mainly appearing after the Financial crisis in 2008 when the zero lower bound limits movements of the spread. The remaining regimes are of the StMAR type, one accommodating eras of low mean and high variability and the other high mean and moderate variability. The former StMAR type regime dominates often when the market possibly anticipates decreases in the FF rate or has increased preference for safety, whereas the latter one mostly prevails when the Fed is arguably not expected to significantly decrease the FF rate target. My findings are consistent with Sarno and Thornton (2003) who found that the FF rate seems to adjust to the TB rate, supporting the hypothesis that the market anticipates movements of the FF rate, moving the TB rate, and hence the spread, in advance.

The rest of this chapter is organized as follows. Section 2.2 first introduces the component processes of the G-StMAR model and then proceeds to define the G-StMAR model and discusses its theoretical properties. Section 2.3 discusses maximum likelihood (ML) estimation of the model parameters and establishes the asymptotic properties of the ML estimator. It is, in particular, discussed how the accompanying R package **uGMAR** (Virolainen, 2018) estimates the model parameters in practice with a two-phase procedure. Section 2.4 describes a simple model selection procedure and discusses numerical consequences of very large degrees of freedom parameter estimates. Section 2.5 presents the empirical application to the interest rate spread and Section 2.6 concludes. Appendix 2.A gives details on the estimation procedure employed by **uGMAR**, Appendix 2.B gives the density functions and some properties of multivariate Gaussian and Student's t-distributions, and Appendix 2.C provides proofs for the stated theorems.

Throughout this chapter, I use the following notation. I write $\boldsymbol{x} = (x_1, ..., x_n)$ for the column vector $\boldsymbol{x}$ where the components x_i may be either scalars or (column) vectors. The notation $\boldsymbol{x} \sim n_d(\boldsymbol{\mu}, \boldsymbol{\Gamma})$ signifies that the random vector $\boldsymbol{x}$ has a d-dimensional Gaussian distribution with mean $\boldsymbol{\mu}$ and (positive definite) covariance matrix $\boldsymbol{\Gamma}$. Similarly, $\boldsymbol{x} \sim t_d(\boldsymbol{\mu}, \boldsymbol{\Gamma}, \nu)$ signifies that $\boldsymbol{x}$ has a d-dimensional t-distribution with mean $\boldsymbol{\mu}$, (positive definite) covariance matrix $\boldsymbol{\Gamma}$, and degrees of freedom ν (assumed to satisfy $\nu > 2$). The vectorization operator vec stacks columns of a matrix on top of each other and, ι_d is the d dimensional vector $(1, 0, ..., 0)$, I_d signifies the identity matrix of dimension d, and $\otimes$ denotes the Kronecker product. Moreover, $\mathbf{1}_d$ and $\mathbf{0}_d$ denote d dimensional vectors of ones and zeros, respectively.

2.2 Models

I consider mixture autoregressive models in which each observation is generated by a mixture component that is randomly selected according to the probabilities given by the mixing weights. The mixture components are either (linear) conditionally homoskedastic Gaussian autoregressions as in the GMAR model (Kalliovirta *et al.*, 2015) or conditionally heteroskedastic Student's t autoregressions as in the StMAR model (Meitz *et al.*, forthcoming). The mixing weights are functions of the past observations constructed in a way that, for a pth order model, leads to ergodicity and full knowledge of the stationary distribution $p+1$ consecutive observations. Moreover, as the mixing weights depend on the full distribution of the past p observations, they allow regime switches to depend on the level, variability, kurtosis, and temporal dependence of the past observations. In this section, I first introduce the component processes of the G-StMAR model and then proceed to define of the G-StMAR model and discuss its properties.

2.2.1 Linear Gaussian and Student's *t* autoregressions

To develop theory and notation, consider first the component processes of the G-StMAR model. For a linear pth order Gaussian or Student t autoregression z_t, we have

$$z_t = \varphi_0 + \sum_{i=1}^{p} \varphi_i z_{t-i} + \sigma_t \varepsilon_t, \quad \varepsilon_t \sim \text{IID}(0,1), \tag{2.2.1}$$

where $\sigma_t > 0$, $\varphi_0 \in \mathbb{R}$, and the autoregressive (AR) parameter $\boldsymbol{\varphi} = (\varphi_1, ..., \varphi_p)$ satisfies the stationarity condition $\boldsymbol{\varphi} \in \mathbb{S}^p$ where

$$\mathbb{S}^p = \{(\varphi_1, ..., \varphi_p) \in \mathbb{R}^p : 1 - \sum_{i=1}^{p} \varphi_i z^i \neq 0 \text{ for } |z| \leq 1\}. \tag{2.2.2}$$

In the case of Gaussian autoregression, the distribution of the error term ε_t is standard normal and σ_t is a constant σ for all t. Denoting $\boldsymbol{z}_t = (z_t, ..., z_{t-p+1})$ and $\mu = \text{E}[z_t]$, $\gamma_j = \text{Cov}(z_t, z_{t-j})$, and $\boldsymbol{\gamma}_p = (\gamma_1, ..., \gamma_p)$, it is well know that the stationary solution to (2.2.1) for the Gaussian autoregression satisfies

$$\boldsymbol{z}_t \sim n_p(\mu \mathbf{1}_p, \boldsymbol{\Gamma}_p), \tag{2.2.3}$$

$$(z_t, \boldsymbol{z}_{t-1}) \sim n_{p+1}(\mu \mathbf{1}_{p+1}, \boldsymbol{\Gamma}_{p+1}), \tag{2.2.4}$$

$$z_t \mid \boldsymbol{z}_{t-1} \sim n_1(\mu + \boldsymbol{\gamma}_p' \boldsymbol{\Gamma}_p^{-1}(\boldsymbol{z}_{t-1} - \mu \mathbf{1}_p), \gamma_0 - \boldsymbol{\gamma}_p' \boldsymbol{\Gamma}_p^{-1} \boldsymbol{\gamma}_p) = n_1(\varphi_0 + \boldsymbol{\varphi}' \boldsymbol{z}_{t-1}, \sigma^2), \tag{2.2.5}$$

where $\mu = \varphi_0/(1 - \boldsymbol{\varphi}' \mathbf{1}_p)$ $\boldsymbol{\gamma}_p = \boldsymbol{\Gamma}_p \boldsymbol{\varphi}$, and the covariance matrices $\boldsymbol{\Gamma}_p$ and $\boldsymbol{\Gamma}_{p+1}$ are Toeplitz matrices given as (see, e.g., Lütkepohl, 2005, Equation (2.1.39))

$$vec(\boldsymbol{\Gamma}_p) = (I_{p^2} - (\Phi \otimes \Phi))^{-1} \iota_{p^2} \sigma^2, \;\; \Phi = \begin{bmatrix} \varphi_1 \cdots \varphi_{p-1} & \varphi_p \\ I_{p-1} & \mathbf{0}_{p-1} \end{bmatrix}, \;\; \boldsymbol{\Gamma}_{p+1} = \begin{bmatrix} \gamma_0 & \boldsymbol{\gamma}_p' \\ \boldsymbol{\gamma}_p & \boldsymbol{\Gamma}_p \end{bmatrix}. \tag{2.2.6}$$

Using the same notation as in (2.2.3)-(2.2.5) for $\boldsymbol{z}_{t-1}$, μ, and $\boldsymbol{\Gamma}_p$, the Student's t autoregressions utilized by Meitz *et al.* (forthcoming) (which have also appeared at least in Spanos, 1994, and Heracleous and Spanos, 2006) are obtained by letting $\varepsilon_t \sim t_1(0, 1, \nu + p)$ with $\nu > 2$ in (2.2.1) and defining

$$\sigma_t^2 = \frac{\nu - 2 + (\boldsymbol{z}_{t-1} - \mu\mathbf{1}_p)'\boldsymbol{\Gamma}_p^{-1}(\boldsymbol{z}_{t-1} - \mu\mathbf{1}_p)}{\nu - 2 + p}\sigma^2. \tag{2.2.7}$$

This definition (which requires the stationarity condition of the AR parameter) guarantees stationarity of the Student's t autoregressions. Distributional properties of such stationary Student's t autoregressions are similar to the Gaussian case, in particular (Meitz *et al.*, forthcoming, Theorem 1),

$$\boldsymbol{z}_t \sim t_p(\mu\mathbf{1}_p, \boldsymbol{\Gamma}_p, \nu), \tag{2.2.8}$$

$$(z_t, \boldsymbol{z}_{t-1}) \sim t_{p+1}(\mu\mathbf{1}_{p+1}, \boldsymbol{\Gamma}_{p+1}, \nu), \tag{2.2.9}$$

$$z_t \mid \boldsymbol{z}_{t-1} \sim t_1(\varphi_0 + \boldsymbol{\varphi}'\boldsymbol{z}_{t-1}, \sigma_t^2, \nu + p). \tag{2.2.10}$$

The aforementioned properties of the component processes are essential in the following discussions and will be exploited implicitly. Gaussian component processes of the G-StMAR model are referred to as *GMAR type* and Student's t component processes as *StMAR type* since they are identical to the component processes of the GMAR model (Kalliovirta *et al.*, 2015) and the StMAR model (Meitz *et al.*, forthcoming), respectively.

2.2.2 Gaussian and Student's *t* mixture autoregressive model

Let y_t ($t = 1, 2, ...$) be the real valued time series of interest, and let $\mathcal{F}_{t-1}$ denote the σ-algebra generated by the random variables $\{y_{t-j}, j > 0\}$. For a G-StMAR model with M mixture components and autoregressive order p, we have

$$y_t = \sum_{m=1}^{M} s_{m,t}(\mu_{m,t} + \sigma_{m,t}\varepsilon_{m,t}), \quad \varepsilon_{m,t} \sim \text{IID}(0, 1), \tag{2.2.11}$$

$$\mu_{m,t} = \varphi_{m,0} + \sum_{i=1}^{p} \varphi_{m,i}y_{t-i}, \quad m = 1, ..., M, \tag{2.2.12}$$

where $\sigma_{m,t} > 0$ are $\mathcal{F}_{t-1}$-measurable, $\varepsilon_{m,t}$ are independent of $\mathcal{F}_{t-1}$, $\varphi_{m,0} \in \mathbb{R}$, $\boldsymbol{\varphi}_m \in \mathbb{S}^p$ (the set $\mathbb{S}^p$ is defined in (2.2.2)), and $s_{1,t}, ..., s_{M,t}$ are unobservable regime variables such that for each t, exactly one of them takes the value one and the others take the value zero. Given the past of y_t, $(s_{1,t}, ..., s_{M,t})$ and $\varepsilon_{m,t}$ are assumed to be conditionally independent, and the conditional probability for regime m occurring at the time t is expressed in terms of the mixing weights $\alpha_{m,t} \equiv \Pr(s_{m,t} = 1 | \mathcal{F}_{t-1})$ that satisfy $\sum_{m=1}^{M} \alpha_{m,t} = 1$ (for all $t = 1, 2, ...$). Each observation is thus generated by a linear autoregression corresponding to some (unobserved)

mixture component m, which is selected randomly according to the probabilities determined by the mixing weights.

The first M_1 mixture components are (linear) Gaussian autoregressions and the rest $M_2 \equiv M - M_1$ are (linear) Student's t autoregressions. Regarding Equation (2.2.11), this means that for $m = 1, ..., M_1$, the terms $\varepsilon_{m,t}$ have standard normal distributions and the variances $\sigma^2_{m,t}$ are constants σ^2_m. For $m = M_1 + 1, ..., M$, the terms $\varepsilon_{m,t}$ follow the t-distribution $t_1(0, 1, \nu_m + p)$ and the variances $\sigma^2_{m,t}$ are as in Equation (2.2.7) except that $\boldsymbol{z}_{t-1}$ is replaced with $\boldsymbol{y}_{t-1} = (y_{t-1}, ..., y_{t-p})$ and the regime specific parameters $\varphi_{m,0}, \boldsymbol{\varphi}_m, \sigma^2_m, \nu_m$ are used to define μ and $\boldsymbol{\Gamma}_p$ therein. The component specific conditional means $\mu_{m,t}$ are defined by Equation (2.2.12) for all the components.

Based on the above specifications, the conditional density function of a G-StMAR model with autoregressive order p is given as

$$f\left(y_t|\mathcal{F}_{t-1}\right) = \sum_{m=1}^{M_1} \alpha_{m,t} n_1(y_t; \mu_{m,t}, \sigma^2_m) + \sum_{m=M_1+1}^{M} \alpha_{m,t} t_1\left(y_t; \mu_{m,t}, \sigma^2_{m,t}, \nu_m + p\right), \quad (2.2.13)$$

where the conditional densities $n_1(y_t; \mu_{m,t}, \sigma^2_m)$ and $t_1\left(y_t; \mu_{m,t}, \sigma^2_{m,t}, \nu_m + p\right)$ are obtained from the properties of the component processes (using the regime specific parameters). The form of the Student's t density function in (2.2.13) is given in Appendix 2.B. The G-StMAR model adds to the class of mixture models introduced by Le, Martin, and Raftery (1996) and further developed by Wong and Li (2000, 2001a,b), Glasbey (2001), Lanne and Saikkonen (2003), and Wong, Chan, and Kam (2009), to name a few.

In order to specify the mixing weights $\alpha_{m,t}$ in (2.2.13), I first define the following function for notational convenience. Let

$$d_m(\boldsymbol{y}; \mu_m \mathbf{1}_p, \boldsymbol{\Gamma}_m, \nu_m) = \begin{cases} n_p(\boldsymbol{y}; \mu_m \mathbf{1}_p, \boldsymbol{\Gamma}_m), & \text{when } m \leq M_1, \\ t_p(\boldsymbol{y}; \mu_m \mathbf{1}_p, \boldsymbol{\Gamma}_m, \nu_m), & \text{when } m > M_1, \end{cases} \quad (2.2.14)$$

where the p-dimensional densities $n_p(\boldsymbol{y}; \mu_m \mathbf{1}_p, \boldsymbol{\Gamma}_m)$ and $t_p(\boldsymbol{y}; \mu_m \mathbf{1}_p, \boldsymbol{\Gamma}_m, \nu_m)$ correspond to the stationary distribution of the mth component process (given in Equations (2.2.3) and (2.2.8)). Denoting $\boldsymbol{y}_{t-1} = (y_{t-1}, ..., y_{t-p})$, the mixing weights of the G-StMAR model are defined as

$$\alpha_{m,t} = \frac{\alpha_m d_m(\boldsymbol{y}_{t-1}; \mu_m \mathbf{1}_p, \boldsymbol{\Gamma}_m, \nu_m)}{\sum_{n=1}^{M} \alpha_n d_n(\boldsymbol{y}_{t-1}; \mu_n \mathbf{1}_p, \boldsymbol{\Gamma}_n, \nu_n)}, \quad (2.2.15)$$

where the parameters $\alpha_1, ..., \alpha_M$ satisfy $\sum_{m=1}^{M} \alpha_m = 1$. The mixing weights are thus weighted ratios of the stationary densities of the component processes corresponding to the p previous observations. This specific definition of the mixing weights is appealing, as it states that the higher the relative weighted likelihood of a regime is, the more likely the process is to generate an observation from it. Moreover, it allows the regime-switching probabilities to depend on the level, variability, kurtosis, and temporal dependence of the past observations. This is a convenient feature for forecasting, and it also facilitates associating statistical characteristics

and economic interpretations to the regimes. It turns out that this formulation of the mixing weights also leads to attractive theoretical properties such as fully known stationary distribution of realizations $(y_t, ..., y_{t-h})$, $h = 0, 1, ..., p$, and ergodicity of the process. These theoretical properties are formally stated in Theorem 2.1 below.

Before stating the theorem, a few notational conventions are provided. I collect the parameters of the G-StMAR model to a $(M(p+3) + M_2 - 1) \times 1$ vector $\boldsymbol{\theta} \equiv (\boldsymbol{\theta}^-, \boldsymbol{\nu})$ where $\boldsymbol{\theta}^- = (\boldsymbol{\vartheta}_1, ..., \boldsymbol{\vartheta}_M, \alpha_1, ..., \alpha_{M-1})$, $\boldsymbol{\vartheta}_m = (\varphi_{m,0}, \boldsymbol{\varphi}_m, \sigma_m^2)$, $\boldsymbol{\varphi}_m = (\varphi_{m,1}, ..., \varphi_{m,p})$, $m = 1, ..., M$, and $\boldsymbol{\nu} = (\nu_{M_1+1}, ..., \nu_M)$. The parameter α_M is omitted because it is obtained from the restriction $\sum_{m=1}^{M} \alpha_m = 1$. The parameter space for the G-StMAR model is

$$\boldsymbol{\Theta} = \left\{ \boldsymbol{\theta} \in \mathbb{R}^{M(2+p)} \times (0,1)^{M-1} \times (2,\infty)^{M_2} : \boldsymbol{\varphi}_m \in \mathbb{S}^p, \sigma_m^2 > 0, \text{ for all } m = 1, ..., M \right\} \tag{2.2.16}$$

where the restriction $\nu_m > 2$ $(m = M_1 + 1, ..., M)$ is made to ensure existence of finite second moments and the set $\mathbb{S}^p$ is as in (2.2.2). A G-StMAR model with autoregressive order p, M_1 GMAR type regimes, and M_2 StMAR type regimes is referred to as the G-StMAR(p, M_1, M_2) model, whenever clarity of the presentation requires.

Theorem 2.1. *Consider the G-StMAR process y_t generated by (2.2.13) and (2.2.15) with $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Then $\boldsymbol{y}_t = (y_t, ..., y_{t-p+1})$ $(t = 1, 2, ...)$ is a Markov chain on $\mathbb{R}^p$ with a stationary distribution characterized by the density*

$$f(\boldsymbol{y}; \boldsymbol{\theta}) = \sum_{m=1}^{M_1} \alpha_m n_p(\boldsymbol{y}; \mu_m \mathbf{1}_p, \boldsymbol{\Gamma}_m) + \sum_{m=M_1+1}^{M} \alpha_m t_p(\boldsymbol{y}; \mu_m \mathbf{1}_p, \boldsymbol{\Gamma}_m, \nu_m). \tag{2.2.17}$$

Moreover, $\boldsymbol{y}_t$ is ergodic.

The stationary distribution of $\boldsymbol{y}_t$ is a mixture of p-dimensional normal and t-distributions with constant mixing weights α_m. By the well known properties of the normal and the t-distribution, all its moments lower than $\min\{\nu_{M_1+1}, ..., \nu_M\}$ exist and are finite. Moreover, as shown in the proof of Theorem 2.1, for any $h = 0, 1, ..., p$, the marginal stationary distribution of the vector $(y_t, .., y_{t-h})$ is also a mixture of normal and t-distributions. This gives the parameters α_m an interpretation as the unconditional probabilities for the observation y_t being generated from the mth component process. Similarly to the GMAR and the StMAR process, the mean, variance, and first p autocovariances of y_t are thus

$$\mathrm{E}[y_t] \equiv \mu_y = \sum_{m=1}^{M} \alpha_m \mu_m, \quad \gamma_j \equiv \sum_{m=1}^{M} \alpha_m \gamma_{m,j} + \sum_{m=1}^{M} \alpha_m (\mu_m - \mu_y)^2, \quad j = 0, 1, ..., p, \tag{2.2.18}$$

where $\gamma_{m,j}$ is the jth autocovariance of the mth component process.

The conditional mean and variance of the G-StMAR process are obtained from the definition of the model as $\mathrm{E}[y_t|\mathcal{F}_{t-1}] = \sum_{m=1}^{M} \alpha_{m,t}\mu_{m,t}$ and

$$\mathrm{Var}(y_t|\mathcal{F}_{t-1}) = \sum_{m=1}^{M_1} \alpha_{m,t}\sigma_m^2 + \sum_{m=M_1+1}^{M} \alpha_{m,t}\sigma_{m,t}^2 + \sum_{m=1}^{M} \alpha_{m,t}\left(\mu_{m,t} - \sum_{n=1}^{M} \alpha_{n,t}\mu_{n,t}\right)^2. \quad (2.2.19)$$

The conditional mean shares a common form with the GMAR model and StMAR model but differs from them in the definition of the mixing weights. The conditional variance includes three components; the first one is related to the conditional variances of the GMAR type components and the second one to the StMAR type components, whereas the third term encapsulates heteroskedasticity caused by variations in the conditional mean.

Notice that the GMAR model (Kalliovirta *et al.*, 2015) can be obtained as a special of the G-StMAR model by setting $M_1 = M$ and $M_2 = 0$, and similarly the StMAR model (Meitz *et al.*, forthcoming) is obtained by setting $M_1 = 0$ and $M_2 = M$. One simply needs to drop the corresponding terms from the formulas, and all the definitions and results stated in this and in the next section also hold for to the GMAR and StMAR models individually. However, some theory developed for the GMAR model, such as geometric ergodicity (Kalliovirta *et al.*, 2015, Theorem A.1), has not been established for the StMAR and G-StMAR models. The GMAR model also requires less (currently) unverified assumptions than the StMAR and G-StMAR models for concluding asymptotic normality of the maximum likelihood estimator (see Kalliovirta *et al.*, 2015, Section 2, Meitz *et al.*, forthcoming, Theorem 3, and Theorem 2.2 of this essay).

2.3 Estimation

Parameters of the G-StMAR model can be estimated with the method of maximum likelihood (ML). Because the stationary distribution of the process is known, the exact log-likelihood function can be used. Suppose the observed time series is $y_{-p+1}, ..., y_0, y_1, ..., y_T$ and that the initial values are stationary. Then the log-likelihood function of the G-StMAR model takes the form

$$L(\boldsymbol{\theta}) = \log\left(\sum_{m=1}^{M_1} \alpha_m n_p(\boldsymbol{y}_0; \mu_m \mathbf{1}_p, \boldsymbol{\Gamma}_m) + \sum_{m=M_1+1}^{M} \alpha_m t_p(\boldsymbol{y}_0; \mu_m \mathbf{1}_p, \boldsymbol{\Gamma}_m, \nu_m)\right) + \sum_{t=1}^{T} l_t(\boldsymbol{\theta}), \quad (2.3.1)$$

where

$$l_t(\boldsymbol{\theta}) = \log\left(\sum_{m=1}^{M_1} \alpha_{m,t} n_1(y_t; \mu_{m,t}, \sigma_m^2) + \sum_{m=M_1+1}^{M} \alpha_{m,t} t_1\left(y_t; \mu_{m,t}, \sigma_{m,t}^2, \nu_m + p\right)\right), \quad (2.3.2)$$

and the density functions $n_d(\cdot;\cdot)$ and $t_d(\cdot;\cdot)$ follow the notation described in Section 2.2.2. If stationarity of the initial values seems unreasonable, one can condition on the initial values

by dropping the first term on the right side of (2.3.1) and base the estimation on the resulting conditional log-likelihood function.

In what follows, I assume estimation based on the conditional log-likelihood function $L_T^{(c)}(\boldsymbol{\theta}) = T^{-1}\sum_{t=1}^{T} l_t(\boldsymbol{\theta})$, i.e., that the ML estimator $\hat{\boldsymbol{\theta}}_T$ maximizes $L_T^{(c)}(\boldsymbol{\theta})$. I have scaled the conditional log-likelihood function with the sample size T so that the notation is consistent with the referred literature.

To investigate the asymptotic properties of the ML estimator $\hat{\boldsymbol{\theta}}_T$, the parameter space $\boldsymbol{\Theta}$ given in (2.2.16) needs to be restricted in a way that guarantees identification of the parameters. This amounts to requiring that components of the G-StMAR model cannot be "relabelled" so that one ends up with the same model with different parameter vector; that is,

$$\begin{aligned}&\alpha_1 > \cdots > \alpha_{M_1} > 0,\ \alpha_{M_1+1} > \cdots > \alpha_M > 0, \text{ and } \boldsymbol{\vartheta}_i = \boldsymbol{\vartheta}_j \text{ only if at least one of the}\\ &\text{conditions } (1)\ 1 \leq i = j \leq M,\ (2)\ i \leq M_1 < j,\ (3)\ i, j > M_1 \text{ and } \nu_i \neq \nu_j \text{ is satisfied.}\end{aligned} \tag{2.3.3}$$

The restrictions required to establish asymptotic properties of the ML estimator are summarized in the following assumption.

Assumption 2.1. *The true parameter value $\boldsymbol{\theta}_0$ is an interior point of $\bar{\boldsymbol{\Theta}}$ which is a compact subset of $\{\boldsymbol{\theta} \in \boldsymbol{\Theta} : (2.3.3)$ holds$\}$.*

Asymptotic properties of the ML estimator under the conventional high-level conditions are stated in the following theorem (which is similar to Theorem 3 in Meitz *et al.*, forthcoming, on the ML estimator of the StMAR model). Denote $\mathcal{I}(\boldsymbol{\theta}) = \mathrm{E}\big[\frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\big]$ and $\mathcal{J}(\boldsymbol{\theta}) = \mathrm{E}\big[\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\big]$.

Theorem 2.2. *Suppose that y_t are generated by the stationary and ergodic G-StMAR process of Theorem 2.1 and that Assumption 2.1 holds. Then $\hat{\boldsymbol{\theta}}_T$ is strongly consistent, i.e., $\hat{\boldsymbol{\theta}}_T \to \boldsymbol{\theta}_0$ almost surely. Suppose further that (i) $T^{1/2}\frac{\partial}{\partial \boldsymbol{\theta}} L_T^{(c)}(\boldsymbol{\theta}_0) \xrightarrow{d} N(0, \mathcal{I}(\boldsymbol{\theta}_0))$ with $\mathcal{I}(\boldsymbol{\theta}_0)$ finite and positive definite, (ii) $\mathcal{J}(\boldsymbol{\theta}_0) = -\mathcal{I}(\boldsymbol{\theta}_0)$, and (iii) $E\big[\sup_{\boldsymbol{\theta}\in\bar{\boldsymbol{\Theta}}_0} \big|\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\big|\big] < \infty$ for some $\bar{\boldsymbol{\Theta}}_0$, compact convex set contained in the interior of $\bar{\boldsymbol{\Theta}}$ that has $\boldsymbol{\theta}_0$ as an interior point. Then $T^{1/2}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \xrightarrow{d} N(0, -\mathcal{J}(\boldsymbol{\theta}_0)^{-1})$.*

If one is willing to assume validity of the conditions (i)-(iii) of Theorem 2.2, the ML estimator $\hat{\boldsymbol{\theta}}_T$ has the conventional limiting distribution, implying that approximative standard errors for the estimates are obtained as usual. Moreover, standard likelihood based tests are applicable as long as the orders M_1 and M_2 are correctly specified. If M_1 or M_2 is chosen too large, some of the parameters are not identified causing the result of Theorem 2.2 to break down. This particularly happens when one tests for the number of regimes as the null hypothesis would imply that some regime is reduced from the model[2] (see the related discussion in Kalliovirta *et al.*, 2015, Section 3.3.2). Similar caution also applies for testing whether a regime is of the GMAR

[2] Meitz and Saikkonen (2021) have, however, recently developed such tests for mixture models with Gaussian conditional densities.

type against the alternative that it is of the StMAR type, as under the null hypothesis $\nu_m = \infty$ for the StMAR type regime m being tested, violating Assumption 2.1. Numerical consequences of the weak identification of very large degrees of freedom parameters are briefly discussed in Section 2.4.

2.3.1 Two-phase maximum likelihood estimation

Finding the ML estimates amounts to maximizing the log-likelihood function (2.3.1) over the high dimensional parameter space (2.2.16) satisfying several constraints. Due to the complexity of the log-likelihood function, finding an analytical solution is infeasible, so numerical optimization methods are required. The EM algorithm (Redner and Walker, 1984) has been a popular choice for estimating mixture models (e.g. Wong and Li, 2000, 2001a,b, and Wong *et al.*, 2009), as it is suitable for problems where all the data relevant to estimation is not observed (for mixture models that is the origin of each observation; in this case, the random variables $s_{1,t}, ..., s_{M,t}$ in (2.2.11)). For the G-StMAR model the EM algorithm is not, however, particularly useful because in each maximization step one faces a new optimization problem that is not much simpler than the original one. This is because in the G-StMAR model the mixing weights also depend on the AR parameters (in a complex way). Conventional gradient based algorithms, on the other hand, tend to converge to some local maximum near the starting point, making them generally insufficient for maximizing multimodal objective functions such as (2.3.1) that require thorough exploration of the parameter space.

Several optimization algorithms capable of escaping from local maxima have been proposed for maximization of complicated multimodal objective functions. Such robust methods, which include simulated annealing and the genetic algorithm (see, e.g., Goffe, Ferrier, and Rogers, 1994 and Dorsey and Mayer, 1995), often perform well but they are computationally heavy and tend to converge slowly when near the global maximum point (see the discussion in Dorsey and Mayer, 1995, Section 3). Following Dorsey and Mayer (1995) (and Meitz, Preve, and Saikkonen, 2018, Meitz *et al.*, forthcoming), I hence suggest employing a hybrid estimation procedure where a genetic algorithm is used to find starting values for a gradient based method, which then often converges to a nearby local maximum or saddle point.

Even with the two-phase estimation procedure, parameters of the G-StMAR model can be challenging to estimate. I have therefore accompanied this essay with the CRAN distributed R package **uGMAR** (Virolainen, 2018) in which the genetic algorithm has been modified to improve its performance.[3] Brief descriptions of the employed genetic algorithm and its modifications are given in Appendix 2.A. After running the genetic algorithm, the estimation is finalized with a variable metric algorithm (Nash, 1990, Algorithm 21, implemented by R Core Team, 2022) using central difference approximation for the gradient of the log-likelihood function. Because of the presence of multiple local maxima, a (sometimes large) number of estimation rounds should be performed to obtain reliable results, for which **uGMAR** makes use of parallel computing to shorten the estimation time.

[3] In addition to the G-StMAR model, **uGMAR** also accommodates the GMAR and StMAR models.

2.4 Building a G-StMAR Model

In empirical applications, building a G-StMAR model amounts to finding a suitable autoregressive order p, the number of GMAR type regimes M_1, and the number of StMAR type regimes M_2. Different strategies for choosing the number of each type of regimes may be considered depending on the application. I propose a simple model selection procedure which takes advantage of the observation that the G-StMAR model is a limiting case of the StMAR model.[4]

It is easy to see that the linear Gaussian autoregression defined in Section 2.2.1 is obtained as a limiting case of the Student's t autoregression with the degrees of freedom parameter tending to infinity. As the mixing weights (2.2.15) are weighted ratios of the component process densities, it then follows that the G-StMAR(p, M_1, M_2) model is obtained as a limiting case of a StMAR(p, M) model with the parameters $\nu_1, ..., \nu_{M_1}$ limiting to infinity. Consequently, if a StMAR(p, M) model is fitted to data generated by a G-StMAR($p, M_1, M - M_1$) process, then asymptotically, the M_1 regimes of the fitted StMAR model are expected to get large degrees of freedom parameter estimates. I therefore suggest building a G-StMAR model by first finding a suitable StMAR model, and then estimating the appropriate G-StMAR model if the fitted StMAR model contains large degrees of freedom parameter estimates. A StMAR model can be specified, for example, by using information criteria together with quantile residual diagnostics (see, e.g., Kalliovirta, 2012).

Overly large degrees of freedom parameter estimates in a StMAR model are redundant, and their weak identification also causes several inconveniences in numerical analysis of the model. They lead to nearly numerically singular Hessian matrix of the log-likelihood function when evaluated at the estimate, making the approximate standard errors often unavailable. Weakly identified degrees of freedom parameters also cause inconvenience in quantile residual based model diagnostics. In particular, the quantile residual tests proposed by Kalliovirta (2012) require a positive definite approximation of the Hessian matrix (evaluated the ML estimate). The tests are thus not applicable for StMAR models with too large degrees of freedom parameter estimates, whereas they are for the corresponding G-StMAR models. Applicability of Kalliovirta's (2012) tests, which take into account the uncertainty caused by estimation of the parameters, might have consequences in model selection when sheer graphical analysis of the quantile residuals fails to reveal inadequacies. I demonstrate such a case in the empirical application.

2.5 Empirical application

I consider the monthly U.S. interest rate spread between the 3-month Treasury bill (TB) secondary market rate and the effective federal funds (FF) rate, covering the period from 1954VII to 2019VII (781 observations). The series is plotted in Figure 2.1 (top left) along with the 3-month TB and FF rates, and with the shaded areas indicating the periods of (NBER based) U.S. recessions. All the data were taken from the Federal Reserve Bank of St. Louis database.

[4] The definition of the StMAR(p, M) model is technically the same as of the G-StMAR($p, 0, M$) model.

Treasury bills are short-term pure discount bonds which are backed by the U.S. government and therefore generally considered to be almost free from default-risk. The effective federal funds rate is the averaged rate at which depository institutes loan federal funds to each other overnight. The overnight FF lending agreements are one of the most liquid financial asset, but unlike TBs, they are subject to a notable default-risk. The relationship between TB and FF rates has been studied, among others, by Simon (1990) and Sarno and Thornton (2003), while Kishor and Marfatia (2013) examine the relationship between TB and FF futures rate.

According to term structure theory, a long-term interest rate should reflect the current and expected future short-term rates, and also perceptions of risk and liquidity in the form of (possibly time-varying) premium. Simon (1990) studied the predictive power of the weekly spread between the 3-month TB and FF rates on the future levels of the FF rate in 1972-1987. He argued that the current and expected future FF rates affect the spread between the TB and FF rates through the repurchase agreement (repo) market[5] because repos are closely linked to the FF rate, and corporations with funds to invest can buy TBs alternatively to investing in consecutive overnight repos. TB rates are linked to the FF rates also because security dealers finance the bulk of their TB inventories in the repo market, which is closely tied to the FF market. Furthermore, when trust in solidity of the banking system weakens, the increased demand for safety lowers TB rates relatively to FF rates. Simon (1990) accounted for this by employing the spread between the 3-month Eurodollar time deposit[6] and TB rates as a risk premium for bank safety. He found that the spread between the 3-month TB and FF rate had significant predictive power on future levels of the FF rate in the volatile nonborrowed reserves operating period (late 1979 - late 1982) but less or none in the other subperiods.

Sarno and Thornton (2003) identified an error correction model (ECM) between the daily 3-month TB and effective FF rate (covering the period from 1974 to 1999) and showed that their ECM, which allows for asymmetries and nonlinearities, outperforms the alternative of a linear ECM. One of their main findings was that the FF rate (which is controlled by the Fed) seems to adjust to the TB rate and not vice versa, supporting the hypothesis that the market anticipates changes in the FF rate, moving the TB rate in advance. Moreover, it appears that the adjustment speed depends on the sign and size of the deviation from the long-run equilibrium. Sarno and Thornton (2003) argued that although there has been a number of procedural changes affecting predictability of the FF rate, their results implicate that the changes have been statistically unimportant. Furthermore, their robustness checks indicate that their findings on the adjustments from disequilibria also hold for monthly data. Variations and asymmetries in the adjustment speed, on the other hand, indicate that the dynamics of the spread between the TB and FF rates might fluctuate along with the level of the spread. This suggests that a mixture model, such as the G-StMAR model, which is able encapsulate such behaviour could be an appropriate choice of model.

[5] In a repo, the *borrower* sells a security to the *lender* and agrees to repurchase it in the future (often in the next day). Effectively, repos function similarly to collateralized loans. See Baklanova, Copeland, and McCaughrin (2015) for an overview of the U.S. repo market.

[6] Eurodollar time deposit is a U.S. dollar-denominated deposit at a bank outside the U.S. with a fixed maturity.

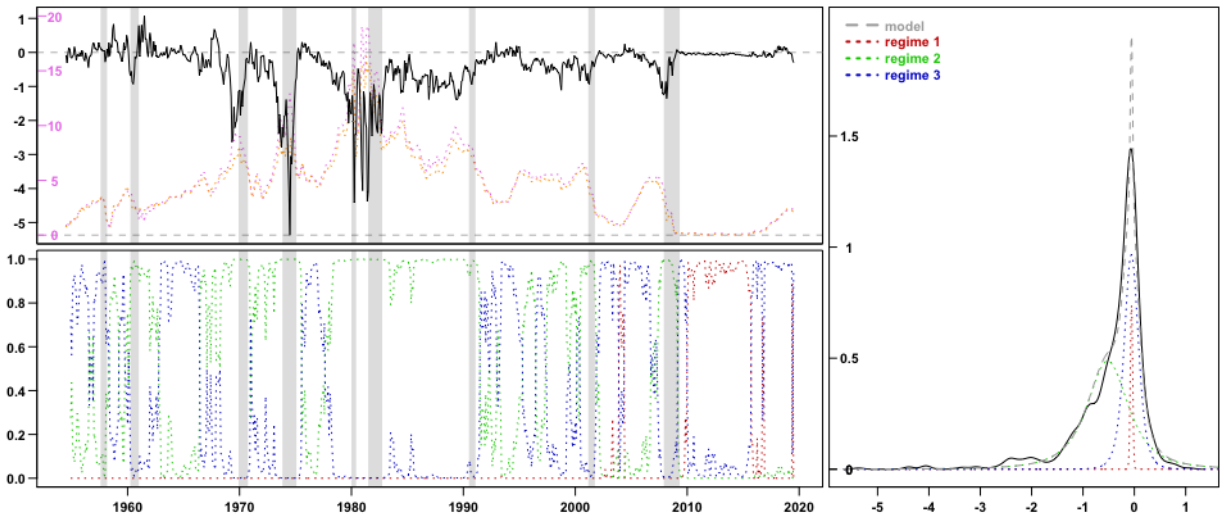

Figure 2.1: On top left, monthly U.S. 3-month Treasury bill secondary market rate minus effective federal funds rate (black solid line), the 3-month Treasury bill secondary market rate (orange dotted line), and the effective federal funds rate (violet dotted line). On bottom left, the mixing weights implied by the G-StMAR$(5, 1, 2)$ model fitted to the interest rate spread series. The shaded areas indicate the periods of (NBER based) U.S. recessions. On right, a Gaussian kernel density estimate of the interest rate spread (black solid line), the mixture density implied by the fitted G-StMAR$(5, 1, 2)$ model (grey dashed line), and the regime densities (blue, green, and red dotted lines).

Kishor and Marfatia (2013) argued that the results in Sarno and Thornton (2003) are not very surprising since the effective FF rate always tends to revert back to the FF target rate, and it does not incorporate markets expectations of the changes in the future FF rate. To get around that, they studied the relationship between the 3-month TB rate and the 1-month FF futures rate which does incorporate information about market's anticipations on the future FF rate. They fitted a linear ECM to a daily series from 1989 to 2008, and found that the TB rate and the FF futures rate both seem to move to correct a short-run disequilibrium.

Interestingly, the spread between the 3-month TB rate and the effective FF rate is most of the time (covered in my sample period) negative. Sarno and Thornton (2003) made a similar observation for their daily series and suggested that only a small fraction of the negative difference could be attributed to the low default-risk of TBs, but that a more plausible explanation is that the interest on TBs is exempt from some local and state taxes. As the smaller taxes have larger effect on paid net interest (relative to interest paid on federal funds) when the interest rates are higher, some movements of the spread could be partially caused by the differences in taxation.

2.5.1 Estimation and model selection

I employ the method of maximum likelihood based on the exact log-likelihood function for estimating the parameters of the considered models. Adequacy of the estimated models is examined using quantile residual diagnostics in the framework presented in Kalliovirta (2012). The quantile residuals of a correctly specified G-StMAR model are asymptotically independent with standard normal distributions (Kalliovirta, 2012, Lemma 2.1), so they can be used for graphical analysis in a similar fashion to conventional Pearson's residuals. In addition to graphical analysis of the quantile residuals, I perform Kalliovirta's (2012) asymptotic tests (which take into account the uncertainty caused by estimation of the parameters) for testing normality, autocorrelation, and conditional heteroskedasticity of the quantile residuals. The estimation, quantile residual diagnostics, and other numerical analysis of the models is conducted using the R package **uGMAR** (Virolainen, 2018) which is available through the CRAN repository.[7] **uGMAR** estimates the model parameters using the two-phase procedure described in Section 2.3.1.

Following the model selection procedure described in Section 2.4, I started by finding a suitable StMAR model. First, I estimated the StMAR(p, M) model with one mixture component, $M = 1$, and autoregressive orders $p = 1, ..., 24$ and found that the order $p = 6$ yields the largest likelihood. Adequacy of the StMAR$(6, 1)$ model was clearly rejected by the quantile residual tests (see Table 2), so I estimated the StMAR(p, M) models with orders $p = 1, ..., 6$ and $M = 2, 3$. The order $(p, M) = (5, 2)$ minimized the Schwarz-Bayesian (BIC) and the Hannan-Quinn (HQIC) information criteria, whereas the Akaike's information criterion (AIC) was minimized by the order $(p, M) = (5, 3)$. Inappropriate estimates extremely near the border of the stationarity region were discarded as they are not solutions of interest (but maximize the likelihood for rather a technical reason, see Chapter 3), so in such cases the next-largest local maximum of the log-likelihood function was considered instead. In both the StMAR$(5, 2)$ and the StMAR$(5, 3)$ model, a very large degrees of freedom estimate for one regime was obtained (approximately 99000 and 95000, respectively), so I estimated the corresponding G-StMAR$(5, 1, 1)$ and G-StMAR$(5, 1, 2)$ models. Removing the weakly identified degrees of freedom parameters by switching to the G-StMAR models enabled me to compute approximate standard errors of the estimates and to calculate Kalliovirta's (2012) test statistics (see Section 2.4). The values of the information criteria are reported in Table 2.2 and the parameter estimates of the G-StMAR models are reported in Table 2.1 with the approximate standard errors for the estimates in parentheses.

Estimates regarding the GMAR type regime are quite similar for the two G-StMAR models, and their standard errors are relatively large. This is because for both of the models the GMAR type regime mainly occurs in the period of near-zero interest rates after 2008 and there are hence only a few observations from that regime (regime 1 in Figure 2.1, bottom left, which displays the mixing weights of the G-StMAR$(5, 1, 2)$ model; the mixing weights of the G-StMAR$(5, 1, 1)$ model are not shown). The three zeros in the variance parameter estimates (and in their standard

[7] There is also MATLAB code available for the StMAR model in the form of **StMAR Toolbox** by Meitz *et al.* (2018).

errors) signify that the estimates (and their standard errors) round to zero in three digits accuracy[8], implying that the GMAR type regime exhibits very low variability (conditionally and unconditionally). The small mixing weight parameter estimates, interpreted as the unconditional probability for the GMAR type regime occurring, reflect the observation that eras of such a low variability have been rare in the sample period. Also, a remarkably large standard error for the second regime's variance parameter sticks out for both of the models. Examination of the profile log-likelihood functions (not shown) does not, however, reveal anything notable.

Since the AR parameter estimates for the G-StMAR$(5, 1, 2)$ model are somewhat similar in all regimes, I estimated a StMAR$(5, 3)$ model with the AR parameters restricted to be the same in all regimes, allowing for changes in the level, variability, and kurtosis only. The degrees of freedom estimate for one regime was very large (approximately 97000), so I estimated the corresponding restricted G-StMAR model which I refer to as the G-StMAR$(5, 1, 2)^r$ model. The parameter estimates of this model are also presented in Table 2.1 with the related statistics, and the values of the information criteria in Table 2.2. The standard errors of the AR parameters are notably smaller than in the non-restricted models because the AR parameters are common for all the regimes.

Figure 2.2 presents the time series, normal quantile-quantile plot, the sample autocorrelation function of the quantile residuals, and the sample autocorrelation function of the squared quantile residuals for the G-StMAR models presented in Table 2.1. Graphical analysis of the quantile residuals does not show significant signs of inadequacy for any of the models. A slightly too fat lower tail in the quantile residuals' distributions and somewhat large, approximately 0.1, sample autocorrelation coefficient at lag 12 sticks out for each of the three models, however.

In order to further study adequacy of the models, I employed Kalliovirta's (2012) tests, and tested for normality, autocorrelation, and conditional heteroskedasticity of the quantile residuals, taking into account $1, 3, 6$, and 12 lags in the autocorrelation and heteroskedasticity tests. The p-values obtained from the tests are reported in Table 2.2. The normality test rejects for all the three models at 1% level of significance, possibly because of the fat lower tails in the quantile residuals' distributions. More interestingly, despite the similarities in the graphical analysis, the autocorrelation tests unambiguously reject adequacy of the G-StMAR$(5, 1, 1)$ model, whereas the p-values are reasonable for the G-StMAR$(5, 1, 2)$ model which also passes the heteroskedasticity tests. The p-values for the autocorrelation tests are rather small also for the restricted G-StMAR$(5, 1, 2)^r$ model, which is preferred by the information criteria, showing some evidence of inadequacy. I therefore prefer the unrestricted G-StMAR$(5, 1, 2)$ model whose overall adequacy seems quite satisfactory. The fact that the restricted model has information criteria values superior to the unrestricted models, however, suggests that imposing the autocorrelation structure to be the same for all regimes would also be a reasonable modelling choice.[9]

[8] More accurate values for the ML estimate of σ_1^2 and its standard error are 3.237×10^{-4} and 6.884×10^{-5} for the G-StMAR$(5, 1, 1)$ model, 3.070×10^{-4} and 6.092×10^{-5} for the G-StMAR$(5, 1, 2)$ model, and 3.593×10^{-4} and 5.552×10^{-5} for the G-StMAR$(5, 1, 2)^r$ model, respectively.

[9] For comparison, I also estimated the GMAR(p, M) model with orders $p = 1, ..., 6$ and $M = 1, ..., 4$. The values of the information criteria were, however, found inferior to my G-StMAR models, with the GMAR$(3, 4)$ model minimizing BIC (-432) and the GMAR$(5, 4)$ model minimizing HQIC (-517) and AIC (-572).

	G-StMAR$(5,1,1)$		G-StMAR$(5,1,2)$		G-StMAR$(5,1,2)^r$	
$\varphi_{1,0}$	–0.011	(0.010)	–0.013	(0.009)	–0.007	(0.002)
$\varphi_{1,1}$	0.587	(0.129)	0.580	(0.124)	0.782	(0.037)
$\varphi_{1,2}$	–0.049	(0.168)	–0.079	(0.163)	–0.058	(0.050)
$\varphi_{1,3}$	0.041	(0.140)	0.042	(0.136)	0.134	(0.050)
$\varphi_{1,4}$	0.006	(0.142)	0.006	(0.141)	–0.040	(0.052)
$\varphi_{1,5}$	0.224	(0.128)	0.209	(0.132)	0.036	(0.042)
σ_1^2	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)
α_1	0.029	(0.021)	0.043	(0.035)	0.035	(0.025)
μ_1	–0.056		–0.055		–0.048	
$\gamma_{1,0}$	0.001		0.001		0.001	
$\varphi_{2,0}$	–0.009	(0.005)	–0.066	(0.025)	–0.079	(0.025)
$\varphi_{2,1}$	0.821	(0.040)	0.845	(0.055)		
$\varphi_{2,2}$	–0.051	(0.053)	–0.038	(0.076)		
$\varphi_{2,3}$	0.153	(0.053)	0.127	(0.075)		
$\varphi_{2,4}$	–0.052	(0.055)	–0.134	(0.077)		
$\varphi_{2,5}$	0.045	(0.042)	0.073	(0.058)		
σ_2^2	4.806	(18.779)	0.541	(2.052)	0.256	(0.374)
ν_2	2.007	(0.026)	2.196	(0.801)	2.499	(0.872)
α_2			0.592	(0.132)	0.600	(0.141)
μ_2	–0.110		–0.519		–0.541	
$\gamma_{2,0}$	24.449		2.109		0.802	
$\varphi_{3,0}$			–0.011	(0.005)	–0.011	(0.005)
$\varphi_{3,1}$			0.720	(0.069)		
$\varphi_{3,2}$			–0.082	(0.090)		
$\varphi_{3,3}$			0.151	(0.090)		
$\varphi_{3,4}$			0.087	(0.098)		
$\varphi_{3,5}$			–0.062	(0.085)		
σ_3^2			0.015	(0.011)	0.015	(0.013)
ν_3			4.320	(2.951)	4.778	(4.511)
μ_3			–0.059		–0.074	
$\gamma_{3,0}$			0.038		0.048	
μ_y	–0.108		–0.331		–0.353	
γ_0	23.744		1.313		0.552	
$L(\hat{\boldsymbol{\theta}})$	309.165		322.121		314.016	

Table 2.1: Maximum likelihood estimates of the G-StMAR$(5,1,1)$, the G-StMAR$(5,1,2)$, and the restricted G-StMAR$(5,1,2)^r$ model based on the exact log-likelihood function, with approximate standard errors for the estimates presented in the brackets. The statistics μ_m and $\gamma_{m,0}$, $m = 1,2,3$, are the stationary mean and variance of each regime, respectively. Likewise, the statistics μ_y and γ_0 are the stationary mean and variance of the process. The maximized log-likelihoods for each model are presented in the bottom row of the table.

	Normality	Autocorrelation				Cond. h.skedasticity				AIC	HQIC	BIC
Number of lags		1	3	6	12	1	3	6	12			
StMAR(6,1)	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	-538	-521	-496
G-StMAR$(5,1,1)$	**0.00**	**0.01**	**0.01**	0.01	**0.00**	0.46	0.38	0.23	**0.00**	-586	-558	-512
G-StMAR$(5,1,2)$	**0.00**	0.18	0.40	0.57	0.16	0.82	0.07	0.18	0.02	-594	-549	-478
G-StMAR$(5,1,2)^r$	**0.00**	0.02	0.07	0.13	0.03	0.68	0.07	0.24	0.03	-598	-571	-528

Table 2.2: The p-values obtained from the Kalliovirta's (2012) quantile residual tests, testing for normality, autocorrelation, and conditional heteroskedasticity of the quantile residuals. The p-values smaller than 0.01 are bolded. In order to improve size properties of the tests, I employed the simulation procedure proposed by Kalliovirta (2012) using samples of length 500000.

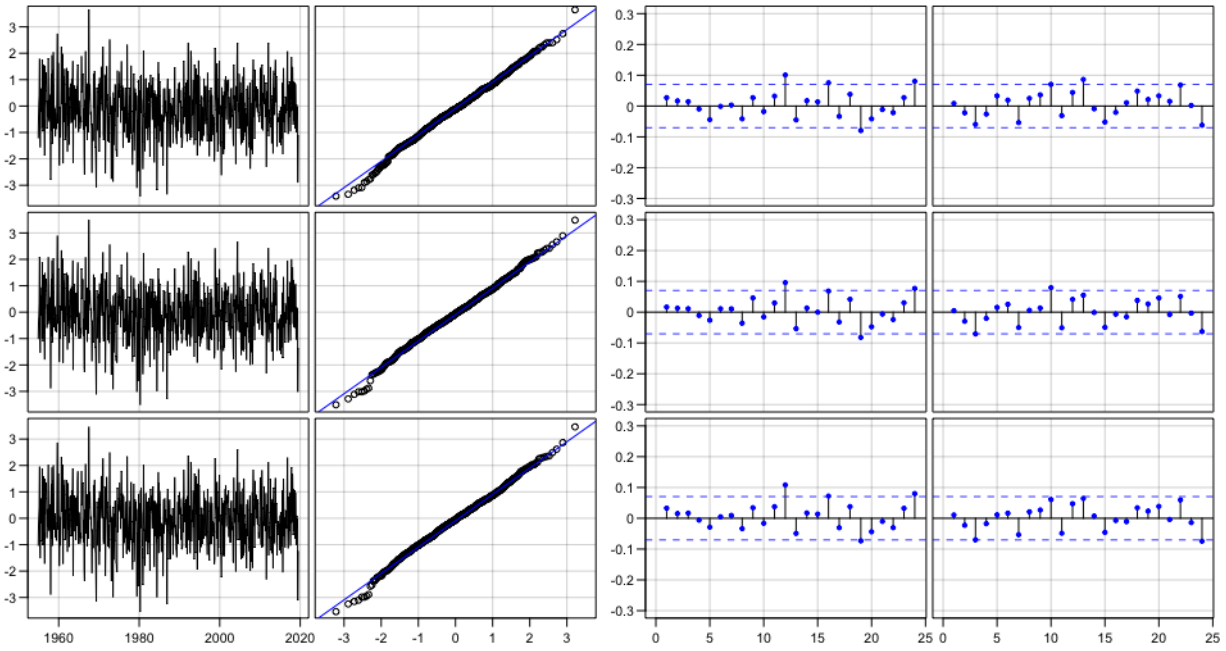

Figure 2.2: Graphical quantile residual diagnostics for the models presented in Table 2.1. The top row is for the G-StMAR$(5,1,1)$ model, the middle row is for the G-StMAR$(5,1,2)$ model, and the bottom row is for the G-StMAR$(5,1,2)^r$ model. The first column presents the time series, the second column the normal quantile-quantile plot, and third column the autocorrelation function of the quantile residuals. The fourth column presents the autocorrelation function of the squared quantile residuals. The blue solid line in the quantile-quantile plots displays the theoretical quantiles, and the blue dashed lines in the autocorrelation function plots are the 95% bounds $\pm 1.96/\sqrt{T}$ ($T = 776$ as the first p values are the initial values) for autocorrelations of an IID sequence, which are presented to give an approximate perception of the magnitude of the sample autocorrelations.

2.5.2 Discussion

My model selection procedure led to the (unrestricted) G-StMAR$(5, 1, 2)$ model which identifies three statistical regimes for the spread between the 3-month TB secondary market rate and the effective FF rate. The mixing weights of the model are presented in Figure 2.1 (bottom left) along with the interest rate spread series (top left). The GMAR type regime (red) dominates the period of near-zero interest rates occurring after 2008, where also the spread stays close to zero and exhibits very low variability. The second regime (green) identifies periods of high variability and low mean, spanning through most of the recessions, whereas the third regime (blue) often occurs[10] after the recessions when the spread moderately varies around zero. These characteristics of the regimes are also highlighted in Figure 2.1 (right) where a kernel density estimate of the spread (black solid line) is presented with the model implied density (grey dashed line) and the regime densities (red, green, and blue dotted lines; regime densities are multiplied by the mixing weight parameter estimates α_m, $m = 1, 2, 3$). The model implied density matches fairly well to the skewed distribution of the observations, but peakiness of the distribution seems a bit exaggerated and the lower tail is not fat enough.

Based on my G-StMAR$(5, 1, 2)$ model, the regime specific unconditional mean of the spread varies from the -0.06 %-units of the first (GMAR type) and third regime to the -0.52 %-units of the second regime, with each regime regularly occurring for several consecutive months. As the second regime dominates during most of the recessions, and also often occurs before the recessions when the interest rates are relatively high, it seems plausible that part of the larger negative mean is explained by expectations of a decrease in the near-future FF rate. The third regime, on the other hand, mostly occurs after the recessions when the interest rates seem relatively low, possibly indicating that the larger mean of the regime could be related to the lack of expected decreases in the FF rate. These findings are consistent with Sarno and Thornton (2003) who found that the FF rate corrects disequilibriums from the long-run relationship, supporting the hypothesis that market's anticipations in the future movements of the FF rate are reflected in the TB rate.

Sarno and Thornton (2003) also found that the adjustment speed of FF rate towards the long-run equilibrium depends on the sign and size of the deviation. Specifically, FF rate below the long-run trend or larger deviation implies faster adjustment, suggesting that too high values of the spread would be corrected faster than too low values. This might partially explain why the low mean second regime usually occurs when the interest rates are declining, but a rise in the FF rate is not always accompanied with a switch to the higher mean third regime. Another possibility is that market's predictions on the future movements of the FF rate are sometimes rather poor or a premium has an increased effect on the opposite direction. During the savings and loan crisis in 1980's and 1990's, increased preference for the safety of TBs would seem like a plausible partial explanation for the moderately negative spread despite of the mainly increasing FF rate from late 1986 to early 1989.

[10] By a regime occurring at a point of time, I mean that according to the estimated mixing weights, the process generated an observation from that regime with a probability close to one.

Overall, the three statistical regimes of my G-StMAR model identify three economic regimes, with the first regime dominating the period in which the movements of the interest rates are limited by the zero lower bound. The second regime arguably occurs often when the market anticipates decreases in the FF rate or possibly has increased preferences for the safety of the almost default-risk free TBs. The third regime seems to mostly occur at times when the Fed is arguably not expected to significantly decrease the FF rate target (because the recession has already passed and the interest rates are relatively low).

2.6 Summary

This essay introduced a mixture autoregressive model which is a combination of the Gaussian mixture autoregressive (GMAR) model (Kalliovirta *et al.*, 2015) and the Student's t mixture autoregressive (StMAR) model (Meitz *et al.*, forthcoming). This model, referred to as the G-StMAR model, has several attractive theoretical and practical properties that are analogous to those of the GMAR and StMAR model. In addition to discussing the properties, it was noted that estimating the parameters of the G-StMAR model can be challenging in practice. Following Dorsey and Mayer (1995) (and Meitz *et al.*, 2018, forthcoming), I proposed using a two-phase estimation procedure where a genetic algorithm is used to find starting values for a gradient based method and accompanied this essay with the R package **uGMAR** (Virolainen, 2018), which implements the two-phase estimation procedure with a modified version of a genetic algorithm.

I stated that the G-StMAR model is a limiting case of a StMAR model with some degrees of freedom parameters tending to infinity, and found that large degrees of freedom estimates in a StMAR model are not only redundant but also cause several inconveniences in numerical analysis of the model. In particular, weak identification of large degrees of freedom parameters was found to lead to numerically nearly singular approximation of the observed information matrix when evaluated at the estimate, making the approximate standard errors for the estimates and Kalliovirta's (2012) diagnostic tests often unavailable. Removing the redundant degrees of freedom parameters by switching to a G-StMAR model was concluded to obviate the problems.

As an empirical application, I considered the monthly U.S. interest rate spread between the 3-month Treasury bill rate and the effective federal funds rate. My G-StMAR model identified three regimes for the spread, with a switch from a StMAR type regime to a GMAR type regime arising from a switch in the economic regime, namely, to a regime where the zero lower bound limits the movements of the interest rates. The two StMAR type regimes accommodate eras of low mean and high variability and high mean and moderate variability. The first StMAR type regime arguably occurs often when the market anticipates decreases in the FF rate or possibly has increased preferences for safety, whereas the second one mostly occurs when the Fed is arguably not expected to significantly decrease the FF rate target. As opposed to modelling the series with a StMAR model containing an overly large degrees of freedom estimate, switching to the more parsimonious G-StMAR model allowed me to numerically compute approximate standard errors for the estimates, and moreover, to perform the Kalliovirta's (2012) quantile

residual tests, which turned out to be useful in the model selection.

2.6. SUMMARY

Bibliography

Baklanova V., Copeland A., McCaughrin R. (2015). "Reference Guide to U.S. Repo and Securities Lending Markets, Staff Report No. 740." *Technical report*, Federal Reserve Bank of New York.

Ding P. (2016). "On the Conditional Distribution of the Multivariate t Distribution." *The American Statistician*, **70**(3), 293–295.

Dorsey R., Mayer W. (1995). "Genetic algorithms for estimation problems with multiple optima, nondifferentiability, and other irregular features." *Journal of Business and Economic Statistics*, **13**(1), 53–66.

Glasbey C. (2001). "Non-linear autoregressive time series with multivariate Gaussian mixtures as marginal distributions." *Journal of Royal Statistical Society: Series C*, **50**(2), 143–154.

Goffe W., Ferrier G., Rogers J. (1994). "Global optimization of statistical functions with simulated annealing." *Journal of Econometrics*, **60**(1-2), 65–99.

Heracleous M., Spanos A. (2006). "Student's t dynamic linear regression: re-examining volatility modeling." In Terrell D., Fomby T.B. (eds.), *Econometric Analysis of Financial and Economic Time Series (Advances in Econometrics, Vol. 20 Part 1)*, pp. 289–319. Emerald Group Publishing Limited, Bingley.

Holzmann H., Munk A., Gneiting T. (2006). "Identifiability of finite mixtures of elliptical distributions." *Scandinavian Journal of Statistics*, **33**(4), 753–763.

Kalliovirta L. (2012). "Misspecification tests based on quantile residuals." *The Econometrics Journal*, **15**(2), 358–393.

Kalliovirta L., Meitz M., Saikkonen P. (2015). "A Gaussian Mixture Autoregressive Model for Univariate Time Series." *Journal of Time Series Analysis*, **36**(2), 247–266.

Kalliovirta L., Meitz M., Saikkonen P. (2016). "Gaussian mixture vector autoregression." *Journal of Econometrics*, **192**(2), 465–498.

Kishor N., Marfatia H. (2013). "Does federal funds futures rate contain information about the treasury bill rate?" *Applied Financial Economics*, **23**(16), 1311–1324.

Lanne M., Saikkonen P. (2003). "Modeling the U.S. Short-Term Interest Rate by Mixture Autoregressive Processes." *Journal of Financial Econometrics*, **1**(1), 96–125.

Le N., Martin R., Raftery A. (1996). "Modeling Flat Stretches, Bursts, and Outliers in Time Series Using Mixture Transition Distribution Models." *Journal of the American Statistical Association*, **91**(436), 1504–1515.

Lütkepohl H. (2005). *New Introduction to Multiple Time Series Analysis*. 1st edition. Springer, Berlin.

Meitz M., Preve D., Saikkonen P. (2018). ***StMAR Toolbox**: A MATLAB Toolbox for Student's t Mixture Autoregressive Models*.

Meitz M., Preve D., Saikkonen P. (forthcoming). "A mixture autoregressive model based on Student's t-distribution." *Communications in Statistics - Theory and Methods*.

Meitz M., Saikkonen P. (2021). "Testing for observation-dependent regime switching in mixture autoregressive models." *Journal of Econometrics*, **222**(1), 601–624.

Meyn S., Tweedie R. (2009). *Markov Chains and Stochastic Stability*. 2nd edition. Cambridge University Press, Cambridge.

Monahan J. (1984). "A note on enforcing stationarity in autoregressive-moving average models." *Biometrika*, **71**(2), 403–404.

Nash J. (1990). *Compact Numerical Methods for Computers. Linear Algebra and Function Minimization*. 2nd edition. Adam Hilger, Bristol and New York.

Newey W., McFadden D. (1994). "Large sample estimation and hyphothesis testing." In Eagle R.F., MacFadden D.L. (eds.), *Handbook of Econometrics*, volume 4, chapter 36. Elsevier Science B.V.

Patnaik L., Srinivas M. (1994). "Adaptive Probabilities of Crossover and Mutation in Genetic Algorithms." *Transactions on Systems, Man and Cybernetics*, **24**(4), 656–667.

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ranga Rao R. (1962). "Relations between Weak and Uniform Convergence of Measures with Applications." *The Annals of Mathematical Statistics*, **33**(2), 659–680.

Redner R., Walker H. (1984). "Mixture Densities, Maximum Likelihood and the Em Algorithm." *Society for Industrial and Applied Mathematics*, **26**(2), 195–239.

Sarno L., Thornton D. (2003). "The dynamic relationship between the federal funds rate and the Treasury bill rate: An empirical investigation." *Journal of Banking & Finance*, **27**(6), 1079–1110.

BIBLIOGRAPHY

Simon D. (1990). "Expectations and the Treasury Bill-Federal Funds Rate Spread over Recent Monetary Policy Regimes." *Journal of Finance*, **45**(2), 567–577.

Smith R., Dike B., Stegmann S. (1995). "Fitness inheritance in genetic algorithms." *Proceedings of the 1995 ACM symbosium on Applied Computing*, pp. 345–350.

Spanos A. (1994). "On modeling heteroskedasticity: the Student's t and elliptical linear regression models." *Econometric Theory*, **10**(2), 286–315.

Virolainen S. (2018). ***uGMAR**: Estimate Univariate Gaussian and Student's t Mixture Autoregressive Models*. R package version 3.4.2 available at CRAN: `https://CRAN.R-project.org/package=uGMAR`.

Virolainen S. (forthcoming). "A mixture autoregressive model based on Gaussian and Student's t-distributions." *Studies in Nonlinear Dynamics & Econometrics*.

Wong C., Chan W., Kam P. (2009). "A Student's t-mixture autoregressive model with applications to heavy-tailed financial data." *Biometrika*, **96**(3), 751–760.

Wong C., Li W. (2000). "On a mixture autoregressive model." *Journal of the Royal Statistical Society*, **62**(1), 95–115.

Wong C., Li W. (2001a). "On a logistic mixture autoregressive model." *Biometrika*, **88**(3), 833–846.

Wong C., Li W. (2001b). "On a Mixture Autoregressive Conditional Heteroskedastic Model." *Journal of the American Statistical Association*, **96**(455), 982–995.

Appendix 2.A Modified genetic algorithm

As discussed in Section 2.3.1, the accompanying R package **uGMAR** (Virolainen, 2018) employs a two-phase procedure for estimating the parameters of the G-StMAR model (and also of the GMAR model Kalliovirta *et al.*, 2015, and the StMAR model Meitz *et al.*, forthcoming). In the first phase, a genetic algorithm is used to find starting values for a gradient based variable metric algorithm (Nash, 1990, Algorithm 21) which then, in the second phase, often converges to a nearby local maximum or saddle point. In this appendix, it is first briefly described how my version of the genetic algorithm functions in general, and then the specific modifications made to enhance estimation of the G-StMAR model are discussed (for more detailed description of a genetic algorithm, see, e.g., Dorsey and Mayer, 1995).

In a genetic algorithm, an initial *population* that consists of different parameter vectors (that are often drawn at random) is first constructed. Then, the genetic algorithm operates iteratively so that in each iteration, referred to as *generation*, the current population consisting of candidate solutions goes through the phases of *selection*, *crossover*, and *mutation*. In the selection phase, parameter vectors are sampled with replacement from the current population to the *reproduction pool* according to probabilities that are based on their *fitness*, that is, on the related log-likelihoods. In the crossover phase, some of the parameter vectors in the reproduction pool are crossed over with each other, with the probabilities of experiencing crossover given by the *crossover rate*. Finally, some of the parameter vectors are mutated in the mutation phase, with the mutation probabilities given by the *mutation rate*. In my version of the genetic algorithm, mutation means that the mutating parameter vector is fully replaced with another parameter vector that is drawn at random (in Dorsey and Mayer, 1995, mutations are drawn for each scalar component of parameter vectors individually). The reproduction pool that has experienced crossovers and mutations is the new population, and the algorithm proceeds to the next generation, *evolving* towards the global maximum a generation after another.

Because the G-StMAR model can be challenging to estimate even with a robust estimation algorithm such as the genetic algorithm, I have made modifications to improve its performance. In particular, a slightly modified version of the individually adaptive crossover rate and mutation rate introduced by Patnaik and Srinivas (1994) is employed in order to force the subaverage solutions to disrupt while protecting the better ones.[11] The fitness inheritance proposed by Smith, Dike, and Stegmann (1995) is deployed to shorten the estimation time by cutting down the number computationally costly evaluations of the log-likelihood function. In order to enhance thorough exploration of the parameter space, the algorithm proposed by Monahan (1984) is used in some random mutations to generate parameter vectors near the boundary of the stationarity region. In the case of a premature convergence, most of the population is mutated so that exploration of the parameter space continues. Moreover, after a large number generations have been run, for faster convergence the random mutations will be targeted to a neighbourhood of the best-so-far parameter vector; I call these *smart mutations*.

[11] I modified the individually adaptive crossover rate to enforce a 40% minimum crossover rate for all individuals in the population

In addition to the modifications described above, I have made further adjustments to care for the special structure of the log-likelihood function. Specifically, the definition of the mixing weights (2.2.15) implies that if a regime has parameter values that fit poorly relative to the other regimes, the mixing weights drop to near zero. The surface of the log-likelihood function thus flattens in the related directions, meaning that the algorithm is unable to converge properly if the proposed parameter vectors do not pose a reasonable fit for all regimes. This problem of unidentified (or redundant) regimes often occurs when the number of mixture components is chosen too large, but it can be present even when the number of mixture components is chosen correctly. In **uGMAR**, I try to resolve this problem by penalizing parameter vectors containing redundant regimes with smaller probabilities to get chosen to the reproduction pool. Moreover, smart mutations are targeted only to the neighbourhood of parameter values that identify all the regimes. If such parameter vectors have not been found (after a large number of generations have been run), combining regimes from different parameter vectors is attempted along with random search.

Appendix 2.B Properties of multivariate Gaussian and Student's t-distribution

Denote a d-dimensional real valued vector by $\boldsymbol{y}$. It is well known that the density function of the d-dimensional multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Gamma}$ is

$$n_d(\boldsymbol{y};\boldsymbol{\mu},\boldsymbol{\Gamma}) = (2\pi)^{-d/2}\det(\boldsymbol{\Gamma})^{-1/2}\exp\left\{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})'\boldsymbol{\Gamma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\right\}. \tag{2.B.1}$$

Similarly to Meitz *et al.* (forthcoming) but differing from the standard form, I parametrize the Student's t-distribution using its covariance matrix as a parameter together with the mean and degrees of freedom. The density function of such a d-dimensional t-distribution with mean $\boldsymbol{\mu}$, covariance matrix $\boldsymbol{\Gamma}$, and $\nu > 2$ degrees of freedom is

$$t_d\left(\boldsymbol{y};\boldsymbol{\mu},\boldsymbol{\Gamma},\nu\right) = C_d(\nu)\text{det}(\boldsymbol{\Gamma})^{-1/2}\left(1+\frac{(\boldsymbol{y}-\boldsymbol{\mu})'\boldsymbol{\Gamma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})}{\nu-2}\right)^{-(d+\nu)/2}, \tag{2.B.2}$$

where

$$C_d(\nu) = \frac{\Gamma\left(\frac{d+\nu}{2}\right)}{\sqrt{\pi^d(\nu-2)^d}\Gamma\left(\frac{\nu}{2}\right)}, \tag{2.B.3}$$

and $\Gamma(\cdot)$ is the gamma function. I assume that the covariance matrix $\boldsymbol{\Gamma}$ is positive definite for both distributions.

Consider a partition $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2)$ of either a normally or t-distributed (with ν degrees of freedom) random vector $\boldsymbol{X}$ such that $\boldsymbol{X}_1$ has dimension $(d_1 \times 1)$ and $\boldsymbol{X}_2$ has dimension $(d_2 \times 1)$. Consider also a corresponding partition of the mean vector $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)$ and the

covariance matrix

$$\boldsymbol{\Gamma} = \begin{bmatrix} \boldsymbol{\Gamma}_{11} & \boldsymbol{\Gamma}_{12} \\ \boldsymbol{\Gamma}_{12}' & \boldsymbol{\Gamma}_{22} \end{bmatrix}, \tag{2.B.4}$$

where, for example, the dimension of $\boldsymbol{\Gamma}_{11}$ is $(d_1 \times d_1)$. Then in the case of normally distributed $\boldsymbol{X}$, $\boldsymbol{X}_1$ has the marginal distribution $n_{d_1}(\boldsymbol{\mu}_1, \boldsymbol{\Gamma}_{11})$ and $\boldsymbol{X}_2$ has the marginal distribution $n_{d_2}(\boldsymbol{\mu}_2, \boldsymbol{\Gamma}_{22})$. In the t-distributed case, the marginal distributions are $t_{d_1}(\boldsymbol{\mu}_1, \boldsymbol{\Gamma}_{11}, \nu)$ and $t_{d_2}(\boldsymbol{\mu}_2, \boldsymbol{\Gamma}_{22}, \nu)$, respectively (see, e.g. Ding, 2016, also in what follows).

In the normally distributed case, the conditional distribution of the random vector $\boldsymbol{X}_1$ given $\boldsymbol{X}_2 = \boldsymbol{x}_2$ is

$$\boldsymbol{X}_1 \mid (\boldsymbol{X}_2 = \boldsymbol{x}_2) \sim n_{d_1}(\boldsymbol{\mu}_{1|2}(\boldsymbol{x}_2), \boldsymbol{\Gamma}_{1|2}(\boldsymbol{x}_2)) \tag{2.B.5}$$

where

$$\boldsymbol{\mu}_{1|2}(\boldsymbol{x}_2) = \boldsymbol{\mu}_1 + \boldsymbol{\Gamma}_{12}\boldsymbol{\Gamma}_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2) \quad \text{and} \tag{2.B.6}$$

$$\boldsymbol{\Gamma}_{1|2}(\boldsymbol{x}_2) = \boldsymbol{\Gamma}_{11} - \boldsymbol{\Gamma}_{12}\boldsymbol{\Gamma}_{22}^{-1}\boldsymbol{\Gamma}_{12}'. \tag{2.B.7}$$

In the t-distributed case, the analogous conditional distribution is

$$\boldsymbol{X}_1 \mid (\boldsymbol{X}_2 = \boldsymbol{x}_2) \sim t_{d_1}(\boldsymbol{\mu}_{1|2}(\boldsymbol{x}_2), \boldsymbol{\Gamma}_{1|2}(\boldsymbol{x}_2), \nu + d_2), \tag{2.B.8}$$

where

$$\boldsymbol{\mu}_{1|2}(\boldsymbol{x}_2) = \boldsymbol{\mu}_1 + \boldsymbol{\Gamma}_{12}\boldsymbol{\Gamma}_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2) \quad \text{and}$$

$$\boldsymbol{\Gamma}_{1|2}(\boldsymbol{x}_2) = \frac{\nu - 2 + (\boldsymbol{x}_2 - \boldsymbol{\mu}_2)'\boldsymbol{\Gamma}_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2)}{\nu - 2 + d_2}(\boldsymbol{\Gamma}_{11} - \boldsymbol{\Gamma}_{12}\boldsymbol{\Gamma}_{22}^{-1}\boldsymbol{\Gamma}_{12}').$$

In particular, we have

$$n_d(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Gamma}) = n_{d_1}(\boldsymbol{x}_1; \boldsymbol{\mu}_{1|2}(\boldsymbol{x}_2), \boldsymbol{\Gamma}_{1|2}(\boldsymbol{x}_2)) n_{d_2}(\boldsymbol{x}_2; \boldsymbol{\mu}_2, \boldsymbol{\Gamma}_{22}) \quad \text{and} \tag{2.B.9}$$

$$t_d(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Gamma}, \nu) = t_{d_1}(\boldsymbol{x}_1; \boldsymbol{\mu}_{1|2}(\boldsymbol{x}_2), \boldsymbol{\Gamma}_{1|2}(\boldsymbol{x}_2), \nu + d_2) t_{d_2}(\boldsymbol{x}_2; \boldsymbol{\mu}_2, \boldsymbol{\Gamma}_{22}, \nu). \tag{2.B.10}$$

Appendix 2.C Proofs

2.C.1 Proof of Theorem 2.1

Suppose $\{y_t\}_{t=1}^{\infty}$ is a G-StMAR process. Then, the process $\boldsymbol{y}_t = (y_t, ..., y_{t-p+1})$ is clearly a Markov chain on $\mathbb{R}^p$. Let $\boldsymbol{y}_0 = (y_0, ..., y_{-p+1})$ be a random vector whose distribution is characterized by the density function

$$f(\boldsymbol{y}_0; \boldsymbol{\theta}) = \sum_{m=1}^{M_1} \alpha_m n_p(\boldsymbol{y}_0; \mu_m \mathbf{1}_p, \boldsymbol{\Gamma}_{m,p}) + \sum_{m=M_1+1}^{M} \alpha_m t_p(\boldsymbol{y}_0; \mu_m \mathbf{1}_p, \boldsymbol{\Gamma}_{m,p}, \nu_m). \tag{2.C.1}$$

Appendix

According to Equations (2.2.3)-(2.2.5), (2.2.8)-(2.2.10), (2.2.13), and (2.2.15), the density of the conditional distribution of y_1 given $\boldsymbol{y}_0$ is

$$\begin{aligned} f(y_1 \mid \boldsymbol{y}_0; \boldsymbol{\theta}) &= \sum_{m=1}^{M_1} \frac{\alpha_m n_p(\boldsymbol{y}_0; \mu_m \mathbf{1}_p, \boldsymbol{\Gamma}_{m,p})}{f(\boldsymbol{y}_0; \boldsymbol{\theta})} n_1(y_1; \mu_{m,1}, \sigma_m^2) \\ &+ \sum_{m=M_1+1}^{M} \frac{\alpha_m t_p(\boldsymbol{y}_0; \mu_m \mathbf{1}_p, \boldsymbol{\Gamma}_{m,p}, \nu_m)}{f(\boldsymbol{y}_0; \boldsymbol{\theta})} t_1(y_1; \mu_{m,1}, \sigma_{m,1}^2, \nu_m + p) \end{aligned} \tag{2.C.2}$$

$$\begin{aligned} &= \sum_{m=1}^{M_1} \frac{\alpha_m}{f(\boldsymbol{y}_0; \boldsymbol{\theta})} n_{p+1}((y_1, \boldsymbol{y}_0); \mu_m \mathbf{1}_{p+1}, \boldsymbol{\Gamma}_{m,p+1}) \\ &+ \sum_{m=M_1+1}^{M} \frac{\alpha_m}{f(\boldsymbol{y}_0; \boldsymbol{\theta})} t_{p+1}((y_1, \boldsymbol{y}_0); \mu_m \mathbf{1}_{p+1}, \boldsymbol{\Gamma}_{m,p+1}, \nu_m). \end{aligned} \tag{2.C.3}$$

The random vector $(y_1, \boldsymbol{y}_0)$ therefore has the density function

$$\begin{aligned} f((y_1, \boldsymbol{y}_0); \boldsymbol{\theta}) &= \sum_{m=1}^{M_1} \alpha_m n_{p+1}((y_1, \boldsymbol{y}_0); \mu_m \mathbf{1}_{p+1}, \boldsymbol{\Gamma}_{m,p+1}) \\ &+ \sum_{m=M_1+1}^{M} \alpha_m t_{p+1}((y_1, \boldsymbol{y}_0); \mu_m \mathbf{1}_{p+1}, \boldsymbol{\Gamma}_{m,p+1}, \nu_m). \end{aligned} \tag{2.C.4}$$

Using the properties of marginal densities of multivariate normal and t-distributions, by integrating y_{-p+1} out, the density of $\boldsymbol{y}_1$ is obtained as[12]

$$f(\boldsymbol{y}_1; \boldsymbol{\theta}) = \sum_{m=1}^{M_1} \alpha_m n_p(\boldsymbol{y}_1; \mu_m \mathbf{1}_p, \boldsymbol{\Gamma}_{m,p}) + \sum_{m=M_1+1}^{M} \alpha_m t_p(\boldsymbol{y}_1; \mu_m \mathbf{1}_p, \boldsymbol{\Gamma}_{m,p}, \nu_m). \tag{2.C.5}$$

Thus, the random vectors $\boldsymbol{y}_0$ and $\boldsymbol{y}_1$ are identically distributed. As the process $\{\boldsymbol{y}_t\}_{t=1}^{\infty}$ is a (time homogeneous) Markov chain, it follows that $\{\boldsymbol{y}_t\}_{t=1}^{\infty}$ has a stationary distribution $\pi_{\boldsymbol{y}}(\cdot)$ characterized by the density (Meyn and Tweedie, 2009, pp. 230-231)

$$f(\cdot; \boldsymbol{\theta}) = \sum_{m=1}^{M_1} \alpha_m n_p(\cdot; \mu_m \mathbf{1}_p, \boldsymbol{\Gamma}_{m,p}) + \sum_{m=M_1+1}^{M} \alpha_m t_p(\cdot; \mu_m \mathbf{1}_p, \boldsymbol{\Gamma}_{m,p}, \nu_m). \tag{2.C.6}$$

For ergodicity, let $P_{\boldsymbol{y}}^p(\boldsymbol{y}, \cdot) = \mathbb{P}(\boldsymbol{y}_p \in \cdot | \boldsymbol{y}_0 = \boldsymbol{y})$ signify the p-step transition probability measure of the process $\boldsymbol{y}_t$. Using the pth order Markov property of y_t, it is easy to check that

[12]Because the covariance matrices $\boldsymbol{\Gamma}_{m,p+1}$ $(m = 1, ..., M)$ have the Toepliz form and $\mu_m \mathbf{1}_p = (\mu_m, ..., \mu_m)$, the marginal densities for random vectors shorter than p are obtained by integrating the desired random variables out, and their distributions are mixtures of normal and t-distributions.

$P^p_{\boldsymbol{y}}(\boldsymbol{y}, \cdot)$ has the density

$$f(\boldsymbol{y}_p|\boldsymbol{y}_0;\boldsymbol{\theta}) = \prod_{t=1}^{p}\left(\sum_{m=1}^{M_1}\alpha_{m,t}n_1(y_t;\mu_{m,t},\sigma_m^2) + \sum_{m=M_1+1}^{M}\alpha_{m,t}t_1(y_t;\mu_{m,t},\sigma_{m,t}^2,\nu_m+p)\right). \tag{2.C.7}$$

Clearly $f(\boldsymbol{y}_p|\boldsymbol{y}_0;\boldsymbol{\theta}) > 0$ for all $\boldsymbol{y}_p \in \mathbb{R}^p$ and all $\boldsymbol{y}_0 \in \mathbb{R}^p$, so it can be concluded that $\boldsymbol{y}_t$ is ergodic in the sense of Meyn and Tweedie (2009, Ch. 13) by using arguments identical to those used in the proof of Theorem 1 in Kalliovirta *et al.* (2015). ■

2.C.2 Proof of Theorem 2.2

First note that $L_T^{(c)}(\boldsymbol{\theta})$ is continuous, and that together with Assumption 2.1, it implies existence of a measurable maximizer $\hat{\boldsymbol{\theta}}_T$. In order to conclude strong consistency of $\hat{\boldsymbol{\theta}}_T$, it needs to be shown that (see, e.g., Newey and McFadden, 1994, Theorem 2.1 and the discussion on page 2122)

(i) the uniform strong law of large numbers holds for the log-likelihood function; that is,

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|L_T^{(c)}(\boldsymbol{\theta}) - \mathrm{E}\left[L_T^{(c)}(\boldsymbol{\theta})\right]\right| \to 0 \quad \text{almost surely as} \quad T\to\infty,$$

(ii) and that the limit of $L_T^{(c)}(\boldsymbol{\theta})$ is uniquely maximized at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

Proof of (i). Because the initial values are assumed to be from the stationary distribution, the process $\boldsymbol{y}_t = (y_t, ..., y_{t-p+1})$, and hence also y_t, is stationary and ergodic, and $\mathrm{E}[L_T^{(c)}(\boldsymbol{\theta})] = \mathrm{E}[l_t(\boldsymbol{\theta})]$. To conclude (i), it thus suffices to show that $\mathrm{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|l_t(\boldsymbol{\theta})|\right] < \infty$ (see Ranga Rao, 1962). This is done by using compactness of the parameter space to derive finite lower and upper bounds for $l_t(\boldsymbol{\theta})$ which is given by

$$l_t(\boldsymbol{\theta}) = \log\left(\sum_{m=1}^{M_1}\alpha_{m,t}n_1(y_t;\mu_{m,t},\sigma_m^2) + \sum_{m=M_1+1}^{M}\alpha_{m,t}t_1\left(y_t;\mu_{m,t},\sigma_{m,t}^2,\nu_m+p\right)\right). \tag{2.C.8}$$

It follows from the structure of the parameter space that $c_1 \leq \sigma_m^2 \leq c_2$ and $c_1 \leq \alpha_m \leq 1-c_1$ for all $m = 1, ..., M$, and $c_3 \leq \nu_m \leq c_2$ for all $m = M_1+1, ..., M$, for some $0 < c_1 < 1$, $c_2 < \infty$ and $c_3 > 2$. Because the exponential function is bounded from above by one on the non-positive real axis, and in addition $c_1 \leq \sigma_m^2$, there exists a constant $U_1 < \infty$ such that

$$n_1(y_t;\mu_{m,t},\sigma_m^2) = \left(2\pi\sigma_m^2\right)^{-1/2}\exp\left(-\frac{(y_t-\mu_{m,t})^2}{2\sigma_m^2}\right) \leq U_1 \tag{2.C.9}$$

for all $m = 1, ..., M_1$.

We also have $c_3 \leq \nu_m + p \leq c_2 + p$ for all $m = M_1 + 1, ..., M$. Combined with the fact that the gamma function is continuous on the positive real axis, this implies that there exist constants $c_4 > 0$ and $c_5 < \infty$ such that

$$c_4 \leq C_1(\nu_m + p) = \frac{\Gamma\left(\frac{1+\nu_m+p}{2}\right)}{\sqrt{\pi(\nu_m + p - 2)}\Gamma\left(\frac{\nu_m+p}{2}\right)} \leq c_5 \tag{2.C.10}$$

for all $m = M_1 + 1, ..., M$. Because $\mathbf{\Gamma}_m$ and hence $\mathbf{\Gamma}_m^{-1}$ is positive definite, $\sigma_m^2 \geq c_1$ and $c_3 \leq \nu_m \leq c_2$, we can find some $c_6 > 0$ such that

$$\sigma_{m,t}^2 = \frac{\nu_m - 2 + (\boldsymbol{y}_{t-1} - \mu_m \mathbf{1}_p)' \mathbf{\Gamma}_m^{-1} (\boldsymbol{y}_{t-1} - \mu_m \mathbf{1}_p)}{\nu_m - 2 + p} \sigma_m^2 \geq c_6 \tag{2.C.11}$$

for all $m = M_1 + 1, ..., M$. Combined with (2.C.10) and (2.C.11), the inequality $-(1 + \nu_m + p)/2 < 0$ implies that there exists a constant $U_2 < \infty$ for which

$$t_1\left(y_t; \mu_{m,t}, \sigma_{m,t}^2, \nu_m + p\right) = \frac{C_1(\nu_m + p)}{\sigma_{m,t}} \left(1 + \frac{(y_t - \mu_{m,t})^2}{(\nu_m + p - 2)\sigma_{m,t}^2}\right)^{-(1+\nu_m+p)/2} \leq U_2. \tag{2.C.12}$$

for all $m = M_1 + 1, ..., M$. According to (2.C.9), (2.C.12) and the restriction $0 \leq \alpha_{m,t} \leq 1$, there exists a constant $U_3 < \infty$ such that

$$l_t(\boldsymbol{\theta}) = \log\left(\sum_{m=1}^{M_1} \alpha_{m,t} n_1(y_t; \mu_{m,t}, \sigma_m^2) + \sum_{m=M_1+1}^{M} \alpha_{m,t} t_1\left(y_t; \mu_{m,t}, \sigma_{m,t}^2, \nu_m + p\right)\right) \leq U_3. \tag{2.C.13}$$

It follows from the compactness of the parameter space that

$$\frac{(y_t - \mu_{m,t})^2}{2\sigma_m^2} \leq c_7(1 + y_t^2 + \boldsymbol{y}_{t-1}'\boldsymbol{y}_{t-1}), \tag{2.C.14}$$

implying

$$\exp\left\{-\frac{(y_t - \mu_{m,t})^2}{2\sigma_m^2}\right\} \geq \exp\left\{-c_7(1 + y_t^2 + \boldsymbol{y}_{t-1}'\boldsymbol{y}_{t-1})\right\}, \tag{2.C.15}$$

for all $m = 1, ..., M_1$, and for some finite constant c_7. By $\sigma_m^2 \leq c_2$ it also holds that $(2\pi\sigma_m^2)^{-1/2} \geq (2\pi c_2)^{-1/2}$, so

$$n_1(y_t; \mu_{m,t}, \sigma_m^2) \geq (2\pi c_2)^{-1/2} \exp\left\{-c_7(1 + y_t^2 + \boldsymbol{y}_{t-1}'\boldsymbol{y}_{t-1})\right\} \tag{2.C.16}$$

for all $m = 1, ..., M_1$.

Accordingly, since $\sigma_{m,t}^2 \geq c_6$ and $\nu_m \geq c_3$, it holds for some $c_8 < \infty$ that

$$1 + \frac{(y_t - \mu_{m,t})^2}{(\nu_m + p - 2)\sigma_{m,t}^2} \leq c_8(1 + y_t^2 + \boldsymbol{y}_{t-1}'\boldsymbol{y}_{t-1}), \;\; m = M_1 + 1, ..., M. \tag{2.C.17}$$

Thus, because $\nu_m \leq c_2$ and the inner functions below take values larger than one, we have

$$\left(1 + \frac{(y_t - \mu_{m,t})^2}{(\nu_m + p - 2)\sigma_{m,t}^2}\right)^{-(1+\nu_m+p)/2} \geq \left(c_8(1 + y_t^2 + \boldsymbol{y}_{t-1}'\boldsymbol{y}_{t-1})\right)^{-(1+c_2+p)/2}. \tag{2.C.18}$$

As Meitz *et al.* (forthcoming) state in the proof of Theorem 3, the quadratic form on the right side of (2.C.11) satisfies

$$(\boldsymbol{y}_{t-1} - \mu_m \mathbf{1}_p)'\boldsymbol{\Gamma}_m^{-1}(\boldsymbol{y}_{t-1} - \mu_m \mathbf{1}_p) \leq c_9(1 + \boldsymbol{y}_{t-1}'\boldsymbol{y}_{t-1}) \tag{2.C.19}$$

for all $m = M_1 + 1, ..., M$, and for some $c_9 < \infty$. Since also $0 < \nu_m - 2 \leq c_2$ and $\sigma_m^2 \leq c_2$, we have $\sigma_{m,t}^2 \leq c_{10}(1 + \boldsymbol{y}_{t-1}'\boldsymbol{y}_{t-1})$ for some finite constant c_{10}. Combining the former inequality with (2.C.10) and (2.C.18) yields a lower bound

$$t_1\left(y_t; \mu_{m,t}, \sigma_{m,t}^2, \nu_m + p\right) \geq \frac{c_4}{(c_{10}(1 + \boldsymbol{y}_{t-1}'\boldsymbol{y}_{t-1}))^{1/2}} \left(c_8(1 + y_t^2 + \boldsymbol{y}_{t-1}'\boldsymbol{y}_{t-1})\right)^{-(1+c_2+p)/2}. \tag{2.C.20}$$

Finally, the restriction $\sum_{m=1}^{M} \alpha_{m,t} = 1$ together with (2.C.16) and (2.C.20) implies

$$\begin{aligned} l_t(\boldsymbol{\theta}) \geq \min \Bigg\{ & -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(c_2) - c_7(1 + y_t^2 + \boldsymbol{y}_{t-1}'\boldsymbol{y}_{t-1}), \\ & \log(c_4) - \frac{1}{2}\log(c_{10}(1 + y_t^2 + \boldsymbol{y}_{t-1}'\boldsymbol{y}_{t-1})) - \frac{1 + c_2 + p}{2}\log\left(c_8(1 + y_t^2 + \boldsymbol{y}_{t-1}'\boldsymbol{y}_{t-1})\right) \Bigg\}. \end{aligned} \tag{2.C.21}$$

As $\mathrm{E}\left[y_t^2 + \boldsymbol{y}_{t-1}'\boldsymbol{y}_{t-1})\right] < \infty$ (because y_t is stationary and has finite second moments), it follows from Jensen's inequality that

$$\mathrm{E}\left[\log\left(c_8(1 + y_t^2 + \boldsymbol{y}_{t-1}'\boldsymbol{y}_{t-1})\right)\right] < \infty \text{ and } \mathrm{E}\left[\log\left(c_{10}(1 + \boldsymbol{y}_{t-1}'\boldsymbol{y}_{t-1})\right)\right] < \infty. \tag{2.C.22}$$

The upper bound (2.C.13) together with (2.C.21) and finiteness of the aforementioned expectations shows that $\mathrm{E}\left[\sup_{(\boldsymbol{\theta},\boldsymbol{\nu})\in\boldsymbol{\Theta}} |l_t(\boldsymbol{\theta})|\right] < \infty$. ■

Proof of (ii). Given that condition (2.3.3) sets a unique order for the mixture components, proving that this identification condition is satisfied amounts to showing that $\mathrm{E}\left[l_t(\boldsymbol{\theta})\right] \leq \mathrm{E}\left[l_t(\boldsymbol{\theta}_0)\right]$, and that the equality $\mathrm{E}\left[l_t(\boldsymbol{\theta})\right] = \mathrm{E}\left[l_t(\boldsymbol{\theta}_0)\right]$ implies

$$\begin{aligned} \boldsymbol{\vartheta}_m = \boldsymbol{\vartheta}_{\tau_1(m),0} \text{ and } \alpha_m = \alpha_{\tau_1(m),0} \text{ when } m = 1, ..., M_1, \text{ and} \\ (\boldsymbol{\vartheta}_m, \nu_m) = (\boldsymbol{\vartheta}_{\tau_2(m),0}, \nu_{\tau_2(m),0}) \text{ and } \alpha_m = \alpha_{\tau_2(m),0} \text{ when } m = M_1 + 1, ..., M, \end{aligned} \tag{2.C.23}$$

for some permutations $\{\tau_1(1), ..., \tau_1(M_1)\}$ and $\{\tau_2(M_1 + 1), ..., \tau_2(M)\}$. For notational clarity, I omit the subscripts from y_t and $\boldsymbol{y}_{t-1}$, and write $\mu_{m,t} = \mu(\boldsymbol{y}; \boldsymbol{\vartheta}_m)$, $\sigma_m^2 = \sigma_m^2(\boldsymbol{\vartheta}_m)$, $\sigma_{m,t}^2 = \sigma_{m,t}^2(\boldsymbol{y}; \boldsymbol{\vartheta}_m, \nu_m)$ for the expressions in (2.C.8) making clear their dependence on the parameter

value. I leave the dependence of $\alpha_{m,t}$ on $\boldsymbol{\theta}$ and $\boldsymbol{y}$ unmarked and denote by $\alpha_{m,0,t}$ mixing weights based on the true parameter value.

Making use of the fact that the density function of $(y_t, \boldsymbol{y}_{t-1})$ has the form (see proof of Theorem 2.1)

$$\begin{aligned} f((y_t, \boldsymbol{y}_{t-1}); \boldsymbol{\theta}) = &\sum_{m=1}^{M_1} \alpha_m n_{p+1}((y_t, \boldsymbol{y}_{t-1})); \mu_m \mathbf{1}_{p+1}, \boldsymbol{\Gamma}_{m,p+1}) \\ &+ \sum_{m=M_1+1}^{M} \alpha_m t_{p+1}((y_t, \boldsymbol{y}_{t-1})); \mu_m \mathbf{1}_{p+1}, \boldsymbol{\Gamma}_{m,p+1}, \nu_m) \end{aligned} \tag{2.C.24}$$

and reasoning based on Kullback-Leibler divergence, arguments analogous to those in Kalliovirta *et al.* (2015, p. 265) can be used to conclude $\mathrm{E}\left[l_t(\boldsymbol{\theta})\right] - \mathrm{E}\left[l_t(\boldsymbol{\theta}_0)\right] \leq 0$ with equality if and only if for almost all $(y, \boldsymbol{y}) \in \mathbb{R}^{p+1}$

$$\begin{aligned} &\sum_{m=1}^{M_1} \alpha_{m,t} n_1(y; \mu(\boldsymbol{y}; \boldsymbol{\vartheta}_m), \sigma_m^2(\boldsymbol{\vartheta}_m)) + \sum_{m=M_1+1}^{M} \alpha_{m,t} t_1(y; \mu(\boldsymbol{y}; \boldsymbol{\vartheta}_m), \sigma_{m,t}^2(\boldsymbol{y}; \boldsymbol{\vartheta}_m, \nu_m)), \nu_m + p) \\ &= \sum_{m=1}^{M_1} \alpha_{m,0,t} n_1(y; \mu(\boldsymbol{y}; \boldsymbol{\vartheta}_{m,0}), \sigma_m^2(\boldsymbol{\vartheta}_{m,0})) \\ &+ \sum_{m=M_1+1}^{M} \alpha_{m,0,t} t_1(y; \mu(\boldsymbol{y}; \boldsymbol{\vartheta}_{m,0}), \sigma_{m,t}^2(\boldsymbol{y}; \boldsymbol{\vartheta}_{m,0}, \nu_{m,0})), \nu_{m,0} + p). \end{aligned} \tag{2.C.25}$$

For each fixed $\boldsymbol{y}$ at a time, the mixing weights, conditional means and variances in (2.C.25) are constants, so the result on identification of finite mixtures of normal and t-distributions in Holzmann, Munk, and Gneiting (2006, Example 1) can be applied (their parametrization of the t-distribution slightly differs from ours, but identification with their parametrization implies identification with my parametrization). For each fixed $\boldsymbol{y}$, there thus exists a permutation $\{\tau_1(1), ..., \tau_1(M_1)\}$ (that may depend on $\boldsymbol{y}$) of the index set $\{1, ..., M_1\}$ such that

$$\alpha_{m,t} = \alpha_{\tau_1(m),0,t}, \;\; \mu(\boldsymbol{y}; \boldsymbol{\vartheta}_m) = \mu(\boldsymbol{y}; \boldsymbol{\vartheta}_{\tau_1(m),0}) \text{ and } \sigma_m^2(\boldsymbol{\vartheta}_m) = \sigma_m^2(\boldsymbol{\vartheta}_{\tau_1(m),0}) \tag{2.C.26}$$

for almost all $y \in \mathbb{R}$ ($m = 1, ..., M_1$). Analogously, for each fixed $\boldsymbol{y}$ there exists a permutation $\{\tau_2(M_1 + 1), ..., \tau_2(M)\}$ (that may depend on $\boldsymbol{y}$) of the index set $\{M_1 + 1, ..., M\}$ such that

$$\begin{aligned} &\nu_m = \nu_{\tau_2(m),0}, \;\; \alpha_{m,t} = \alpha_{\tau_2(m),0,t}, \;\; \mu(\boldsymbol{y}; \boldsymbol{\vartheta_m}) = \mu(\boldsymbol{y}; \boldsymbol{\vartheta}_{\tau_2(m),0}) \text{ and} \\ &\sigma_{m,t}^2(\boldsymbol{y}; \boldsymbol{\vartheta}_m, \nu_m) = \sigma_{m,t}^2(\boldsymbol{y}; \boldsymbol{\vartheta}_{\tau_2(m),0}, \nu_{\tau_2(m),0}), \end{aligned} \tag{2.C.27}$$

for almost all $y \in \mathbb{R}$ ($m = M_1 + 1, ..., M$).

As argued by Kalliovirta *et al.* (2015, pp. 265-266) for the GMAR type components, it follows from (2.C.26) that $\boldsymbol{\vartheta}_m = \boldsymbol{\vartheta}_{\tau_1(m),0}$ and $\alpha_m = \alpha_{\tau_1(m),0}$ for $m = 1, ..., M_1$. Accordingly,

Meitz *et al.* (forthcoming) showed that (2.C.27) implies $\boldsymbol{\vartheta}_m = \boldsymbol{\vartheta}_{\tau_2(m),0}$, $\nu_m = \nu_{\tau_2(m),0}$ and $\alpha_m = \alpha_{\tau_2(m),0}$ for $m = M_1 + 1, ..., M$, completing the proof of strong consistency.

Given consistency and assumptions of the theorem, asymptotic normality of the ML estimator can now be concluded using standard arguments. The required steps can be found, for example, in Kalliovirta, Meitz, and Saikkonen (2016, proof of Theorem 3). I omit the details for brevity. ■

Chapter 3

uGMAR: a family of mixture autoregressive models in R

3.1 Introduction

A popular method for modelling univariate time series is to employ a linear autoregressive (AR) model that assumes the process to be generated by a weighted sum of the preceding p observations, an intercept term, and a random error. The error process is often assumed to be serially uncorrelated with zero mean and constant variance. This encompasses conditionally homoskedastic processes, such as independent and identically distributed (IID) processes, as well as conditionally heteroskedastic processes, such as autoregressive conditional heteroskedasticity (ARCH) processes (Engle, 1982) and generalized autoregressive conditional heteroskedasticity (GARCH) processes (Bollerslev, 1986).

Several R packages accommodate linear AR modelling with various types of error processes. The R package **forecast** (Hyndman, Athanasopoulos, Bergmeir, Caceres, Chhay, O'Hara-Wild, Petropoulos, Razbash, Wang, and Yasmeen, 2021), for instance, accommodates estimation of AR models with seasonal components. The R package **fGarch** (Wuertz, Setz, Chalabi, Boudt, Chausse, and Miklovac, 2020), on the other hand, facilitates estimation of AR models with ARCH and GARCH errors following various distributions, including normal, Student's t-, and generalized error distributions and their skewed versions. A more comprehensive set of error processes are provided in the popular R package **rugarch** (Ghalanos, 2020). It accommodates a rich set of different GARCH processes with several error distributions, including the regular and skewed versions of normal, t-, and generalized error distributions, as well as generalized hyperbolic normal and inverse Gaussian distributions, to name a few.

A linear AR model with potentially skewed GARCH errors can often filter the autocorrelation and conditional heteroskedasticity from the series very well. But in some cases, it cannot adequately capture all the relevant characteristics of the series, including shifts in the mean or volatility, and changes in the dynamics of the process. Such nonlinear features frequently occur in economic time series when the underlying data generating dynamics vary in time, for

example, depending on the specific state of the economy.

Various types of time series models capable of capturing such regime-switching behavior have been proposed, one of them being the class of mixture models introduced by Le, Martin, and Raftery (1996) and further developed by, among others, Wong and Li (2000, 2001a,b), Glasbey (2001), Lanne and Saikkonen (2003), Kalliovirta, Meitz, and Saikkonen (2015), Meitz, Preve, and Saikkonen (forthcoming), and Virolainen (forthcoming, also Chapter 2 of this dissertation). Following the recent developments by Kalliovirta *et al.* (2015), Meitz *et al.* (forthcoming), and Virolainen (forthcoming), I consider the Gaussian mixture autoregressive (GMAR) model, the Student's t mixture autoregressive (StMAR) model, and the Gaussian and Student's t mixture autoregressive (G-StMAR) model. These three models constitute an appealing family of (univariate) mixture autoregressive models that I call the GSMAR models.

A GSMAR process generates each observation from one of its mixture components, which are either conditionally homoskedastic linear Gaussian autoregressions or conditionally heteroskedastic linear Student's t autoregressions. The mixture component that generates each observation is randomly selected according to the probabilities determined by the mixing weights that, for a pth order model, depend on the full distribution of the previous p observations. Consequently, the regime-switching probabilities may depend on the level, variability, kurtosis, and temporal dependence of the past observations. The specific formulation of the mixing weights also leads to attractive theoretical properties such as ergodicity and full knowledge of the stationary distribution of $p + 1$ consecutive observations.

This essay describes the R package **uGMAR** providing a comprehensive set of easy-to-use tools for GSMAR modelling, including unconstrained and constrained maximum likelihood (ML) estimation of the model parameters, quantile residual based model diagnostics, simulation from the processes, and forecasting. The emphasis is on estimation, as it can, in my experience, be rather tricky. In particular, due to the endogenously determined mixing weights, the log-likelihood function has a large number of modes, and in large areas of the parameter space, the log-likelihood function is flat in multiple directions. The log-likelihood function's global maximum point is also frequently located very near the boundary of the parameter space. It turns out, however, that such near-the-boundary estimates often maximize the log-likelihood function for rather a technical reason, and it might be more appropriate to prefer an alternative estimate based on the largest local maximum point that is clearly in the interior of the parameter space.

The model parameters are estimated by running multiple rounds of a two-phase estimation procedure in which a modified genetic algorithm is used to find starting values for a gradient based variable metric algorithm. Because of the multimodality of the log-likelihood function, some of the estimation rounds may end up in different local maximum points, thereby enabling the researcher to build models not only based on the global maximum point but also on the local ones. The estimated models can be conveniently examined with the `summary` and `plot` methods. For evaluating their adequacy, **uGMAR** utilizes quantile residual diagnostics in the framework presented in Kalliovirta (2012), including graphical diagnostics as well as Kalliovirta's (2012) diagnostic tests that take into account uncertainty about the true parameter value. Following

Kalliovirta *et al.* (2015) and Meitz *et al.* (forthcoming), forecasting is based on a Monte Carlo simulation method.

Other statistical software implementing the GSMAR models include the **StMAR Toolbox** for MATLAB (Meitz, Preve, and Saikkonen, 2018). It currently (version 1.0.0) covers the StMAR model of autoregressive orders $p = 1, 2, 3, 4$ and $M = 1, 2, 3$ mixture components, and it contains tools for maximum likelihood estimation, calculation of quantile residuals, simulation, and forecasting. Also the **StMAR Toolbox** estimates the model parameters by using a genetic algorithm to find starting values for a gradient based method, but **uGMAR** takes the procedure of Meitz *et al.* (2018, forthcoming) further by modifying a genetic algorithm for more efficient estimation. **uGMAR** also has the advantage that it does not impose restrictions on the order of the model and it provides a wider variety of tools for analyzing the estimated models; for instance, functions for calculating quantile residual diagnostic tests (Kalliovirta, 2012) and plotting the graphs of the profile log-likelihood functions about the estimate.

The R package **gmvarkit** (Virolainen, 2018) functions similarly to **uGMAR** and accommodates multivariate versions of the GSMAR models, including structural models with statistically identified shocks. These models include the (structural) Gaussian mixture vector autoregressive model (Kalliovirta, Meitz, and Saikkonen, 2016, and Chapter 4 of this dissertation), the (structural) Student's t mixture vector autoregressive model (see Chapter 5), and the (structural) Gaussian and Student's t mixture vector autoregressive model (see Chapter 5). The R package **mixAR** (Boshnakov and Ravagli, 2021), in turn, allows frequentist and Bayesian estimation of mixture (vector) autoregressive models with constant mixing weights (e.g., Fong, Li, Yau, and Wong, 2007, Wong and Li, 2000) and various error distributions.

The remainder of this chapter is organized as follows. Section 3.2 introduces the GSMAR models and discusses some of their properties. Section 3.3 discusses estimation of the model parameters and model selection. It also illustrates how the GSMAR models can be estimated and examined with **uGMAR**, and how parameter constraints can be tested. In Section 3.4, I describe quantile residuals and demonstrate how they can be utilized to evaluate model adequacy in **uGMAR**. Section 3.5 shows how the GSMAR models can be built with given parameter values. In Section 3.6, I first show how to simulate observations from a GSMAR process, and then I illustrate how to forecast future values of a GSMAR process with a simulation-based Monte Carlo method. Section 3.7 concludes and collects some useful functions in **uGMAR** to a single table for convenience. Appendix 3.A explains why some maximum likelihood estimates, that are very near the boundary of the parameter space, might be inappropriate and demonstrates that a local maximum point that is clearly in the interior of the parameter space can often be a more reasonable estimate. Finally, Appendix 3.B derives closed form expressions for the quantile residuals of the GSMAR models.

Throughout this paper, I use the monthly U.S. interest rate spread between the 10-year and 1-year Treasury rates for the empirical illustrations. I deploy the notation $n_d(\boldsymbol{\mu}, \boldsymbol{\Gamma})$ for the d-dimensional normal distribution with mean $\boldsymbol{\mu}$ and (positive definite) covariance matrix $\boldsymbol{\Gamma}$, and $t_d(\boldsymbol{\mu}, \boldsymbol{\Gamma}, \nu)$ for the d-dimensional t-distribution with mean $\boldsymbol{\mu}$, (positive definite) covariance matrix $\boldsymbol{\Gamma}$, and $\nu > 2$ degrees of freedom. The corresponding density functions are denoted as

$n_d(\cdot; \boldsymbol{\mu}, \boldsymbol{\Gamma})$ and $t_d(\cdot; \boldsymbol{\mu}, \boldsymbol{\Gamma}, \nu)$, respectively. By $\mathbf{1}_p = (1, ..., 1)$ $(p \times 1)$, I denote p-dimensional vector of ones.

3.2 Models

This section introduces the GMAR model (Kalliovirta *et al.*, 2015), the StMAR model (Meitz *et al.*, forthcoming), and the G-StMAR model (Virolainen, forthcoming, also Chapter 2), a family of mixture autoregressive models that I call the GSMAR models. First, I consider the models in a general framework and then proceed to their specific definitions. For brevity, I only give the definition of the more general G-StMAR model but explain how the GMAR and StMAR models are obtained as special cases of it, namely, by taking all the component models to be of either Gaussian or Student's t type.

3.2.1 Mixture autoregressive models

Let y_t, $t = 1, 2, ...$, be the real valued time series of interest, and let $\mathcal{F}_{t-1}$ denote the σ-algebra generated by the random variables $\{y_{t-j}, j > 0\}$. For a GSMAR model with autoregressive order p and M mixture components, we have

$$y_t = \sum_{m=1}^{M} s_{m,t}(\mu_{m,t} + \sigma_{m,t}\varepsilon_{m,t}), \quad \varepsilon_{m,t} \sim \text{IID}(0, 1), \tag{3.2.1}$$

$$\mu_{m,t} = \varphi_{m,0} + \sum_{i=1}^{p} \varphi_{m,i} y_{t-i}, \quad m = 1, ..., M, \tag{3.2.2}$$

where $\sigma_{m,t} > 0$ are $\mathcal{F}_{t-1}$-measurable, $\varepsilon_{m,t}$ are independent of $\mathcal{F}_{t-1}$, $\varphi_{m,0} \in \mathbb{R}$, and $s_{1,t}, ..., s_{M,t}$ are unobservable regime variables such that for each t, exactly one of them takes the value one and the others take the value zero. Given the past of y_t, $(s_{1,t}, ..., s_{M,t})$ and $\varepsilon_{m,t}$ are assumed to be conditionally independent, and the conditional probability for an observation to be generated from the mth regime at time t is expressed in terms of ($\mathcal{F}_{t-1}$-measurable) mixing weights $\alpha_{m,t} \equiv \mathrm{P}(s_{m,t} = 1 | \mathcal{F}_{t-1})$ that satisfy $\sum_{m=1}^{M} \alpha_{m,t} = 1$. Furthermore, for each component model, the autoregressive parameters satisfy the usual stationarity condition, $1 - \sum_{i=1}^{p} \varphi_{m,i} z^i \neq 0$ for $|z| \leq 1$, which guarantees stationarity of the GSMAR models (Theorem 2.1 in Chapter 2).

The definition (3.2.1) and (3.2.2) implies that at each t, the observation is generated by a linear autoregression corresponding to some randomly selected (unobserved) mixture component m, and that $\mu_{m,t}$ and $\sigma^2_{m,t}$ can be interpreted as the conditional mean and variance of this component process. In the GMAR model (Kalliovirta *et al.*, 2015), the mixture components are conditionally homoskedastic Gaussian autoregressions, whereas in the StMAR model (Meitz *et al.*, forthcoming), they are conditionally heteroskedastic Students t autoregressions, while the G-StMAR model (Chapter 2) combines both types of mixture components. The mixing weights are functions of the preceding p observations.

3.2.2 The Gaussian and Student's *t* mixture autoregressive model

In the G-StMAR model, for $m = 1, ..., M_1$ in (3.2.1), the terms $\varepsilon_{m,t}$ have standard normal distributions and the conditional variances $\sigma^2_{m,t}$ are constants σ^2_m. For $m = M_1 + 1, ..., M$, the terms $\varepsilon_{m,t}$ follow the t-distribution $t_1(0, 1, \nu_m + p)$ and the conditional variances $\sigma^2_{m,t}$ are defined as

$$\sigma^2_{m,t} = \frac{\nu_m - 2 + (\boldsymbol{y}_{t-1} - \mu_m \mathbf{1}_p)' \boldsymbol{\Gamma}_m^{-1} (\boldsymbol{y}_{t-1} - \mu_m \mathbf{1}_p)}{\nu_m - 2 + p} \sigma^2_m, \tag{3.2.3}$$

where $\boldsymbol{y}_{t-1} = (y_{t-1}, ..., y_{t-p})$ $(p \times 1)$, $\nu_m > 2$ is a degrees of freedom parameter, $\sigma^2_m > 0$ is a variance parameter, $\mu_m = \varphi_0 / (1 - \sum_{i=1}^p \varphi_{m,i})$ is the stationary mean, and $\boldsymbol{\Gamma}_m$ is the stationary $(p \times p)$ covariance matrix of the mth component process (see Section 2.2.1 in Chapter 2).

This specification leads to a model in which the conditional density function of y_t given its past, $f(\cdot | \mathcal{F}_{t-1})$, is

$$f(y_t | \mathcal{F}_{t-1}) = \sum_{m=1}^{M_1} \alpha_{m,t} n_1(y_t; \mu_{m,t}, \sigma^2_m) + \sum_{m=M_1+1}^{M} \alpha_{m,t} t_1\left(y_t; \mu_{m,t}, \sigma^2_{m,t}, \nu_m + p\right). \tag{3.2.4}$$

That is, the first M_1 component processes of the G-StMAR model are homoskedastic Gaussian autoregressions, and the remaining $M_2 \equiv M - M_1$ component processes are heteroskedastic Student's t autoregressions.

In the GMAR model (Kalliovirta *et al.*, 2015), all M component processes are Gaussian autoregressions, so its conditional density function is obtained by setting $M_1 = M$ and dropping the second sum in (3.2.4). In the StMAR model (Meitz *et al.*, forthcoming), all M component processes are Student's t autoregressions, so its conditional density function is obtained by setting $M_1 = 0$ and dropping the first sum in (3.2.4). As the component processes of the G-StMAR model coincide with those of the GMAR model and the StMAR model, I often refer to them as GMAR type or StMAR type, accordingly.

In order to specify the mixing weights, I first define the following function for notational convenience. Let

$$d_m(\boldsymbol{y}; \mu_m \mathbf{1}_p, \boldsymbol{\Gamma}_m, \nu_m) = \begin{cases} n_p(\boldsymbol{y}; \mu_m \mathbf{1}_p, \boldsymbol{\Gamma}_m), & \text{when } m \leq M_1, \\ t_p(\boldsymbol{y}; \mu_m \mathbf{1}_p, \boldsymbol{\Gamma}_m, \nu_m), & \text{when } m > M_1, \end{cases} \tag{3.2.5}$$

where the p-dimensional densities $n_p(\boldsymbol{y}; \mu_m \mathbf{1}_p, \boldsymbol{\Gamma}_m)$ and $t_p(\boldsymbol{y}; \mu_m \mathbf{1}_p, \boldsymbol{\Gamma}_m, \nu_m)$ correspond to the stationary distribution of the mth component process (given in Equations (2.2.3) and (2.2.8) in Chapter 2). The mixing weights of the G-StMAR model are defined as

$$\alpha_{m,t} = \frac{\alpha_m d_m(\boldsymbol{y}_{t-1}; \mu_m \mathbf{1}_p, \boldsymbol{\Gamma}_m, \nu_m)}{\sum_{n=1}^M \alpha_n d_n(\boldsymbol{y}_{t-1}; \mu_n \mathbf{1}_p, \boldsymbol{\Gamma}_n, \nu_n)}, \tag{3.2.6}$$

where the parameters $\alpha_1, ..., \alpha_M$ satisfy $\sum_{m=1}^M \alpha_m = 1$. The mixing weights of the GMAR model are obtained from (3.2.5) and (3.2.6) by setting $M_1 = M$, whereas the mixing weights of the StMAR model are obtained by setting $M_1 = 0$.

Because the mixing weights are weighted ratios of the stationary densities corresponding to the previous p observations, the greater the relative weighted likelihood of a regime is, the more likely the process is to generate an observation from it. Moreover, as the mixing weights depend on the full distribution of the previous p observations, the regime-switching probabilities may depend on the level, variability, kurtosis, and temporal dependence of the past observations. This is a convenient property for forecasting, and it also facilitates associating statistical characteristics and economic interpretations to the regimes.

The specific formulation of the mixing weights also leads to attractive theoretical properties. Specifically, the G-StMAR process $\boldsymbol{y}_t = (y_t, ..., y_{t-p+1})$ $(p \times 1)$, $t = 1, 2, ...$, is ergodic, and it has fully known marginal stationary distribution that is characterized by the density (Theorem 2.1 in Chapter 2; see the proof of this theorem for the stationary distribution of $1, ..., p+1$ consecutive observations)

$$f(\boldsymbol{y}) = \sum_{m=1}^{M_1} \alpha_m n_p(\boldsymbol{y}; \mu_m \mathbf{1}_p, \boldsymbol{\Gamma}_m) + \sum_{m=M_1+1}^{M_2} \alpha_m t_p(\boldsymbol{y}; \mu_m \mathbf{1}_p, \boldsymbol{\Gamma}_m, \nu_m). \tag{3.2.7}$$

That is, the stationary distribution is a mixture of M_1 p-dimensional Gaussian distributions and M_2 p-dimensional Student's t-distributions with constant mixing weights α_m, $m = 1, ..., M$. For $h = 0, ..., p$, the marginal stationary distribution of $(y_t, ..., y_{t-h})$ is also a mixture of Gaussian and Student's t distributions with constant mixing weights α_m, so the mixing weights parameters α_m can be interpreted as the unconditional probabilities of an observation being generated from the mth component process.

In **uGMAR**, the parameters of the GSMAR models are collected to a $(M(p+3)+M_2-1\times 1)$ vector $\boldsymbol{\theta} \equiv (\boldsymbol{\vartheta}_1, ..., \boldsymbol{\vartheta}_M, \alpha_1, ..., \alpha_{M-1}, \boldsymbol{\nu})$, where $\boldsymbol{\vartheta}_m = (\varphi_{m,0}, \boldsymbol{\varphi}_m, \sigma_m^2)$, $\boldsymbol{\varphi}_m = (\varphi_{m,1}, ..., \varphi_{m,p})$, $m = 1, ..., M$, and $\boldsymbol{\nu} = (\nu_{M_1+1}, ..., \nu_M)$. The parameter α_M is omitted because it is obtained from the restriction $\sum_{m=1}^{M} \alpha_m = 1$, and in the GMAR model, the vector $\boldsymbol{\nu}$ is omitted, as the model does not contain degrees of freedom parameters. The knowledge of the parameter vector is particularly required for building models with given parameter values, which is discussed in Section 3.5.

3.3 Estimation and model selection

3.3.1 Log-likelihood function

uGMAR employs the method of maximum likelihood (ML) for estimating the parameters of the GSMAR models. Suppose the observed time series is $y_{-p+1}, ..., y_0, y_1, ..., y_T$ and that the initial values are stationary. Then, the log-likelihood function of the G-StMAR model takes the form

$$L(\boldsymbol{\theta}) = \log\left(\sum_{m=1}^{M} \alpha_m d_m(\boldsymbol{y}_0; \mu_m \mathbf{1}_p, \boldsymbol{\Gamma}_m, \nu_m)\right) + \sum_{t=1}^{T} l_t(\boldsymbol{\theta}), \tag{3.3.1}$$

where $d_m(\cdot; \mu_m \mathbf{1}_p, \mathbf{\Gamma}_m, \nu_m)$ is defined in (3.2.5),

$$l_t(\boldsymbol{\theta}) = \log\left(\sum_{m=1}^{M_1} \alpha_{m,t} n_1(y_t; \mu_{m,t}, \sigma_m^2) + \sum_{m=M_1+1}^{M} \alpha_{m,t} t_1\left(y_t; \mu_{m,t}, \sigma_{m,t}^2, \nu_m + p\right)\right), \quad (3.3.2)$$

and the density functions $n_d(\cdot;\cdot)$ and $t_d(\cdot;\cdot)$ follow the notation described in Section 3.2.2. Log-likelihood functions of the GMAR model and the StMAR model can be obtained as special cases by setting $M_1 = M$ or $M_1 = 0$, respectively, and dropping the redundant sums.

If stationarity of the initial values seems unreasonable, one can condition on the initial values by dropping the first term on the right side of (3.3.1) and base the estimation on the resulting conditional log-likelihood function. The ML estimator of a stationary GSMAR model is strongly consistent and has the conventional limiting distribution under the conventional high level conditions as is given in Kalliovirta *et al.* (2015, pp.254-255), Meitz *et al.* (forthcoming, Theorem 3), and Theorem 2.2 in Chapter 2.

3.3.2 Two-phase estimation procedure

Finding the ML estimate amounts to maximizing the log-likelihood function (3.3.1) over a high dimensional parameter space satisfying several constraints. Due to the complexity of the log-likelihood function, finding an analytical solution is infeasible, so numerical optimization methods are required. Following Dorsey and Mayer (1995) and Meitz *et al.* (2018, forthcoming), **uGMAR** employs a two-phase estimation procedure in which a genetic algorithm is used to find starting values for a gradient based method, which then often converges to a nearby local maximum or saddle point. Because of the presence of multiple local maxima, a (sometimes large) number of estimation rounds should be performed to obtain reliable results, for which **uGMAR** makes use of parallel computing to shorten the estimation time.

The genetic algorithm in **uGMAR** is, at core, mostly based on the description by Dorsey and Mayer (1995) but several modifications have been deployed to improve its performance. The modifications include the ones proposed by Patnaik and Srinivas (1994) and Smith, Dike, and Stegmann (1995) as well as further adjustments that take into account model specific issues related to the mixing weights' dependence on the preceding observations. For a more detailed description of the genetic algorithm and its modifications, see Appendix 2.A in Chapter 2. After running the genetic algorithm, the estimation is finalized with a variable metric algorithm (Nash, 1990, Algorithm 21, implemented by R Core Team, 2022) using central difference approximation for the gradient of the log-likelihood function.

3.3.3 Model selection

Before illustrating with examples how the GSMAR models can be estimated with **uGMAR**, it is helpful to first briefly discuss the problem of model selection. Finding a suitable GSMAR model involves several selections: one needs to choose the type of the model (GMAR, StMAR,

or G-StMAR), the autoregressive order p, and the number of mixture components M (in the G-StMAR model, the number of GMAR type regimes M_1 and the number of StMAR type regimes M_2). Following Kalliovirta *et al.* (2015, Section 3.1), I suggest starting the model selection by first considering linear AR models, and then building up to the more complex regime-switching models if the linear models are found inadequate. After finding a suitable GSMAR model, simplifications obtained by parameter restrictions can be considered (constrained estimation is discussed in Section 3.3.6, testing the constraints in Section 3.3.7, and diagnostics checks for evaluating the adequacy of the model in Section 3.4).

When selecting the type of the GSMAR model, it is useful to take into account the features of the different types of models. The GMAR model incorporates linear Gaussian AR processes as its mixture components and can flexibly model changes in the conditional mean. But as its component processes are conditionally homoskedastic, it can capture changes in the conditional variance only through the regime-switching dynamics. The StMAR model, on the other hand, incorporates ARCH type conditional heteroskedasticity within each regime with the conditional variance (3.2.3), and can thereby account for stronger forms of conditional heteroskedasticity.

In the StMAR model, the autoregressive order p is also the lag order of the ARCH type conditional variance. The conditional variance depends on the past observations through the same parameters as the conditional mean (3.2.2), which can be restrictive when the regime-specific conditional mean is strong but conditional variance is weak[1] (or vice versa). It may therefore be worthwhile to first try whether the simpler GMAR model can adequately capture the characteristics of the series.

If the conditional variance is constant in some regimes but time-varying in other regimes, the G-StMAR model can be employed, as it contains both conditionally homoskedastic GMAR type regimes and conditionally heteroskedastic StMAR type regimes. For choosing the number of GMAR and StMAR type regimes in the G-StMAR model, I propose following the strategy described in Section 2.4 in Chapter 2 and first finding a suitable StMAR model. If the estimated StMAR model contains overly large degrees of freedom parameter estimates, those regimes should be switched to GMAR type by estimating the appropriate G-StMAR model (this is discussed in more detail Section 3.3.4).

For the illustrations, I use the monthly U.S. interest rate spread between the 10-year and 1-year Treasury constant maturity rates, covering the period from 1982 January to 2020 December (468 observations). The series was retrieved from the Federal Reserve Bank of St. Louis database. After installing **uGMAR**, the data can be loaded with the following lines of code:

```
R> library("uGMAR")
R> data("M10Y1Y", package = "uGMAR")
```

For finding the suitable type and order of the model, it is often useful to plot several figures illustrating the statistical properties of the series. A time series plot can be examined to obtain an overall perception of series, and to investigate whether there seem to be apparent changes in

[1] By strong (weak) conditional variance or mean, I mean strong (weak) dependence on the preceding observations.

the dynamics of series, or shifts in the mean or volatility that would indicate a possible presence of multiple regimes.

The time series plot of the interest rate spread `M10Y1Y` is shown in the top left panel of Figure 3.2 (in Section 3.3.5). It shows that the process consistently produces consecutive observations of the same magnitude, which are then followed by a transition to another magnitude. There thus appears to be shifts in the mean of the process and the changes are occasionally rapid.

A non-parametric estimate of the density function, such as a kernel density estimate, can be examined to evaluate whether the marginal density of a linear AR model can adequately describe it, and if not, what might be the correct number of regimes. Multiple modes in the marginal distribution can be accounted for by accommodating each one of them with a regime in the GSMAR model. Skewness and many other forms of non-Gaussianity can also be accommodated with a mixture of normal or t-distributions, but it is less straightforward to determine the correct number of regimes. One should, nevertheless, be conservative with the choice of M, because with too many regimes in the model, some of the parameters are not identified (see Kalliovirta *et al.*, 2015, Sections 3.1 and 3.2.2 and the references therein).

A kernel density estimate of the interest rate spread is depicted in the right panel of Figure 3.2 (black solid line). There are two visible modes in the density function, so a linear model (with unimodal error distribution) is clearly inadequate to describe it, while a two-regime mixture model could be appropriate. Even a three-regime model could be considered in order to explain the hump shape in the right tail of the distribution.

Examining the sample partial autocorrelation function (PACF) of the series can help in selecting the correct autoregresive order p, as for a pth order AR process, there should be a visible break in the PACF after the lag p. If the series is not autocorrelated, the sample partial autocorrelation function of the squared series may similarly help to detect the order of ARCH type conditional heteroskedasticity. In the case of an autocorrelated series, it might be useful to first fit an AR model with a suitable autoregressive order, and then examine the PACF of the squared residuals. The sample partial autocorrelation function of the series `M10Y1Y` (calculated using the function `pacf` from the package **stats**, R Core Team, 2022) is presented in the left panel of Figure 3.1.

Figure 3.1 shows that the PACF of the series has very large partial autocorrelation coefficient (PACC) at the first lag, relatively large PACCs at the second and fourth lags, and visibly smaller PACCs after the fourth lag. The autoregressive order $p = 4$ thereby seems a reasonable candidate for a parsimonious AR model.[2] Hence, I fitted a Gaussian AR(4) model to the series and examined the PACF of its residuals and squared residuals, which are depicted in the middle and right panels of Figure 3.1, respectively.

The PACF of the AR(4) model's residuals shows that there is not much autocorrelation left in the residuals, so the autoregressive order $p = 4$ seems sufficient for capturing the autocorrelation structure of the series. The PACF of the AR(4) model's squared residuals shows PACCs outside the 95% critical bounds at lags 1, 3, and 8. Thereby the order 4 could be somewhat sufficient

[2] It turns out that the order $p = 4$ also minimizes the Akaike information criterion among the Gaussian AR(p) models, $p = 1, ..., 24$, based on the exact log-likelihood function (not shown).

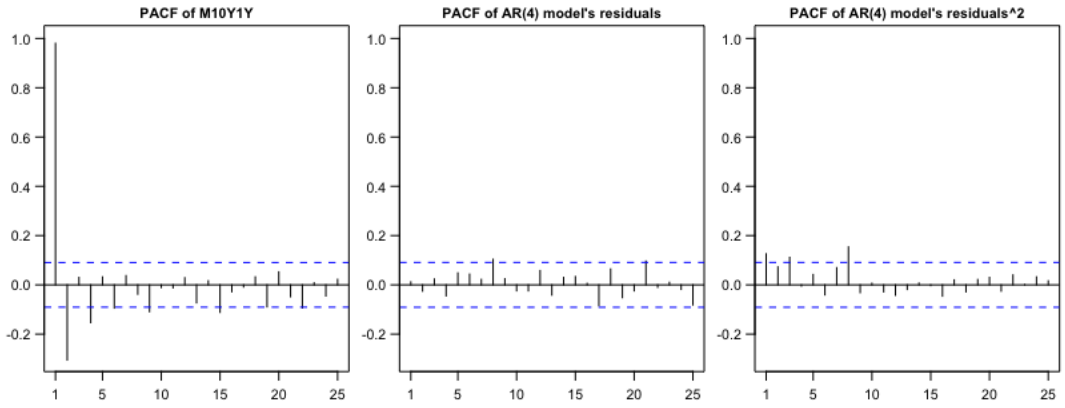

Figure 3.1: The sample partial autocorrelation function of the series `M10Y1Y` for the lags $1, ..., 25$ (on the right). The sample partial autocorrelation function of the Pearson residuals of a Gaussian AR(4) model (on the middle) and of the squared residuals (on the right) for the lags $1, ..., 25$. The blue dashed lines are the 95% critical bounds for partial autocorrelation coefficients of an IID process.

for modelling the (potentially present) ARCH type conditional heteroskedasticity, but the order 9 could also be considered for a less parsimonious model.[3] However, as the lag 8 PACC of the residuals is moderate, the lag 8 PACC of the squared residuals could be related to the unmodelled autocorrelation rather than conditional heteroskedasticity.[4] A StMAR model might, therefore, be appropriate with the autoregressive order $p = 4$, although it may not be sufficient for modelling the conditional heteroskedasticity at larger lags. As discussed above, the two modes in the kernel density estimate of the series, in turn, indicate that two regimes seems like a good starting point for building the model.

If the candidate model is found inadequate, one may try to use a different autoregressive order p or to add a regime to the model (or switch to the StMAR model, if a GMAR model is found inadequate). Note that while with linear AR models increasing the autoregressive order typically improves the fit, this is not necessarily the case with GSMAR models, as the autoregressive order affects the regime-switching dynamics. In particular, because the mixing weights (3.2.6) are calculated using the whole joint distribution of the previous p observations, with a small p, the regime-switching probabilities react more sensitively to individual observations than with a large p. It may hence be useful to also try to decrease the autoregressive order rather than just increase it.

In addition to comparing model adequacy (or forecasting accuracy, for example), informa-

[3] Since the residuals are calculated from the difference between the current observation and a linear function of the preceding p observations, unmodelled autocorrelation (or ARCH type conditional heteroskedasticity) at lag q may show up at the lag $q - 1$ PACC of the (squared) residuals.

[4] For completeness, I fitted a Gaussian AR(9) model to the series and studied the PACF of its residuals and squared residuals. I found the lag 8 PACC of the residuals small but the lag 8 PACC of the squared residuals large, suggesting the possible presence of ARCH type conditional heteroskedasticity at the lag 8.

tion criteria can be utilized in the selection of the GSMAR model. **uGMAR** calculates the Akaike (AIC), Hannan-Quinn (HQIC), and Schwarz-Bayesian (BIC) information criteria. The values of the information criteria are not directly comparable for models with different autoregressive orders if estimation is based on the conditional log-likelihood function, as the numbers of observations used in estimation are different due to the different number of initial values. With the conditional log-likelihood function, the values of the information criteria can be divided by the number of observations used in the estimation (that is, the length of the series minus p) to obtain more comparable statistics. However, as the conditional estimation with each order p is based on slightly different observations, the comparison should be done with caution. The exact log-likelihood function, in contrast, employs the full series in estimation and thereby yields comparable values of information criteria for models with different orders p.

3.3.4 Examples of unconstrained estimation

In this section, I demonstrate how to estimate GSMAR models with **uGMAR** and provide several examples in order to illustrate various frequently occurring situations. In addition to the ordinary estimation, I particularly show how a GSMAR model can be built based on a local-only maximum point when the ML estimate seems unreasonable (see Appendix 3.A). I also consider the estimation of the appropriate G-StMAR model when the estimated StMAR model contains overly large degrees of freedom estimates (see Virolainen, forthcoming, Section 4).

In **uGMAR**, the GSMAR models are defined as class `gsmar` S3 objects, which can be created with given parameter values using the constructor function `GSMAR` (see Section 3.5) or by using the estimation function `fitGSMAR`, which estimates the parameters and then builds the model. For estimation, `fitGSMAR` needs to be supplied with a univariate time series and the arguments specifying the model. The necessary arguments for specifying the model include the autoregressive order `p`, the number of mixture components `M`, and `model`, which should be either `"GMAR"`, `"StMAR"`, or `"G-StMAR"`. For GMAR and StMAR models, the argument `M` is a positive integer, whereas for the G-StMAR model it is a length two numeric vector specifying the number of GMAR type regimes in the first element and the number of StMAR type regimes in the second.

Additional arguments may be supplied to `fitGSMAR` in order to specify, for example, whether the exact log-likelihood function should be used instead of the conditional one (`conditional`), whether the model should be parametrized with the intercepts $\varphi_{m,0}$ or the regimewise unconditional means μ_m (`parametrization`), how many estimation rounds should be performed (`ncalls`), and how many central processing unit (CPU) cores should be used in the estimation (`ncores`). Some of the estimation rounds may end up in local-only maximum points or saddle points, but reliability of the estimation results can be improved by increasing the number of estimation rounds. A large number of estimation rounds may be required particularly when the number of mixture components is large, as the surface of the log-likelihood function becomes increasingly more challenging. It is also possible to adjust the settings of the genetic algorithm that is used to find the starting values. The available options are listed in the documentation of the function `GAfit` to which the arguments adjusting the settings will be passed.

Section 3.3.3 concluded that a StMAR model with autoregressive order $p = 4$ and $M = 2$ mixture components seems like a reasonable candidate for modelling the monthly interest rate spread `M10Y1Y`. The following code fits this model to the series using the conditional log-likelihood function and performing 12 estimation rounds with eight CPU cores. The argument `seeds` supplies the seeds that initialize the random number generator at the beginning of each call to the genetic algorithm, thereby yielding reproducible results.

```
R> fit42t <- fitGSMAR(M10Y1Y, p = 4, M = 2, model = "StMAR",
+    conditional = TRUE, ncalls = 12, ncores = 8, seeds = 4:15)

Using 8 cores for 12 estimation rounds...
Optimizing with a genetic algorithm...
  |++++++++++++++++++++++++++++++++++++++++++++++++++| 100% elapsed=11s
Results from the genetic algorithm:
The lowest loglik:  143.403
The mean loglik:    159.237
The largest loglik: 172.344
Optimizing with a variable metric algorithm...
  |++++++++++++++++++++++++++++++++++++++++++++++++++| 100% elapsed=02s
Results from the variable metric algorithm:
The lowest loglik:  167.572
The mean loglik:    179.117
The largest loglik: 182.353
Finished!
Warning message:
In warn_dfs(p = p, M = M, params = params, model = model) :
  The model contains overly large degrees of freedom parameter values.
  Consider switching to a G-StMAR model by setting the corresponding regimes
  to be GMAR type with the function 'stmar_to_gstmar'.
```

The progression of the estimation process is reported with a progress bar giving an estimate of the remaining estimation time. Also statistics on the spread of the log-likelihoods are printed after each estimation phase. The progress bars are generated during parallel computing with the package **pbapply** (Solymos and Zawadzki, 2020).

The function throws a warning in the above example, because the model contains at least one very large degrees of freedom parameter estimate. Such estimates are warned about, because very large degrees of freedom parameters are redundant in the model and their weak identification might lead to numerical problems (see Section 2.4 in Chapter 2). Specifically, overly large degrees of freedom parameter estimates may induce a nearly numerically singular Hessian matrix of the log-likelihood function when evaluated at the estimate, making the approximate standard errors and Kalliovirta's (2012) quantile residual tests often unavailable.

The estimates can be examined with the `print` method:

```
R> fit42t

Model:
 StMAR, p = 4, M = 2, #parameters = 15, #observations = 468,
```

```
 conditional, intercept parametrization, not restricted, no constraints.

Regime 1
Mix weight: 0.81
Reg mean: 1.87
Var param: 0.04
Df param: 9.75

y = [0.06] + [1.28]y.1 + [-0.36]y.2 + [0.20]y.3 + [-0.15]y.4 + [sigma_mt]eps

Regime 2
Mix weight: 0.19
Reg mean: 0.55
Var param: 0.01
Df param: 9348.94

y = [0.04] + [1.34]y.1 + [-0.59]y.2 + [0.54]y.3 + [-0.36]y.4 + [sigma_mt]eps
```

The parameter estimates are reported for each mixture component separately so that the estimates can be easily interpreted. Each regime's autoregressive formula is presented in the form

$$y_t = \varphi_{m,0} + \varphi_{m,1}y_{t-1} + ... + \varphi_{m,p}y_{t-p} + \sigma_{m,t}\varepsilon_{m,t}. \tag{3.3.3}$$

The other statistics are listed above the formula, including the mixing weight parameter α_m, the unconditional mean μ_m, the variance parameter σ_m^2, and the degrees freedom parameter ν_m. For GMAR type regimes (if any), $\sigma_{m,t} = \sigma_m$ so the estimate of the variance parameter σ_m^2 is reported directly in the autoregressive formula.

The above printout shows that the second regime's degrees of freedom parameter estimate is very large, which might induce numerical problems. However, since a StMAR model with some degrees of freedom parameters tending to infinity coincides with the G-StMAR model with the corresponding regimes switched to GMAR type, one may avoid the problems by switching to the appropriate G-StMAR model (see Section 2.4 in Chapter 2). Switching to the appropriate G-StMAR model is recommended also because it removes the redundant degrees of freedom parameters from the model, thereby reducing its complexity. The function `stmar_to_gstmar` does this switch automatically by first removing the large degrees of freedom parameters and then estimating the G-StMAR model with a variable metric algorithm (Nash, 1990, Algorithm 21) using the induced parameter vector as the initial value.

To exemplify, the following code switches all the regimes of the StMAR model `fit42t` with a degrees of freedom parameter estimate larger than 100 to GMAR type, and then estimates the corresponding G-StMAR model.

```
R> fit42gs <- stmar_to_gstmar(fit42t, maxdf = 100)
```

The `summary` method can be used to obtain a more detailed printout of the estimated the G-StMAR model:

```
R> summary(fit42gs, digits = 2)
```

```
Model:
 G-StMAR, p = 4, M1 = 1, M2 = 1, #parameters = 14, #observations = 468,
 conditional, intercept parametrization, not restricted, no constraints.

 log-likelihood: 182.35, AIC: -336.71, HQIC: -313.89, BIC: -278.75

Regime 1 (GMAR type)
Moduli of AR poly roots: 1.16, 1.45, 1.45, 1.16
Mix weight: 0.19 (0.09)
Reg mean: 0.55
Reg var:  0.14

y = [0.04] + [1.34]y.1 + [-0.59]y.2 + [0.54]y.3 + [-0.36]y.4 + sqrt[0.01]eps
    (0.01)   (0.10)       (0.20)      (0.19)       (0.12)          (0.00)

Regime 2 (StMAR type)
Moduli of AR poly roots: 1.07, 2.02, 2.02, 1.51
Mix weight: 0.81
Reg mean: 1.87
Var param: 0.04 (0.01)
Df param: 9.75 (4.17)
Reg var:  1.01

y = [0.06] + [1.28]y.1 + [-0.36]y.2 + [0.20]y.3 + [-0.15]y.4 + [sigma_mt]eps
    (0.02)   (0.05)       (0.09)      (0.09)       (0.06)

Process mean: 1.62
Process var:  1.11
First p autocors: 0.98 0.96 0.93 0.89
```

In the G-StMAR model, estimates for GMAR type regimes are reported before StMAR type regimes, in a decreasing order according to the mixing weight parameter estimates. As shown above, the model `fit42gs` incorporates one GMAR type regime and one StMAR type regime. The mixing weight parameter estimate 0.19 of the GMAR type regime indicates that in the long run, roughly 19% of the observations are generated from this regime. Estimates of the unconditional mean and variance (0.55 and 0.14, respectively) are visibly smaller in the GMAR type regime than in the StMAR type regime (1.87 and 1.01, respectively). Hence, the GMAR type seems to mostly account for the periods when the series takes smaller values and is less volatile, while the StMAR type regime covers the more volatile periods of larger values. Interestingly, the AR parameters are somewhat similar in both regimes, implying that it could be appropriate to restrict them to be identical (this will be tested in Section 3.3.7).

Approximate standard errors are given in parentheses under or next to the related estimates. Note that the last mixing weight parameter estimate does not have an approximate standard error because it is not parametrized. Likewise, there is no standard error for the intercepts if mean parametrization is used (by setting `parametrization = "mean"` in `fitGSMAR`) and vice versa. In order to obtain standard errors for the regimewise unconditional means or intercepts, one can easily swap between the mean and intercept parametrizations with the function

`swap_parametrization`.

Missing values are reported when **uGMAR** is not able to calculate the standard error. This typically happens either because there is an overly large degrees of freedom parameter estimate in the model (as discussed above) or because the estimation algorithm did not stop a local maximum. In the latter case, the observed information matrix is not necessarily positive definite, implying that the diagonal entries of its inverse might not all be positive. Consequently, when extracting the approximate standard errors by taking the square roots of the diagonal entries from the inverse of the observed information matrix, the possibly present negative entries will lead to missing values.

Section 3.3.5 discusses how to use the tools in **uGMAR** to evaluate whether the estimate is a local maximum (and how to improve the reliability of it being the global maximum). If the estimate is not a local maximum, one may try running more iterations of the variable metric algorithm with the function `iterate_more`. However, often when the algorithm does not stop a local maximum, it stopped to an unreasonable point very near the boundary of the parameter space. As will be discussed next, in such a case it might be more appropriate to consider an alternative estimate that is clearly in the interior of the parameter space.

Other statistics reported in the summary printout include the log-likelihood and values of the information criteria, the first and second moments of the process, as well as regime-specific unconditional means, unconditional variances, and moduli of the roots of the AR polynomials $1-\sum_{i=1}^{p}\varphi_{m,i}z^i,\ m=1,...,M$. If some of the moduli are very close to one, the related estimates are near the boundary of the stationarity region. I demonstrate in Appendix 3.A that when such solutions are accompanied with a very small variance parameter estimate, they might not be reasonable estimates and possibly maximize the log-likelihood function for a technical reason only. Consequently, the estimate related to the next-largest local maximum could be considered.

This is possible in **uGMAR**, because the estimation function `fitGSMAR` stores the estimates from all the estimation rounds so that a GSMAR model can be built based on any one of them, most conveniently with the function `alt_gsmar`. The desired estimation round can be specified either with the argument `which_round` or `which_largest`. The former specifies the round in the estimation order, whereas the latter specifies it in a decreasing order of the log-likelihoods.

To give an example of a case where the estimates are very close the boundary of the stationarity region, I estimate the G-StMAR model directly with the following code.

```
R> fit42gs2 <- fitGSMAR(M10Y1Y, p = 4, M = c(1, 1), model = "G-StMAR",
+    conditional = TRUE, ncalls = 16, ncores = 8, seeds = 72:87)

Using 8 cores for 16 estimation rounds...
Optimizing with a genetic algorithm...
  |++++++++++++++++++++++++++++++++++++++++++++++++++| 100% elapsed=12s
Results from the genetic algorithm:
The lowest loglik:  140.441
The mean loglik:    155.421
The largest loglik: 167.858
Optimizing with a variable metric algorithm...
  |++++++++++++++++++++++++++++++++++++++++++++++++++| 100% elapsed=02s
```

```
Results from the variable metric algorithm:
The lowest loglik:  152.034
The mean loglik:    174.794
The largest loglik: 192.43
Finished!
Warning message:
In warn_ar_roots(ret) :
  Regime 1 has near-unit-roots! Consider building a model from the next-
  largest local maximum with the function 'alt_gsmar' by adjusting its
  argument 'which_largest'.
```

The function throws a warning, because the largest found maximum point incorporates a regime that is very close to the boundary of the stationarity region, indicating that the estimate might be inappropriate. The `summary` method produces the following printout for the model:

```
R> summary(fit42gs2, digits = 2)

Model:
 G-StMAR, p = 4, M1 = 1, M2 = 1, #parameters = 14, #observations = 468,
 conditional, intercept parametrization, not restricted, no constraints.

 log-likelihood: 192.43, AIC: -356.86, HQIC: -334.05, BIC: -298.90

Regime 1 (GMAR type)
Moduli of AR poly roots: 1.00, 1.00, 1.00, 1.00
Mix weight: 0.02 (0.03)
Reg mean: 2.65
Reg var:  0.13

y = [3.77] + [1.19]y.1 + [-1.81]y.2 + [1.19]y.3 + [-1.00]y.4 + sqrt[0.00]eps
    (0.02)   (0.01)       (0.01)      (0.01)       (0.00)          (0.00)

Regime 2 (StMAR type)
Moduli of AR poly roots: 1.04, 1.93, 1.93, 1.48
Mix weight: 0.98
Reg mean: 0.89
Var param: 0.04 (0.01)
Df param: 4.98 (1.67)
Reg var:  1.75

y = [0.02] + [1.30]y.1 + [-0.36]y.2 + [0.21]y.3 + [-0.17]y.4 + [sigma_mt]eps
    (0.01)   (0.05)       (0.08)      (0.08)       (0.05)

Process mean: 0.92
Process var:  1.78
First p autocors: 0.99 0.97 0.95 0.93
```

The summary statistics reveal that there are four near-unit-roots in the GMAR type regime and the variance parameter estimate is very small. Such estimates often occur when there are several regimes in the model and the estimation algorithm is ran a large number of times.

If the estimate is deemed inappropriate, it is easy to build a model based on the second-largest maximum point that was found in the estimation procedure. Below, the first line of the code builds the model based on the second-largest maximum point, and the second line calls the `summary` method to produce a detailed printout of the model.

```
R> fit42gs3 <- alt_gsmar(fit42gs2, which_largest = 2)
R> summary(fit42gs3, digits = 2)
```

```
Model:
 G-StMAR, p = 4, M1 = 1, M2 = 1, #parameters = 14, #observations = 468,
 conditional, intercept parametrization, not restricted, no constraints.

 log-likelihood: 182.35, AIC: -336.71, HQIC: -313.89, BIC: -278.75

Regime 1 (GMAR type)
Moduli of AR poly roots: 1.16, 1.45, 1.45, 1.16
Mix weight: 0.19 (0.09)
Reg mean: 0.55
Reg var:  0.14

y = [0.04] + [1.34]y.1 + [-0.59]y.2 + [0.54]y.3 + [-0.36]y.4 + sqrt[0.01]eps
    (0.01)   (0.10)       (0.20)      (0.19)       (0.12)          (0.00)

Regime 2 (StMAR type)
Moduli of AR poly roots: 1.07, 2.02, 2.02, 1.51
Mix weight: 0.81
Reg mean: 1.87
Var param: 0.04 (0.01)
Df param: 9.75 (4.14)
Reg var:  1.01

y = [0.06] + [1.28]y.1 + [-0.36]y.2 + [0.20]y.3 + [-0.15]y.4 + [sigma_mt]eps
    (0.02)   (0.05)       (0.09)      (0.09)       (0.06)

Process mean: 1.62
Process var:  1.11
First p autocors: 0.98 0.96 0.93 0.89
```

The above printout shows that the estimates related to the second-largest local maximum are the same as of the model `fit42gs` (which was estimated based on a StMAR model with a very large degrees of freedom parameter estimate) and that they are clearly inside the stationarity region for all regimes. If also the second-largest maximum point seems unreasonable, a GSMAR model can be built based on the next-largest maximum point by adjusting the argument `which_largest` in the function `alt_gsmar` accordingly.

3.3.5 Further examination of the estimates

In addition to examining the summary printout, it is often useful to visualize the model by plotting the mixing weights together with the time series and the model's (marginal) stationary

density together with a kernel density estimate of the time series. That is exactly what the plot method for GSMAR models does. For instance, the following command creates Figure 3.2:

```
R> plot(fit42gs)
```

As Figure 3.2 (the top and bottom left panels) shows, the first regime prevails when the spread takes small values, while the second regime mainly dominates when the spread takes large values. The graph of the model's marginal stationary density (the right panel), on the other hand, shows that the two regimes capture the two modes in the marginal distribution of the spread. The hump shape in the right tail of the kernel density estimate is not captured by the mixture of the two distributions, but a third regime could be added for the purpose (the three regime model is not studied for brevity).

It is also sometimes interesting to examine the time series of (one-step) conditional means and variances of the process along with the time series the model was fitted to. This can be done conveniently with the function `cond_moment_plot`, where the argument `which_moment` should be specified with `"mean"` or `"variance"` accordingly. In addition to the conditional moment of the process, `cond_moment_plot` also displays the conditional means or variances of the regimes multiplied by the mixing weights. Note, however, that the conditional variance of the process is not generally the same as the weighted sum of regimewise conditional variances, as it includes a component that encapsulates heteroskedasticity caused by variation in the conditional mean (see Equation (2.2.19) in Chapter 2).

The variable metric algorithm employed in the final estimation does not necessarily stop at a local maximum point. The algorithm might also stop at a saddle point or near a local maximum, when the algorithm is not able to increase the log-likelihood, or at any point, when the maximum number of iterations has been reached. In the latter case, the estimation function throws a warning, but saddle points and inaccurate estimates need to be detected by the researcher.

It is well known that in a local maximum point, the gradient of the log-likelihood function is zero, and the eigenvalues of the Hessian matrix are all negative. In a local minimum, the eigenvalues of the Hessian matrix are all positive, whereas in a saddle point, some of them are positive and some negative. Nearly numerically singular Hessian matrices occur when the surface of the log-likelihood function is very flat about the estimate in some directions. This particularly happens when the model contains overly large degrees of freedom parameter estimates or the mixing weights $\alpha_{m,t}$ are estimated close to zero for all $t = 1, ..., T$ for some regime m.

uGMAR provides several functions for evaluating whether the estimate is a local maximum point. The function `get_foc` returns the (numerically approximated) gradient of the log-likelihood function evaluated at the estimate, and the function `get_soc` returns eigenvalues of the (numerically approximated) Hessian matrix of the log-likelihood function evaluated at the estimate. The numerical derivatives are calculated using the central difference approximation

$$\frac{\partial L(\boldsymbol{\theta})}{\partial \theta_i} \approx \frac{f(\boldsymbol{\theta} + \boldsymbol{h}^{(i)}) - f(\boldsymbol{\theta} - \boldsymbol{h}^{(i)})}{2h}, \quad h > 0, \tag{3.3.4}$$

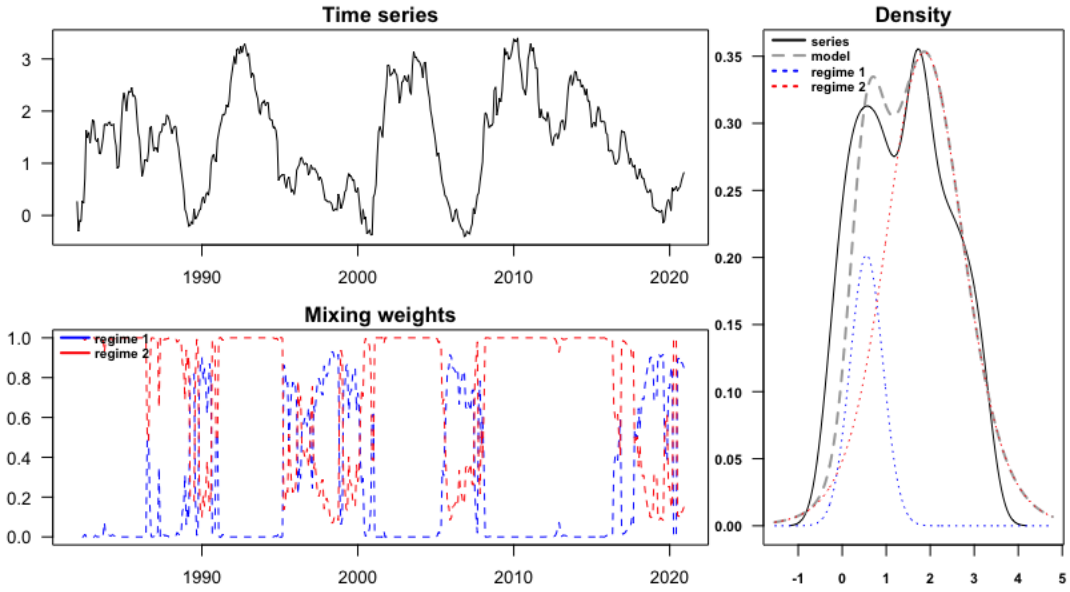

Figure 3.2: The figure produced by the command `plot(fit42gs)`. On the top left, the monthly spread between the 10-year and 1-year Treasury constant maturity rates, covering the period from 1982 January to 2020 December. On the bottom left, the estimated mixing weights of the G-StMAR model (`fit42gs`) fitted to the interest rate spread (blue dashed line for the first regime and red dashed line for the second regime). On the right, the one-dimensional marginal stationary density of the estimated G-StMAR model (grey dashed line) along with a kernel density estimate of the spread (black solid line) and marginal stationary densities of the regimes multiplied by the mixing weight parameter estimates (blue and red dotted lines).

where θ_i is the ith element of $\boldsymbol{\theta}$ and $\boldsymbol{h}^{(i)} = (0, ..., 0, h, 0, ..., 0)$ contains h as its ith element. By default, the difference $h = 6 \cdot 10^{-6}$ is used for all parameters except for overly large degrees of freedom parameters, whose partial derivatives are approximated using larger differences. The difference is increased for large degrees of freedom parameters, because the limited precision of the float point presentation induces artificially rugged surfaces to the their profile log-likelihood functions, and the increased differences diminish the related numerical error. On the other hand, as the surface of the profile log-likelihood function is very flat about a large degrees of freedom parameter estimate, large differences work well for the approximation.

For example, the following code calculates the first order condition for the G-StMAR model `fit42gs`:

```
R> get_foc(fit42gs)

 [1]  0.0576396128 -0.0364233988 -0.0242331476 -0.0144442609 -0.0161249574
 [6]  0.0411603528 -0.0171471584 -0.0490156277 -0.0659635759 -0.0587742714
[11] -0.0635655297  0.0686981920 -0.0374653647  0.0002778317
```

and the following code calculates the second order condition:

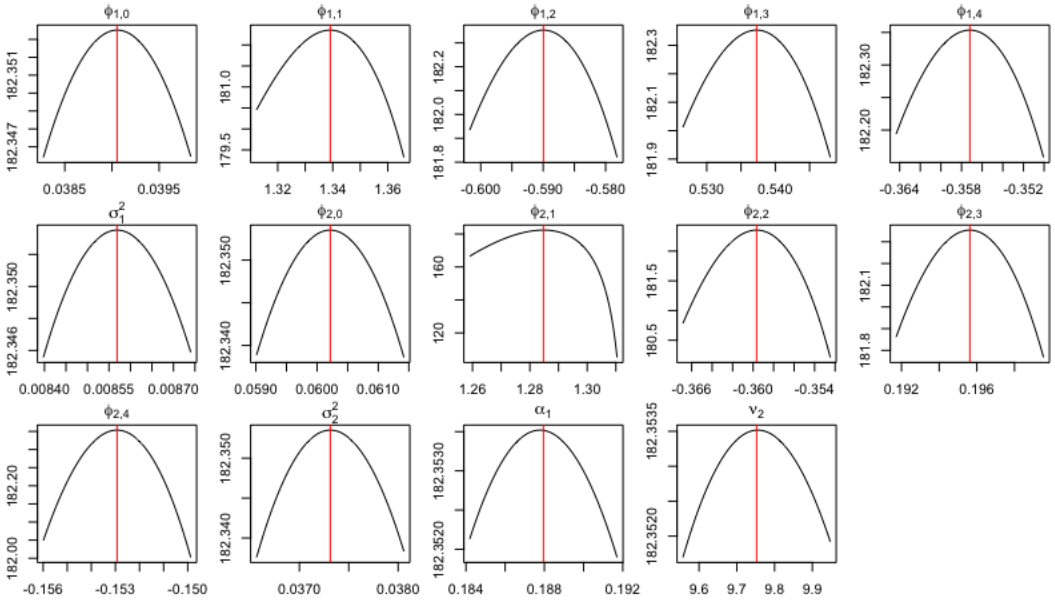

Figure 3.3: The figure produced by the command `profile_logliks(fit42gs)`. Graphs of the profile log-likelihood functions of the estimated G-StMAR model `fit42gs` with the red vertical lines pointing the estimates.

```
R> get_soc(fit42gs)

 [1] -5.753554e-02 -1.354508e+01 -4.394382e+01 -6.467642e+01 -1.204519e+02
 [6] -1.672692e+02 -2.619181e+02 -8.869383e+02 -2.045380e+03 -4.862797e+03
[11] -4.355348e+04 -5.455077e+04 -2.727695e+05 -5.564824e+05
```

All eigenvalues of the Hessian matrix are negative, which points to a local maximum, but the gradient of the log-likelihood function seems to somewhat deviate from zero. The gradient might be inaccurate, because it is based on a numerical approximation. It is also possible that the estimate is inaccurate, because it is based on approximative numerical estimation, and the estimates are therefore not expected to be exactly accurate. Whether the estimate is a local maximum point with accuracy that is reasonable enough, can be evaluated by plotting the graphs of the profile log-likelihood functions about the estimate. In **uGMAR**, this can be done conveniently with the function `profile_logliks`.

The exemplify, the following command plots the graphs of profile log-likelihood functions of the estimated G-StMAR model `fit42gs`:

```
R> profile_logliks(fit42gs, scale = 0.02, precision = 200)
```

The output is displayed in Figure 3.3, showing that the estimate's accuracy is reasonable, as changing any individual parameter value marginally would not visibly increase the log-likelihood. The argument `scale` can be adjusted to shorten or lengthen the interval shown in the horizontal

axis. If one zooms in enough by setting `scale` to a very small number, it can be seen that the estimate is not exactly at the local maximum, but it is so close that moving there would not increase the log-likelihood notably. The argument `precision` can be adjusted to increase the number of points the graph is based on. For faster plotting, it can be decreased, and for more precision, it can be increased.

I have discussed tools that can be utilized to evaluate whether the found estimate is a local maximum with a reasonable accuracy. It is, however, more difficult to establish that the estimate is the global maximum. With **uGMAR**, the best way to increase the reliability that the found estimate is the global maximum, is to run more estimation rounds by adjusting the argument `ncalls` of the estimation function `fitGSMAR`. When a large number of estimation rounds is run (and $M > 1$), `fitGSMAR` often finds peculiar near-the-boundary estimates that have extremely spiky profile log-likelihood functions for some parameters and are thus difficult to find (see Appendix 3.A). Therefore, it seems plausible that `fitGSMAR` also finds a reasonable ML estimate with a good reliability.

3.3.6 Examples of constrained estimation

Alternatively to the unconstrained estimation, linear constraints can be imposed on the autoregressive (AR) parameters of the model; that is, on $\varphi_{m,1}, ..., \varphi_{m,p}$, $m = 1, ..., M$. **uGMAR** deploys two types of constraints: the AR parameters can be restricted to be identical in all regimes and linear constraints can be applied to each regime separately. In order to impose the former type of constraints, the estimation function simply needs to be supplied with the argument `restricted = TRUE`.

For instance, the G-StMAR, $p = 4$, $M_1 = 1$, $M_2 = 1$ model (`fit42gs`) estimated in Section 3.3.4 obtained somewhat similar estimates for the AR parameters in both regimes. The following code estimates a version of this model such that the AR parameters are restricted to be identical in both regimes. Note that this model still allows for shifts in the mean, as the intercept parameters can vary across the regimes. The argument `print_res = FALSE` tells `fitGSMAR` not to the print the spread of the log-likelihoods obtained from each phase of estimation.

```
R> fit42gsr <- fitGSMAR(M10Y1Y, p = 4, M = c(1, 1), model = "G-StMAR",
+    restricted = TRUE, ncalls = 12, ncores = 8, seeds = 1:12,
+    print_res = FALSE)

Using 8 cores for 12 estimation rounds...
Optimizing with a genetic algorithm...
  |++++++++++++++++++++++++++++++++++++++++++++++++++| 100% elapsed=07s
Optimizing with a variable metric algorithm...
  |++++++++++++++++++++++++++++++++++++++++++++++++++| 100% elapsed=01s
Finished!
```

The summary printout of the model shows the AR parameter estimates are the same in both regimes:

```
R> summary(fit42gsr)

Model:
 G-StMAR, p = 4, M1 = 1, M2 = 1, #parameters = 10, #observations = 468,
 conditional, intercept parametrization, AR parameters restricted, no
 constraints.

 log-likelihood: 180.02, AIC: -340.04, HQIC: -323.74, BIC: -298.64

Regime 1 (GMAR type)
Moduli of AR poly roots: 1.21, 1.83, 1.83, 1.21
Mix weight: 0.51 (0.17)
Reg mean: 2.13
Reg var:  0.46

y = [0.13] + [1.29]y.1 + [-0.40]y.2 + [0.25]y.3 + [-0.20]y.4 + sqrt[0.03]eps
    (0.03)   (0.05)       (0.08)      (0.08)       (0.05)          (0.00)

Regime 2 (StMAR type)
Moduli of AR poly roots: 1.21, 1.83, 1.83, 1.21
Mix weight: 0.49
Reg mean: 0.54
Var param: 0.05 (0.06)
Df param: 2.76 (1.18)
Reg var:  0.83

y = [0.03] + [1.29]y.1 + [-0.40]y.2 + [0.25]y.3 + [-0.20]y.4 + [sigma_mt]eps
    (0.01)   (0.05)       (0.08)      (0.08)       (0.05)

Process mean: 1.35
Process var:  1.27
First p autocors: 0.98 0.95 0.91 0.87
```

In constrast to the unrestricted model, this model has larger regimewise unconditonal mean in the GMAR type regime than in the StMAR type regime. According to the unconditional regimewise variances, the StMAR type regime is the more volatilite regime in this model as well.

Whether imposing the constraints is reasonable, can be evaluated by employing a statistical test, comparing values of the information criteria, or examining the model adequacy, for example. As the summary printout shows, the information criteria values all decreased as opposed to the unrestricted model, implying that the constraints could be appropriate. Discussion on testing the constraints is postponed to Section 3.3.7, whereas diagnostics checks for evaluating the model adequacy are covered in Section 3.4.

The other type of constraints in **uGMAR** are of the form

$$\boldsymbol{\varphi}_m = \boldsymbol{C}_m\boldsymbol{\psi}_m, \quad m = 1, ..., M, \tag{3.3.5}$$

where $\boldsymbol{C}_m$ is a known $(p \times q_m)$ constraint matrix with full column rank, $\boldsymbol{\psi}_m$ is a $(q_m \times 1)$ parameter vector, and $\boldsymbol{\varphi}_m = (\varphi_{m,1}, ..., \varphi_{m,p})$ contains the AR coefficients of the mth regime.

In order to apply the constraints, the estimation function should be supplied with the argument `constraints` containing a list of the constraint matrices $\boldsymbol{C}_m$, $m = 1, ..., M$.

To exemplify, consider a GMAR model with autoregressive order $p = 3$ and $M = 2$ mixture components. To constrain the third AR coefficient of the second regime ($\varphi_{2,3}$) to zero but leaving the first regime unconstrained, I deploy the following list of constraint matrices:

```
R> C_list <- list(diag(3), matrix(c(1, 0, 0, 0, 1, 0), nrow = 3))
R> C_list

[[1]]
     [,1] [,2] [,3]
[1,]    1    0    0
[2,]    0    1    0
[3,]    0    0    1

[[2]]
     [,1] [,2]
[1,]    1    0
[2,]    0    1
[3,]    0    0
```

After setting up the constraints, the constrained model can be estimated as follows:

```
R> fit32c <- fitGSMAR(M10Y1Y, p = 3, M = 2, model = "GMAR",
+    constraints = C_list, ncalls = 12, ncores = 8, seeds = 1:12,
+    print_res = FALSE)

Using 8 cores for 12 estimation rounds...
Optimizing with a genetic algorithm...
  |++++++++++++++++++++++++++++++++++++++++++++++++++| 100% elapsed=05s
Optimizing with a variable metric algorithm...
  |++++++++++++++++++++++++++++++++++++++++++++++++++| 100% elapsed=01s
Finished!
```

Printout of the model shows that the third AR parameter estimate of the second regime is zero:

```
R> fit32c

Model:
 GMAR, p = 3, M = 2, #parameters = 10, #observations = 468,
 conditional, intercept parametrization, not restricted, linear constraints
 imposed.

Regime 1
Mix weight: 0.56
Reg mean: 1.26

y = [0.02] + [1.25]y.1 + [-0.19]y.2 + [-0.07]y.3 + sqrt[0.01]eps
```

```
Regime 2
Mix weight: 0.44
Reg mean: 1.72

y = [0.07] + [1.27]y.1 + [-0.32]y.2 + [0.00]y.3 + sqrt[0.05]eps
```

Notice that even when the pth AR coefficient is restricted to zero, the pth lag of that regime is accounted for in the mixing weights (3.2.6) and in the case of a StMAR type regime also in the conditional variance (3.2.3).

If both types of constraints are applied at the same time, only a single constraint matrix should be supplied (not in a list). Consider a GSMAR model with $p = 2$ and $M = 2$, for example, and suppose the AR coefficients should be restricted to be identical in both regimes and the second AR coefficient ($\varphi_{m,2}$) should be constrained to be the negative of the first coefficient ($\varphi_{m,1}$). Then, the estimation function should be supplied with the arguments `restricted = TRUE` and `constraints = matrix(c(1, -1), nrow = 2)`. As demonstrated above, **uGMAR**'s implementation for applying linear constraints is not the most general one, but it makes applying some of the most typical constraints convenient, as the constraint matrices remain small.

3.3.7 Testing parameter constraints

One way to asses the validity of the imposed constraints is to compare the values of information criteria of the constrained and unconstrained models. **uGMAR**, nonetheless, also provides functions for testing the constraints with the likelihood ratio test and the Wald test, which are applicable as the ML estimator of a GSMAR model has the conventional asymptotic distribution (as long as the model is correctly specified and one is willing to assume the validity of the required unverified assumptions, see Kalliovirta *et al.*, 2015, pp. 254-255, Meitz *et al.*, forthcoming, Theorem 3, and Chapter 2, Theorem 2.2). For a discussion on the likelihood ratio and Wald tests, see Buse (1982) and the references therein, for example.

The likelihood ratio test considers the null hypothesis that the true parameter value $\boldsymbol{\theta}_0$ satisfies some constraints imposed on these parameters (such that the constrained parameter space is a subset of the parameter space, which is presented in Equation (2.2.16) in Chapter 2 for the GSMAR models). Denoting by $\hat{L}_U$ and $\hat{L}_C$ the (maximized) log-likelihoods based on the unconstrained and constrained ML estimates, respectively, the test statistic takes the form

$$LR = 2(\hat{L}_U - \hat{L}_C). \tag{3.3.6}$$

Under the null, the test statistic is asymptotically χ^2-distributed with the degrees of freedom given by the difference in the dimensions of the unconstrained and constrained parameter spaces.

With **uGMAR**, the likelihood ratio test can be calculated with the function `LR_test`, which takes the unconstrained model (a class `gsmar` object) as its first argument and the constrained model as the second argument. For instance, in Section 3.3.6, I estimated a G-StMAR, $p = 4$, $M_1 = 1$, $M_2 = 1$ model such that the AR parameters are restricted to be identical in both

regimes (the model `fit42gsr`), i.e., $\boldsymbol{\varphi}_1 = \boldsymbol{\varphi}_2$. The following code tests those constraints against the unconstrained model `fit42gs` with the likelihood ratio test and prints the results.

```
R> LR_test(fit42gs, fit42gsr)

        Likelihood ratio test

data:  fit42gs and fit42gsr
LR = 4.6695, df = 4, p-value = 0.3229
alternative hypothesis: the true parameter does not satisfy the constraints
imposed in fit42gsr
```

The large p-value indicates that I cannot reject the constraints at any conventional level of significance, and it might thereby be reasonable to consider the constrained model if it is found adequate.

uGMAR implements the Wald test of the null hypothesis

$$A\boldsymbol{\theta}_0 = c, \tag{3.3.7}$$

where A is a $(k \times d)$ matrix with full row rank, c is a $(k \times 1)$ vector, $\boldsymbol{\theta}_0$ is the true parameter value, d is the dimension of the parameter space, and k is the number of constraints. The Wald test statistic takes the form

$$W = (A\hat{\boldsymbol{\theta}} - c)'[A\mathcal{J}(\hat{\boldsymbol{\theta}})^{-1}A']^{-1}(A\hat{\boldsymbol{\theta}} - c), \tag{3.3.8}$$

where $\mathcal{J}(\hat{\boldsymbol{\theta}})$ is the observed information matrix evaluated at the ML estimate $\hat{\boldsymbol{\theta}}$. Under the null, the test statistic is asymptotically χ^2-distributed with k degrees of freedom (which is the difference in the dimensions of the constrained and unconstrained parameter spaces).

With **uGMAR**, the Wald test can be calculated with function `Wald_test`, which takes the estimated unconstrained model (as a class `gsmar` object) as the first argument, the matrix A as the second argument, and the vector c as the third argument. To exemplify, I test whether the AR parameters and intercepts are identical in both regimes of the G-StMAR, $p = 4$, $M_1 = 1$, $M_2 = 1$ model, i.e., the null hypothesis $(\varphi_{1,0}, \boldsymbol{\varphi}_1) = (\varphi_{2,0}, \boldsymbol{\varphi}_2)$. The $(d \times 1)$ parameter vector $\boldsymbol{\theta}$ (which is presented at the end of Section 3.2.2 and again in Section 3.5) contains the intercept and AR parameters of the first regime in the entries $1, ..., 5$ and the intercept and AR parameters of the second regime in the entries $7, ..., 11$. The appropriate matrix A and vector c that state the hypothesis are set in the first two lines of the following code, and the third line calculates the test.

```
R> c <- rep(0, times = 5)
R> A <- cbind(diag(5), c, -diag(5), c, c)
Wald_test(fit42gs, A = A, c = c)

        Wald test

data:  fit42gs, A, c
W = 15.107, df = 5, p-value = 0.009916
alternative hypothesis: the true parameter theta does not satisfy
A%*%theta = c
```

As the above printout shows, the p-value is small enough to reject the null at the 1% level of significance, even though the null hypothesis that the AR parameters are equal in both regimes could not be rejected by the likelihood ratio test. Using the model `fit42gsr` to calculate a Wald test, that tests equality of the intercepts conditional on the constraint that the AR parameters are identical in both regimes, produces the p-value 0.00025 (not shown for brevity). Thus, the intercepts are not likely equal if the AR parameters are identical in both regimes.[5] As is demonstrated above, the Wald test has the benefit that it does not require estimation of the constrained model, and it is, therefore, not limited to the type of constraints **uGMAR** accommodates. The likelihood ratio test, on the other hand, is more conveniently calculated once the constrained model has been estimated.

Note that the standard tests are not applicable if the number of GMAR or StMAR type regimes is chosen too large, as then some of the parameters are not identified, causing the result of the asymptotic normality of the ML estimator to break down. This particularly happens when one tests for the number of regimes in the model, as under the null some of the regimes are reduced from the model[6] (see the related discussion in Kalliovirta *et al.*, 2015, Section 3.3.2). Similar caution applies for testing whether a regime is of the GMAR type against the alternative that it is of the StMAR type. Then $\nu_m = \infty$ under the null for the regime m to be tested, which violates the assumption that the parameter value is in the interior of a compact subset of the parameter space (see Theorem 2.2 and Assumption 2.1 in Chapter 2)

3.4 Quantile residual based model diagnostics

In the GSMAR models, the empirical counterparts of the error terms $\varepsilon_{m,t}$ in (3.2.1) cannot be calculated, because the regime that generated each observation is unknown, making the conventional residual based diagnostics unavailable. Therefore, **uGMAR** utilizes so called *quantile residuals*, which are suitable for evaluating adequacy of the GSMAR models. Deploying the framework presented in Kalliovirta (2012), quantile residuals are defined as

$$R_t = \Phi^{-1}(F(y_t|\mathcal{F}_{t-1})), \quad t = 1, 2, ..., T, \tag{3.4.1}$$

where $\Phi^{-1}(\cdot)$ is the standard normal quantile function and $F(\cdot|\mathcal{F}_{t-1})$ is the conditional cumulative distribution function of the considered GSMAR process (conditional on the previous observations). Closed form expressions for the quantile residuals of the GSMAR processes are derived in Appendix 3.B.

[5] The test results do not, however, allow to infer that the process is likely bimodal, because GSMAR processes incorporating component processes with distinct means can have unimodal skewed marginal distributions. Moreover, one cannot infer about the (in)equality of the means of the component processes based on the (in)equality of the intercepts if the AR parameters are allowed vary freely. In particular, our null hypothesis $(\varphi_{1,0}, \boldsymbol{\varphi}_1) = (\varphi_{2,0}, \boldsymbol{\varphi}_2)$ does not test whether the component processes have identical means, as identical means can be obtained also with various other constraints. Identicality of the means can, however, be tested directly by switching to the mean parametrization (with the function `swap_parametrization`) and calculating the appropriate Wald test.

[6] Meitz and Saikkonen (2021) have, however, recently developed such tests for mixture models with Gaussian conditional densities.

The empirical counterparts of the quantile residuals are calculated by using the parameter estimate and the observed data in (3.4.1). For a correctly specified GSMAR model, the empirical counterparts of the quantile residuals based on the ML estimator are asymptotically independent with standard normal distributions (Kalliovirta, 2012, Lemma 2.1). Hence, quantile residuals can be used for graphical analysis similarly to the conventional Pearson residuals.

In **uGMAR**, quantile residuals can be analyzed graphically with the function `diagnostic_plot`, which plots the quantile residual time series, normal quantile-quantile plot, and sample autocorrelation functions of the quantile residuals and squared quantile residuals. If one sets `plot_indstats = TRUE` in the function arguments, `diagnostic_plot` also plots the standardized individual statistics discussed in Kalliovirta (2012, pp. 369-370) with their approximate 95% critical bounds.

The individual statistics, which test for remaining autocorrelation or heteroskedasticity in specific lags, can be calculated either based on the observed data or based on the simulation procedure proposed by Kalliovirta (2012). In the simulation procedure, the individual statistics' approximate standard errors are based on a sample simulated from the estimated process. According to Kalliovirta's (2012) Monte Carlo study, the simulation procedure may improve size properties of the related tests, but it makes calculation of the statistics computationally more demanding - particularly if the simulated sample is very large.

The likelihood ratio test calculated in Section 3.3.7 accepted hypothesis that the AR coefficients of the G-StMAR $p = 4$, $M_1 = 1$, $M_2 = 1$ model are identical in both regimes. In order to evaluate whether this constrained model (`fit42gsr`) can adequately capture the autocorrelation structure, conditional heteroskedasticity, and distribution of the series, I create a diagnostic plot with the following code. I include Kalliovirta's (2012) individual statistic to the figure based on the observed data and calculated for the first 20 lags.

```
R> diagnostic_plot(fit42gsr, nlags = 20, plot_indstats = TRUE)
```

The resulting plot is presented in Figure 3.4. The quantile residual time series (the top left panel) has a period when it takes several consecutive negative values (roughly the observations $260, ..., 300$ with also some positive observations in between), but other than that it seems to somewhat resemble an IID standard normal process. The normal quantile-quantile plot (the top right panel) shows that the quantile residuals' distribution has too fat right tail. This is possibly due to the inability to explain the hump shape in the right tail of the series' distribution with a mixture of one normal and one t-distribution, when the two modes are accounted for.

The quantile residuals' sample autocorrelation function (the middle left panel) shows that there are no particularly large autocorrelation coefficients in the lags $1, ..., 20$. Moreover, as all Kalliovirta's (2012) autocorrelation statistics fall inside the asymptotic 95% critical bounds, the model seems to adequately describe the autocorrelation structure of the series. The sample autocorrelation function of the squared quantile residuals (the middle right panel), on the other hand, has a relatively large coefficient at the lag eight. Kalliovirta's (2012) conditional heteroskedasticity statistics (the bottom right panel) fall outside the asymptotic 95% critical bounds at the lags four and six, but at the lag eight the statistic is inside the bounds. Overall, it ap-

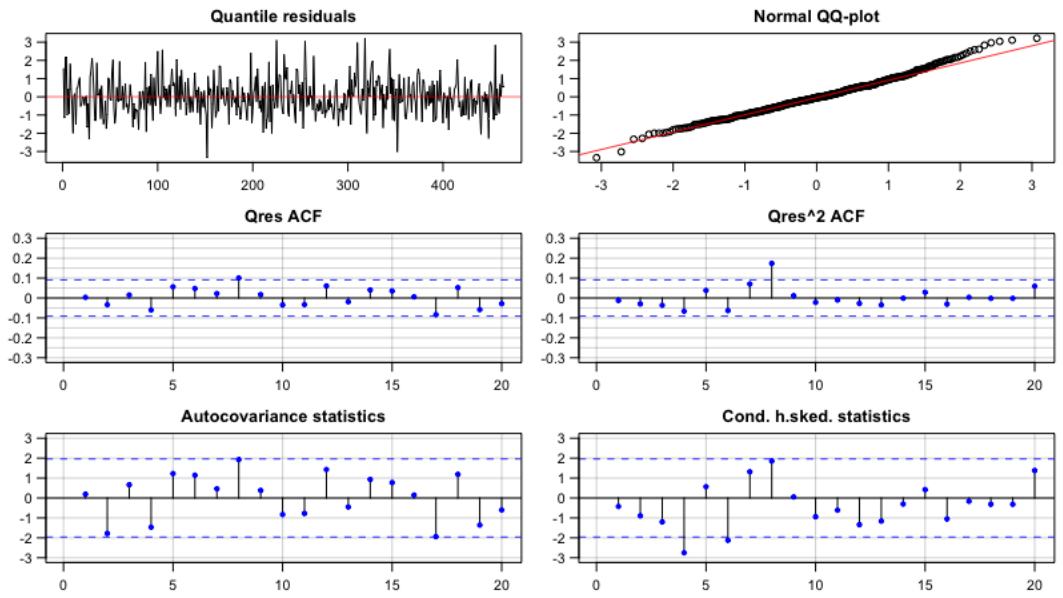

Figure 3.4: Diagnostic plot for the fitted model `fit42gsr` created using the function `diagnostic_plot`. The quantile residual time series (top left), normal quantile-quantile plot (top right), sample autocorrelation functions of the quantile residuals (middle left) and squared quantile residuals (middle right), and the individual autocorrelation (bottom left) and heteroskedasticity (bottom right) statistics discussed in Kalliovirta (2012, pp. 369-370). The blue dashed lines in the sample autocorrelation figures are the $1.96T^{-1/2}$ lines denoting 95% critical bounds for IID-observations, whereas for Kalliovirta's (2012) individual statistics they are the approximate 95% critical bounds.

pears that in addition to the distribution, the model might not adequately explain the conditional heteroskedasticity of the series.

In order to employ the simulation procedure for calculating the individual statistics, one needs to set the length of the simulated sample with the argument `nsimu`. If `nsimu` is not larger than the length of the observed data, the statistics will be based on the observed data. In addition to `diagnostic_plot`, quantile residuals can be graphically examined with the function `quantile_residual_plot`, which plots the quantile residual time series and a histogram.

Analyzing quantile residuals graphically gives an overview of the model's adequacy, but it is often appealing to also carry out a formal testing procedure. Kalliovirta (2012) proposes three specific tests for testing normality, autocorrelation, and conditional heteroskedasticity of the quantile residuals. Kalliovirta's (2012) tests take into account the uncertainty caused by estimation of the parameters and they are shown to perform well in a simulation study (Kalliovirta, 2012, Section 4).

In **uGMAR**, the quantile residual tests can be calculated with the function `quantile_residual_tests`, whose arguments include the model and the numbers of lags to be included in the autocorrelation (`lags_ac`) and heteroskedasticity (`lags_ch`) tests. Similarly to the individual statistics discussed

in the context of the diagnostic plot, the tests can be based either on the observed data or on the simulation procedure. The simulation procedure can be deployed by setting the argument `nsimu` to be larger than the number of observations.

The following code calculates the quantile residual tests for the restricted G-StMAR model `fit42gsr` by deploying the simulation procedure based on a simulated sample of length 10000 and taking into account $1, 3, 6$, and 12 lags in the autocorrelation and heteroskedasticity tests. By default, the lags for the heteroskedasticity tests are the same as for the autocorrelation tests, so it is enough to set the autocorrelation test lags with the argument `lags_ac`.

```
R> set.seed(1)
R> qrtr <- quantile_residual_tests(fit42gsr, lags_ac = c(1, 3, 6, 12),
+    nsimu = 10000)

Normality test p-value: 0.018

Autocorrelation tests:
lags | p-value
  1  |  0.849
  3  |  0.084
  6  |  0.488
  12 |  0.213

Conditional heteroskedasticity tests:
lags | p-value
  1  |  0.713
  3  |  0.299
  6  |  0.017
  12 |  0.000
```

The test results reveal that the model does not seem to adequately capture the conditional heteroskedasticity in the series when taking into account 12 lags. Also, the normality test and the heteroskedasticity test with six lags pass only at 1% level of significance. The rest of the tests, including all the autocorrelation tests, pass at 5% level of significance, confirming the findings from examining the diagnostic plot: the model seem to adequately explain the autocorrelation structure of the series but struggles in capturing the distribution and conditional heteroskedasticity. Nevertheless, the inadequacies do not seem particularly serious.

Because the restricted model was found somewhat inadequate, I run the quantile residual tests for the unrestricted model as well in order to evaluate whether it captures the statistical properties of the series more adequately. The following code runs the same diagnostics tests for the unrestricted model `fit42gs`.

```
R> set.seed(1)
R> qrt <- quantile_residual_tests(fit42gs, lags_ac = c(1, 3, 6, 12),
+    nsimu = 10000)

Normality test p-value: 0.087
```

```
Autocorrelation tests:
lags | p-value
  1  |  0.475
  3  |  0.020
  6  |  0.289
  12 |  0.077

Conditional heteroskedasticity tests:
lags | p-value
  1  |  0.579
  3  |  0.137
  6  |  0.002
  12 |  0.000
```

As the p-values show, relaxing the restrictions improved the model's capability to capture the distribution of the series, but according to the test results, the unrestricted model does not explain conditional heteroskedasticity as well as the restricted one when taking into account six lags (since the test now rejects at 1% level of significance). Also, the autocorrelation test with three lags only passes at 1% level of significance. It thereby appears that the more parsimonious restricted model could be more appropriate. Adding a third regime to the model or trying a different autoregressive order could also be considered for potentially improving the adequacy.

uGMAR often fails to calculate the quantile residual tests for GSMAR models with very large degrees of freedom parameter estimates, but the problem can be avoided by switching to the appropriate G-StMAR model with the function `stmar_to_gstmar`, which removes the redundant degrees of freedom parameters (see Section 3.3.4 of this essay and Section 2.4 in Chapter 2). Calculation of the tests may also fail when the estimate is very close to the boundary of the parameter space in which case it might be appropriate to consider an estimate from the next-largest local maximum point of the log-likelihood function. To that end, the function `alt_gsmar` can be used as demonstrated in Section 3.3.4 and in Appendix 3.A.

3.5 Building a GSMAR model with specific parameter values

The function `GSMAR` facilitates building GSMAR models without estimation, for instance, in order to simulate observations from a GSMAR process with specific parameter values. The parameter vector (of length $M(p+3)+M_2-1$ for unconstrained models) has the form $\boldsymbol{\theta} = (\boldsymbol{\vartheta}_1, ..., \boldsymbol{\vartheta}_M, \alpha_1, ..., \alpha_{M-1}, \boldsymbol{\nu})$ where

$$\boldsymbol{\vartheta}_m = (\varphi_{m,0}, \varphi_{m,1}, ..., \varphi_{m,p}, \sigma_m^2), \;\; m = 1, ..., M, \text{ and} \tag{3.5.1}$$

$$\boldsymbol{\nu} = (\nu_{M_1+1}, ..., \nu_M). \tag{3.5.2}$$

In the GMAR model (when $M_1 = M$), the vector $\boldsymbol{\nu}$ is omitted, as the GMAR model does not contain degrees of freedom parameters. For models with constraints on the autoregressive parameters, the parameter vectors are expressed in a different way. For brevity, they are only

presented in the package documentation, because the hand-specified parameter values can be set to satisfy any constraints as is.

In addition to the parameter vector, `GSMAR` should be supplied with arguments `p` and `M` specifying the order of the model similarly to the estimation function `fitGSMAR` discussed in Sections 3.3.4 and 3.3.6. If one wishes to parametrize the model with the regimewise unconditional means (μ_m) instead of the intercepts ($\varphi_{m,0}$), the argument `parametrization` should be set to `"mean"` in which case the intercept parameters $\varphi_{m,0}$ are replaced with μ_m in the parameter vector. By default, **uGMAR** uses intercept parametrization.

To exemplify, I build the GMAR $p = 2$, $M = 2$ model that is used in the simulation experiment in Appendix 3.A. The model has intercept parametrization and parameter values $\boldsymbol{\vartheta}_1 = (0.9, 0.4, 0.2, 0.5)$, $\boldsymbol{\vartheta}_2 = (0.7, 0.5, -0.2, 0.7)$, and $\alpha_1 = 0.7$. After building the model, I use the `print` method to examine it:

```
R> params22 <- c(0.9, 0.4, 0.2, 0.5, 0.7, 0.5, -0.2, 0.7, 0.7)
R> mod22 <- GSMAR(p = 2, M = 2, params = params22, model = "GMAR")
R> mod22

Model:
 GMAR, p = 2, M = 2, #parameters = 9,
 conditional, intercept parametrization, not restricted, no constraints.

Regime 1
Mix weight: 0.70
Reg mean: 2.25

y = [0.90] + [0.40]y.1 + [0.20]y.2 + sqrt[0.50]eps

Regime 2
Mix weight: 0.30
Reg mean: 1.00

y = [0.70] + [0.50]y.1 + [-0.20]y.2 + sqrt[0.70]eps
```

It is possible to include data in the models built with `GSMAR` by either providing the data in the argument `data` when creating the model or by adding the data afterwards with the function `add_data`. When the model is supplied with data, the mixing weights, one-step conditional means and variances, and quantile residuals can be calculated and included in the model. The function `add_data` can also be used to update data to an estimated GSMAR model without re-estimating the model.

3.6 Simulation and forecasting

3.6.1 Simulation

uGMAR implements the S3 method `simulate` for simulating observations from GSMAR processes. The method requires the process to be given as a class `gsmar` object, which are typically

created either by estimating a model with the function `fitGSMAR` or by specifying the parameter values by hand and building the model with the constructor function `GSMAR`. The initial values required to simulate the first p observations can be either set by hand (with the argument `init_values`) or drawn from the stationary distribution of the process (by default). The argument `nsim` sets the length of the sample path to be simulated.

To give an example, the following code sets the random number generator seed to one and simulates the 500 observations long sample path that is used in the simulation experiment in Appendix 3.A from the GMAR process built in Section 3.5:

```
R> mysim <- simulate(mod22, nsim = 500, seed = 1)
```

Our implementation of `simulate` returns a list containing the simulated sample path in `$sample`, the mixture component that generated each observation in `$component`, and the mixing weights in `$mixing_weights`.

3.6.2 Simulation based forecasting

Deriving multiple-steps-ahead point predictions and prediction intervals analytically for the GSMAR models is very complicated, so **uGMAR** employs the following simulation-based method. By using the last p observations of the data up to the date of forecasting as initial values, a large number of sample paths for the future values of the process are simulated. Then, sample quantiles from the simulated sample paths are calculated to obtain prediction intervals, and the median or mean is used for point predictions. A similar procedure is also applied to forecast future values of the mixing weights, which might be of interest because the researcher can often associate specific characteristics to different regimes.

Forecasting is most conveniently done with the `predict` method. The available arguments include the number of steps ahead to be predicted (`n_ahead`), the number sample paths the forecast is based on (`nsimu`), possibly multiple confidence levels for prediction intervals (`pi`), prediction type (`pred_type`), and prediction interval type (`pi_type`). The prediction type can be either `median`, `mean`, or for one-step-ahead forecasts also the exact conditional mean, `cond_mean`. The prediction interval type can be any of `"two-sided"`, `"upper"`, `"lower"`, or `"none"`.

As an example, I use the unrestricted G-StMAR $p = 4, M_1 = 1, M_2 = 1$ model fitted to the monthly interest rate spread in Section 3.3.4 to forecast the spread 12 months ahead, i.e., for the year 2021. The point prediction is based on median and 10000 simulated future sample paths, and the two-sided prediction intervals are calculated for the confidence levels 0.95 and 0.80.

```
R> set.seed(1)
R> mypred <- predict(fit42gs, n_ahead = 12, nsimu = 10000,
+    pi = c(0.95, 0.8), pred_type = "median", pi_type = "two-sided")
R> mypred

Prediction by median, two-sided prediction intervals with levels 0.95, 0.8.
Forecast 12 steps ahead, based on 10000 simulations.
```

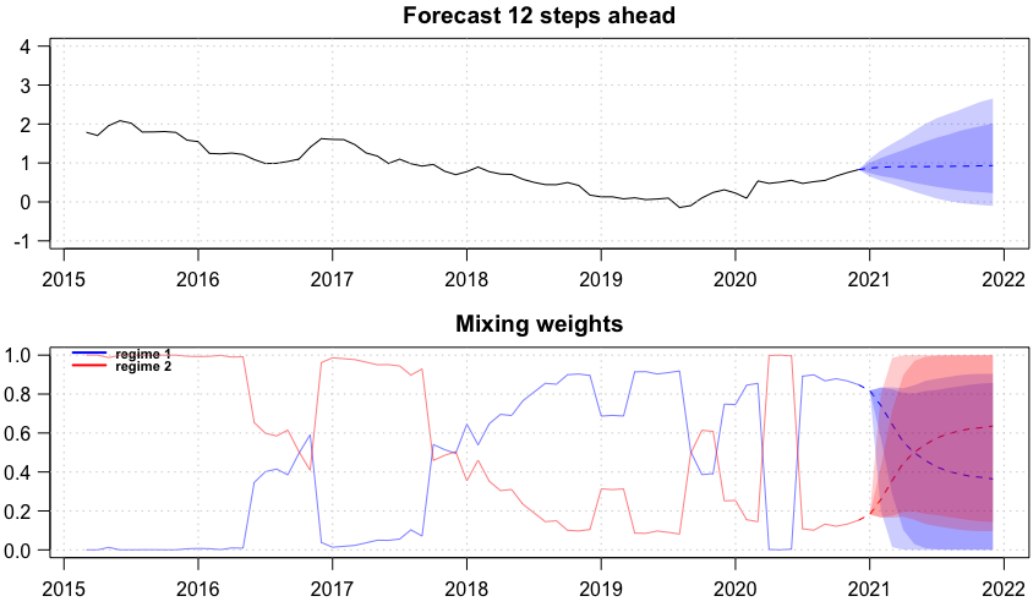

Figure 3.5: The figure created by the `predict` method for the G-StMAR model `fit42gs`. Twelve-months-ahead point prediction for the monthly interest rate spread (top) and the model's mixing weights (bottom) together with several preceding observations and prediction intervals with confidence levels 0.95 (outer interval) and 0.80 (inner interval).

```
   0.025  0.1 median  0.9 0.975
1   0.66 0.74   0.87 1.00  1.11
2   0.55 0.66   0.89 1.13  1.32
3   0.46 0.62   0.90 1.23  1.49
4   0.36 0.56   0.91 1.33  1.65
5   0.26 0.49   0.91 1.45  1.83
6   0.17 0.44   0.91 1.55  2.01
7   0.09 0.38   0.91 1.65  2.15
8   0.02 0.34   0.91 1.73  2.26
9  -0.02 0.30   0.92 1.82  2.36
10 -0.05 0.27   0.92 1.89  2.47
11 -0.08 0.25   0.93 1.95  2.58
12 -0.10 0.23   0.93 2.02  2.65

Point forecasts and prediction intervals for mixing weights can be obtained
with $mix_pred and $mix_pred_ints, respectively.
```

The `predict` method plots the results by default but this can be also avoided by setting `plot_res = FALSE` in the arguments. The results can be plotted afterwards by using the `plot` method for the class `gsmarpred` objects that the `predict` method returns.

The figure created by the above example is presented in Figure 3.5. The point forecast does not predict any significant movements for the spread, but the prediction intervals appear to be

skewed to the right. A possible explanation to the skewed prediction intervals is that at the time of forecasting, the spread takes a value that is closer to the mean of the low-mean first regime than to the mean of the high-mean second regime. Hence, even if the process proceeds in the first regime, it does not (on average) move much lower, but switching to the second regime would (on average) lead to notably larger observations. Also, the forecast for the mixing weights reveals that after a few months, the high-mean second regime is predicted to become more probable than than the low-mean first regime, thus, explaining the skewed prediction intervals.

3.7 Summary

Mixture autoregressive models are useful for analyzing time series that exhibit nonlinear, regime-switching features. The GMAR model, the StMAR model, and the G-StMAR model constitute an appealing family of such models, the GSMAR models, with attractive theoretical and practical properties. This essay introduced the R package **uGMAR** providing a comprehensive set of easy-to-use tools for GSMAR modelling, including unconstrained and constrained maximum likelihood estimation of the model parameters, quantile residual based model diagnostics, simulation, forecasting, and more. For convenience, I have collected some useful functions in **uGMAR** to Table 3.1.

The model parameters are estimated with the method of maximum likelihood by employing a two-phase procedure, which uses a genetic algorithm to find starting values for a variable metric algorithm. Notably, due to the endogenously determined mixing weights, the maximum likelihood estimate is often found very close to the boundary of the stationarity region of some regimes. I explained in Appendix 3.A why such estimates might be inappropriate and showed how a GSMAR model can be built based on an alternative estimate related to the next-largest local maximum point.

Computational details

The results in this essay were obtained using R 4.1.2 and **uGMAR** 3.4.1 package running on MacBook Pro 14", 2021, with Apple M1 Pro processor, 16 Gt of unified RAM, and macOS Monterey 12.1 operating system.

uGMAR takes use of the R package **Brobdingnag** (Hankin, 2007) to handle values extremely close to zero in the evaluation of the first term of the exact log-likelihood function (3.3.1). The package **gsl** (Hankin, 2006) is utilized to calculate some of the quantile residuals (3.4.1) with a hypergeometric function. In order to improve computational efficiency in the numerical estimation procedure, the formula proposed by Galbraith and Galbraith (1974) is utilized to directly compute the inverses of the covariance matrices $\mathbf{\Gamma}_m$, $m = 1, ..., M$, (which appear in (3.2.3), (3.2.5), (3.2.6), and in the first term of (3.3.1)), as only the inverses are required for calculating the quantities in the log-likelihood function. Finally, the algorithm proposed by

Related to	Name	Description
Estimation	`fitGSMAR`	Estimate a GSMAR model.
	`alt_gsmar`	Build a GSMAR model based on results from any estimation round.
	`stmar_to_gstmar`	Estimate a G-StMAR model based on a StMAR (or G-StMAR) model with large degrees of freedom parameters.
	`iterate_more`	Run more iterations of the variable metric algorithm for a preliminary estimated GSMAR model.
Estimates	`summary` (method)	Detailed printout of the estimates.
	`plot` (method)	Plot the series with the estimated mixing weights and a kernel density estimate of the series with the stationary density of the model.
	`get_foc`	Calculate numerically approximated gradient of the log-likelihood function evaluated at the estimate.
	`get_soc`	Calculate eigenvalues of numerically approximated Hessian of the log-likelihood function evaluated at the estimate.
	`profile_logliks`	Plot the graphs of the profile log-likelihood functions.
	`cond_moment_plot`	Plot the model implied one-step conditional means or variances.
Diagnostics	`quantile_residual_tests`	Calculate quantile residual tests.
	`diagnostic_plot`	Plot quantile residual diagnostics.
	`quantile_residual_plot`	Plot quantile residual time series and histogram.
Forecasting	`predict` (method)	Forecast future observations and mixing weights of the process.
Simulation	`simulate` (method)	Simulate from a GSMAR process.
Create model	`GSMAR`	Construct a GSMAR model based on specific parameter values.
Hypothesis testing	`LR_test`	Calculate likelihood ratio test.
	`Wald_test`	Calculate Wald test.
Other	`add_data`	Add data to a GSMAR model.
	`swap_parametrization`	Swap between mean and intercept parametrizations.

Table 3.1: Some useful functions in **uGMAR** sorted according to their usage. The note "method" in parentheses after the name of a function signifies that it is an S3 method for a class `gsmar` object.

Monahan (1984) is employed to generate random stationary autoregressive coefficients in the genetic algorithm.

Some of the estimation results (and thereby everything that is calculated based on the estimates) may vary slightly when running the code on different computers. This is due to a small numerical error in the gradient of the log-likelihood function caused by the limited precision of the floating-point representation. The negligible numerical error accumulates in each iteration of the variable metric algorithm, which hence advances in slightly different paths on different computers (with given initial values). After a large number of iterations, the algorithm might therefore end up in slightly different points. This particularly occurs when there are StMAR type regimes in the model, possibly because there are often many different pairs of degrees of freedom and variance parameter values that are relatively close to each other and yield almost the same log-likelihoods.

Bibliography

Aomoto K., Kita M. (2011). *Theory of Hypergeometric Functions*. 1st edition. Springer-Verlag, Tokyo.

Bollerslev T. (1986). "Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics*, **31**(3), 307–327.

Boshnakov G. N., Ravagli D. (2021). ***mixAR**: Mixture Autoregressive Models*. R package version 0.22.5.

Buse A. (1982). "The Likelihood Ratio, Wald, and Lagrange Multiplier Tests: An Expository Note." *The American Statistician*, **36**(3a), 153–157.

Dorsey R., Mayer W. (1995). "Genetic algorithms for estimation problems with multiple optima, nondifferentiability, and other irregular features." *Journal of Business and Economic Statistics*, **13**(1), 53–66.

Engle R. F. (1982). "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation." *Econometrica*, **50**(4), 987–1007.

Fong P., Li W., Yau C., Wong C. (2007). "On a mixture vector autoregressive model." *The Canadian Journal of Statistics*, **35**(1), 135–150.

Galbraith J., Galbraith R. (1974). "On the inverses of some patterned matrices arising in the theory of stationary time series." *Journal of Applies Probability*, **11**(1), 63–71.

Ghalanos A. (2020). ***rugarch**: Univariate GARCH Models.* R package version 1.4-4.

Glasbey C. (2001). "Non-linear autoregressive time series with multivariate Gaussian mixtures as marginal distributions." *Journal of Royal Statistical Society: Series C*, **50**(2), 143–154.

Hankin R. (2006). "Special Functions in R: Introducing the **gsl** Package." *R News*, **6**(4).

Hankin R. (2007). "Very Large Numbers in R: Introducing Package **Brobdingnag**." *R News*, **7**(3).

Hyndman R., Athanasopoulos G., Bergmeir C., Caceres G., Chhay L., O'Hara-Wild M., Petropoulos F., Razbash S., Wang E., Yasmeen F. (2021). ***forecast**: Forecasting Functions for Time Series and Linear Models*. R package version 8.15.

Kalliovirta L. (2012). "Misspecification tests based on quantile residuals." *The Econometrics Journal*, **15**(2), 358–393.

Kalliovirta L., Meitz M., Saikkonen P. (2015). "A Gaussian Mixture Autoregressive Model for Univariate Time Series." *Journal of Time Series Analysis*, **36**(2), 247–266.

Kalliovirta L., Meitz M., Saikkonen P. (2016). "Gaussian mixture vector autoregression." *Journal of Econometrics*, **192**(2), 465–498.

Lanne M., Saikkonen P. (2003). "Modeling the U.S. Short-Term Interest Rate by Mixture Autoregressive Processes." *Journal of Financial Econometrics*, **1**(1), 96–125.

Le N., Martin R., Raftery A. (1996). "Modeling Flat Stretches, Bursts, and Outliers in Time Series Using Mixture Transition Distribution Models." *Journal of the American Statistical Association*, **91**(436), 1504–1515.

Meitz M., Preve D., Saikkonen P. (2018). ***StMAR Toolbox**: A MATLAB Toolbox for Student's t Mixture Autoregressive Models*.

Meitz M., Preve D., Saikkonen P. (forthcoming). "A mixture autoregressive model based on Student's t-distribution." *Communications in Statistics - Theory and Methods*.

Meitz M., Saikkonen P. (2021). "Testing for observation-dependent regime switching in mixture autoregressive models." *Journal of Econometrics*, **222**(1), 601–624.

Monahan J. (1984). "A note on enforcing stationarity in autoregressive-moving average models." *Biometrika*, **71**(2), 403–404.

Nash J. (1990). *Compact Numerical Methods for Computers. Linear Algebra and Function Minimization*. 2nd edition. Adam Hilger, Bristol and New York.

Patnaik L., Srinivas M. (1994). "Adaptive Probabilities of Crossover and Mutation in Genetic Algorithms." *Transactions on Systems, Man and Cybernetics*, **24**(4), 656–667.

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Smith R., Dike B., Stegmann S. (1995). "Fitness inheritance in genetic algorithms." *Proceedings of the 1995 ACM symbosium on Applied Computing*, pp. 345–350.

Solymos P., Zawadzki Z. (2020). ***pbapply**: Adding Progress Bar to '*apply' Functions*. R package version 1.4-3.

Virolainen S. (2018). ***gmvarkit**: Estimate Gaussian and Student's t Mixture Vector Autoregressive Models*. R package version 2.0.3 available at CRAN: `https://CRAN.R-project.org/package=gmvarkit`.

Virolainen S. (forthcoming). "A mixture autoregressive model based on Gaussian and Student's t-distributions." *Studies in Nonlinear Dynamics & Econometrics*.

Wong C., Li W. (2000). "On a mixture autoregressive model." *Journal of the Royal Statistical Society*, **62**(1), 95–115.

Wong C., Li W. (2001a). "On a logistic mixture autoregressive model." *Biometrika*, **88**(3), 833–846.

Wong C., Li W. (2001b). "On a Mixture Autoregressive Conditional Heteroskedastic Model." *Journal of the American Statistical Association*, **96**(455), 982–995.

Wuertz D., Setz T., Chalabi Y., Boudt C., Chausse P., Miklovac M. (2020). ***fGarch**: Rmetrics - Autoregressive Conditional Heteroskedastic Modelling*. R package version 3042.83.2.

Appendix 3.A Simulation experiment

This simulation experiment demonstrates why the log-likelihood function's global maximum point, that is found very near the boundary of the parameter space, might not be a reasonable estimate and why it might be more appropriate to consider a local-only maximum point that is clearly in the interior of the parameter space. I generated 500 observations from a GMAR $p = 2$, $M = 2$ process with the parameter values given in the first row of Table 3.2 ($\boldsymbol{\theta}$) and initial values generated from the stationary distribution of the process. This model is built with **uGMAR** as an example in Section 3.5, and the sample path is generated as an example in Section 3.6.1.

I estimated a GMAR $p = 2$, $M = 2$ model to the generated sample based on the exact log-likelihood function by performing 100 estimation round using the following code (output is omitted for brevity):

```
R> fit22 <- fitGSMAR(mysim$sample, p = 2, M = 2, model = "GMAR",
+    conditional = FALSE, ncalls = 100, ncores = 8, seeds = 1:100)
```

The obtained estimates are reported on the second row of Table 3.2 ($\hat{\boldsymbol{\theta}}_1$) together with the moduli of each regime's AR polynomial's ($1 - \sum_{i=1}^{p} \varphi_{m,i} z^i$) roots. The modulus of the ith root in the mth regime is denoted by the symbol $\xi_{m,i}$. The stationarity condition requires that all the moduli are strictly greater than one, so the second regime is very close to the boundary of the stationarity region (both roots are approximately 1.000011). Also the variance parameter σ_2^2 is close to its lower bound zero (it is approximately $9 \cdot 10^{-6}$).

These estimates produce a large log-likelihood, because the second regime's very small conditional variance makes the related density function in the term $l_t(\boldsymbol{\theta})$ (3.3.2) to take large values near its mean, and the strong conditional mean targets individual observations there. This is illustrated in Figure 3.6 (bottom panel), where the terms $l_t(\boldsymbol{\theta})$ are presented (green solid line) together with the second regime's related weighted densities $\alpha_{2,t} n_1(y_t; \mu_{2,t}, \sigma_2^2)$ (red dotted line). The black "X"-symbols denote the points where the second regime's conditional mean deviates from the corresponding observation by less than 0.005. Evidently, the second regime contributes to the log-likelihood function only in the individual points where both, the terms $l_t(\boldsymbol{\theta})$ and the scaled densities $\alpha_{2,t} n_1(y_t; \mu_{2,t}, \sigma_2^2)$, take large values due to the observation being close to the mean of the second regime's spikelike conditional density function. Because the scaled densities take large enough values in those individual points, the log-likelihood is larger for this kind of estimate than for a reasonable estimate.

The top panel of Figure 3.6 presents the true mixing weights of the GMAR process's second regime (black solid line) together with the mixing weights based on the estimate $\hat{\boldsymbol{\theta}}_1$ (red dashed line). As the figure shows, the estimated mixing weights are spiky and have no resemblance to the true mixing weights. Although the true mixing weights can be spiky for some GSMAR processes, spiking mixing weights are also typical for potentially inappropriate near-the-boundary estimates.

This kind of near-the-boundary estimates are often found when a subset of the regimes explains the variation in the series reasonably well, leaving some of the regimes available for targeting individual observations with very small conditional variance and very strong conditional

	$\varphi_{1,0}$	$\varphi_{1,1}$	$\varphi_{1,2}$	σ_1^2	$\varphi_{2,0}$	$\varphi_{2,1}$	$\varphi_{2,2}$	σ_2^2	α_1	$\xi_{1,1}$	$\xi_{1,2}$	$\xi_{2,1}$	$\xi_{2,2}$
$\boldsymbol{\theta}$	0.90	0.40	0.20	0.50	0.70	0.50	−0.20	0.70	0.70	1.45	3.45	2.24	2.24
$\hat{\boldsymbol{\theta}}_1$	0.58	0.56	0.10	0.61	7.85	−1.67	−1.00	0.00	0.99	1.42	6.86	1.00	1.00
$\hat{\boldsymbol{\theta}}_2$	1.16	0.39	0.08	0.54	0.77	0.35	−0.17	0.53	0.63	1.86	6.90	2.42	2.42

Table 3.2: On the first row, the true parameter values of the GMAR $p = 2$, $M = 2$ process that generated the sample path used in the simulation experiment. On the second row, the estimates that maximized the log-likelihood function based 100 estimation rounds. On the third row, the estimates from the largest such log-likelihood function's maximum point that is not very near the boundary of the stationarity region. In each row after the estimates or parameter values, the moduli of the related AR polynomial's roots are presented.

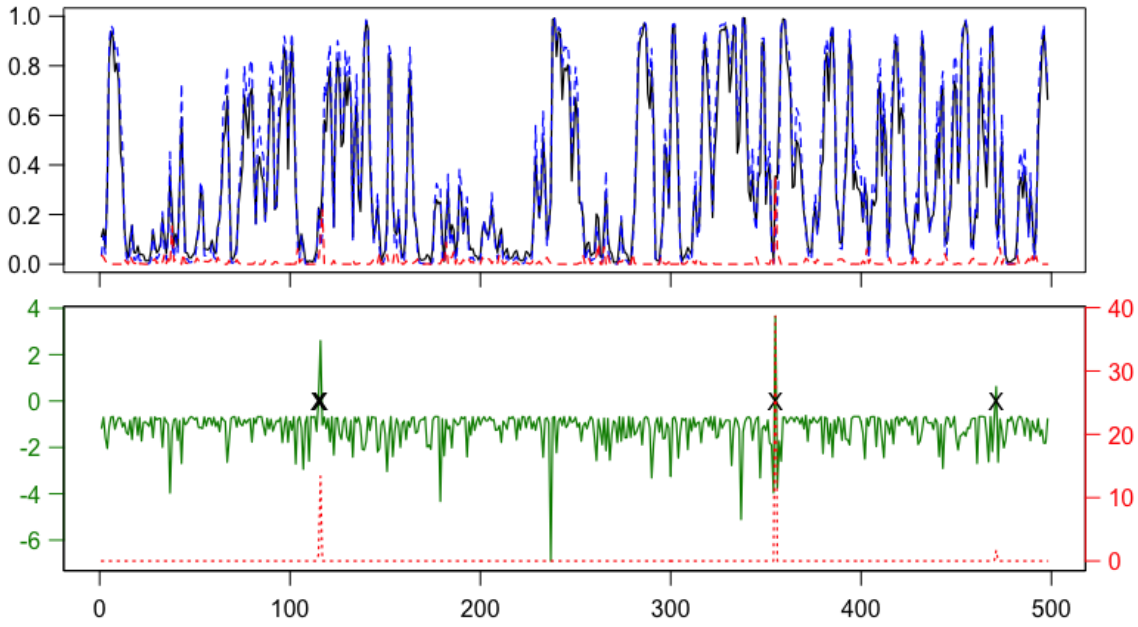

Figure 3.6: On the top, the GMAR $p = 2$, $M = 2$ process's second regime's true mixing weights (black solid line), the mixing weights based on the estimate $\hat{\boldsymbol{\theta}}_1$ in the second row of Table 3.2 (red dashed line), and the mixing weights based on the estimate $\hat{\boldsymbol{\theta}}_2$ in the third row of Table 3.2 (blue dashed line). On the bottom, the terms (3.3.2) from the second term of the log-likelihood function (3.3.1) (green solid line) and the second regime's densities in the terms (3.3.2) multiplied by the estimated mixing weights (blue dotted line), i.e., $\alpha_{2,t}n_1(y_t; \mu_{2,t}, \sigma_2^2)$, both based on the estimate $\hat{\boldsymbol{\theta}}_1$. The "X"-symbols denote the points where the second regime's conditional mean for the model based on estimate $\hat{\boldsymbol{\theta}}_1$ deviates from the corresponding observation by less than 0.005.

mean. As such estimates seem to maximize the log-likelihood function for a technical reason, and not necessarily because they represent a good guess for the true parameter value, it might be appropriate to consider an alternative estimate related to the next-largest local maximum point. To exemplify, I build a model based on the largest local maximum point that is clearly in the interior of the parameter space. In my estimation based on 100 rounds of the two-phase procedure, such an estimate is found at the point that induced the third largest log-likelihood, and it is obtained as follows:

```
R> fit22_alt <- alt_gsmar(fit22, which_largest = 3)
```

The corresponding estimate is presented on the third row of Table 3.2 ($\hat{\boldsymbol{\theta}}_2$). This local maximum point is substantially closer to the true parameter value in the second regime. The resemblance to the true parameter value is also highlighted in Figure 3.6 (top panel), where the second regime's estimated mixing weights (blue dashed line) are presented together with the true mixing weights (black solid line).

Finally, observe that the estimate $\hat{\boldsymbol{\theta}}_1$ presented in Table 3.2 is not the accurate maximum likelihood estimate, which can be noticed by examining graphs of the related profile log-likelihood functions with the command `profile_logliks(fit22)` (not shown). The numerical estimation using numerical approximation for the gradient of the log-likelihood function can be inaccurate near the boundary of a multidimensional parameter space subject to several constraints. Consequently, other similar near-the-boundary points that induce larger log-likelihood than $\hat{\boldsymbol{\theta}}_1$ can be found by running more estimation rounds. It should also be noted that sometimes the estimate is near the boundary of the stationarity region because the series is very persistent, and being near the boundary does not hence necessarily imply that the MLE is inappropriate.

Appendix 3.B Closed form expressions of quantile residuals

This section derives closed form expressions for the quantile residuals utilized by **uGMAR** and discussed in Section 3.4. For the GSMAR models, the quantile residuals are defined as

$$R_t = \Phi^{-1}(F(y_t|\mathcal{F}_{t-1})), \quad t = 1, 2, ..., T, \tag{3.B.1}$$

where $\Phi^{-1}(\cdot)$ is the standard normal quantile function,

$$F(y_t|\mathcal{F}_{t-1}) = \sum_{m=1}^{M} \alpha_{m,t} \int_{-\infty}^{y_t} f_m(u_t|\mathcal{F}_{t-1}) du_t \tag{3.B.2}$$

is the conditional cumulative distribution function of the considered GSMAR process (conditional on the previous observations), and $f_m(\cdot|\mathcal{F}_{t-1})$ is the conditional density function of the mth component process. To find a closed form expression for the quantile residuals defined in (3.B.1) and (3.B.2), it therefore suffices to solve the integrals $\int_{-\infty}^{y_t} f_m(u_t|\mathcal{F}_{t-1}) du_t$, $m = 1, ..., M$, for GMAR type and StMAR type mixture components.

Appendix

In the case of a GMAR type component, the conditional density function is the Gaussian density function with mean $\mu_{m,t}$ and variance σ_m^2. For $m \leq M_1$ in (3.B.2), I therefore have

$$\int_{-\infty}^{y_t} f_m(u_t|\mathcal{F}_{t-1})du_t = \int_{-\infty}^{y_t} n_1(u_t;\mu_{m,t},\sigma_m^2)du_t = \Phi\left(\frac{u_t-\mu_{m,t}}{\sigma_m}\right), \tag{3.B.3}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function.

In the case of a StMAR type component, the conditional density function is the Student's t density function with mean $\mu_{m,t}$, variance $\sigma_{m,t}^2$, and $\nu_m + p$ degrees of freedom given as (Meitz *et al.*, forthcoming, Appendix A)

$$t_1(u_t;\mu_{m,t},\sigma_{m,t}^2,\nu_m{+}p) = \frac{\Gamma\left(\frac{1+\nu_m+p}{2}\right)}{\sqrt{\pi(\nu_m+p-2)}\Gamma\left(\frac{\nu_m+p}{2}\right)}\sigma_{m,t}^{-1}\left(1+\frac{(u_t-\mu_{m,t})^2}{(\nu_m+p-2)\sigma_{m,t}^2}\right)^{-(1+\nu_m+p)/2} \tag{3.B.4}$$

where $\Gamma(\cdot)$ is the gamma function. Taking use of the symmetry of the Student's t distribution about its mean $\mu_{m,t}$, I obtain

$$\int_{-\infty}^{y_t} f_m(u_t|\mathcal{F}_{t-1})du_t = \frac{1}{2} + \int_{\mu_{m,t}}^{y_t} t_1(u_t;\mu_{m,t},\sigma_{m,t}^2,\nu_m+p)du_t. \tag{3.B.5}$$

By applying the change of variables $\tilde{u}_{m,t} \equiv u_t - \mu_{m,t}$ in the integral, the right side of (3.B.5) can be expressed as

$$\frac{1}{2} + \frac{\Gamma\left(\frac{1+\nu_m+p}{2}\right)}{\sqrt{\pi(\nu_m+p-2)}\Gamma\left(\frac{\nu_m+p}{2}\right)}\sigma_{m,t}^{-1}\int_0^{\tilde{y}_{m,t}}\left(1+\frac{\tilde{u}_{m,t}^2}{a_{m,t}}\right)^{-b_m}d\tilde{u}_{m,t}, \tag{3.B.6}$$

where $\tilde{y}_{m,t} \equiv y_t - \mu_{m,t}$, $a_{m,t} \equiv (\nu_m + p - 2)\sigma_{m,t}^2$, and $b_m \equiv (1+\nu_m+p)/2$. Then, by applying the change of variables $z_{m,t} \equiv \tilde{u}_{m,t}^2/\tilde{y}_{m,t}$, the integral in the expression (3.B.6) can be expressed as

$$\int_0^{\tilde{y}_{m,t}}\left(1+\frac{\tilde{u}_{m,t}^2}{a_{m,t}}\right)^{-b_m}d\tilde{u}_{m,t} = \frac{1}{2}\int_0^{\tilde{y}_{m,t}}\left(\frac{\tilde{y}_{m,t}}{z_{m,t}}\right)^{1/2}\left(1+\frac{z_{m,t}\tilde{y}_{m,t}}{a_{m,t}}\right)^{-b_m}dz_{m,t}. \tag{3.B.7}$$

By applying the third change of variables $x_{m,t} \equiv z_{m,t}/\tilde{y}_{m,t}$ and using the properties of the gamma function, the right side of (3.B.7) can be expressed using a hypergeometric function as

$$\frac{\tilde{y}_{m,t}}{2}\int_0^1 x_{m,t}^{-1/2}\left(1-x_{m,t}\left(-\frac{\tilde{y}_{m,t}^2}{a_{m,t}}\right)\right)^{-b_m}dx_{m,t} = \tilde{y}_{m,t}\times{}_2\mathrm{F}_1\left(\frac{1}{2},b_m,\frac{3}{2};-\frac{\tilde{y}_{m,t}^2}{a_{m,t}}\right), \tag{3.B.8}$$

where the hypergeometric function is defined as (Aomoto and Kita, 2011, Section 1.3.1)

$${}_2\mathrm{F}_1(a,b,c;x) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(c-a)}\int_0^1 s^{a-1}(1-s)^{c-a-1}(1-sx)^{-b}ds, \tag{3.B.9}$$

when $|x| < 1$, $a > 0$, and $c - a > 0$ (when $a, c \in \mathbb{R}$).

Using the above result, we have

$$\int_{-\infty}^{y_t} f_m(u_t|\mathcal{F}_{t-1}) = \frac{1}{2} + \frac{\Gamma\left(\frac{1+\nu_m+p}{2}\right)}{\sqrt{\pi(\nu_m+p-2)}\Gamma\left(\frac{\nu_m+p}{2}\right)}\sigma_{m,t}^{-1}\tilde{y}_{m,t} \times {}_2\mathrm{F}_1\left(\frac{1}{2}, b_m, \frac{3}{2}; -\frac{\tilde{y}_{m,t}^2}{a_{m,t}}\right) \tag{3.B.10}$$

for $m > M_1$, whenever $\left|-\frac{\tilde{y}_{m,t}^2}{a_{m,t}}\right| < 1$. That is, the closed form expression (3.B.10) exists when

$$|y_t - \mu_{m,t}| < \sqrt{(\nu_m + p - 2)\sigma_{m,t}^2}. \tag{3.B.11}$$

If this condition does not hold, **uGMAR** calculates the quantile residual by numerically integrating the conditional density function $f_m(\cdot|\mathcal{F}_{t-1})$.

Chapter 4

Structural Gaussian mixture vector autoregressive model with application to the asymmetric effects of monetary policy shocks

4.1 Introduction

Tracing out the effects of an economic shock is a major task in econometrics. A popular approach is to consider a set of key variables and utilize a structural vector autoregressive (SVAR) or structural error correction (SVEC) model for the purpose. They have well established theoretical grounds (see Kilian and Lütkepohl, 2017, and the references therein) and are accommodated by many of the popular statistical software packages. Linear SVAR and SVEC models are not, however, suitable for modelling series in which the underlying data generating dynamics are nonlinear or the shocks have asymmetric effects in different states of the economy. Models capable of capturing such features include mixture models, such as the mixture vector autoregressive model (Fong, Li, Yau, and Wong, 2007), the mixture periodic vector autoregressive model (Bentarzi and Djeddou, 2014), the Gaussian mixture vector autoregressive (GMVAR) model (Kalliovirta, Meitz, and Saikkonen, 2016), and the logit mixture vector autoregressive model (Burgard, Neuenkirch, and Nöckel, 2019).

This essay introduces a structural version of the GMVAR model. In the structural GMVAR (SGMVAR) model of autoregressive order p, the regime-switching dynamics are endogenously determined by the full distribution of the previous p observations. Specifically, the greater the relative weighted likelihood of a regime is, the more likely the process is to generate an observation from it. This facilitates associating statistical characteristics and economic interpretations to the regimes. The specific formulation of the mixing weights also leads to attractive theoretical properties, such as ergodicity and fully known stationary distribution of $p + 1$ consecutive observations.

The effects of the structural shocks depend on the initial values of the included variables, and they are also allowed to vary according to the sign and size of the shock due to possibly resulting regime-switches. Consequently, the (generalized) impulse response functions reflect the prevailing macroeconomic conditions that are transmitted to the regime-switching probabilities through the level, variability, and temporal as well as contemporaneous dependence of the past observations. Because the shocks may have asymmetric effects with respect to their size, the conditional heteroskedasticity of the reduced form error needs to be controlled for. Therefore, the impact matrix of the SGMVAR model is time-varying and constructed so that it captures the conditional heteroskedasticity of the reduced form error, thereby enabling to standardize the conditional variance of each structural shock to a constant. The initial effects of a constant-sized structural shock are, hence, amplified according to the conditional variance of the reduced form error, also reflecting the prevailing state of the economy.

Identification of the shocks requires that they are simultaneously orthogonalized in all regimes. I show that together with any constant standardization of the structural shock's conditional variance, this condition generally leads to a unique identification of the impact matrix up to ordering of its columns and changing all signs in a column. Thus, as long as one is willing to impose the assumption of a single (time-varying) impact matrix, the columns of the impact matrix unambiguously characterize the estimated impact effects of the shocks without further constraints. The identification does not, however, reveal which column of the impact matrix is related to which shock. Since the impact matrix is also subject to estimation error, further constraints may be needed for labelling the shocks. The constraints are testable, as they are overidentifying.

In order to formulate the impact matrix and the identification conditions, it is convenient to utilize the well known matrix decomposition (Muirhead, 1982, Theorem A9.9) proposed by Lanne and Lütkepohl (2010) and Lanne, Lütkepohl, and Maciejowsla (2010) for a similar identification problem. Lanne and Lütkepohl (2010) assume that the reduced form error covariance matrices admit this decomposition, then show that the shocks are statistically identified, and finally test conventional zero constraints that lead to economically interpretable shocks. Lanne *et al.* (2010), in turn, note that the shocks are readily identified when the matrix decomposition is imposed to the reduced form error covariance matrices. My approach differs from them in that I obtain locally identified structural shocks by directly investigating the properties of the impact matrix. I also provide a general set of conditions for identifying any subset of the shocks that allows for using sign constraints alone or together with zero constraints. Moreover, I (partially) relax a technical condition required for statistical identification of the model and allow identification of a subset of the shocks when the model is only partially identified.

My empirical application studies asymmetries in the expected effects of monetary policy shocks in the U.S. using a quarterly series covering the period from 1954Q3 to 2021Q4. My SGMVAR model identifies two regimes: a stable inflation regime and an unstable inflation regime. The unstable inflation regime is characterized by high or volatile inflation, and it mainly prevails in the 1970's, early 1980's, during the Financial crisis, and in the COVID-19 crisis from 2020Q3 onwards. The stable inflation regime, in turn, is characterized by moderate inflation, and it prevails when the unstable inflation regime does not. I find the effects of the monetary

policy shock relatively symmetric in the unstable inflation regime, as it rarely causes a switch to the stable inflation regime. A contractionary (expansionary) monetary policy shock appears to first increase (decrease) inflation after which the inflation significantly decreases (increases) for several years. The strong contraction (expansion) in the cyclical component of the GDP lasts for roughly three years and is followed by a small short-term expansion (contraction) before the response decays to zero.

In the stable inflation regime, the (generalized) impulse responses are strongly asymmetric with the respect to the sign and size of the monetary policy shock as well as to the initial state of the economy. A contractionary shock causes, on average, roughly a three-year hump-shaped contraction of the GDP, but it also seems to increase inflation by driving the economy towards the unstable inflation regime. A small expansionary shock does not move prices much on average, but a large expansionary shock often drives the economy towards the unstable inflation regime and propagates high and persistent inflation. The high inflation is followed by a significant monetary policy tightening and persistent contraction of the GDP after the initial expansion. On average, the real effects of the monetary policy shock are found somewhat stronger in the stable inflation regime than in the unstable inflation regime.

The GMVAR model has been previously applied in impulse response analysis by Kalliovirta and Malinen (2020), who identify the shocks by constraining the reduced form error covariance matrices, and allow the impact responses of the variables to vary relative to each other across the regimes. My assumption of a common (time-varying) impact matrix for all the regimes constraints the relative magnitudes of the impact responses of the variables to be time-invariant (for each shock), but it leads to flexible identification conditions and enables to test the validity of the identifying constraints. Kalliovirta and Malinen (2020) estimate the impulse response functions for each regime of the GMVAR model separately as if each of them was a linear VAR. In contrast, I allow the regime to switch as a result of a shock and estimate the true (generalized) impulse response functions of the non-linear VAR.

Structural mixture VARs, in general, have been previously applied for studying to the effects of monetary policy shocks at least by Burgard *et al.* (2019), who proposed a mixture VAR with logistic mixing weights and Cholesky identified shocks. As opposed to Burgard *et al.* (2019), my identification scheme is more flexible in the sense that it does not require many (or necessarily any) zero constraints on the impact effects of the shocks. Moreover, in my model the regime-switching probabilities depend on the full distribution of the preceding p observations instead of just on the level of the switching-variables.

The rest of this chapter is organized as follows. Section 4.2 defines the reduced form GMVAR model. In Section 4.3, the structural GMVAR model is first introduced. Then, identification of the shocks and estimation of the model parameters are discussed. Section 4.4 discusses impulse response analysis and describes the generalized impulse response function (GIRF) (Koop, Pesaran, and Potter, 1996). Section 4.5 presents the empirical application and Section 4.6 summarizes. Appendices provide proofs for the stated lemma and propositions, a Monte Carlo algorithm for estimating the GIRF, and details on the empirical application. Finally, I have accompanied this essay with the CRAN distributed R package **gmvarkit** (Virolainen,

2018a), which is comprehensively documented and provides a comprehensive set of tools for numerical analysis of the model.

4.2 Reduced form GMVAR model

To build theory and notation, consider first the reduced form GMVAR model introduced by Kalliovirta *et al.* (2016). Let y_t ($t = 1, 2, ...$) be the d-dimensional time series of interest and $\mathcal{F}_{t-1}$ denote the σ-algebra generated by the random vectors $\{y_{t-j}, j > 0\}$. For a GMVAR model with M mixture components and autoregressive order p, we have

$$y_t = \sum_{m=1}^{M} s_{m,t}(\mu_{m,t} + u_{m,t}), \quad u_{m,t} \sim NID(0, \Omega_m) \tag{4.2.1}$$

$$\mu_{m,t} = \phi_{m,0} + \sum_{i=1}^{p} A_{m,i} y_{t-i}, \quad m = 1, ..., M, \tag{4.2.2}$$

where $\phi_{m,0} \in \mathbb{R}^d$ are intercept parameters, Ω_m are positive definite covariance matrices, and for each m, the coefficient matrices $A_{m,i}$, $i = 1, ..., p$, are assumed to satisfy the usual stability condition

$$\det\left(I_d - \sum_{i=1}^{p} A_{m,i} z^i\right) \neq 0 \text{ for } |z| \leq 1, \ \ m = 1, ..., M, \tag{4.2.3}$$

which guarantees stationarity of the component processes. The unobservable regime variables $s_{1,t}, ..., s_{M,t}$ are such that at each t, exactly one of them takes the value one and the others take the value zero according to the conditional probabilities $\mathrm{P}(s_{m,1} = 1|\mathcal{F}_{t-1}) \equiv \alpha_{m,t}$ that satisfy $\sum_{m=1}^{M} \alpha_{m,t} = 1$. The normally and independently distributed (NID) errors $u_{m,t}$ are assumed independent of $\mathcal{F}_{t-1}$, and conditional on $\mathcal{F}_{t-1}$, $(s_{1,t}, ..., s_{M,t})$ and $u_{m,t}$ are independent.

The definition (4.2.1)-(4.2.2) implies that at each t, the process generates an observation from one of its mixture components, a linear VAR process, that is randomly selected according to the probabilities given by the mixing weights $\alpha_{m,t}$. Denoting $\boldsymbol{y}_{t-1} = (y_{t-1}, ..., y_{t-p})$, the mixing weights are defined as (Kalliovirta *et al.*, 2016, Equation (7))

$$\alpha_{m,t} = \frac{\alpha_m n_{dp}(\boldsymbol{y}_{t-1}; \mathbf{1}_p \otimes \mu_m, \boldsymbol{\Sigma}_m)}{\sum_{n=1}^{M} \alpha_n n_{dp}(\boldsymbol{y}_{t-1}; \mathbf{1}_p \otimes \mu_n, \boldsymbol{\Sigma}_n)}, \ \ m = 1, ..., M, \tag{4.2.4}$$

where $\alpha_1, ..., \alpha_M$ are mixing weight parameters that satisfy $\sum_{m=1}^{M} \alpha_m = 1$ and $n_{dp}(\cdot; \mathbf{1}_p \otimes \mu_m, \boldsymbol{\Sigma}_m)$ is the density function of the dp-dimensional normal distribution with mean $\mathbf{1}_p \otimes \mu_m$ and covariance matrix $\boldsymbol{\Sigma}_m$. The symbol $\mathbf{1}_p$ denotes a p-dimensional vector of ones, $\otimes$ is Kronecker product, $\mu_m = (I_d - \sum_{i=1}^{p} A_{m,i})^{-1}\phi_{m,0}$, and the covariance matrix $\boldsymbol{\Sigma}_m$ is given in Lütkepohl (2005, Equation (2.1.39)) but using the parameters of the mth component process. That is, $n_{dp}(\cdot; \mathbf{1}_p \otimes \mu_m, \boldsymbol{\Sigma}_m)$ corresponds to the density function of the stationary distribution of the mth component process.

The mixing weights are thus weighted ratios of the the component process stationary densities corresponding to the previous p observations. Consequently, the researcher can associate specific characteristics or give economic interpretations to the regimes. In addition to the (generalized) impulse response functions of the observable variables, the responses of the mixing weights may therefore be of interest. The definition of the mixing weights also leads to attractive theoretical properties such as ergodicity and knowledge of the stationary distribution of $p+1$ consecutive observations (Kalliovirta *et al.*, 2016, Theorem 1, see the proof of Theorem 1 for the stationary distribution of $p+1$ consecutive observations). Specifically, the stationary distribution of the process $\boldsymbol{y}_t = (y_t, ..., y_{t-p+1})$ is a mixture of dp-dimensional normal distributions that is characterized by the density

$$f(\boldsymbol{y}) = \sum_{m=1}^{M} \alpha_m n_{dp}(\boldsymbol{y}; \mathbf{1}_p \otimes \mu_m, \boldsymbol{\Sigma}_m). \tag{4.2.5}$$

The knowledge of the stationary distribution is taken advantage of in the impulse response analysis in Section 4.4.

4.3 Structural GMVAR model

4.3.1 The model setup

Consider the GMVAR model defined in (4.2.1)-(4.2.2). I focus on the "B-model" setup and write the structural GMVAR model as

$$y_t = \sum_{m=1}^{M} s_{m,t} \left(\phi_{m,0} + \sum_{i=1}^{p} A_{m,i} y_{t-i} \right) + B_t e_t, \tag{4.3.1}$$

and

$$u_t \equiv B_t e_t = \begin{cases} u_{1,t} \sim N(0, \Omega_1) & \text{if} \quad s_{1,t} = 1 \quad (\text{with probability } \alpha_{1,t}) \\ u_{2,t} \sim N(0, \Omega_2) & \text{if} \quad s_{2,t} = 1 \quad (\text{with probability } \alpha_{2,t}) \\ \quad \vdots & \\ u_{M,t} \sim N(0, \Omega_M) & \text{if} \quad s_{M,t} = 1 \quad (\text{with probability } \alpha_{M,t}) \end{cases} \tag{4.3.2}$$

where the probabilities are expressed conditionally on $\mathcal{F}_{t-1}$ and e_t is an orthogonal structural error. Unlike in the conventional SVAR analysis, the invertible $(d \times d)$ "B-matrix" (or impact matrix) B_t, which governs the contemporaneous relations of the shocks, is time-varying and a function of $y_{t-1}, ..., y_{t-p}$. This enables to amplify a constant-sized structural shock according to the conditional variance of the reduced form error, which varies according to the mixing weights. Appropriate modelling of conditional heteroskedasticity in the B-matrix is of interest, because the (generalized) impulse response functions may be asymmetric with respect to the size of the shock.

We have $\Omega_{u,t} \equiv \text{Cov}(u_t|\mathcal{F}_{t-1}) = \sum_{m=1}^{M} \alpha_{m,t}\Omega_m$, while the conditional covariance matrix of the structural errors $e_t = B_t^{-1}u_t$ (which have a mixture normal distribution and are not IID but martingale differences and therefore uncorrelated) is obtained as

$$\text{Cov}(e_t|\mathcal{F}_{t-1}) = \sum_{m=1}^{M} \alpha_{m,t}B_t^{-1}\Omega_m B_t^{\prime -1}. \tag{4.3.3}$$

The B-matrix B_t should therefore be chosen so that the structural shocks are orthogonal regardless of which regime they come from. I will next discuss the properties of any such B-matrix that solves the diagonalization problem. Then, I present a locally unique solution under a constant normalization of the structural error's conditional variance. After that, in the following two subsections, I will discuss global identification of the shocks, allowing also only partial identification of the model.

Specifically, I show that my model readily identifies the B-matrix up to ordering of its columns and changing all signs in a column, but it is not revealed which column of the B-matrix is related to which shock. The identification follows from the assumption $e_t = B_t^{-1}u_t$ (and Assumption 4.1 below), which (as I show in this section) implies that for each shock the relative magnitudes of the impact responses of the variables stay constant over time.[1] This is different to the conventional SVAR setup, where the identification of the B-matrix requires further constraints to be imposed on the model. Conventionally, the shocks are often identified, for instance, by placing economically motivated zero constraints on the impact or the long-run effects of the shocks (e.g., Kilian and Lütkepohl, 2017, Chapters 8 and 10). Sign constraints, in turn, are commonly used to obtain a set identification with less restrictive or economically more plausible constraints (e.g., Kilian and Lütkepohl, 2017, Chapter 13).

In Section 4.3.2, also I make use of zero and sign constraints, but I do it in order to formally label the already locally identified columns of the B-matrix by the shocks of interest. The required conditions are, nevertheless, flexible, and allow for using sign constraints alone or together with zero constraints. Some of the constraints are also testable, as they are overidentifying. Section 4.3.3 additionally takes advantage of zero constraints to identify the shock of interest when the condition for identification through conditional heteroskedasticity fails.

It turns out that any invertible B-matrix that simultaneously diagonalizes the covariance matrices has linearly independent eigenvectors of the matrix $\Omega_m\Omega_1^{-1}$ as its columns. If $M > 2$, the matrices $\Omega_m\Omega_1^{-1}$, $m = 2, ..., M$, thus need to share the common eigenvectors in B_t, which restricts the parameter space for the covariance matrices. In this case, the existence of such B-matrix can be tested with a likelihood ratio test, for example. Denoting the eigenvalues of $\Omega_m\Omega_1^{-1}$ as λ_{mi}, the B-matrix is also unique up to scalar multiples and ordering of its columns if none of the pairs of λ_{mi}, $i = 1, ..., d$, is identical for all $m = 2, ..., M$. These results are formalized in the following assumption and lemma.

[1] See Kilian and Lütkepohl (2017, Chapter 14) for a discussion on identification by heteroskedasticity in a linear VAR model.

Assumption 4.1. *Consider M positive definite $(d \times d)$ covariance matrices Ω_m, $m = 1, ..., M$, and denote the strictly positive eigenvalues of the matrices $\Omega_m\Omega_1^{-1}$ as λ_{mi}, $i = 1, ..., d$, $m = 2, ..., M$. Suppose that for all $i \neq j \in \{1, ..., d\}$, there exists an $m \in \{2, ..., M\}$ such that $\lambda_{mi} \neq \lambda_{mj}$.*

Lemma 4.1. *Consider M positive definite $(d \times d)$ covariance matrices Ω_m, $m = 1, ..., M$, and an invertible $(d \times d)$ matrix B_t such that $B_t^{-1}\Omega_m B_t'^{-1}$ are diagonal matrices with strictly positive diagonal elements. Then, B_t has eigenvectors of $\Omega_m\Omega_1^{-1}$ as its columns. Moreover, B_t is unique up to scalar multiples and ordering of its columns if Assumption 4.1 holds.*

Under Assumption 4.1, the columns of B_t are unique up to scalar multiples and ordering, implying that the shocks are identified up to sign, size, and ordering. Normalizing the conditional covariance matrix of the structural error to a constant diagonal matrix then identifies the B-matrix up to sign and ordering of the shocks. This is formalized in the following proposition.

Proposition 4.1. *Consider M positive definite $(d \times d)$ covariance matrices, Ω_m, $m = 1, ..., M$, and an invertible $(d \times d)$ matrix B_t such that $B_t^{-1}\Omega_m B_t'^{-1}$ are diagonal matrices with strictly positive diagonal elements. Suppose that Assumption 4.1 holds. Then, if the conditional covariance matrix of the structural error, $Cov(e_t|\mathcal{F}_{t-1}) = \sum_{m=1}^{M} \alpha_{m,t} B_t^{-1}\Omega_m B_t'^{-1}$, is normalized to a constant diagonal matrix with strictly positive diagonal entries, the B-matrix B_t is unique up to ordering of its columns and changing all signs in a column.*

That is, by fixing an ordering and signs for the columns of the B-matrix, the solution to the diagonalization problem is unique for any given (constant) normalization of the structural error's conditional covariance matrix, say, an identity matrix. In order to find the related B-matrix, it is then convenient to utilize the following matrix decomposition for the reduced form error covariance matrices, which was also employed by Lanne and Lütkepohl (2010) and Lanne *et al.* (2010) to solve a similar identification problem. My specification of the B-matrix differs from Lanne *et al.* (2010) who assume that the instantaneous effects of the shocks are time-invariant, but it extends the one in Lanne and Lütkepohl (2010) to accommodate time-varying mixing weights.

I decompose the reduced form error covariance matrices as

$$\Omega_1 = WW' \quad \text{and} \quad \Omega_m = W\Lambda_m W', \quad m = 2, ..., M, \tag{4.3.4}$$

where the diagonal of $\Lambda_m = \text{diag}(\lambda_{m1}, ..., \lambda_{md})$, $\lambda_{mi} > 0$ $(i = 1, ..., d)$, contains the eigenvalues of the matrix $\Omega_m\Omega_1^{-1}$ and the columns of the nonsingular W are the related eigenvectors (that are the same for all m by construction). When $M = 2$, the decomposition (4.3.4) always exists (Muirhead, 1982, Theorem A9.9), but for $M > 2$ its existence requires that the matrices $\Omega_m\Omega_1^{-1}$ share the common eigenvectors in W. This is, however, testable and relates to the earlier discussion on the existence of a B-matrix that simultaneously diagonalizes the reduced form error covariance matrices.

Any scalar multiples of linearly independent eigenvectors of $\Omega_m\Omega_1^{-1}$ comprise an appropriate B-matrix, but only specific scalar multiples comprise the locally unique B-matrix associated

with a given normalization of structural error's conditional covariance matrix. Direct calculation shows that the B-matrix associated with the normalization $\text{Cov}(e_t|\mathcal{F}_{t-1}) = I_d$ is obtained as

$$B_t = W(\alpha_{1,t}I_d + \sum_{m=2}^{M} \alpha_{m,t}\Lambda_m)^{1/2}, \tag{4.3.5}$$

where $B_tB_t' = \Omega_{u,t}$. Since $B_t^{-1}\Omega_m B_t'^{-1} = \Lambda_m(\sum_{n=1}^{M} \alpha_{n,t}\Lambda_n)^{-1}$ where $\Lambda_1 \equiv I_d$, the B-matrix (4.3.5) simultaneously diagonalizes $\Omega_1, ..., \Omega_M$, and $\Omega_{u,t}$ for each t so that the structural error's conditional covariance matrix is normalized to an identity matrix:

$$\text{Cov}(e_t|\mathcal{F}_{t-1}) = \sum_{m=1}^{M} \alpha_{m,t}\Lambda_m \left(\sum_{n=1}^{M} \alpha_{n,t}\Lambda_n\right)^{-1} = I_d. \tag{4.3.6}$$

The SGMVAR model assumes a single B-matrix that varies continuously in time according to the conditional covariance matrix of the reduced form error, which in turn varies according to the mixing weights. I established that under Assumption 4.1 and a normalization of the structural error's conditional variance, the B-matrix is unique up to ordering of its columns and switching all signs in a column. Hence, as long as one is willing to impose the assumption of a single (time-varying) B-matrix, the columns of the B-matrix unambiguously characterize the estimated impact effects of the shocks, but they do not reveal which column is related to which shock. Since the impact matrix is also subject to estimation error, further constraints may be needed for labelling the shocks.[2]

4.3.2 Identification of the shocks

I derived a locally unique solution for the B-matrix (4.3.5) under Assumption 4.1. However, global identification requires fixing the signs and the ordering of its the columns. The signs can be fixed by placing a single strict sign constraint in each of the columns of W, whereas the ordering of the columns can be fixed by fixing an ordering for the eigenvalues λ_{mi} in the diagonals of Λ_m. This leads to statistical identification of the model with any arbitrary ordering, but it does not reveal which column of the B-matrix is related to which shock.

A structural shock relates to an economic shock through the specific constraints in the corresponding column of W (or equally of the B-matrix) that only the shock of interest satisfies. If such constraints are readily satisfied in the (unrestricted) estimate of W, the identification amounts to labelling the structural shocks by the appropriate economic shocks, as long as the constraints are strong enough to pin down a unique ordering for the columns of W (this argument will be formalized in Proposition 4.2 below). If the unrestricted estimate of W is such

[2] As opposed to my single B-matrix, an alternative specification of the structural model would incorporate a separate B-matrix for each of the regimes. My B-matrix allows the magnitude of the impact effects of a constant sized shock to vary according to the mixing weights, but unlike my model, the alternative specification would allow variation in the impact effects also relative to the other variables. The alternative specification would, however, require other identification constraints, which may also restrict variation of the impact effects.

that the shocks of interest cannot be uniquely associated to it, the appropriate constraints can be placed for their identification.[3]

As in practice the interest is often in identifying only some specific shock or shocks, it is of interest to consider only partial identification of the B-matrix as well. Specifically, the jth structural shock is uniquely identified if the jth column of the B-matrix (4.3.5) is unique for given mixing weights $\alpha_{1,t}, ..., \alpha_{M,t}$. This requires that the jth columns of W and Λ_m, $m = 2, .., M$, are unique. The following proposition gives sufficient conditions for global identification of the last d_1 shocks when the related pairs of λ_{mi} are distinct for some m (which is always the case under Assumption 4.1 but does not require Assumption 4.1 if $d_1 < d$).

Proposition 4.2. *Suppose $\Omega_1 = WW'$ and $\Omega_m = W\Lambda_m W'$, $m = 2, ..., M$, where $\Lambda_m = diag(\lambda_{m1}, ..., \lambda_{md})$, $\lambda_{mi} > 0$ $(i = 1, ..., d)$, contains the eigenvalues of $\Omega_m\Omega_1^{-1}$ in the diagonal and the columns of the nonsingular W are the related eigenvectors. Then, the last d_1 structural shocks are uniquely identified if*

(1) for all $j > d - d_1$ and $i \neq j$ there exists an $m \in \{2, ..., M\}$ such that $\lambda_{mi} \neq \lambda_{mj}$,

(2) the columns of W are constrained in a way that for all $i \neq j > d - d_1$, the ith column cannot satisfy the constraints of the jth column as is nor after changing all signs in the ith column, and

(3) there is at least one (strict) sign constraint in each of the last d_1 columns of W.

Condition (3) of Proposition 4.2 fixes the signs in the last d_1 columns of W and therefore the signs of the instantaneous effects of the corresponding shocks. Changing the signs of the columns is effectively the same as changing the signs of the corresponding shocks, so Condition (3) is not restrictive, however (as the structural shock has a distribution that is symmetric about zero). The assumption that the identified shocks are the last d_1 shocks is neither restrictive as one may always reorder the structural shocks accordingly.

For example, if $d = 3$, $\lambda_{m1} \neq \lambda_{m3}$ for some m, and $\lambda_{m2} \neq \lambda_{m3}$ for some m, the third structural shock can be identified with the following constraints:

$$B_t = \begin{bmatrix} * & * & * \\ + & + & - \\ + & + & + \end{bmatrix} \text{ or } \begin{bmatrix} - & * & + \\ - & + & - \\ * & + & + \end{bmatrix} \text{ or } \begin{bmatrix} + & + & 0 \\ * & * & * \\ * & * & + \end{bmatrix} \tag{4.3.7}$$

and so on, where "$*$" signifies that the element is not constrained, "$+$" denotes a strict positive and "$-$" a strict negative sign constraint, and "0" means that the element is constrained to zero. In the first example, Condition (2) is satisfied because the last shock is assumed to move the last two variables to the opposite directions when the first two shocks are assumed to move them to the same direction, implying that the first two shocks cannot satisfy the constraint imposed on

[3] For a more thorough discussion on economic shocks and their identification, see Ramey (2016) and Uhlig (2017), for example.

the last shock (as is nor after changing all signs of the impact responses). Similarly in the second example, the last shock moves to opposite directions the variables that the first two shocks move to the same direction. The last example imposes a zero constraint for the impact response of the first variable to the last shock, while the first two shocks impose strict sign constraints. Since the non-zero impact responses of the first two shocks cannot satisfy the zero constraint of the last shock, Condition (2) is satisfied. By using sign and zero constraints in this manner, it is easy to produce further examples that lead to the identification of the last shock.

Imposing sign or zero constraints on W equals to placing them on B_t, so they can be justified economically. Under Assumption 4.1, the model is statistically identified prior to imposing the constraints, making the parameter constraints required in Condition (2) also testable. This different to the conventional SVAR setup in which the identifying constraints cannot be validated statistically (e.g., Kilian and Lütkepohl, 2017, Chapters 8 and 10). Similarly to the conventional SVAR model, labelling the shocks formally with the economic shocks of interest, however, requires the identification constraints to be economically motivated. As Proposition 4.2 shows and the examples in (4.3.7) demonstrate, my method facilitates finding economically plausible identification constraints by flexibly using sign constraints alone or in combination with zero constraints. A point identification can be obtained even with only sign constraints, while in the conventional SVAR setup, sign constraints alone lead to a set identification only (e.g., Kilian and Lütkepohl, 2017, Chapter 13). If Assumption 4.1 fails, the structural GMVAR model is not fully identified and the problem of testing the parameter constraints is non-standard, which is briefly addressed in the next section.

4.3.3 Identification of the shocks under partial identification of the model

If Assumption 4.1 is violated and the structural GMVAR model is thus not statistically identified, the shocks of interest can still be identified with Proposition 4.2 if Condition (1) is satisfied. When the shocks of interest do not satisfy Condition (1), their identification requires stronger constraints than in Proposition 4.2. Therefore, I present the following proposition that provides sufficient criteria for global identification of the last d_1 shocks when Condition (1) fails; specifically, when exactly one of the eigenvalues λ_{mi} with $i \neq j > d - d_1$ is identical to λ_{mj} for all m. For simplicity, I assume that only one of the shocks with identical eigenvalues is to be identified, i.e., $i \leq d - d_1$ above.

Proposition 4.3. *Let $d_1 < d$. Consider the matrix decomposition of Proposition 4.2 and further suppose that for $j = d - d_1 + 1$ and some $i \leq d - d_1$, we have $\lambda_{mi} = \lambda_{mj}$ for all m, but for all $l \notin \{i, j\}$, $\lambda_{ml} \neq \lambda_{mj}$ for some m. Then, the last d_1 structural shocks are uniquely identified if Conditions (1)-(3) of Proposition 4.2 are otherwise satisfied, and in addition*

(4) the column $i \leq d - d_1$ of W such that $\lambda_{mi} = \lambda_{mj}$ for all m has at least one (strict) sign constraint and the jth column has a zero constraint where the ith column has the (strict) sign constraint.

Note that the assumption $j = d - d_1 + 1$ is made without loss of generality because the structural shocks can always be reordered accordingly by also reordering the columns of W (including the constraints) and the eigenvalues λ_{mi} correspondingly.

To exemplify, if $d = 4$, $\lambda_{m1} \neq \lambda_{m4}$ for some m, $\lambda_{m2} \neq \lambda_{m4}$ for some m, and $\lambda_{m3} = \lambda_{m4}$ for all m, the following constraints lead to global identification of last shock:

$$B_t = \begin{bmatrix} * & * & * & * \\ * & * & + & 0 \\ + & + & * & - \\ + & + & * & + \end{bmatrix} \text{ or } \begin{bmatrix} * & * & - & 0 \\ * & * & * & * \\ + & - & * & + \\ - & + & * & + \end{bmatrix} \text{ or } \begin{bmatrix} + & - & - & 0 \\ * & * & * & * \\ * & * & * & * \\ * & * & * & + \end{bmatrix} \tag{4.3.8}$$

and so on. Condition (4) is satisfied in each of the above examples, because the third shock has a strict sign constraint for the variable that the last shock imposes a strict zero constraint. As is demonstrated above, the structural shocks can often be identified with flexible constraints even when some of the eigenvalues are identical for all regimes.

Under Conditions (1) and (3) of Proposition 4.2, the additional constraints on W were stated testable because they are overidentifying and statistical identification of the model can always be achieved by fixing the order of the eigenvalues λ_{mi}, as long as none of the pairs of λ_{mi}, $i = 1, ..., d$, is identical for all $m = 2, ..., M$. In the setup of Proposition 4.3, however, when $\lambda_{mi} = \lambda_{mj}$ for all m and some $i \neq j$, the model is not generally identified even when one fixes a unique ordering for the eigenvalues and the columns of W. Also, even if Condition (4) of Proposition 4.3 is satisfied, only partial identification of the B-matrix is obtained since nothing guarantees unique identification of the ith column of W, which would require stronger conditions. Consequently, the model is not identified under the null nor the alternative hypothesis when testing for the constraints in Conditions (2) and (4), making the testing problem nonstandard and the conventional asymptotic distributions of the likelihood ratio and Wald test statistics unreliable. The same applies when one tests the equality of the eigenvalues in order to asses the validity of Condition (1) of Proposition 4.2, as the model is not identified under the null. Deriving formal tests under no identification is, however, a major task and beyond the scope of this essay.[4]

If more than two eigenvalues are identical for all $m = 2, ..., M$ but they are not all identical, it may still be possible to find flexible conditions for identification of the shocks. Specifically, the idea utilized in the proof of Proposition 4.3 (presented in Appendix 4.A) can be applied to larger numbers of identical eigenvalues. If all the eigenvalues are identical for all covariance matrices, then $\Omega_m = \lambda_{m1}\Omega_1$ and the identification condition is the same as for the conventional

[4] Lütkepohl, Meitz, Netšunajev, and Saikkonen (2021) discussed a related testing problem under no identification and developed an asymptotic Wald type test for testing equality of the λ_{mi} parameters in the context of a linear SVAR model incorporating two volatility regimes with a known change point and (reduced form) shocks arriving from a class of elliptical distributions. Meitz and Saikkonen (2021), on the other hand, studied the asymptotic properties of a likelihood ratio test statistic under no identification when testing for the number of regimes in mixture models with Gaussian conditional densities. One of the studied models is the GMAR model (Kalliovirta, Meitz, and Saikkonen, 2015), which is the univariate counterpart of the GMVAR model (Kalliovirta *et al.*, 2016). See Lewis (2022) for a discussion on weak identification in models identified by heteroskedasticity.

SVAR model (which is given, for example, in Lütkepohl, 2005, Section 9.1.2 for the B-model). As a general remark, observe that constraining an element of B_t to be any constant other than zero is infeasible, because all elements on the right side of (4.3.5) are either zero or time-varying due to the time-varying mixing weights.[5]

4.3.4 Maximum likelihood estimation

The parameters of the reduced form GMVAR model are collected to the vector $\boldsymbol{\theta} = (\boldsymbol{\vartheta}_1, ..., \boldsymbol{\vartheta}_M, \alpha_1, ..., \alpha_{M-1})$ $((M(d^2p+d+d(d+1)/2)+1)\times 1)$, where $\boldsymbol{\vartheta}_m = (\phi_{m,0}, \text{vec}(A_{m,1}), ..., \text{vec}(A_{m,p}), \text{vech}(\Omega_m))$, vec is a vectorization operator that stacks the columns of a matrix on top of each other, and vech stacks the columns of a matrix from the main diagonal downwards (including the main diagonal). The last mixing weight parameter α_M is omitted because it is obtained from the constraint $\sum_{m=1}^{M} \alpha_m = 1$.

Using the notation described in Section 4.2, indexing the observed data as $y_{-p+1}, ..., y_0, y_1, ..., y_T$, and assuming that the initial values $\boldsymbol{y}_0 = (y_{-p+1}, ..., y_0)$ are stationary, the exact log-likelihood function of the reduced form GMVAR model takes the form (Kalliovirta *et al.*, 2016, Equations (9) and (10))

$$L_t(\boldsymbol{\theta}) = \log\left(\sum_{m=1}^{M} \alpha_m n_{dp}(\boldsymbol{y}_0; \mathbf{1}_p \otimes \mu_m, \boldsymbol{\Sigma}_m)\right) + \sum_{t=1}^{T} l_t(\boldsymbol{\theta}), \tag{4.3.9}$$

where

$$l_t(\boldsymbol{\theta}) = \log\left(\sum_{m=1}^{M} \alpha_{m,t}(2\pi)^{-d/2}\det(\Omega_m)^{-1/2}\exp\left\{-\frac{1}{2}(y_t - \mu_{m,t})'\Omega_m^{-1}(y_t - \mu_{m,t})\right\}\right). \tag{4.3.10}$$

If it does not seem reasonable to assume that the initial values are stationary, one may condition on them and base the estimation on the conditional log-likelihood function, which is obtained by dropping the first term on the right side of (4.3.9).

The reduced form GMVAR model can be estimated by maximizing the exact or conditional likelihood function in (4.3.9) and (4.3.10) with respect to the parameter $\boldsymbol{\theta}$. To ensure identifica-

[5] I have focused on the B-model, where the structure is imposed on the contemporaneous relations of the shocks. Alternatively, one may consider the "A-model" setup in which the structure is placed on the contemporaneous relations of the observable variables governed by the "A-matrix" (see, e.g., Lütkepohl, 2005, Section 9.1.1). The A-model is obtained implicitly from the B-model (4.3.1) and (4.3.2) by defining the A-matrix as $A_t \equiv B_t^{-1}$, where B_t is given by (4.3.5). In this case, the structural model Equation (4.3.1) becomes

$$A_t y_t = \sum_{m=1}^{M}(A_t\phi_{m,0} + \sum_{i=1}^{p} A_t A_{m,i} y_{t-i}) + e_t,$$

thereby incorporating continuously varying intercepts and coefficient matrices due to the constant covariance normalization of the structural shocks. In practice, however, one needs to carefully derive how any specific constraint on A_t can be imposed by restricting W.

tion, the parameter space should be constrained so that the mixture components cannot be 'relabelled', for instance, by assuming that the mixing weight parameters are in a decreasing order, $\alpha_M > \cdots \alpha_1 > 0$, and $\boldsymbol{\vartheta}_i = \boldsymbol{\vartheta}_j$ only if $i = j$ (Kalliovirta *et al.*, 2016, Equation (11)). If $M = 2$, the structural GMVAR model is then obtained by simultaneously diagonalizing the reduced form error covariance matrices as discussed Section 4.3.1. However, should overidentifying restrictions be imposed on B_t through W or if $M \geq 3$, it is more convenient to reparametrize the model with W and Λ_m, $m = 2, ..., M$, instead of $\Omega_1, ..., \Omega_M$ and maximize the log-likelihood function subject to the new set of parameters and constraints. In this case, the decomposition (4.3.4) is plugged in to the log-likelihood function and the $\text{vech}(\Omega_1), ..., \text{vech}(\Omega_M)$ are replaced with $\text{vec}(W), \lambda_2, ..., \lambda_M$, where $\lambda_m = (\lambda_{m1}, ..., \lambda_{md})$, in the parameter vector $\boldsymbol{\theta}$.

Maximizing the complex and highly multimodal log-likelihood function can be challenging in practice, particularly if there are more than two regimes. Following Dorsey and Mayer (1995), Meitz, Preve, and Saikkonen (2018, forthcoming), and Virolainen (2018b, forthcoming), I employ a two-phase estimation procedure where, in the first phase, a genetic algorithm is used to find starting values for a gradient based method which then, in the second phase, often converges to a nearby local maximum or saddle point. The genetic algorithm in the accompanying R package **gmvarkit** (Virolainen, 2018a) has been modified to improve its performance significantly, and it functions similarly to the one described in Chapter 2 for the univariate GMAR (Kalliovirta *et al.*, 2015), StMAR (Meitz *et al.*, forthcoming), and G-StMAR (Virolainen, forthcoming, also Chapter 2) models. In order to obtain reliable results, a (sometimes very large) number of estimation rounds should be performed, for which **gmvarkit** makes use of parallel computing.

4.4 Impulse response analysis

The expected effects of the structural shocks in the SGMVAR model generally depend on the initial values as well as on the sign and size of the shock, which makes the conventional way of calculating impulse responses unsuitable (see, e.g., Kilian and Lütkepohl, 2017, Chapter 4). Following Koop *et al.* (1996) and Kilian and Lütkepohl (2017, Section 18.2.2), I therefore consider the generalized impulse response function (GIRF) defined as

$$\text{GIRF}(h, \delta_j, \mathcal{F}_{t-1}) = \text{E}[y_{t+h}|\delta_j, \mathcal{F}_{t-1}] - \text{E}[y_{t+h}|\mathcal{F}_{t-1}], \tag{4.4.1}$$

where h is the chosen horizon and $\mathcal{F}_{t-1} = \sigma\{y_{t-j}, j > 0\}$ as before. The first term on the right side is the expected realization of the process at time $t + h$ conditionally on a structural shock of size $\delta_j \in \mathbb{R}$ in the jth element at time t and the previous observations. The latter term on the right side is the expected realization of the process conditionally on the previous observations only. The GIRF thus expresses the expected difference in the future outcomes when the structural shock of size δ_j in the jth element hits the system at time t as opposed to all shocks being random.

It is easy to see that the SGMVAR model has a p-step Markov property, so conditioning on (the σ-algebra generated by) the p previous observations $\boldsymbol{y}_{t-1} = (y_{t-1}, ..., y_{t-p})$ is effectively the

same as conditioning on $\mathcal{F}_{t-1}$ at time t and later. The history $\boldsymbol{y}_{t-1}$ can be either fixed or random, but with random history the GIRF becomes a random vector, however. Using fixed $\boldsymbol{y}_{t-1}$ makes sense when one is interested in the effects of the shock at a particular point of time, whereas more general results are obtained by assuming that $\boldsymbol{y}_{t-1}$ follows the stationary distribution of the process. If one is, on the other hand, interested in a specific regime, $\boldsymbol{y}_{t-1}$ can be assumed to follow the stationary distribution of the corresponding component model.

The GIRF and its distributional properties can be approximated with a Monte Carlo algorithm that generates (partial) realizations of the process and then takes the sample mean for point estimate. If $\boldsymbol{y}_{t-1}$ is random and follows the distribution G, the GIRF should be estimated for different values of $\boldsymbol{y}_{t-1}$ generated from G, and then the sample mean and sample quantiles can be taken to obtain the point estimate and confidence intervals that reflect the uncertainty about the initial value. Such an algorithm, adapted from Koop *et al.* (1996, pp. 135-136) and Kilian and Lütkepohl (2017, pp. 601-602), is given in Appendix 4.B.

Because the SGMVAR model facilitates associating statistical characteristics and economic interpretations to the regimes, and because asymmetries in the GIRFs are caused by regime-switches, it may be of interest to also examine the effects of a structural shock to the mixing weights $\alpha_{m,t}$, $m = 1, ..., M$. We then consider the related GIRFs

$$\text{GIRF}_{\alpha_m}(h, \delta_j, \mathcal{F}_{t-1}) = \text{E}[\alpha_{m,t+h}|\delta_j, \mathcal{F}_{t-1}] - \text{E}[\alpha_{m,t+h}|\mathcal{F}_{t-1}] \tag{4.4.2}$$

for which point estimates and confidence intervals can be constructed similarly to (4.4.1).

4.5 Empirical application

My empirical application studies asymmetries in the effects of U.S. monetary policy shocks. Asymmetric effects of U.S. monetary policy shocks have been studied, among others, by Weise (1999), Garcia and Schaller (2002), Lo and Piger (2005), and Höppner, Melzer, and Neumann (2008), who all found the effects of monetary policy shocks to production stronger during recessions (or low growth periods) than booms (or high growth periods). Weise (1999) also found evidence in favor of large and small shocks having different effects, and large positive and negative monetary shocks having different effects. Höppner *et al.* (2008) concluded that the real effects of monetary policy shocks have decreased over their sample period from 1962 to 2002. Tenreyro and Thwaites (2016), on the other hand, found the effects of U.S. monetary policy shocks less powerful in recessions. Primiceri (2005) found evidence of time-variation in the U.S. monetary policy.

I consider the quarterly U.S. data covering the period from 1954Q3 to 2021Q4 (270 observations) and consisting of four variables: real GDP, GDP implicit price deflator, producer price index (all commodities), and an interest rate variable. My policy variable is the interest rate variable, which is the effective federal funds (FF) rate from 1954Q3 to 2008Q2. After that I replaced it with the Wu and Xia (2016) shadow rate, which is not constrained by the zero lower bound and also quantifies unconventional monetary policy measures. The Wu and Xia (2016)

shadow rate series was retrieved from the Federal Reserve Bank of Atlanta's website and the rest of the data were retrieved from the Federal Reserve Bank of St. Louis database.

The logarithms of real GDP, GDP deflator, and producer price index are clearly nonstationary (the series are not shown) and hence need to be detrended before fitting a GMVAR model to the series. I detrended the logarithm of the real GDP by separating its cyclical component from the trend with the one-sided Hodrick-Prescott (HP) filter and then considering the cyclical component.[6] I thereby implicitly assume that the monetary policy shock does not have permanent effects on real output. I detrended the logarithms of the price variables by taking the first difference and multiplying it by hundred, so that the resulting series approximate the percentage growth rates. The interest rate variable is treated as stationary.

The series are presented in the first four top panels of Figure 4.1 with the shaded areas indicating the periods of NBER based U.S. recessions. Throughout, I refer to the variables as GDP (output), GDPDEF (prices), PPI (commodity prices), and RATE (interest rate) without making it explicit that some of them are detrended. The (S)GMVAR model of autoregressive order p and M mixture components is referred to as (S)GMVAR(p, M) model.

I select the order of my GMVAR model by first finding a suitable autoregressive order for a linear Gaussian VAR; that is, a GMVAR($p, 1$) model. The AIC is minimized by the order $p = 3$, suggesting that this might be the appropriate lag order for modelling autocorrelation. So I estimate a GMVAR($3, 2$) model, which I find superior to the linear VAR. Graphical quantile residual diagnostics reveal that my GMVAR($3, 2$) model adequately captures the autocorrelation structure of the series, but some of the conditional heteroskedasticity and excess kurtosis is not captured. In my view, the overall adequacy of the model is, nevertheless, reasonable enough for further analysis. Details on the model selection and quantile residual diagnostics are given in Appendix 4.C.

The estimated mixing weights of the two regimes are presented in the bottom panel of Figure 4.1. The second regime (red) mainly dominates during periods of high inflation and interest rate in the 1970's and 1980's, after the collapse of Lehman Brothers in the Financial crisis until the end of 2009, and finally during the COVID-19 crisis from the third quarter of 2020 onwards. I refer to this regime as *the unstable inflation regime*, as it generally exhibits high or volatile inflation. The first regime (blue) prevails when the second one does not: before 1970's, short periods during 1970's, and from the mid 1980's onwards but excluding the Financial crisis and the COVID-19 crisis (but including the first two quarters of 2020). I refer to this regime as *the stable inflation regime*, as it is characterized by moderate inflation. Details about the characteristics of the regimes are provided in Appendix 4.C.

[6] The one-sided HP filter was obtained from the two-sided HP filter by applying the filter up to horizon t, taking the last observation, and repeating this procedure for the full sample $t = 1, ..., T$. In order to allow the series to start from any phase of the cycle, I applied the one-sided filter to the full available sample from 1947Q1 to 2021Q4 before extracting my sample period from it. I computed the two-sided HP filter with the R package **lpirfs** (Adämmer, 2021) by using the standard smoothing parameter value of 1600. For robustness, I also considered the first differences and the linear projection filter proposed by Hamilton (2018). But they led to generalized impulse response functions where supposedly contractionary monetary policy shocks seemed to have significant, persistent expansionary effects on real GDP (not shown). Hence, I preferred the HP filter.

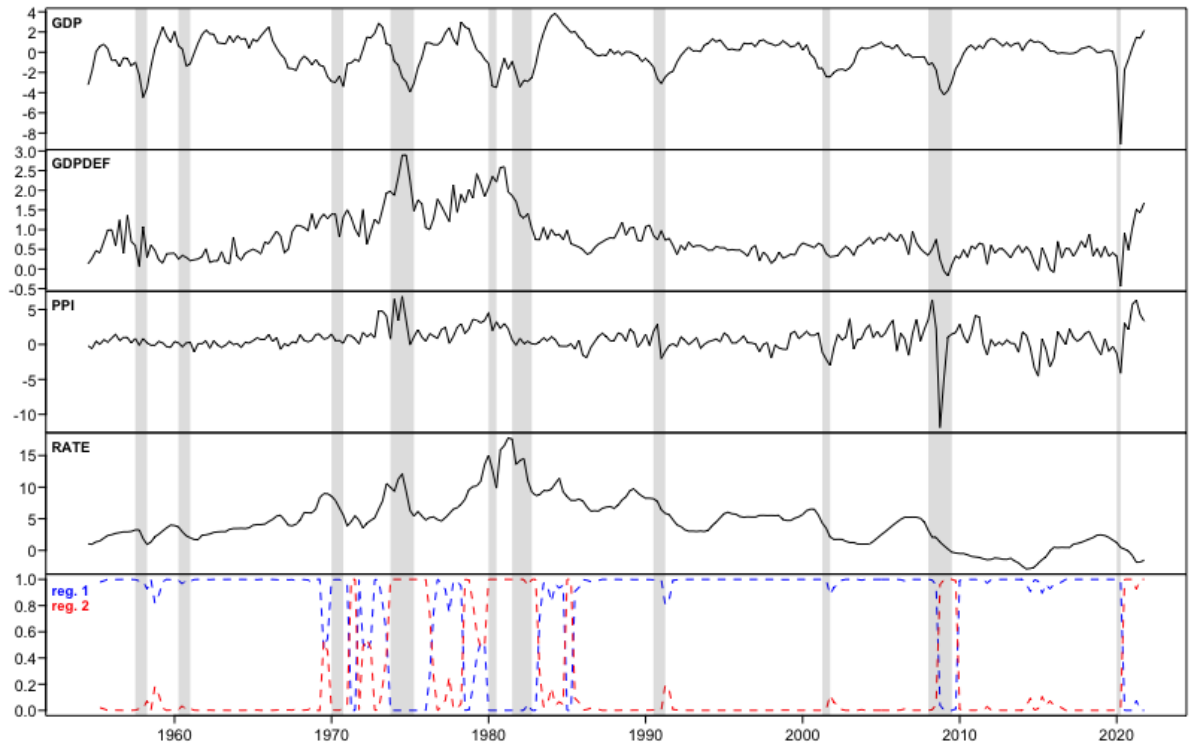

Figure 4.1: Quarterly U.S. series covering the period from 1954Q3 to 2021Q4. The top panel presents the cyclical component of real GDP (GDP) which I separated from the trend using the one-sided Hodrick-Prescott filter. The second and third panels present the log-differences of GDP implicit price deflator (GDPDEF) and producer price index (PPI) multiplied by hundred. The fourth panel presents an interest rate variable, which is the effective federal funds from 1954Q3 to 2008Q2 and the Wu and Xia (2016) shadow rate from 2008Q3 to 2021Q4. The bottom panel shows the estimated mixing weights of the fitted GMVAR$(3, 2)$ model. The shaded areas indicate the NBER based U.S. recessions.

4.5.1 Identification of the monetary policy shock

Decomposing the covariance matrices of the reduced form GMVAR$(3, 2)$ model as in (4.3.4) gives the following estimates for the structural parameters:

$$\hat{W} = \begin{bmatrix} \mathbf{0.14}\ (0.054) & \mathbf{0.22}\ (0.065) & \mathbf{0.44}\ (0.053) & -0.13\ (0.136) \\ \mathbf{-0.20}\ (0.014) & -0.05\ (0.028) & \mathbf{0.07}\ (0.012) & -0.00\ (0.022) \\ 0.00\ (0.168) & \mathbf{-1.03}\ (0.078) & \mathbf{0.47}\ (0.110) & -0.06\ (0.158) \\ 0.03\ (0.029) & 0.03\ (0.041) & 0.18\ (0.100) & \mathbf{0.30}\ (0.055) \end{bmatrix}, \quad \hat{\lambda}_2 = \begin{bmatrix} \mathbf{1.08}\ (0.227) \\ \mathbf{3.02}\ (0.636) \\ \mathbf{11.05}\ (2.473) \\ \mathbf{18.20}\ (3.601) \end{bmatrix}, \tag{4.5.1}$$

where the ordering of the variables is $y_t = (\text{GDP}_t, \text{GDPDEF}_t, \text{PPI}_t, \text{RATE}_t)$, the estimates $\hat{\lambda}_{2i}$ are in an increasing order (which fixes an arbitrary ordering for the columns of $\hat{W}$), and approximate standard errors are given in parentheses next to the estimates. The estimates that deviate

from zero by more than two times their approximate standard error are bolded. I proceed by assuming that all the λ_{2i}, $i = 1, ..., 4$, are different to each other, i.e., that Assumption 4.1 holds, which leads to statistical identification of the model. After identifying the monetary policy shock, the robustness of my identification with respect to this unjustified assumption is discussed.[7]

Based on the estimates and their standard errors in (4.5.1), the first shock moves GDP and inflation to the opposite directions, whereas the second shock moves GDP and commodity price inflation to the opposite directions. Since the instantaneous movements of the interest rate variable are insignificant, these two shocks do not seem plausible candidates for the monetary policy shock. The third shock moves the interest rate variable on impact more significantly than the first two shocks, but since production and both prices move significantly to the same direction, its characteristics appear similar to an aggregate demand shock and not a monetary policy shock. The last shock moves the interest rate variable significantly, while it also moves GDP, inflation, and commodity price inflation to the opposite direction, which is consistent with many of the standard the economic theories (e.g., Galí, 2015). The impact effects of GDP, inflation, and commodity price inflation are, however, statistically insignificant and the response of inflation is very weak. Nevertheless, among the four structural shocks obtained for the model, the characteristics of the last shock mostly resemble those of a monetary policy shock, so I deem it as the monetary policy shock.

Identifying the monetary policy shock formally by Proposition 4.2 requires such constraints to be imposed on W that it can be unambiguously distinguished from the other shocks. I assume that the monetary policy shock moves the GDP and commodity price inflation to the opposite direction from the interest rate variable. In addition, I impose a zero constraint on the instantaneous movement of inflation, as the unrestricted estimated is very close to zero compared to the approximate standard error, and it allows to avoid making restrictive assumptions about the first and third shocks. The Wald test produces the p-value 0.92 for the zero constraint, so it is not rejected.

To distinguish the monetary policy shock from the other shocks, I assume that the first and third shocks move inflation at impact. This is not economically restrictive (since the responses can be very small), but it is a statistically reasonable assumption, as the Wald test rejects the hypotheses that the impact responses are zero (jointly or individually) with p-values less than 10^{-7}. The Wald test produces the p-value 0.056 for the hypothesis that the second shock does not move inflation at impact, so it is not rejected. Therefore, I assume that the second shock moves the GDP and commodity price inflation to the opposite directions, as the corresponding estimates are large compared to their approximate standard errors making the constraints statistically sensible.

The above-described identification produced the following estimates for the GMVAR$(3, 2)$

[7] The approximate standard errors cannot be used to infer about the (in)equality of λ_{2i}, $i = 1, ..., 4$, without considerably complex examinations, as they are (asymptotically) valid only if the λ_{2i} are different to each other in the first place (see Lütkepohl *et al.*, 2021, for a related discussion).

model:

$$\hat{W}_2 = \begin{bmatrix} \mathbf{0.14}\ (0.054) & \mathbf{0.22}\ (0.065) & \mathbf{0.45}\ (0.037) & -0.12\ (0.065) \\ \mathbf{-0.20}\ (0.014) & -0.05\ (0.028) & \mathbf{0.07}\ (0.012) & 0 \\ 0.00\ (0.168) & \mathbf{-1.03}\ (0.078) & \mathbf{0.47}\ (0.109) & -0.05\ (0.068) \\ 0.03\ (0.028) & 0.03\ (0.041) & \mathbf{0.17}\ (0.044) & \mathbf{0.30}\ (0.026) \end{bmatrix}, \quad \hat{\lambda}_2 = \begin{bmatrix} \mathbf{1.08}\ (0.227) \\ \mathbf{3.02}\ (0.640) \\ \mathbf{11.05}\ (2.583) \\ \mathbf{18.20}\ (4.398) \end{bmatrix} \tag{4.5.2}$$

The estimates changed only slightly from the unrestricted ones in (4.5.1), and the negative impact responses of output and commodity prices in the fourth column remain statistically insignificant. The estimates for λ_{23} and λ_{24} are somewhat close to each other relative to their standard errors, but due to my zero constraint on the inflation, Proposition 4.3 identifies the monetary policy shock even if $\lambda_{23} = \lambda_{24}$ (or $\lambda_{21} = \lambda_{24}$). The monetary policy shock is identified also if additionally $\lambda_{2i} = \lambda_{2j}$ for any $i, j = 1, 2, 3$, so my identification is not particularly sensitive to the validity of the unjustified Assumption 4.1 (while the approximate standard errors and the Wald test results are invalid if the assumption fails).

4.5.2 Generalized impulse response functions

Due to the endogenously determined regime-switching probabilities and the fact that I allow the regime to switch as a result of a shock, there are multiple types of possible asymmetries: the impulse responses can vary depending on the initial value as well as on the sign and size of the shock. I study the state-dependence of the (generalized) impulse response functions by drawing initial values from the stationary distribution of each regime separately. Then, I calculate the 90% confidence intervals that reflect uncertainty about the initial value within the given regime as is described in Section 4.4 and Appendix 4.B. Asymmetries related to the sign and size of the shock are studied by estimating the GIRFs for positive (contractionary) and negative (expansionary) one-standard-error (small) and two-standard-error (large) shocks. After estimating the GIRFs, they are scaled so that the peak effect of the interest variable is 25 basis points within the first four quarters, making the responses to shocks of different sign and size comparable.[8]

Figure 4.2 presents the GIRFs $h = 0, 1, ..., 32$ quarters ahead estimated for the identified monetary policy shock.[9] The GIRFs of inflation rate and commodity price inflation rate are not accumulated to levels. From top to bottom, the responses of GDP, inflation rate, commodity price inflation rate, interest rate, and the first regime's mixing weights are depicted in each row, respectively. The first [third] column shows the responses to small contractionary (blue solid line) and expansionary (red dashed line) shocks with the initial values generated from the stationary distribution of the first [second] regime. The second [fourth] column shows the responses to large contractionary and expansionary shocks with the initial values generated from

[8] The GIRFs are scaled based on the peak response instead of the initial response, because the peak response is much higher compared to the initial response in the first regime than in the second regime. Scaling the GIRFs based on the initial response would then shift the response of the interest rate variable significantly higher in the first regime than in the second regime.

[9] I use $R_1 = R_2 = 2500$ in the Monte Carlo algorithm, i.e., for each regime, size, and sign of the shock I draw 2500 initial values, and for each of those initial values the GIRF is estimated based on 2500 different sample paths.

the first [second] regime. The shaded areas are the 90% confidence intervals that reflect uncertainty about the initial value within the given regime. The responses of the second regime's mixing weights are not depicted because they are the negative of those of the first regime.

In the first regime (the stable inflation regime; the first and second columns of Figure 4.2), contractionary (expansionary) monetary policy shock causes a significant contraction (expansion) in the GDP, with the peak effect occurring after three to four quarters. On average, the response is hump-shaped and decays to zero roughly after three years from impact.[10] As the confidence bounds show, the response switches sign from some of the starting values, while some of the starting values display a persistent contraction. Particularly large shocks seem to cause a delayed expansion (contraction) from many of the starting values - shortly after the impact when the shock is contractionary and later when the shock is expansionary.

The prices seem to mostly rise in response to a contractionary monetary policy shock, although one would often expect contractionary monetary policy shocks to decrease inflation due to the decreased demand. This is often referred to as the price puzzle, and it has been discussed recently, for instance, in Ramey (2016, Section 3.3.2, and the references therein).[11] As is explained in Appendix 4.C.3, the prices rise because the shock drives the economy towards the unstable inflation regime, which has higher long-run inflation.

On average, inflation does not move much in response to a small expansionary monetary policy shock, while the interest rate stays low relatively persistently and is accompanied with a roughly three years long expansion of the GDP. A large expansionary shock drives the economy towards the unstable inflation regime relatively more than a small expansionary shock, as the responses of the first regime's mixing weights show. Consequently, the inflation mostly increases and the interest rate variable increases relatively fast towards zero, and as the confidence bounds show, from many of the starting values the interest rate variable overshoots significantly. It is shown in Appendix 4.C.3 that these GIRFs are the ones that also display particularly high peak inflation, and that the significant monetary policy tightening is accompanied with a persistent contraction of the GDP after the initial expansion.

In the second regime (the unstable inflation regime; the third and fourth columns of Fig-

[10] By zero, I mean the expected observation if all the shocks were random. Accordingly, by positive I mean expected observations larger than that and by negative expected observations smaller than that.

[11] A popular explanation is that the Fed uses more information in predicting the future inflation than the autoregressive system of the variables included in the model (Sims, 1992). Consequently, the identified monetary policy shock also contains a component that incorporates the Fed's endogenous response to the prediction of the future inflation that is not captured by the autoregressive system of the included variables. If the endogenous response is not strong enough to offset the predicted inflation, the impulse responses may then display a rise in the inflation. Another explanation proposes that the prices increase due to the cost-push effect of the monetary policy shock: an increase in the nominal interest rate increases the marginal cost of production of the firms who operate on borrowed money, and thereby decreases the aggregate supply and increases the price level (e.g., Barth and Ramey, 2001, Ravenna and Walsh, 2006). Several authors have, however, argued that the cost-channel is not likely strong enough to cause a price puzzle even in the short-run (e.g., Castelnuovo, 2012, Kaufmann and Scharler, 2009, Rabanal, 2007). Nonetheless, I find my empirical results interesting, as the long-run price puzzle arises only from some of the starting values, while its occurrence is also sensitive to the sign and size of the shock. Moreover, as is discussed in Appendix 4.C.1, enforcing linearity to the autoregressive dynamics makes the price puzzle worse.

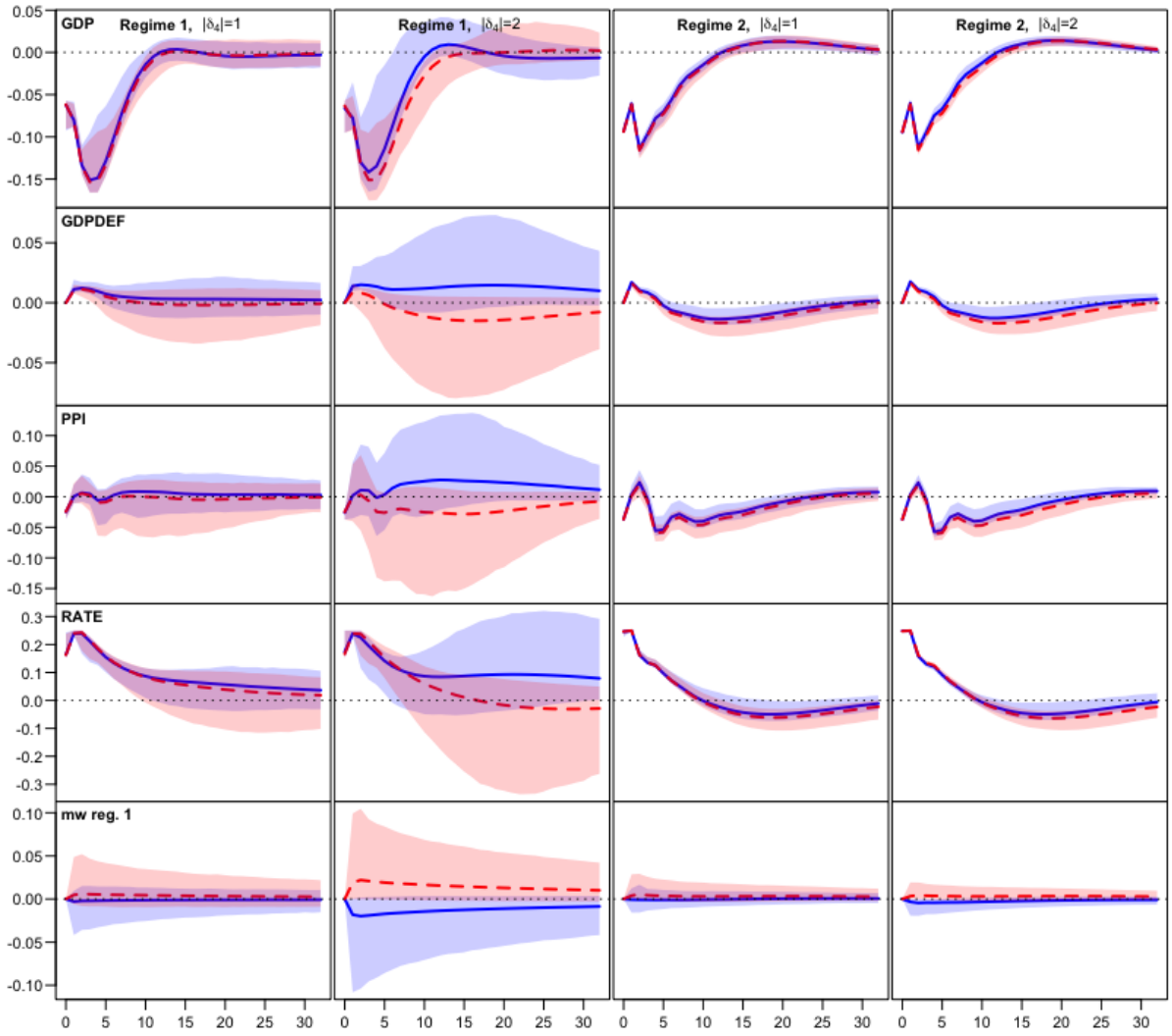

Figure 4.2: Generalized impulse response functions $h = 0, 1, ..., 32$ quarters ahead estimated for the monetary policy shock identified in Section 4.5.1 using $R_1 = R_2 = 2500$ in the Monte Carlo algorithm presented in Appendix 4.B. From the top to bottom, the responses of production, prices, commodity prices, the interest rate variable, and the first regime's mixing weights are depicted in each row, respectively. The GIRFs are not accumulated for the prices variables, i.e., to responses are for inflation rates. The first [third] column shows the responses to a positive (blue solid line) and negative (red dashed line) one-standard-error shocks with the initial values generated from the stationary distribution of the first [second] regime. The second [fourth] column shows the responses to a positive and negative two-standard-error shocks with the initial values generated from the first [second] regime. All GIRFs have been scaled so that peak effect of the interest rate variable is 25 basis points during the first four quarters. The shaded areas represent the 90% confidence intervals that reflect uncertainty about the initial state within the given regime. Responses of the second regime's mixing weights are omitted because they are the negative of the first regime's mixing weights' responses.

ure 4.2), contractionary (expansionary) monetary policy shock causes a strong contraction (expansion) of the GDP, with the peak effect occurring after two quarters. After roughly three years, from most of the starting values the response overshoots and becomes expansionary (contractionary) before decaying to zero. Inflation rate and commodity price inflation rate rise after the impact and then decrease significantly for several years before the price levels stabilize.

The interest rate decays towards zero for roughly two years after which, on average, it decreases (increases) slightly below (above) zero before returning to zero. The response of the interest rate switches its sign less significantly when the shock is contractionary, when also the inflationary effects of the shock are slightly weaker. The scaled GIRFs are almost identical for small and large shocks, so there does not appear to be much asymmetries with respect to the size of the shock. The asymmetries are weak, because the monetary policy shock has relatively weak effect on the regime-switching probabilities, as the scaled responses of the first regime's mixing weights show (the third and fourth bottom panels of Figure 4.2).

Overall, expansionary and contractionary monetary policy shocks seem to both mostly increase the probability of the unstable inflation regime, significantly more so if the economy is in the stable inflation regime when the shock arrives (the bottom panels in Figure 4.2). In the stable inflation regime, a large shock increases the probability of entering the unstable inflation regime relatively more than a small shock, and often propagates high and persistent inflation, which is followed by a significant monetary policy tightening and a persistent contraction of the GDP. On average, the real effects of monetary monetary policy shocks are somewhat stronger in the stable inflation regime than in the unstable inflation regime, but the (average) effects die out equally fast.

4.6 Summary

I introduced a structural version of the Gaussian mixture vector autoregressive model (Kalliovirta *et al.*, 2016) that incorporates endogenously determined mixing weights and a time-varying B-matrix. I showed that my model generally identifies the structural shocks up to ordering and sign, but does not reveal which column of the B-matrix is related to which shock. Since the B-matrix is also subject to an estimation error, I made use of the matrix decomposition proposed by Lanne and Lütkepohl (2010) and Lanne *et al.* (2010) and derived general conditions for formally identifying any subset of the shocks. This led to flexible identification conditions, and some of the constraints are also testable. For impulse response analysis, I utilized the generalized impulse response function (Koop *et al.*, 1996) and proposed a Monte Carlo algorithm for its estimation by making use of the known stationary distribution of the SGMVAR process. This essay is accompanied with the CRAN distributed R package **gmvarkit** (Virolainen, 2018a), which provides a comprehensive set of tools for numerical analysis of the model.

My empirical application studied asymmetries in the expected effects of monetary policy shocks in the U.S. using a quarterly series covering the period from 1954Q3 to 2021Q4. My SGMVAR model identified two regimes: a stable inflation regime and an unstable inflation regime. The unstable inflation regime is characterized by high or volatile inflation, and it mainly

prevails in the 1970's, early 1980's, during the Financial crisis, and in the COVID-19 crisis from 2020Q3 onwards. The stable inflation regime, in turn, is characterized by moderate inflation, and it prevails when the stable inflation regime does not. I found the effects of the monetary policy shock relatively symmetric in the unstable inflation regime, as it rarely causes a switch to the stable inflation regime. A contractionary (expansionary) monetary policy shock appears to first increase (decrease) inflation after which the inflation significantly decreases (increases) for several years. The strong contraction (expansion) in the cyclical component of GDP lasts for roughly three years and is followed by a relatively mild expansion (contraction) along with the interest rate variable overshooting to the negative (positive) side.

The effects of the monetary policy shock were found strongly asymmetric in the stable inflation regime with respect to the initial state of the economy as well as to the sign and size of the shock. A large shock often causes relatively stronger inflationary effects than a small shock, while both contractionary and expansionary shocks seem to increase inflation by driving the economy towards the unstable inflation regime. A small expansionary shock does not move prices much on average, but a large expansionary shock often drives the economy towards the unstable inflation regime and propagates high and persistent inflation. The high inflation is followed by a significant monetary policy tightening and a persistent contraction of the GDP after the initial expansion. On average, the real effects of the monetary policy shock were found somewhat stronger in the stable inflation regime than in the unstable inflation regime.

Bibliography

Adämmer P. (2021). ***lpirfs**: Local Projections Impulse Response Functions*. R package version: 0.2.0.

Anderson H., Vahid F. (1998). "Testing multiple equation systems for common nonlinear components." *Journal of Econometrics*, **84**(1), 1–36.

Barth M. J., Ramey V. A. (2001). "The Cost Channel of Monetary Transmission." In Bernanke B. S., Rogoff K. (eds.), *NBER Macroeconomics Annual*, volume 16, pp. 199–239. MIT Press, Cambridge.

Bentarzi M., Djeddou L. (2014). "On Mixture Periodic Vector Autoregressive Models." *Communications in Statistics - Simulation and Computation*, **43**(10), 2325–2352.

Burgard J., Neuenkirch M., Nöckel M. (2019). "State-Dependent Transmission of Monetary Policy in the Euro Area." *Journal of Money, Credit and Banking*, **51**(7), 2053–2070.

Castelnuovo E. (2012). "Testing the Structural Interpretation of the Price Puzzle with a Cost-Channel Model." *Oxford Bulletin of Economics and Statistics*, **74**(3), 425–452.

Dorsey R., Mayer W. (1995). "Genetic algorithms for estimation problems with multiple optima, nondifferentiability, and other irregular features." *Journal of Business and Economic Statistics*, **13**(1), 53–66.

Fong P., Li W., Yau C., Wong C. (2007). "On a mixture vector autoregressive model." *The Canadian Journal of Statistics*, **35**(1), 135–150.

Galí J. (2015). *Monetary Policy, Inflation, and the Business Cycle*. 2nd edition. Princeton University Press, Princeton and Oxford.

Garcia R., Schaller H. (2002). "Are the Effects of Monetary Policy Asymmetric?" *Economic Inquiry*, **40**(1), 102–119.

Hamilton J. D. (2018). "WHY YOU SHOULD NEVER USE THE HODRICK-PRESCOTT FILTER." *The Review of Economics and Statistics*, **100**(5), 831–843.

Höppner F., Melzer C., Neumann T. (2008). "Changing effects of monetary policy in the US – evidence from a time-varying coefficient VAR." *Applied Economics*, **40**(18), 2353–2360.

Kalliovirta L. (2012). "Misspecification tests based on quantile residuals." *The Econometrics Journal*, **15**(2), 358–393.

Kalliovirta L., Malinen T. (2020). "Non-Linearity and Cross-Country Dependence of Income Inequality." *The Review of Income and Wealth*, **66**(1), 227–249.

Kalliovirta L., Meitz M., Saikkonen P. (2015). "A Gaussian Mixture Autoregressive Model for Univariate Time Series." *Journal of Time Series Analysis*, **36**(2), 247–266.

Kalliovirta L., Meitz M., Saikkonen P. (2016). "Gaussian mixture vector autoregression." *Journal of Econometrics*, **192**(2), 465–498.

Kalliovirta L., Saikkonen P. (2010). "Reliable Residuals for Multivariate Nonlinear Time Series Models." *Unpublished revision of HECER discussion paper No. 247*.

Kaufmann S., Scharler J. (2009). "Financial systems and the cost channel transmission of monetary policy shocks." *Economic Modelling*, **26**(1), 40–46.

Kilian L., Lütkepohl H. (2017). *Structural Vector Autoregressive Analysis*. 1st edition. Cambridge University Press, Cambridge.

Koop G., Pesaran M., Potter S. (1996). "Impulse response analysis in nonlinear multivariate models." *Journal of Econometrics*, **74**(1), 119–147.

Lanne M., Lütkepohl H. (2010). "Structural Vector Autoregressions With Nonnormal Residuals." *Journal of Business & Economic Statistics*, **28**(1), 159–168.

Lanne M., Lütkepohl H., Maciejowsla K. (2010). "Structural vector autoregressions with Markov switching." *Journal of Economic Dynamics and Control*, **34**(2), 121–131.

Lewis D. (2022). "Robust Inference in Models Identified via Heteroskedasticity." *The Review of Economics and Statistics*, **104**(3), 510–524.

Lo M., Piger J. (2005). "Is the Response of Output to Monetary Policy Asymmetric? Evidence from a Regime- Switching Coefficients Model." *Journal of Money, Credit and Banking*, **37**(5), 865–886.

Lütkepohl H. (2005). *New Introduction to Multiple Time Series Analysis*. 1st edition. Springer, Berlin.

Lütkepohl H., Meitz M., Netšunajev A., Saikkonen P. (2021). "Testing identification via heteroskedasticity in structural vector autoregressive models." *The Econometrics Journal*, **24**(1), 1–22.

Meitz M., Preve D., Saikkonen P. (2018). ***StMAR Toolbox**: A MATLAB Toolbox for Student's t Mixture Autoregressive Models*.

Meitz M., Preve D., Saikkonen P. (forthcoming). "A mixture autoregressive model based on Student's t-distribution." *Communications in Statistics - Theory and Methods*.

Meitz M., Saikkonen P. (2021). "Testing for observation-dependent regime switching in mixture autoregressive models." *Journal of Econometrics*, **222**(1), 601–624.

Muirhead R. (1982). *Aspects of Multivariate Statistical Theory*. 1st edition. John Wiley & Sons, Hoboken, New Jersey.

Primiceri G. (2005). "Time Varying Structural Vector Autoregressions and Monetary Policy." *The Review of Economic Studies*, **72**(3), 821–852.

Rabanal P. (2007). "Does inflation increase after a monetary policy tightening? Answers based on an estimated DSGE model." *Journal of Economic Dynamics and Control*, **31**(3), 906–937.

Ramey V. A. (2016). "Macroeconomic Shocks and Their Propagation." In Taylor J. B., Uhlig H. (eds.), *Handbook of Macroeconomics*, volume 2, chapter 2. Elsevier Science B.V.

Ravenna F., Walsh C. E. (2006). "Optimal monetary policy with the cost channel." *Journal of Monetary Economics*, **53**(2), 199–216.

Sims A. (1992). "Interpreting the macroeconomic time series facts." *European economic review*, **36**(5), 975–1000.

Tenreyro S., Thwaites G. (2016). "Pushing on a String: US Monetary Policy Is Less Powerful in Recessions." *American Economic Journal: Macroeconomics*, **8**(4), 43–74.

Tsay R. (1998). "Testing and Modeling Multivariate Threshold Models." *Journal of the American Statistical Association*, **93**(443), 1188–1202.

Uhlig H. (2017). "Shocks, Sign Restrictions, and Identification." In Honoré B., Pakes A., Piazzesi M., Samuelson L. (eds.), *Advances in Economics and Econometrics*, volume 2, chapter 4. Cambridge University Press.

Virolainen S. (2018a). ***gmvarkit**: Estimate Gaussian and Student's t Mixture Vector Autoregressive Models*. R package version 2.0.3 available at CRAN: `https://CRAN.R-project.org/package=gmvarkit`.

Virolainen S. (2018b). ***uGMAR**: Estimate Univariate Gaussian and Student's t Mixture Autoregressive Models*. R package version 3.4.2 available at CRAN: `https://CRAN.R-project.org/package=uGMAR`.

Virolainen S. (forthcoming). "A mixture autoregressive model based on Gaussian and Student's t-distributions." *Studies in Nonlinear Dynamics & Econometrics*.

Weise C. (1999). "The Asymmetric Effects of Monetary Policy: A Nonlinear Vector Autoregression Approach." *Journal of Money, Credit and Banking*, **31**(1), 85–108.

Wu J., Xia F. (2016). "Measuring the Macroeconomic Impact of Monetary Policy at the Zero Lower Bound." *Journal of Money, Credit and Banking*, **48**(2-3), 253–291.

Appendix 4.A Proofs

4.A.1 Proof of Lemma 4.1

Consider M positive definite $(d \times d)$ covariance matrices Ω_m, $m = 1, ..., M$ and suppose B is any invertible $(d \times d)$ matrix such that $B^{-1}\Omega_m B'^{-1}$ are diagonal matrices with strictly positive diagonal entries. It follows that $B^{-1}\Omega_1 B'^{-1} = \Lambda_m^{-1} B^{-1}\Omega_m B'^{-1}$ for some $(d \times d)$ diagonal matrices Λ_m^{-1}, $m = 2, ..., M$, that have strictly positive diagonal entries. Elementary matrix algebra then shows that these identities are equivalent to $B\Lambda_m = \Omega_m\Omega_1^{-1}B$, $m = 2, ..., M$. Thus, the matrices B and Λ_m solve the eigenvalue problem of $\Omega_m\Omega_1^{-1}$ with the diagonal of $\Lambda_m = \text{diag}(\lambda_{m1}, ..., \lambda_{md})$ containing the strictly positive eigenvalues and the columns of B being the related eigenvectors. Since this holds for any invertible $(d \times d)$ matrix B that simultaneously diagonalizes the covariance matrices, it is also a necessary property of a time-varying B-matrix B_t.

Suppose also BA solves the eigenvalue problems of $\Omega_m\Omega_1^{-1}$, $m = 2, ..., M$, for some invertible $(d \times d)$ matrix A. That is, $BA\Lambda_m = \Omega_m\Omega_1^{-1}BA$ which is equivalent to $A\Lambda_m A^{-1} = B^{-1}\Omega_m\Omega_1^{-1}B$. But since $B^{-1}\Omega_m\Omega_1^{-1}B = \Lambda_m$, this implies that $A\Lambda_m A^{-1} = \Lambda_m$, which is equivalent to $A\Lambda_m = \Lambda_m A$. Thus, $\lambda_{mi}a_{ij} = \lambda_{mj}a_{ij}$ where a_{ij} is the ijth element of A. It follows that $a_{ij} = 0$ if $\lambda_{mi} \neq \lambda_{mj}$ for some m, implying that A is diagonal matrix under Assumption 4.1, and BA multiplies each of the columns of B by a scalar. It is well known that eigenvalues of a matrix are unique (up to order), but since the diagonal elements of Λ_m can be in any order, so can the related eigenvectors that are the columns of B. That is, B is unique up to scalar multiples and ordering of its columns. Since the above holds for any appropriate B-matrices B and BA, it holds also for a time-varying B-matrix B_t at each t.■

4.A.2 Proof of Proposition 4.1

Lemma 4.1 shows that the B-matrix B_t is unique up to scalar multiples and reordering of its columns. Suppose that the conditional covariance matrix of the structural error is normalized to a constant diagonal matrix with strictly positive diagonal entries, say C. That is, $\sum_{m=1}^{M} \alpha_{m,t} B_t^{-1}\Omega_m B_t'^{-1} = C$, which is equivalent to $\sum_{m=1}^{M} \alpha_{m,t}\Omega_m = B_t C B_t'$. Suppose that this identity also holds with another B-matrix, $B_t E_t$, where E_t is a possibly time-varying, invertible $(d \times d)$ matrix. We have $\sum_{m=1}^{M} \alpha_{m,t}(B_t E_t)^{-1}\Omega_m(B_t E_t)'^{-1} = C$, which is equivalent to $\sum_{m=1}^{M} \alpha_{m,t}\Omega_m = (B_t E_t)C(B_t E_t)'$. Thus, $B_t C B_t' = (B_t E_t)C(B_t E_t)'$. By Lemma 4.1, the B-matrix is unique up to scalar multiples and reordering of its columns, so with a given ordering of the columns, E_t is a diagonal matrix. It then follows from $B_t C B_t' = (B_t E_t)C(B_t E_t)'$ that $C = E_t C E_t$, which in turn implies $c_i = e_{t,i}^2 c_i$, where c_i and $e_{t,i}$ are the ith diagonal elements of C and E_t, respectively. Therefore, $e_{t,i} = \pm 1$, implying that with a given ordering of the columns, (for each t) B_t is unique up to changing all signs in a column. Therefore, B_t is unique up ordering of its columns and changing all signs in a column.■

4.A.3 Proof of Proposition 4.2

Let $\Omega_1, ..., \Omega_M$ be positive definite covariance matrices and consider the decomposition $\Omega_1 = WW'$ and $\Omega_m = W\Lambda_m W', \ \ m = 2, ..., M,$, where $\Lambda_m = \text{diag}(\lambda_{m1}, ..., \lambda_{md})$, $\lambda_{mi} > 0$ ($i = 1, .., d$), contains the eigenvalues of $\Omega_m\Omega_1^{-1}$ in the diagonal and the columns of the nonsingular W are the related eigenvectors. The decomposition always exists when $M = 2$ (see, e.g., Muirhead, 1982, Theorem A9.9) but not necessarily when $M \geq 3$. In the following, I assume the covariance matrices satisfy the decomposition.

Repeating some of the proof in Lanne *et al.* (2010, p. 130, see also the proof of Theorem A9.9 in Muirhead, 1982) for convenience, suppose that we also have $\Omega_1 = DD'$ and $\Omega_m = D\Lambda_m D'$, $m = 2, ..., M$, for some nonsingular $(d \times d)$ matrix D. Because $D^{-1}WW'D'^{-1} = D^{-1}\Omega_1 D'^{-1} = I_d$, the matrix $Q' \equiv D^{-1}W$ is orthogonal, and hence, $D = WQ$ and $\Lambda_m Q = Q\Lambda_m$. It follows that $\lambda_{mi}q_{ij} = \lambda_{mj}q_{ij}$ where q_{ij} is the ijth element of Q. Thus, $q_{ij} = 0$ if $\lambda_{mi} \neq \lambda_{mj}$ for some m. Assuming that this (Condition (1) of Proposition 4.2) is satisfied by the last $d_1 \in \{1, ..., d\}$ eigenvalues, it follows that Q is a block-diagonal matrix with two blocks in the diagonal. Denoting $d_0 \equiv d - d_1$, the first block is a $(d_0 \times d_0)$ matrix and the second one is a $(d_1 \times d_1)$ diagonal matrix with $q_{d_0+1,d_0+1}, ..., q_{d,d}$ in the diagonal (if $d_1 = d$, Q simply reduces to a diagonal matrix).

As the blocks in the diagonal of an orthogonal block-diagonal matrix are orthogonal and the real eigenvalues of a diagonal orthogonal matrix are ± 1, it follows that the real eigenvalues of the second block in the diagonal of Q are ± 1. Then, because the eigenvalues of a block-diagonal matrix are the eigenvalues of the blocks in the diagonal, and eigenvalues of a diagonal matrix are its diagonal elements (and Q is real), it must be that $q_{d_0+1,d_0+1}, ..., q_{d,d}$ are ± 1.

Thus, because $D = WQ$, the last d_1 columns of W are unique up to changing all signs in a column for given Λ_m, $m = 2, ..., M$. Since Λ_m are unique up to ordering of the diagonal elements and Condition (2) fixes a unique ordering for the last d_1 columns of W and hence also for the related eigenvalues $\lambda_{mi}, i > d_0$, the last d_1 columns of the B-matrix (4.3.5) are uniquely identified up to changing all signs in a column. Finally, Condition (3) fixes the signs in the last d_1 columns of W and consequently of B_t, implying that the last d_1 columns of the B-matrix are (globally) unique for given mixing weights $\alpha_{1,t}, ..., \alpha_{M,t}$. Moreover, if $d_1 = d$, the decomposition (4.3.4) of $\Omega_1, ..., \Omega_M$ is (globally) unique.■

4.A.4 Proof of Proposition 4.3

Consider the matrix decomposition of Ω_m, $m = 1, ..., M$, of Proposition 4.2. It is shown in the proof of Proposition 4.2 that any $(d \times d)$ matrix D that also satisfies $\Omega_1 = DD'$ and $\Omega_m = D\Lambda_m D'$, $m = 2, ..., M$, can be presented as $D = WQ$ where Q is orthogonal and $q_{ij} = 0$ when $\lambda_{mi} \neq \lambda_{mj}$ for some m. Then observe that the jth column of WQ is a linear combination of the columns of W, with the multiplier of the ith column given by q_{ij}. Denoting $d_0 \equiv d - d_1$, it follows that if $\lambda_{mi} = \lambda_{mj}$ for $i \neq j > d_0$ and all m, but for all $l \notin \{i, j\}$, $\lambda_{ml} \neq \lambda_{mj}$ for some m, the jth column of WQ is a linear combination of the ith and jth columns of W. But if the jth column (of W and WQ) obeys a zero constraint where the ith column obeys a strict sign

constraint (Condition (4) of Proposition 4.3), the multiplier q_{ij} must be zero. That is, under the conditions of Proposition 4.3, with $j = d_0 + 1$ and $i < j$, we have $q_{l,d_0+1} = 0$ for all $l \neq d_0 + 1$ and $q_{lk} = 0$ for all $l, k = d_0 + 2, .., d$ such that $l \neq k$.

By the above discussion, when $d_1 > 1$, Q is a block-diagonal matrix with two blocks in the diagonal: the first one being a $(d_0 + 1 \times d_0 + 1)$ matrix

$$\tilde{Q} \equiv \begin{bmatrix} q_{1,1} & \cdots & q_{1,d_0} & 0 \\ \vdots & \ddots & \vdots & \vdots \\ q_{d_0,1} & \cdots & q_{d_0,d_0} & 0 \\ q_{d_0+1,1} & \cdots & q_{d_0+1,d_0} & q_{d_0+1,d_0+1} \end{bmatrix} \tag{4.A.1}$$

and the second one a $(d_1 - 1 \times d_1 - 1)$ diagonal matrix with $q_{d_0+2,d_0+2}, ..., q_{d,d}$ in the diagonal. When $d_1 = 1$, we simply have $Q = \tilde{Q}$ where $\tilde{Q}$ is as in (4.A.1). Consequently, for $k > d_0$ the kth column of WQ equals to the kth column of W multiplied by $q_{k,k}$. It then remains to show that $q_{k,k} = \pm 1$ for all $k = d_0 + 1, ..., d$, after which one may conclude global uniqueness of the last d_1 columns of the B-matrix (4.3.5) with arguments similar to the proof of Proposition 4.2.

Because only the last element of the last column of $\tilde{Q}$ is nonzero, the minors of the elements $q_{d_0+1,1}, ..., q_{d_0+1,d_0}$ are singular. Therefore, it follows from the cofactor presentation of the inverse of $\tilde{Q}$ (e.g., Muirhead, 1982, Appendices A4 and A5) that only the last element in the last column of the inverse of $\tilde{Q}$ is nonzero. Since $\tilde{Q}$ is orthogonal, as it is the upper-left block of the block-diagonal orthogonal matrix Q, its transpose is also its inverse. Hence, only the last element in the last column of the transpose of $\tilde{Q}$ is nonzero. Also, by the definition of $\tilde{Q}$, only the last element in last row of the transpose of $\tilde{Q}$ is nonzero. That is, the transpose is of the form

$$\tilde{Q}' = \begin{bmatrix} q_{1,1} & \cdots & q_{d_0,1} & 0 \\ \vdots & \ddots & \vdots & \vdots \\ q_{1,d_0} & \cdots & q_{d_0,d_0} & 0 \\ 0 & \cdots & 0 & q_{d_0+1,d_0+1} \end{bmatrix}, \tag{4.A.2}$$

implying that

$$\tilde{Q} = \begin{bmatrix} q_{1,1} & \cdots & q_{1,d_0} & 0 \\ \vdots & \ddots & \vdots & \vdots \\ q_{d_0,1} & \cdots & q_{d_0,d_0} & 0 \\ 0 & \cdots & 0 & q_{d_0+1,d_0+1} \end{bmatrix}. \tag{4.A.3}$$

The matrix Q is therefore an orthogonal block-diagonal matrix with two blocks in the diagonal. The first block is the upper-left $(d_0 \times d_0)$ submatrix of $\tilde{Q}$ and the second block is the $(d_1 \times d_1)$ diagonal matrix with $q_{d_0+1,d_0+1}, ..., q_{d,d}$ in the diagonal. Now reasoning similar to the proof of Proposition 4.2 shows that $q_{k,k} = \pm 1$ for $k = d_0 + 1, ..., d$.■

Appendix 4.B Monte Carlo algorithm

I present a Monte Carlo algorithm that produces point estimates and with random initial value $\boldsymbol{y}_{t-1} = (y_{t-1}, ..., y_{t-p})$ also confidence intervals for the generalized impulse response function defined in (4.4.1). My algorithm is adapted from Koop *et al.* (1996, pp. 135-136) and Kilian and Lütkepohl (2017, pp. 601-602). I assume that the history $\boldsymbol{y}_{t-1}$ follows a known distribution G, which may be such that it produces a single outcome with probability one (corresponding to a fixed $\boldsymbol{y}_{t-1}$), or it can be the stationary distribution of the process or of a specific regime. In the following, $y^{(i)}_{t+h}(\delta_j, \boldsymbol{y}_{t-1})$ denotes a realization of the process at time $t+h$ conditional on the structural shock of magnitude δ_j in the jth element of e_t hitting the system at time t and on the p observations $\boldsymbol{y}_{t-1} = (y_{t-1}, ..., y_{t-p})$ preceding the time t, whereas $y^{(i)}_{t+h}(\boldsymbol{y}_{t-1})$ denotes an alternative realization conditional on the history $\boldsymbol{y}_{t-1}$ only.

The algorithm proceeds with the following steps.

0. Decide the horizon H, the numbers of repetitions R_1 and R_2, and the size δ_j for the jth structural shock that is of interest.

1. Draw an initial value $\boldsymbol{y}_{t-1}$ from G.

2. Draw $H+1$ independent realizations of a shock ε_t from $N(0, I_d)$. Also, draw an initial regime $m \in \{1, ..., M\}$ according to the probabilities given by the mixing weights $\alpha_{1,t}, ..., \alpha_{M,t}$ and compute the reduced form shock $u_t = W\Lambda_m^{1/2}\varepsilon_t$, where $\Lambda_1 = I_d$. Then, compute the structural shock $e_t = B_t^{-1}u_t$ and impose the size δ_j on its jth element to obtain e_t^*. Finally, calculate the modified reduced form shock $u_t^* = B_t e_t^*$.[12]

3. Use the modified reduced form shock u_t^* and the rest H standard normal shocks ε_t obtained from Step 2 to compute realizations $y^{(i)}_{t+h}(\delta_j, \boldsymbol{y}_{t-1})$ for $h = 0, 1, ..., H$, iterating forward so that in each iteration the regime m that generates the observation is first drawn according to the probabilities given by the mixing weights. At $h = 0$, the initial regime and the modified reduced form shock u_t^* calculated from the structural shock in Step 2 is used, and from $h = 1$ onwards the $h+1$th shock ε_t is used to calculate the reduced form shock $u_{t+h} = W\Lambda_m^{1/2}\varepsilon_{t+h}$, where $\Lambda_1 = I_d$ and m is the selected regime.

4. Use the reduced form shock u_t and the rest H standard normal shocks ε_t obtained from Step 2 to compute realizations $y^{(i)}_{t+h}(\boldsymbol{y}_{t-1})$ for $h = 0, 1, ..., H$, so that the reduced form shock u_t (calculated in Step 2) is used to compute the time $h = 0$ realization. Otherwise proceed similarly to the previous step.

5. Calculate $y^{(i)}_{t+h}(\delta_j, \boldsymbol{y}_{t-1}) - y^{(i)}_{t+h}(\boldsymbol{y}_{t-1})$.

[12] The independent standard normal shocks ε_t are introduced here to control random variation across the two sample paths $y^{(i)}_{t+n}(\delta_j, \boldsymbol{y}_{t-1})$ and $y^{(i)}_{t+n}(\boldsymbol{y}_{t-1})$.

6. Repeat Steps 2-5 R_1 times and calculate the sample mean of $y_{t+h}^{(i)}(\delta_j, \boldsymbol{y}_{t-1}) - y_{t+n}^{(i)}(\boldsymbol{y}_{t-1})$ for $h = 0, 1, ..., H$ to obtain an estimate of the GIRF$(h, \delta_j, \boldsymbol{y}_{t-1})$.

7. Repeat Steps 1-6 R_2 times to obtain estimates of GIRF$(h, \delta_j, \boldsymbol{y}_{t-1})$ with different starting values $\boldsymbol{y}_{t-1}$ generated from the distribution G. Then take the sample mean and sample quantiles over the estimates to obtain point estimate and confidence intervals for the GIRF with random initial value.

Notice that if a fixed initial value $\boldsymbol{y}_{t-1}$ is used, Step 7 is redundant.

Appendix 4.C Details on the empirical application

4.C.1 Model selection

The maximum likelihood (ML) estimation of the models, quantile residual diagnostics, estimation of generalized impulse response functions, and other numerical analysis are carried out with the CRAN distributed R package **gmvarkit** (Virolainen, 2018a) that accompanies this essay. The R package **gmvarkit** also contains the dataset studied in the empirical application, thus facilitating reproduction of my results. The estimation is based on the exact log-likelihood function. For evaluating the adequacy of the models, I employ quantile residual diagnostics in the framework of Kalliovirta and Saikkonen (2010) (see also the related paper by Kalliovirta, 2012, for discussion on quantile residual based model diagnostics in a univariate setting). For a correctly specified GMVAR model, the empirical counterparts of the quantile residuals are asymptotically independent with multivariate standard normal distributions and can hence be used for graphical analysis in a similar manner to the conventional Pearson residuals (Kalliovirta and Saikkonen, 2010, Lemma 3).[13]

I started by estimating linear Gaussian VARs with the autoregressive orders $p = 1, ..., 12$, i.e., GMVAR$(p, 1)$ models. BIC was minimized by the order $p = 1$, HQIC by the order $p = 2$, and AIC by the order $p = 3$, suggesting that the appropriate autoregressive order is likely relatively small. Hence, I then estimated the two-regime GMVAR$(p, 2)$ models with $p = 1, ..., 4$. BIC was minimized by the order $p = 1$ and HQIC and AIC by the order $p = 2$. Graphical quantile residual diagnostics revealed the $p = 1$ was clearly inadequate to capture the autocorrelation structure of the series, while also the $p = 2$ was somewhat inadequate (not shown). I therefore considered the order $p = 3$, which I found adequate to capture the autocorrelation structure of

[13] Kalliovirta and Saikkonen (2010) also propose formal diagnostic tests for testing normality, autocorrelation, and conditional heteroskedasticity of the quantile residuals. The tests take into account the uncertainty about the true parameter value and can be calculated based on the observed data or by employing a simulation procedure for better size properties. I found these tests very forgiving without the simulation procedure and quite conservative without it. For instance, when taking into account the first four lags in the autocorrelation test, without the simulation procedure my GMVAR$(3, 2)$ model obtains the p-value 0.999, while with the simulation procedure, using sample of length 10000, the p-value is 0.000. I therefore rather employ graphical diagnostics and compare the statistical properties of the quantile residuals to the ones of four-variate IID standard normal process.

Model	Log-lik	BIC	HQIC	AIC
GMVAR$(1,1)$	-4.404	9.430	9.191	9.030
GMVAR$(2,1)$	-4.245	9.444	9.077	8.831
GMVAR$(3,1)$	-4.160	9.606	9.112	8.780
GMVAR$(1,2)$	-3.558	8.381	7.894	7.568
GMVAR$(2,2)$	-3.276	8.481	7.740	7.242
GMVAR$(3,2)$	-3.183	8.958	7.961	7.292
GMVAR$(4,2)$	-3.113	9.482	8.230	7.390

Table 4.1: The log-likelihoods and values of the information criteria divided by the number of observations for the discussed GMVAR(p, M) models.

the series (see Section 4.C.2). I also considered the order $p = 4$, but since it increased AIC from $p = 3$, which was already found adequate to explain the autocorrelation structure, and the order $p = 3$ was found suitable for the linear VAR as well, I preferred the more parsimonious GMVAR$(3, 2)$ model. The log-likelihoods and values of the information criteria are presented in Table 4.1 for the discussed models.

Table 4.1 shows that the GMVAR$(3, 2)$ model has significantly smaller BIC, HQIC, and AIC than all of the linear VARs. According to graphical quantile residual diagnostics, the GMVAR$(3, 2)$ model also explains the statistical characteristics of the data more adequately than the linear VARs (see Section 4.C.2; graphical diagnostics of the linear VARs are not shown for brevity). Hence, I find my GMVAR$(3, 2)$ model superior to the linear VARs.

It is possible that the superior fitness is due to the accommodation of time-varying covariance matrix or intercepts and cannot be attributed to the time-varying autoregression (AR) matrices. To test whether this is the case, I estimated two additional GMVAR$(3, 2)$ models. In the first one, I constrained the AR matrices and intercept parameters to be identical in both regimes, thereby allowing for time-varying covariance matrix only. In the second one, I constrained the AR matrices to be identical in both regimes, thereby allowing for time-varying intercepts and covariance matrix only. Because the constrained models are nested to the GMVAR$(3, 2)$ model and the maximum likelihood estimator has the conventional asymptotic distribution under the conventional assumptions (Kalliovirta *et al.*, 2016, Theorem 3), the validity of the constraints can be tested with the likelihood ratio test. The likelihood ratio test produces the p-value 0.011 for the former type of constraint and the p-value 0.007 for the latter type of constraints, thus, rejecting both constraints the 5% level of significance and the latter type of constraints with the 1% level of significance.[14]

[14]For robustness, I also estimated the constrained models with autoregressive orders $p = 1, 2, 4$ (and $M = 2$) and tested the validity of the constraints in these models with the likelihood ratio test. Both types of constraints were rejected for the $p = 2, 4$ models with p-values less than 0.0008, but the GMVAR$(1, 2)$ model accepted constraining the AR matrices to be identical in both regimes with the p-value 0.3 and rejected constraining both AR matrices and intercepts with the p-value 0.005. The rejection of the constraints does not, hence, seem particularly sensitive to the choice of p. Notably, the validity of the likelihood ratio test requires that the unrestricted model is correctly

According to the small p-values, it seems likely that also the AR matrices vary in time. But since the p-values were not particularly small, I studied the generalized impulse response functions of the constrained GMVAR$(3, 2)$ models as well. I found that if the AR matrices and intercepts are constrained identical in both regimes (a linear VAR with two volatility regimes), small and large as well as contractionary and expansionary monetary policy shocks induce a long-run price puzzle in both regimes. If only the AR matrices are constrained to be identical in both regimes, small and large contractionary monetary policy shocks induce a medium- to long-run price puzzle and expansionary shocks medium-run price puzzle in the unstable inflation regime. But the GIRFs did not change much in the stable inflation regime (and thus display a long-run price puzzle for contractionary shocks in this regime as well). The GIRFs of the constrained models are not shown for brevity.[15]

For comparison, I also considered a Cholesky identified Gaussian SVAR model with the autoregressive order $p = 3$ (as suggested by AIC) and the interest rate variable ordered last. This model did not only display a long-run price puzzle but it also displayed a short-run output puzzle, i.e., the response of GDP was positive (in the point estimate) before it became negative as a response to a contractionary monetary policy shock. A Cholesky SVAR with the order $p = 4$ shows permanent decrease in the price level after roughly 9 years from the impact but the response of GDP still has the wrong sign in the period after the impact (not shown).

4.C.2 Adequacy and characteristics of the selected model

In order to study the adequacy of my GMVAR$(3, 2)$ model, I examine the quantile residual time series, sample auto- and crosscorrelation functions of the quantile residuals and squared quantile residuals, and normal quantile-to-quantile plots. The sample auto- and crosscorrelation functions (presented in Figure 4.3) show that there is not much auto- or crosscorrelation in the quantile residuals. The GDP deflator has some moderate sized autocorrelation coefficients (ACC) at small lags, but they are not very large. There are also moderate sized coefficients at larger lags in the crosscorrelation function of GDP and PPI as well as in the autocorrelation function of the interest rate variable. Nonetheless, given that in total of 316 correlation coefficients are presented, some of them are expected to be moderate sized for an IID process as well.[16]

The sample auto- and crosscorrelation functions of the squared quantile residuals are pre-

specified, so the tests that do not use the correct autoregressive order p do not produce reliable result. If one of the orders $M = 2$ and $p = 2, 3, 4$ is correct, then the constraints are, nevertheless, rejected.

[15] For the constrained models, the monetary policy shock was identified with the constraints discussed in Section 4.5.1 similarly to the unconstrained model, as the unrestricted impact effects were similar in sign and magnitude to the unconstrained model. Also the estimated mixing weights were quite similar: one of the regimes prevailed in the 1970's, 1980's, the Financial crisis, and COVID-19 crisis, but also short periods during other times. For the ease of communication, I hence refer to them as stable inflation regime and unstable inflation regime similarly to the unconstrained model.

[16] Increasing the autoregressive order to $p = 4$ reduces some of the ACCs of the GDP deflator's quantile residuals, but it does not help with the price puzzle.

sented in Figure 4.3. GDP, GDP deflator, and PPI each have at least one exceedingly large ACC in their autocorrelation functions, but ACCs of the interest rate variable are reasonable. There are also two exceedingly large coefficients at large lags in the crosscorrelation function of the interest rate variable and GDP deflator. My GMVAR$(3, 2)$ model is therefore clearly inadequate to capture the conditional heteroskedasticity in the series.

The quantile residual time series (the top panels of Figure 4.5) also show some heteroskedasticity and several outliers in the quantile residuals. There is a particularly large (marginal) quantile residual of the GDP in the beginning of the COVID-19 crisis, when the COVID-19 lockdown caused a fast and vast drop in the cycle. I do not view this large negative quantile residual of the GDP as an inadequacy, however, as the COVID-19 drop is known to be caused by an exceptionally large exogenous shock, and therefore a large (quantile) residual is expected for a correctly specified model. The normal quantile-quantile-plots (the bottom panels of Figure 4.5) show that the marginal quantile residual distributions have excess kurtosis but are quite symmetric. The quantile residuals of GDP deflator seem slightly skewed to the right and GDP slightly to the left, however.

In my view, the overall adequacy of the model is decent enough for further analysis, particularly since autocorrelation structure of the data is captured reasonably well. Some of the conditional heteroskedasticity in the data remains unmodelled, which is not completely innocent because the mixing weights may depend on the volatility of the series. The unmodelled conditionally heteroskedasticy is not very extreme though: there is a single ACC of roughly the size 0.3 in the autocorrelation functions of the squared quantile residuals of GDP, GDPDEF, and PPI, while almost all of the crosscrorrelation coefficients are of reasonable size. The excess kurtosis in the marginal distribution of the quantile residuals, in turn, does not seem particularly severe. Accommodating stronger forms of conditional heteroskedasticity and excess kurtosis by utilizing Student's t distributed error terms similarly to Meitz *et al.* (2018) in the univariate setting is beyond the scope of this essay and addressed in Chapter 5.[17]

The estimated mixing weights of the two regimes are presented in the bottom panel of Figure 4.1. The second regime (red) mainly dominates during the periods of high inflation and interest rate in the 1970's and 1980's, after the collapse of Lehman Brothers in the Financial crisis until the end of 2009, and finally during the COVID-19 crisis from the third quarter of 2020 onwards. The first regime (blue) prevails when the second one does not: before 1970's, short periods during 1970's, and from the mid 1980's onwards but excluding the Financial crisis and the COVID-19 crisis (but including the first two quarters of 2020). Therefore, it appears that the second regime is mainly dominant when inflation has been high or volatile, while the first regime is dominant in more stable times.

[17]The GMVAR model's capability to capture the conditional heteroskedasticity and marginal distribution of the series can be improved by adding a third regime. With $p = 3$, however, the number of parameters in each regime is rather high: 62 plus a mixing weight parameter for all but the last regime. In the most rare regime, this may be too much compared to the number of observations from that regime for any meaningful inference based on the estimates to take place. With smaller p, on the other hand, the model's capability to capture the autocorrelation structure of the series at larger lags might not be adequate. Because the three regime models are also tedious to estimate in practice, I focus on the two regime models.

		GDP		GDPDEF		PPI		RATE	
	$\hat{\alpha}_m$	$\hat{\mu}_{m,1}$	$\hat{\sigma}^2_{m,1}$	$\hat{\mu}_{m,2}$	$\hat{\sigma}^2_{m,2}$	$\hat{\mu}_{m,3}$	$\hat{\sigma}^2_{m,3}$	$\hat{\mu}_{m,4}$	$\hat{\sigma}^2_{m,4}$
Regime 1	0.56	-0.20	1.92	0.71	0.20	0.72	1.87	4.44	12.12
Regime 2	0.44	-0.07	7.83	1.45	0.95	1.75	10.96	7.57	34.28

Table 4.2: Mixing weight parameter estimates ($\hat{\alpha}_m$) and marginal stationary means ($\hat{\mu}_{m,i}$) and variances ($\hat{\sigma}^2_{m,i}$) of the component series implied by the fitted GMVAR$(3, 2)$ model for each of the regimes.

The mixing weight parameters have the interpretation of being the unconditional probabilities for an observation being generated from each regime. For a correctly specified model, they should hence approximately reflect the proportions of observations generated from each regime. The first regime has a mixing weight parameter estimate 0.56 (shown in Table 4.2), and it covers approximately 81% of the series (approximated as the mean of the estimated mixing weights), whereas the second regime has the implied mixing weight parameter estimate 0.44 and it covers approximately 19% of the series. The mixing weight parameter estimates are therefore somewhat disproportionate to the relative number of observations from each regime. This can be attributed to estimation error (rather than misspecification), however, as the approximate standard error for the first regime's mixing weight parameter estimate is as high as 0.213. Nonetheless, the mixing weight parameter estimates seem reasonable enough not to distort the generalized impulse response functions too much.

Based on the model implied marginal stationary means and variances presented in Table 4.2, neither of the regimes is particularly recessionary or expansionary, but the first regime has lower unconditional mean for the GDP, while the second one has much higher unconditional variance. Both regimes also prevail during recessions and expansions (see Figure 4.1). In the first regime, the GDP deflator has the (estimated) unconditional mean 0.71, which implies long-run yearly inflation of approximately 2.9%, and unconditional variance 0.20. In the second regime, the GDP deflator has the (estimated) unconditional mean 1.45, which implies long-run yearly inflation of approximately 5.9%, and unconditional variance 0.95. That is, the estimated long-run inflation is quite reasonable in the first regime, while it is excessive and volatile in the second regime. Also the commodity price inflation and the interest rate variable have much higher unconditional mean and variance the second regime than in the first regime. Based on the significantly higher unconditional means and variances of the inflation, commodity price inflation, and the interest rate variable, as well as on the timing of the dominance of the regimes (see Figure 4.1 and the discussion above), I refer to the second regime as an unstable inflation regime. Accordingly, I refer to the first regime as the stable inflation regime.

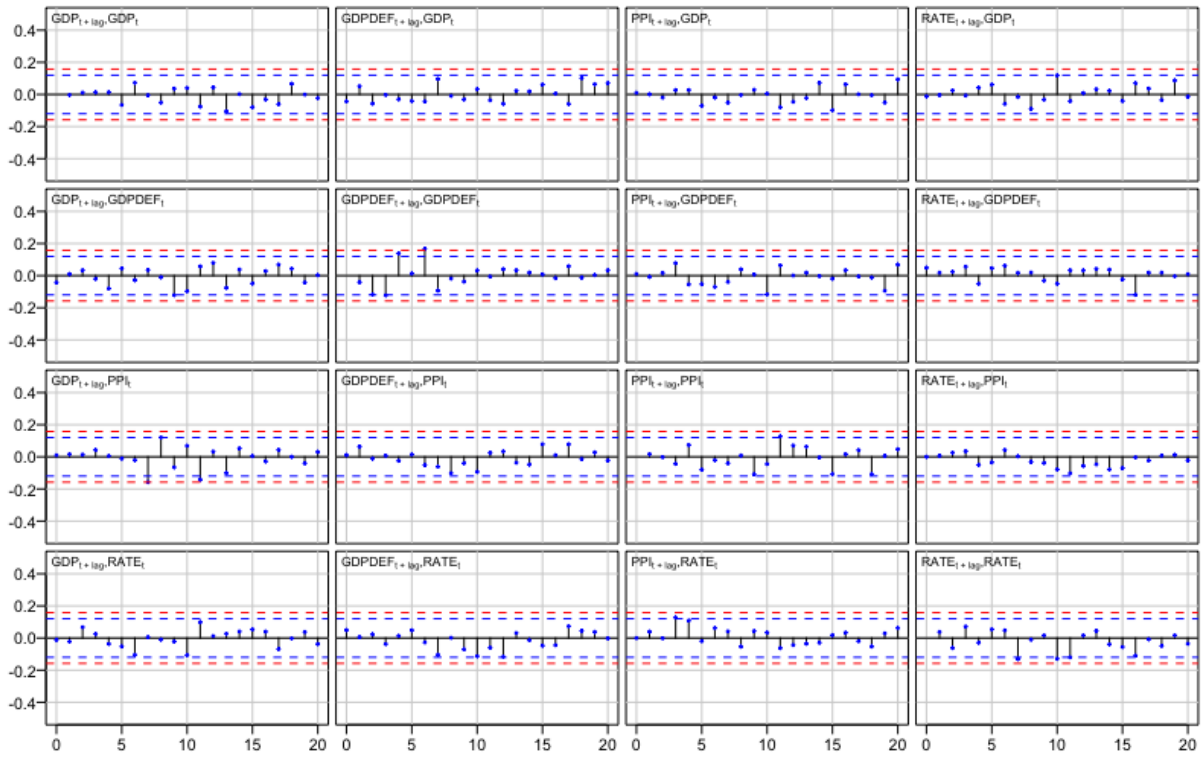

Figure 4.3: Auto- and crosscorrelation functions of the quantile residuals of the fitted GMVAR$(3, 2)$ model for the lags $0, 1, ..., 20$. The lag zero autocorrelation coefficients are omitted, as they are one by convention. The blue dashed lines are the 95% bounds $\pm 1.96/\sqrt{T}$ ($T = 267$ as the first $p = 3$ observations were used as the initial values) for autocorrelations of IID observations, whereas the red dashed lines are the corresponding 99% bounds $\pm 2.58/\sqrt{T}$. These bounds are presented to give an approximate perception on the magnitude of the correlation coefficients.

Appendix

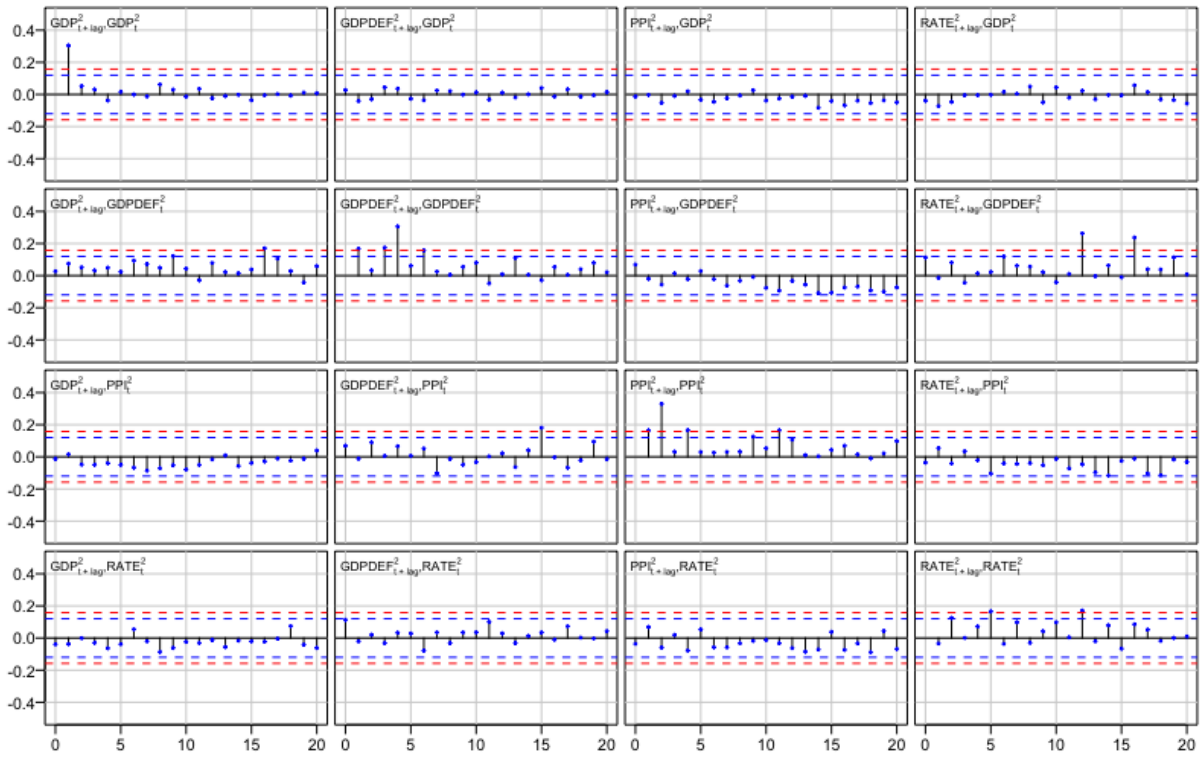

Figure 4.4: Auto- and crosscorrelation functions of the squared quantile residuals of the fitted GMVAR$(3, 2)$ model for the lags $0, 1, ..., 20$. The lag zero autocorrelation coefficients are omitted, as they are one by convention. The blue dashed lines are the 95% bounds $\pm 1.96/\sqrt{T}$ ($T = 267$ as the first $p = 3$ observations were used as the initial values) for autocorrelations of IID observations, whereas the red dashed lines are the corresponding 99% bounds $\pm 2.58/\sqrt{T}$. These bounds are presented to give an approximate perception on the magnitude of the correlation coefficients.

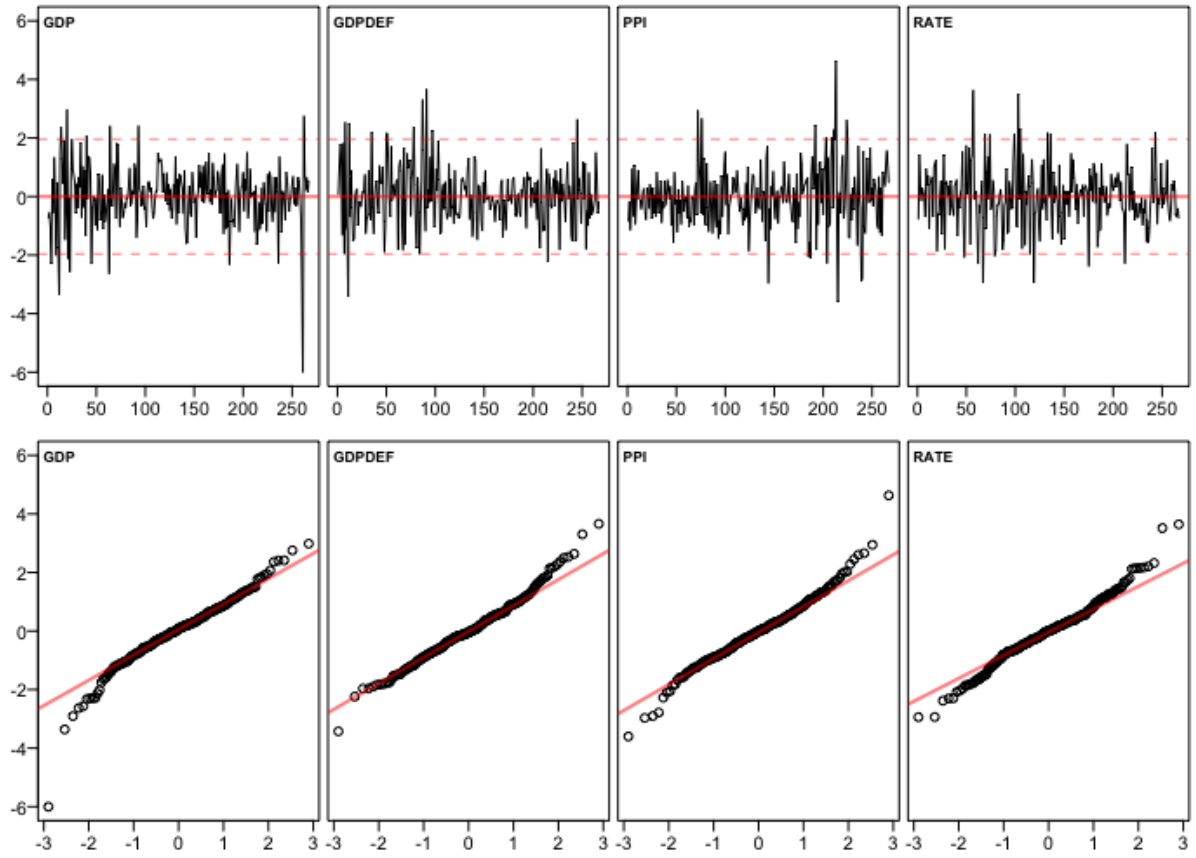

Figure 4.5: Quantile residual time series and normal quantile-quantile-plots of the fitted GMVAR$(3, 2)$ model.

4.C.3 Individual GIRFs in the stable inflation regime

The confidence bounds of the GIRFs are relatively wide in the stable inflation regime (the first and second columns of Figure 4.2), and they display some unexpected results such as prices rising in response to a contractionary monetary policy shock, the interest variable overshooting significantly as a response to an expansionary monetary policy shock, and counterproductive response of the GDP after the initial expansion (or contraction). In order to investigate how these results appear in the model dynamics, and to what extend they might be economically sensible, I have depicted 500 individual GIRFs (each estimated based on 2500 Monte Carlo repetitions) in each column of Figure 4.6 with the starting values generated from the stable inflation regime. The first (third) column presents the GIRFs to a one-standard-error contractionary (expansionary) monetary policy shock and the second (fourth) column presents the GIRFs to a two-standard-error contractionary (expansionary) monetary policy shock. After estimating the GIRFs, they were scaled to correspond to a 25 basis point increase (decrease) of the interest rate variable. The GIRFs that display (scaled) peak inflation greater than 5 basis points for one-

standard-error shocks and 10 basis points for two-standard-error shocks are colored red and the rest blue.

One of the unexpected observations in Section 4.5.2 is that (in the stable inflation regime) the prices seem to often rise in response to both contractionary and expansionary monetary policy shocks, particularly if the shock is large. From the perspective of the model dynamics, the reason is that from many of the starting values the monetary policy shock drives the economy towards the unstable inflation regime, as I next explain. The bottom row of Figure 4.6 shows that in the high inflation red GIRFs, the probability of the unstable inflation regime increases sharply in the period after the impact (while at impact the mixing weights are predetermined). This implies that the impact responses of the observable variables induce a greater probability of the unstable inflation regime, which then moves the observable variables in the following periods accordingly. Thus, the monetary policy shock drives the economy towards the unstable inflation regime (that has high long-run inflation), which in part causes an increase in inflation (and not the vice versa).

The red GIRFs in each column of Figure 4.6 show that the GIRFs exhibiting particularly high increase in inflation (and commodity price inflation) also display a persistent increase in the interest rate variable. Given the movements of the prices, the response of the interest rate variable is economically sensible when the Fed's endogenous response to high inflation is tight monetary policy. When the interest rate rises, inflation starts to finally decrease after several years from impact but so does the GDP.

If the shock is contractionary (the first and second columns of Figure 4.6), the GDP temporarily recovers in the red GIRFs relatively fast before persistently decreasing along with the rising interest rate. The temporary recovery of the GDP might be related to the higher unconditional mean of the GDP in the unstable inflation regime. Since there are only approximately 49 observations from the unstable inflation regime (estimated as the sum of the mixing weights), the temporary recovery can possibly be attributed to estimation error. Also, particularly when the contractionary shock is large, from some of the initial values the GDP's response overshoots significantly to the positive side without the delayed contraction that the red GIRFs display. Further investigation revealed that these GIRFs are mostly the ones that display positive peak deflation and a response of the interest rate variable that overshoots to the negative side, and therefore, the expansion has an economic explanation through the expansionary monetary policy (not shown).

If the shock is expansionary (the third and fourth columns of Figure 4.6), both high inflation red GIRFs and low inflation blue GIRFs display roughly same length expansions of the GDP. The high inflation red GIRFs in which the interest rate significantly overshoots to the positive side, however, display a delayed contraction of the GDP after the initial expansion. That is, particularly a large expansionary (but also a contractionary) shock drives the economy towards the unstable inflation regime, in part causing the high inflation. This results in a significant monetary policy tightening, which is accompanied with a persistent contraction of the GDP.

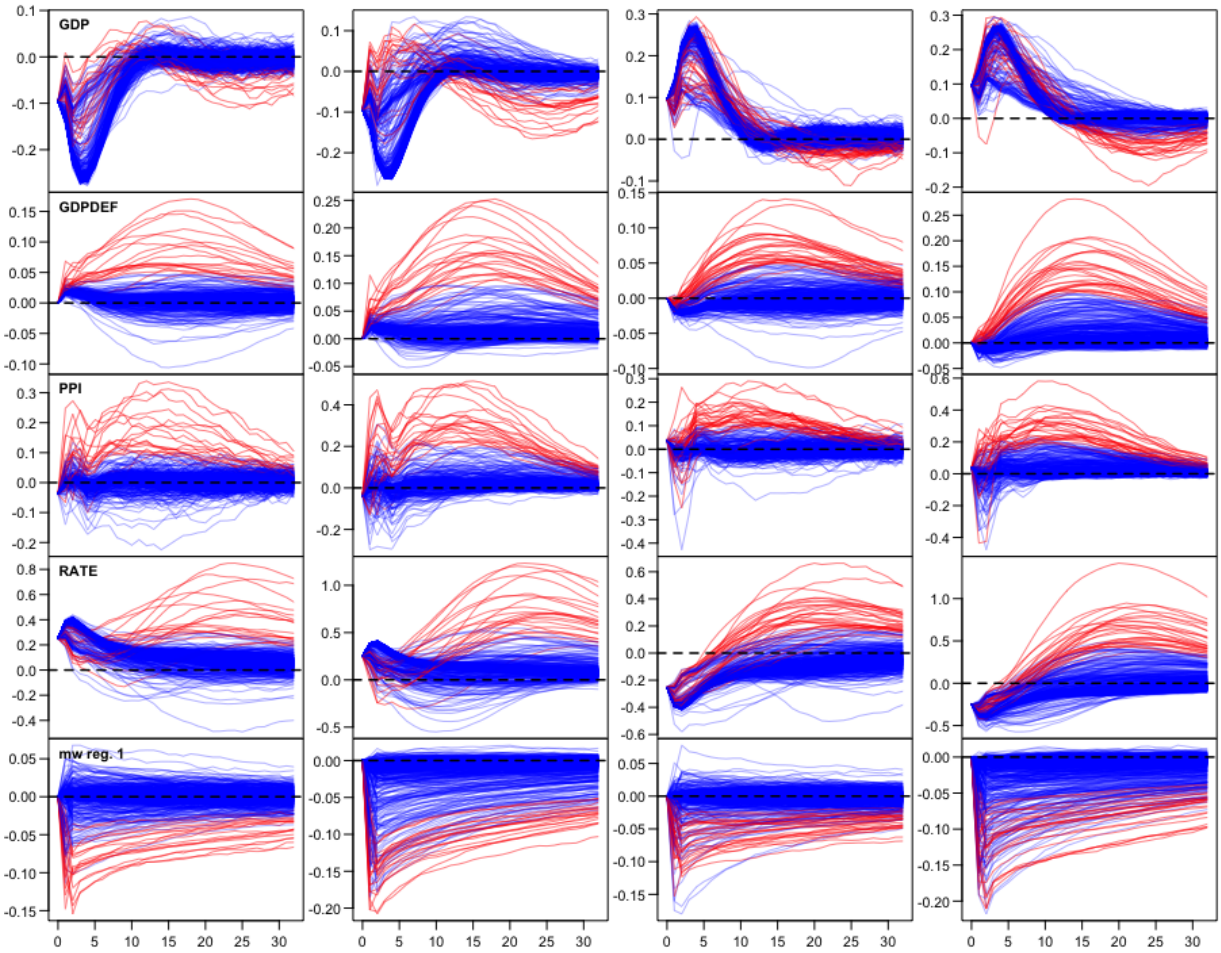

Figure 4.6: Generalized impulse response functions of the identified monetary policy shock of the fitted GMVAR$(3, 2)$ model. Each column presents 500 GIRFs, each based on a random starting value drawn from the stationary distribution of the first regime and 2500 Monte Carlo repetitions. The first (third) column present GIRFs to one-standard-error contractionary (expansionary) shocks and the second (fourth) column to two-standard-error contractionary (expansionary) shocks. From the top, the first panels show the response of the cyclical component of GDP, the second panels show the response of log-differenced implicit GDP deflator, the third panels show the response of log-differenced producer price index (all commodities), the fourth panels show the response of the interest rate variable, and the bottom panels show the response of the first regime's mixing weights. The GIRFs are scaled to correspond to a 25 basis point instantaneous increase (decrease) of the interest rate variable. The GIRFs that display (scaled) peak inflation greater than 5 basis points for one-standard-error shocks and 10 basis for two-standard-error shocks are colored red and the rest blue.

Chapter 5

Gaussian and Student's *t* mixture vector autoregressive model with application to the asymmetric effects of monetary policy shocks in the Euro area

5.1 Introduction

Mixture autoregressive models are useful for modelling series in which the data generating dynamics vary in time. Such variation may arise due to wars, crises, business cycle fluctuations, or policy shifts, for example. Mixture autoregressive models can be described as collections of (typically linear) autoregressive models, which are called mixture components, components processes, or regimes. At each time point, the process generates an observation from one of its mixture components that is randomly selected according to the probabilities given by the mixing weights.

Several new mixture autoregressive models have been introduced recently. Kalliovirta, Meitz, and Saikkonen (2015) introduced the Gaussian mixture autoregressive (GMAR) model, which incorporates linear Gaussian autoregressions as its mixture components and mixing weights that, for a pth order model, depend on the full distribution of the previous p observations. The specific definition of the mixing weights leads to attractive theoretical and practical properties, such as ergodicity and full knowledge of the stationary distribution of $p+1$ consecutive observations. Kalliovirta, Meitz, and Saikkonen (2016) introduced a multivariate version of this model, the Gaussian mixture vector autoregressive (GMVAR) model, which employs linear Gaussian vector autoregressions (VAR) as its mixture components and has analogous properties to the GMAR model. Burgard, Neuenkirch, and Nöckel (2019), on the other hand, proposed a model with linear Gaussian VARs as mixture components, and mixing weights that depend on switching variables through a logistic function. Meitz, Preve, and Saikkonen (forthcoming) introduced the Student's t mixture autoregressive (StMAR) model with analogous properties to the GMAR

model, but incorporating conditionally heteroskedastic mixture components based on Student's t-distribution. Chapter 2 (also Virolainen, forthcoming) introduced the Gaussian and Student's t mixture autoregressive (G-StMAR) model, where some of the mixture components are based on a Gaussian distribution and some on a t-distribution.

This essay introduces a multivariate version of the G-StMAR model, and as a special case also a multivariate version of the StMAR model. The Gaussian and Student's t mixture vector autoregressive (G-StMVAR) model accommodates conditionally homoskedastic linear Gaussian VARs and conditionally heteroskedastic linear Student's t VARs as its mixture components. Both types of mixture components have the same form for the conditional mean, a linear function of the preceding p observations, but the conditional covariance matrices are different. The linear Gaussian VARs have constant conditional covariance matrices. The conditional covariance matrices of the linear Student's t VARs, in turn, consist of a constant covariance matrix multiplied by a time-varying scalar that depends on the quadratic form of the previous p observations. In this sense, the conditional covariance is of ARCH (autoregressive conditional heteroskedasticity) type. But since it is just a time-varying scalar multiplying the constant covariance matrix, it is not as general as the conventional multivariate ARCH process that allows the entries of the conditional covariance matrix to vary relative to each other (e.g., Lütkepohl, 2005, Section 16.3). The specific formulation of the conditional covariance matrix is, nonetheless, convenient for establishing stationary properties similar to the linear Gaussian VARs. My specification of the conditional covariance is also parsimonious, as it only depends on the degrees of freedom and the autoregressive parameters (in addition to the parameters in the constant covariance matrix). This is particularly advantegeous in the context of mixture VARs, as the large number of parameters may often be a problem even without an ARCH component.

For a pth order G-StMVAR model, the mixing weights are defined as weighted ratios of the components process stationary densities corresponding to the previous p observations. This formulation is appealing, as it states that the greater the relative weighted likelihood of a regime is, the more likely the process is to generate an observation from it. Moreover, it facilitates associating statistical characteristics and economic interpretations to the regimes. It turns out that the specific formulation of the mixing weights also leads to attractive theoretical properties, such as ergodicity and full knowledge of the stationary distribution of $p + 1$ consecutive observations. In contrast to the GMVAR model, my model is able to capture excess kurtosis and conditional heteroskedasticity within the regimes. If all of the regimes are assumed to be linear Student's t VARs, a multivariate version of the StMAR model is obtained as a special case. I refer to this model as the StMVAR model.

It turns out that the G-StMVAR model is a limiting case of the StMVAR model with the degrees of freedom parameters of some of the regimes tending to infinity. The GMVAR model is obtained if the degrees of freedom parameters of all the regimes tend to infinity. Hence, if a StMVAR model is fitted to a series generated by a process in which some of the regimes are linear Gaussian VARs, the degrees of freedom parameters of these regimes are (asymptotically) expected to get large estimates. In empirical applications, the numbers of Gaussian and Student's t regimes can, therefore, be selected by first finding a suitable StMVAR model. Then, if some of

the regimes obtain a large degrees of freedom parameter estimate, they should be accommodated by switching to the appropriate G-StMVAR model. As opposed to a StMVAR model with very large degrees of freedom parameter estimates, the G-StMVAR model avoids the numerical problems caused by weak identification of these parameters.

In addition to the reduced form model, I propose a structural version of the G-StMVAR model that generalizes the SGMVAR model introduced in Chapter 4 to accommodate conditionally heteroskedastic Student's t regimes. The SG-StMVAR model incorporates a time-varying impact matrix that varies according to the conditional variance of the reduced form error. As a consequence of a single impact matrix, identification of the shocks requires that the error term covariance matrices are simultaneously diagonalized in all regimes. Together with a constant normalization of the structural error's conditional covariance matrix, this condition generally leads to uniquely identified shocks up to ordering and sign. Hence, as long as one is willing to assume a single (time-varying) impact matrix, its columns characterize the estimated impact effects of the shocks, but it is not revealed which column is related to which shock. Because the impact matrix is also subject to estimation error, further constraints may be needed for labelling the shocks. The identification conditions are the same as in Chapter 4, however, and I repeat some of them for convenience.

The empirical application studies asymmetries in the expected effects of the monetary policy shock in the Euro area and considers a monthly data covering the period from January 1999 to December 2021. My StMVAR model identifies two regimes: a low-growth regime and a high-growth regime. The low-growth regime is characterized by a negative (but volatile) output gap, and it mainly prevails after the collapse of Lehman Brothers in the Financial crisis but obtains large mixing weights also during and before the early 2000's recession. The high-growth regime is characterized by a positive output gap and it mainly dominates before the Financial crisis.

I find strong asymmetries with respect to the initial state of the economy and sign of the shock, but asymmetries with respect to the size of the shock are weak. The real effects are less enduring for an expansionary shock than for a contractionary shock. Particularly in the high-growth regime, a contractionary shock persistently drives the economy towards the low-growth regime, which translates to a very persistent decrease in the output gap. The inflationary effects of the monetary policy shock are stronger in the high-growth regime than in the low-growth regime, and in the latter the price level does not move much on average.

The rest of this chapter is organized as follows. Section 5.2 introduces the linear Student's t VAR and establishes its stationary properties. Section 5.3 introduces the G-StMVAR model and discusses its properties. Section 5.4 introduces the structural G-StMVAR model and briefly discusses identification of the shocks. Section 5.5 discusses estimation of the model parameters by the method of maximum likelihood (ML) and establishes the asymptotic properties of the ML estimator. Section 5.6 discusses a strategy for building a G-StMVAR model, and Section 5.7 presents the empirical application to the asymmetric effects of the Euro area monetary policy shock. Appendix 5.A provides the density functions and some properties of the Gaussian and Student's t distributions, Appendix 5.B gives proofs for the stated theorems, Appendix 5.C provides details on the empirical application, and Appendix 5.D derives a closed form expression

for the quantile residual of the G-StMVAR model. Finally, this essay is accompanied with the CRAN distributed R package **gmvarkit** (Virolainen, 2018) that provides tools for estimation and other numerical analysis of the models.

Throughout this chapter, I use the following notation. I write $x = (x_1, ..., x_n)$ for the column vector x where the components x_i may be either scalars or (column) vectors. The notation $x \sim n_d(\mu, \Sigma)$ signifies that the random vector x has a d-dimensional Gaussian distribution with mean μ and (positive definite) covariance matrix Σ, and $n_d(\cdot; \mu, \Sigma)$ denotes the corresponding density function. Similarly, $x \sim t_d(\mu, \Sigma, \nu)$ signifies that x has a d-dimensional t-distribution with mean μ, (positive definite) covariance matrix Σ, and degrees of freedom ν (assumed to satisfy $\nu > 2$), and $t_d(\cdot; \mu, \Sigma, \nu)$ denotes the corresponding density function. The vectorization operator vec stacks columns of a matrix on top of each other and $vech$ stacks them from the main diagonal downwards (including the main diagonal). I_d signifies the identity matrix of dimension d, $\otimes$ denotes the Kronecker product, and $\mathbf{1}_d$ denotes a d-dimensional vectors of ones.

5.2 Linear Gaussian and Student's *t* vector autoregressions

The G-StMVAR model accommodates two types of mixture components: conditionally homoskedastic linear Gaussian vector autoregressions and conditionally heteroskedastic linear Student's t vector autoregressions. In this section, I define these linear autoregressions and establish their stationary properties. Consider the d-dimensional linear VAR model defined as

$$z_t = \phi_0 + \sum_{i=1}^{p} A_i z_{t-1} + \Omega_t^{1/2} \varepsilon_t, \tag{5.2.1}$$

where the error process ε_t identically and independently distributed (IID), $\Omega_t^{1/2}$ is a symmetric square root matrix of the positive definite $(d \times d)$ covariance matrix Ω_t for all t, and $\phi_0 \in \mathbb{R}^d$. The $(d \times d)$ autoregression matrices are assumed to satisfy $\boldsymbol{A}_p \equiv [A_1 : ... : A_p] \in \mathbb{S}^{d \times dp}$, where

$$\mathbb{S}^{d \times dp} = \{[A_1 : ... : A_p] \in \mathbb{R}^{d \times dp} : \det(I_d - \sum_{i=1}^{p} A_i z^i) \neq 0 \text{ for } |z| \leq 1\} \tag{5.2.2}$$

defines the usual stability condition of a linear VAR. The linear Gaussian VAR is obtained from (5.2.1) by assuming that ε_t follows the d-dimensional standard normal distribution and that the conditional covariance matrix is a constant, $\Omega_t = \Omega$. I will first establish the stationary properties of the linear Gaussian VAR, and by making use of the introduced notation, I then introduce the linear Student's t VAR.

Under the stability condition, the linear Gaussian VAR is stationary, and the following properties are obtained. Denoting $\boldsymbol{z}_t = (z_t, ..., z_{t-p+1})$ and $\boldsymbol{z}_t^+ = (z_t, \boldsymbol{z}_{t-1})$, it is well known that the

stationary solution to (5.2.1) satisfies

$$\begin{aligned}
\boldsymbol{z}_t &\sim n_{dp}(\mathbf{1}_p \otimes \mu, \Sigma_p) \\
\boldsymbol{z}_t^+ &\sim n_{d(p+1)}(\mathbf{1}_{p+1} \otimes \mu, \Sigma_{p+1}) \\
z_t|\boldsymbol{z}_{t-1} &\sim n_d(\mu + \Sigma_{1p}\Sigma_p^{-1}(\boldsymbol{z}_{t-1} - \mathbf{1}_p \otimes \mu), \Sigma_1 - \Sigma_{1p}\Sigma_p^{-1}\Sigma_{1p}') = n_d(\phi_0 + \boldsymbol{A}_p\boldsymbol{z}_{t-1}, \Omega),
\end{aligned} \tag{5.2.3}$$

where the last line defines the conditional distribution of z_t given $\boldsymbol{z}_{t-1}$. Denoting by $\Sigma(h)$, $h = 0, \pm1, \pm2, ...$, the lag h autocovariance matrix of z_t, the quantities $\mu, \Sigma_p, \Sigma_1, \Sigma_{1p}, \Sigma_{p+1}$ are given as (see, e.g., Lütkepohl, 2005, pp. 23, 28-29)

$$\begin{aligned}
\mu =& (I_d - \sum_{i=1}^{p} A_i)^{-1}\phi_0 && (d \times 1) \\
\text{vec}(\Sigma_p) =& (I_{(dp)^2} - \boldsymbol{A} \otimes \boldsymbol{A})^{-1}\text{vec}(\boldsymbol{\Omega}) && ((dp)^2 \times 1) \\
\Sigma_1 =& \Sigma(0) && (d \times d) \\
\Sigma(p) =& A_1\Sigma(p-1) + \cdots + A_p\Sigma(0) && (d \times d) \\
\Sigma_{1p} =& [\Sigma(1) : ... : \Sigma(p-1) : \Sigma(p)] = \boldsymbol{A}_p\Sigma_p && (d \times dp) \\
\Sigma_{p+1} =& \begin{bmatrix} \Sigma_1 & \Sigma_{1p} \\ \Sigma_{1p}' & \Sigma_p \end{bmatrix} && (d(p+1) \times d(p+1))
\end{aligned} \tag{5.2.4}$$

where

$$\begin{aligned}
\Sigma_p =& \underset{(dp \times dp)}{\begin{bmatrix} \Sigma(0) & \Sigma(1) & \cdots & \Sigma(p-1) \\ \Sigma(-1) & \Sigma(0) & \cdots & \Sigma(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma(-p+1) & \Sigma(-p+2) & \cdots & \Sigma(0) \end{bmatrix}}, \\
\boldsymbol{A} =& \underset{(dp \times dp)}{\begin{bmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ I_d & 0 & \cdots & 0 & 0 \\ 0 & I_d & & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I_d & 0 \end{bmatrix}}, \text{ and } \boldsymbol{\Omega} = \underset{(dp \times dp)}{\begin{bmatrix} \Omega & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}}.
\end{aligned} \tag{5.2.5}$$

In order to construct a linear Student's t VAR with stationary properties analogous to (5.2.3), the appropriate marginal distribution of $p+1$ consecutive observations is considered. Then, a connection to the VAR (5.2.1) is made through the conditional distribution, and finally this process and its stationary properties are formally established. Suppose that for a random vector in $\mathbb{R}^{d(p+1)}$ it holds that $(z, \boldsymbol{z}) \sim t_{d(p+1)}(\mathbf{1}_{p+1} \otimes \mu, \Sigma_{p+1}, \nu)$, where $\nu > 2$. By the properties of a multivariate Student's t-distribution (given in Appendix 5.A), the conditional distribution of z

given $\boldsymbol{z}$ is $z|\boldsymbol{z} \sim t_d(\mu(\boldsymbol{z}), \Omega(\boldsymbol{z}), \nu + dp)$, where

$$\mu(\boldsymbol{z}) = \phi_0 + \boldsymbol{A}_p \boldsymbol{z} \tag{5.2.6}$$

$$\Omega(\boldsymbol{z}) = \frac{\nu - 2 + (\boldsymbol{z} - \mathbf{1}_p \otimes \mu)' \Sigma_p^{-1} (\boldsymbol{z} - \mathbf{1}_p \otimes \mu)}{\nu - 2 + dp} \Omega. \tag{5.2.7}$$

It is easy to see that a VAR of the form (5.2.1) that has the above-described conditional Student's t-distribution is obtained by assuming that $\varepsilon_t \sim t_d(0, I_d, \nu + dp)$ and that the conditional covariance matrix Ω_t is of the form (5.2.7). The following theorem then formally establishes this Student's t VAR and its stationary properties (which is analogous to Theorem 1 in Meitz *et al.*, forthcoming, considering a univariate version of the Student's t autoregression).

Theorem 5.1. *Suppose $\phi_0 \in \mathbb{R}^d$, $[A_1 : ... : A_p] \in \mathbb{S}^{d \times dp}$,$\Omega \in \mathbb{R}^{d \times d}$ is positive definite, and that $\nu > 2$. Then, there exists a process $\boldsymbol{z}_t = (z_t, ..., z_{t-p+1})$ $(t = 0, 1, 2, ...)$ with the following properties.*

(i) The process $\boldsymbol{z}_t$ is a Markov chain on $\mathbb{R}^{dp}$ with a stationary distribution characterized by the density function $t_{dp}(\mathbf{1}_p \otimes \mu, \Sigma_p, \nu)$. When $\boldsymbol{z}_0 \sim t_{dp}(\mathbf{1}_p \otimes \mu, \Sigma_p, \nu)$, we have, for $t = 1, 2, ...$, that $\boldsymbol{z}_t^+ \sim t_{d(p+1)}(\mathbf{1}_{p+1} \otimes \mu, \Sigma_{p+1}, \nu)$ and the conditional distribution of z_t given $\boldsymbol{z}_{t-1}$ is

$$z_t|\boldsymbol{z}_{t-1} \sim t_d(\mu(\boldsymbol{z}_{t-1}), \Omega(\boldsymbol{z}_{t-1}), \nu + dp). \tag{5.2.8}$$

(ii) Furthermore, for $t = 1, 2, ...$, the process z_t has the representation

$$z_t = \phi_0 + \sum_{i=1}^{p} A_i z_{t-i} + \Omega_t^{1/2} \varepsilon_t \tag{5.2.9}$$

where $\Omega_t = \Omega(\boldsymbol{z}_{t-1})$ is the conditional covariance matrix (see (5.2.7)), $\varepsilon_t \sim$ IID $t_d(0, I_d, \nu + dp)$, and ε_t are independent of $\{z_{t-j}, j > 0\}$ for all t.

Analogously to the univariate linear Student's t autoregression discussed in Meitz *et al.* (forthcoming), the results (i) and (ii) in Theorem 5.1 are comparable to properties (5.2.3) and (5.2.1) of the Gaussian counterpart. Part (i) shows that both the stationary and conditional distributions of z_t are t–distributions, whereas part (ii) clarifies the connection to the standard VAR model.

My Student's t VAR has a conditional mean identical to the Gaussian VAR, but unlike the Gaussian VAR, it is conditionally heteroskedastic. Specifically, the conditional variance (5.2.7) consists of a constant covariance matrix that is multiplied by a time-varying scalar that depends on the quadratic form of the preceding p observations through the autoregressive parameters. In this sense, the model has a 'VAR(p)–ARCH(p)' representation, but the ARCH type conditional variance is not as general as in the conventional multivariate ARCH process (e.g., Lütkepohl, 2005, Section 16.3) that allows the entries of the conditional covariance matrix to vary relative

to each other. My model is, however, more parsimonious than the conventional VAR-ARCH model, as the conditional covariance depends only on the degrees of freedom and autoregressive parameters (in addition to the parameters in the constant covariance matrix). Student's t VARs similar to mine have previously appeared at least in Heracleous (2003) and Poudyal (2012).

5.3 The Gaussian and Student's *t* mixture vector autoregressive model

The G-StMVAR model can be described as a collection of linear autoregressive models that are the linear Gaussian VARs or the linear Student's t VARs defined in Section 5.2. At each time point, the process generates an observation from one of its mixture components that is randomly selected according to the probabilities given by the mixing weights. This definition is formalized next.

Let y_t $(t = 1, 2, ...)$ be the real valued d-dimensional time series of interest, and let $\mathcal{F}_{t-1}$ denote σ-algebra generated by the random vectors $\{y_s, s < t\}$. In a G-StMVAR model with autoregressive order p and M mixture components (or regimes), the observations y_t are assumed to be generated by

$$y_t = \sum_{m=1}^{M} s_{m,t}(\mu_{m,t} + \Omega_{m,t}^{1/2}\varepsilon_{m,t}), \tag{5.3.1}$$

$$\mu_{m,t} = \phi_{m,0} + \sum_{i=1}^{p} A_{m,i}y_{t-i}, \tag{5.3.2}$$

where the following conditions hold (which are similar to Condition 1 in Kalliovirta *et al.*, 2016).

Condition 5.1.

(a) For $m = 1, ..., M_1 \leq M$, the random vectors $\varepsilon_{m,t}$ are IID $n_d(0, I_d)$ distributed, and for $m = M_1 + 1, ..., M$, they are IID $t_d(0, I_d, \nu_m + dp)$ distributed. For all m, $\varepsilon_{m,t}$ are independent of $\mathcal{F}_{t-1}$.

(b) For each $m = 1, ..., M$, $\phi_{m,0} \in \mathbb{R}^d$, $\boldsymbol{A}_{m,p} \equiv [A_{m,1} : ... : A_{m,p}] \in \mathbb{S}^{d\times dp}$ (the set $\mathbb{S}^{d\times dp}$ is defined in (5.2.2)), and Ω_m is positive definite. For $m = 1, ..., M_1$, the conditional covariance matrices are constants, $\Omega_{m,t} = \Omega_m$. For $m = M_1 + 1, ..., M$, the conditional covariance matrices $\Omega_{m,t}$ are as in (5.2.7), except that $\boldsymbol{z}$ is replaced with $\boldsymbol{y}_{t-1} = (y_{t-1}, ..., y_{t-p})$ and the regime specific parameters $\phi_{m,0}$, $\boldsymbol{A}_{m,p}$,Ω_m,ν_m are used to define the quantities therein. For $m = M_1 + 1, ..., M$, also $\nu_m > 2$.

(c) The unobservable regime variables $s_{1,t}, ..., s_{M,t}$ are such that at each t, exactly one of them takes the value one and the others take the value zero according to the conditional probabilities expressed in terms of the ($\mathcal{F}_{t-1}$-measurable) mixing weights $\alpha_{m,t} \equiv \mathrm{P}(s_{m,t} = 1|\mathcal{F}_{t-1})$ that satisfy $\sum_{m=1}^{M} \alpha_{m,t} = 1$.

(d) Conditionally on $\mathcal{F}_{t-1}$, $(s_{1,t}, ..., s_{M,t})$ and $\varepsilon_{m,t}$ are assumed independent.

The conditions $\nu_m > 2$ in (b) are made to ensure the existence of second moments. This definition implies that the G-StMVAR model generates each observation from one of its mixture components, a linear Gaussian or Student's t vector autoregression discussed in Section 5.2, and that the mixture component is selected randomly according to the probabilities given by the mixing weights $\alpha_{m,t}$.

The first M_1 mixture components are assumed to be linear Gaussian VARs, and the last $M_2 \equiv M - M_1$ mixture components are assumed to be linear Student's t VARs. If all the component processes are Gaussian VARs ($M_1 = M$), the G-StMVAR model reduces to the GMVAR model of Kalliovirta *et al.* (2016). If all the component processes are Student's t VARs ($M_1 = 0$), I refer to the model as the StMVAR model.

Equations (5.3.1) and (5.3.2) and Condition 5.1 lead to a model in which the conditional density function of y_t conditional on its past, $\mathcal{F}_{t-1}$, is given as

$$f(y_t|\mathcal{F}_{t-1}) = \sum_{m=1}^{M_1} \alpha_{m,t} n_d(y_t; \mu_{m,t}, \Omega_m) + \sum_{m=M_1+1}^{M} \alpha_{m,t} t_d(y_t; \mu_{m,t}, \Omega_{m,t}, \nu_m + dp). \quad (5.3.3)$$

The conditional densities $n_d(y_t; \mu_{m,t}, \Omega_{m,t})$ are obtained from (5.2.3), whereas $t_d(y_t; \mu_{m,t}, \Omega_{m,t}, \nu_m + dp)$ are obtained from Theorem 5.1. The explicit expressions of the density functions are given in Appendix 5.A. To fully define the G-StMVAR model it is then left to specify the mixing weights $\alpha_{m,t}$.

Analogously to Kalliovirta *et al.* (2015), Kalliovirta *et al.* (2016), Meitz *et al.* (forthcoming), and Chapter 2, I define the mixing weights as weighted ratios of the component process stationary densities corresponding to the previous p observations. In order to formally specify the mixing weights, the following function is first defined for notational convenience. Let

$$d_{m,dp}(\boldsymbol{y}; \mathbf{1}_p \otimes \mu_m, \Sigma_{m,p}, \nu_m) = \begin{cases} n_{dp}(\boldsymbol{y}; \mathbf{1}_p \otimes \mu_m, \Sigma_{m,p}), & \text{when } m \leq M_1, \\ t_{dp}(\boldsymbol{y}; \mathbf{1}_p \otimes \mu_m, \Sigma_{m,p}, \nu_m), & \text{when } m > M_1, \end{cases} \quad (5.3.4)$$

where the dp-dimensional densities $n_{dp}(\boldsymbol{y}; \mathbf{1}_p \otimes \mu_m, \Sigma_{m,p})$ and $t_{dp}(\boldsymbol{y}; \mathbf{1}_p \otimes \mu_m, \Sigma_{m,p}, \nu_m)$ correspond to the stationary distribution of the mth component process (given in Equation (5.2.3) for the Gaussian regimes and in Theorem 5.1 for the Student's t regimes). Denoting $\boldsymbol{y}_{t-1} = (y_{t-1}, ..., y_{t-p})$, the mixing weights of the G-StMVAR model are defined as

$$\alpha_{m,t} = \frac{\alpha_m d_{m,dp}(\boldsymbol{y}_{t-1}; \mathbf{1}_p \otimes \mu_m, \Sigma_{m,p}, \nu_m)}{\sum_{n=1}^{M} \alpha_n d_{n,dp}(\boldsymbol{y}_{t-1}; \mathbf{1}_p \otimes \mu_n, \Sigma_{n,p}, \nu_n)}, \quad (5.3.5)$$

where $\alpha_m \in (0, 1)$, $m = 1, ..., M$, are mixing weights parameters assumed to satisfy $\sum_{m=1}^{M} \alpha_m = 1$, $\mu_m = (I_d - \sum_{i=1}^{p} A_{m,i})^{-1} \phi_{m,0}$, and covariance matrix $\Sigma_{m,p}$ is given in (5.2.4) and (5.2.5) but using the regime specific parameters to define the quantities therein.

Because the mixing weights are weighted ratios of the component process stationary densities corresponding to the previous p observations, the greater the relative weighted likelihood of

a regime is, the more likely the process generates an observation from it. This is a convenient feature for forecasting, and it also facilitates associating statistical characteristics and economic interpretations to the regimes. Moreover, it turns out that this specific formulation of the mixing weights leads to attractive theoretical properties such as full knowledge of the stationary distribution of $p+1$ consecutive observations and ergodicity of the process. These properties are summarized in the following theorem.

Before stating the theorem, a few notational conventions are provided. The parameters of a G-StMVAR model are collected to the $((M(d+d^2p+d(d+1)/2+2)-M_1-1)\times 1)$ vector $\boldsymbol{\theta}=(\boldsymbol{\vartheta}_1,...,\boldsymbol{\vartheta}_M,\alpha_1,...,\alpha_{M-1},\boldsymbol{\nu})$, where $\boldsymbol{\vartheta}_m=(\phi_{m,0},vec(\boldsymbol{A}_{m,p}),vech(\Omega_m))$ and $\boldsymbol{\nu}=(\nu_{M_1+1},...,\nu_M)$. The last mixing weight parameter α_M is not parametrized because it is obtained from the restriction $\sum_{m=1}^{M}\alpha_m=1$. The G-StMVAR model with autoregressive order p, and M_1 Gaussian and M_2 Student's t mixture components is referred to as the G-StMVAR(p,M_1,M_2) model, whenever the order of the model needs to be emphasized.

Theorem 5.2. *Consider the G-StMVAR process y_t generated by (5.3.1), (5.3.2), and (5.3.5) with Condition 5.1 satisfied. Then, $\boldsymbol{y}_t=(y_t,...,y_{t-p+1})$ is a Markov chain on $\mathbb{R}^{dp}$ with stationary distribution characterized by the density*

$$f(\boldsymbol{y};\boldsymbol{\theta})=\sum_{m=1}^{M_1}\alpha_m n_{dp}(\boldsymbol{y};\mathbf{1}_p\otimes\mu_m,\Sigma_{m,p})+\sum_{m=M_1+1}^{M}\alpha_m t_{dp}(\boldsymbol{y};\mathbf{1}_p\otimes\mu_m,\Sigma_{m,p},\nu_m). \tag{5.3.6}$$

Moreover, $\boldsymbol{y}_t$ is ergodic.

The stationary distribution is a mixture of M_1 dp-dimensional Gaussian distributions and M_2 dp-dimensional t-distributions with constant mixing weights α_m. The proof of Theorem 5.2 in Appendix 5.B shows that the marginal stationary distributions of $1,...,p+1$ consecutive observations are likewise mixtures of Gaussian and t-distributions. This gives the mixing weight parameters α_m, $m=1,..,M$, interpretation as the unconditional probabilities of an observation being generated from the mth component process. The unconditional mean, covariance, and first p autocovariances are hence obtained as $E[y_t]=\sum_{m=1}^{M}\alpha_m\mu_m$ and

$$Cov(y_t,y_{t-j})=\sum_{m=1}^{M}\alpha_m\Sigma_m(j)+\sum_{m=1}^{M}\alpha_m\left(\mu_m-E[y_t]\right)\left(\mu_m-E[y_t]\right)', \tag{5.3.7}$$

where $j=0,1,...,p$ and $\Sigma_m(j)$ is the jth autocovariance matrix of the mth component process.

The conditional mean of the G-StMVAR process can be expressed as $E[y_t|\mathcal{F}_{t-1}]=\sum_{m=1}^{M}\alpha_{m,t}\mu_{m,t}$ and the conditional covariance matrix as

$$\begin{aligned}Cov(y_t|\mathcal{F}_{t-1})=&\sum_{m=1}^{M_1}\alpha_{m,t}\Omega_m+\sum_{m=M_1+1}^{M}\alpha_{m,t}\Omega_{m,t}\\&+\sum_{m=1}^{M}\alpha_{m,t}\left(\mu_{m,t}-E[y_t|\mathcal{F}_{t-1}]\right)\left(\mu_{m,t}-E[y_t|\mathcal{F}_{t-1}]\right)'.\end{aligned} \tag{5.3.8}$$

That is, the conditional mean is a weighted sum of the component processes' conditional means with the weights given by the time-varying mixing weights $\alpha_{m,t}$. The conditional variance consists of three terms. The first term is a weighted sum of the Gaussian component processes' conditional covariance matrices, and the second term is a weighted sum of the Student's t component processes' conditional covariance matrices with the weights given by the time-varying mixing weights, while the third term captures conditional heteroskedasticity caused by variations in the conditional mean.

By construction, the StMVAR model does not generally filter out autocorrelation as well as its Gaussian counterpart, the GMVAR model (Kalliovirta *et al.*, 2016), because the autoregressive parameters are also the coefficients for ARCH type conditional heteroskedasticity. This property arises from the utilization of the multivariate Student's t-distribution as the stationary distribution of the component processes. The utilization of the t-distribution allows for parsimonious modelling of series that display fat tails and conditional heteroskedasticity within the regimes. This is particularly advantegeous in the context of mixture VARs, as the large number of parameters may often be a problem even without an ARCH component. Appropriate modelling of kurtosis and conditional heteroskedasticity is important, since they may affect the endogenously determined regime-switching probabilities. Ignoring the modelling of kurtosis and conditional heteroskedasticity would leave out potentially important dynamics that may affect the outcome of an empirical investigation. If some of the regimes have a constant conditional covariance matrix and zero excess kurtosis, they are allowed to be conditionally homoskedastic linear Gaussian VARs, which leads to the G-StMVAR model.

5.4 Structural G-StMVAR model

5.4.1 The model setup

The G-StMVAR model can be extended to a structural version similarly to the structural GMVAR model discussed in Chapter 4.[1] Consider the G-StMVAR model defined by (5.3.1), (5.3.2), and (5.3.5) with Condition 5.1 satisfied. I write the structural G-StMVAR model as

$$y_t = \sum_{m=1}^{M} s_{m,t}(\phi_{m,0} + \sum_{i=1}^{p} A_{m,i}y_{t-i}) + B_t e_t \tag{5.4.1}$$

[1] The structural GMVAR model introduced in Chapter 4 is obtained as special case of my model by selecting $M_1 = M$, i.e., that all the regimes are of the GMVAR type.

and

$$u_t \equiv B_t e_t = \begin{cases} u_{1,t} \sim n_d(0, \Omega_{1,t}) & \text{if} \quad s_{1,t} = 1 \quad \text{(with probability } \alpha_{1,t}\text{)} \\ \vdots & \\ u_{M_1,t} \sim n_d(0, \Omega_{M_1,t}) & \text{if} \quad s_{M_1,t} = 1 \quad \text{(with probability } \alpha_{M_1,t}\text{)} \\ u_{M_1+1,t} \sim t_d(0, \Omega_{M_1+1,t}, \nu_{M_1+1} + dp) & \text{if} \quad s_{M_1+1,t} = 1 \quad \text{(with probability } \alpha_{M_1+1,t}\text{)} \\ \vdots & \\ u_{M,t} \sim t_d(0, \Omega_{M,t}, \nu_M + dp) & \text{if} \quad s_{M,t} = 1 \quad \text{(with probability } \alpha_{M,t}\text{)} \end{cases} \tag{5.4.2}$$

where the probabilities are expressed conditionally on $\mathcal{F}_{t-1}$ and e_t $(d \times 1)$ is an orthogonal structural error. This definition is similar to Equations (4.3.1) and (4.3.2) in Chapter 4 but with Student's t regimes in addition to the Gaussian ones.

For the Gaussian regimes ($m = 1, ..., M_1$), $\Omega_{m,t} = \Omega_m$. For the Student's t regimes ($m = M_1 + 1, ..., M$), $\Omega_{m,t} = \omega_{m,t}\Omega_m$, where

$$\omega_{m,t} = \frac{\nu_m - 2 + (\boldsymbol{y}_{t-1} - \mathbf{1}_p \otimes \mu_m)' \Sigma_{m,p}^{-1} (\boldsymbol{y}_{t-1} - \mathbf{1}_p \otimes \mu_m)}{\nu_m - 2 + dp}. \tag{5.4.3}$$

The invertible $(d \times d)$ "B-matrix" B_t, which governs the contemporaneous relationships of the shocks, is time-varying and a function of $y_{t-1}, ..., y_{t-p}$. I will define the B-matrix so that it captures the conditional heteroskedasticity of the reduced form error, and thereby amplifies a constant-sized structural shock accordingly. Appropriate modelling of conditional heteroskedasticity in the B-matrix is of interest because the (generalized) impulse response functions may be asymmetric with respect to the size of the shock.

We have $\Omega_{u,t} \equiv \text{Cov}(u_t|\mathcal{F}_{t-1}) = \sum_{m=1}^{M_1} \alpha_{m,t}\Omega_m + \sum_{m=M_1+1}^{M} \alpha_{m,t}\omega_{m,t}\Omega_m$, while the conditional covariance matrix of the structural error $e_t = B_t^{-1}u_t$ (which are not IID but martingale differences and thereby uncorrelated) is obtained as

$$\text{Cov}(e_t|\mathcal{F}_{t-1}) = \sum_{m=1}^{M_1} \alpha_{m,t} B_t^{-1} \Omega_m B_t'^{-1} + \sum_{m=M_1+1}^{M} \alpha_{m,t}\omega_{m,t} B_t^{-1} \Omega_m B_t'^{-1}. \tag{5.4.4}$$

Therefore, the B-matrix should be chosen so that the structural shocks are orthogonal regardless of which regime they come from. Chapter 4 shows that any such B-matrix has (linearly independent) eigenvectors of the matrix $\Omega_m\Omega_1^{-1}$ as its columns. Moreover, it is shown that under the following assumption and a constant normalization of the structural error's conditional variance, say, $\Omega_{u,t} = I_d$, the B-matrix is unique up to ordering of its columns and changing all signs in a column.[2]

Assumption 5.1. *Consider M positive definite $(d \times d)$ covariance matrices Ω_m, $m = 1, ..., M$, and denote the strictly positive eigenvalues of the matrices $\Omega_m\Omega_1^{-1}$ as λ_{mi}, $i = 1, ..., d$, $m =$*

[2] Chapter 4 shows the uniqueness of the B-matrix for the structural GMVAR model, but the results apply to my structural G-StMVAR model as well.

$2, ..., M$. *Suppose that for all* $i \neq j \in \{1, ..., d\}$, *there exists an* $m \in \{2, ..., M\}$ *such that* $\lambda_{mi} \neq \lambda_{mj}$.

Thus, as long as one is willing to assume a single (time-varying) B-matrix, its columns generally characterize the estimated impact effects of the shocks, but it is not revealed which column is related to which shock. Since the B-matrix is also subject to estimation error, further constraints may be needed for labelling the shocks.

Following Chapter 4 (and Lanne and Lütkepohl (2010) and Lanne, Lütkepohl, and Maciejowsla (2010)), I utilize the following matrix decomposition that is convenient for specifying the B-matrix and deriving the identification conditions. I decompose the error term covariance matrices as

$$\Omega_1 = WW' \quad \text{and} \quad \Omega_m = W\Lambda_m W', \quad m = 2, ..., M, \tag{5.4.5}$$

where the diagonal of $\Lambda_m = \text{diag}(\lambda_{m1}, ..., \lambda_{md})$, $\lambda_{mi} > 0$ $(i = 1, ..., d)$, contains the eigenvalues of the matrix $\Omega_m\Omega_1^{-1}$ and the columns of the nonsingular W are the related eigenvectors (that are the same for all m by construction). When $M = 2$, decomposition (5.4.5) always exists (Muirhead, 1982, Theorem A9.9), but for $M \geq 3$ its existence requires that the matrices $\Omega_m\Omega_1^{-1}$ share the common eigenvectors in W. If this is not the case, the B-matrix does not exist (see Section 4.3.1 in Chapter 4), but its existence is, however, testable.

Similarly to Chapter 4, any scalar multiples of W's columns comprise an appropriate B-matrix, but only specific scalar multiples comprise the locally unique B-matrix associated with a given normalization of the structural error's conditional covariance matrix. Direct calculation shows that the B-matrix associated with the normalization $\text{Cov}(e_t|\mathcal{F}_{t-1}) = I_d$ is obtained as

$$B_t = W(\sum_{m=1}^{M_1} \alpha_{m,t}\Lambda_m + \sum_{m=M_1+1}^{M} \alpha_{m,t}\omega_{m,t}\Lambda_m)^{1/2}, \tag{5.4.6}$$

where $B_tB_t' = \Omega_{u,t}$. Since $B_t^{-1}\Omega_m B_t'^{-1} = \Lambda_m(\sum_{n=1}^{M_1} \alpha_{n,t}\Lambda_n + \sum_{n=M_1+1}^{M} \alpha_{n,t}\omega_{n,t}\Lambda_n)^{-1}$, the B-matrix (5.4.6) simultaneously diagonalizes $\Omega_1, ..., \Omega_M$, and $\Omega_{u,t}$ (and thereby also $\Omega_{1,t}, ..., \Omega_{M,t}$) for each t so that $\text{Cov}(e_t|\mathcal{F}_{t-1}) = I_d$.

5.4.2 Identification of the shocks

I have established that in the model defined by Equations (5.4.1) and (5.4.2) with the normalization $\text{Cov}(e_t|\mathcal{F}_{t-1}) = I_d$, the structural shocks are identified up ordering and sign under Assumption 5.1. Global statistical identification of the shocks is therefore obtained by fixing the signs and ordering of the columns of B_t. The ordering of the columns can be fixed by fixing an arbitrary ordering for the eigenvalues in the diagonals of Λ_m, $m = 2, .., M$. The signs, in turn, can be normalized by placing a single strict sign constraint in each column of B_t.

However, the interest is often in identifying some specific shock (or shocks). To that end, the correct structural shock needs to be uniquely related to the shock of interest through constraints on the B-matrix (or equally W) that only the shock of interest satisfies. Proposition 4.2 in

Chapter 4 gives formal conditions for global identification of any subset of the shocks when the relevant pairs eigenvalues λ_{mi} are distinct in some regime. Chapter 4 also derives conditions for globally identifying some of the shocks when one of the relevant pairs of the eigenvalues is identical in all regimes (Proposition 4.3). For convenience, the conditions are repeated in the former case below, but in the latter case (as well as for the proof of the proposition below), I refer to Chapter 4.

Proposition 5.1. *Suppose* $\Omega_1 = WW'$ *and* $\Omega_m = W\Lambda_m W', \quad m = 2, ..., M$, *where* $\Lambda_m = diag(\lambda_{m1}, ..., \lambda_{md})$, $\lambda_{mi} > 0$ $(i = 1, ..., d)$, *contains the eigenvalues of* $\Omega_m\Omega_1^{-1}$ *in the diagonal and the columns of the nonsingular* W *are the related eigenvectors. Then, the last* d_1 *structural shocks are uniquely identified if*

(1) for all $j > d - d_1$ *and* $i \neq j$ *there exists an* $m \in \{2, ..., M\}$ *such that* $\lambda_{mi} \neq \lambda_{mj}$,

(2) the columns of W *are constrained in a way that for all* $i \neq j > d - d_1$, *the* i*th column cannot satisfy the constraints of the* j*th column as is nor after changing all signs in the* i*th column, and*

(3) there is at least one (strict) sign constraint in each of the last d_1 *columns of* W.

Condition (3) fixes the signs in the last d_1 columns of W, and therefore the signs of the instantaneous effects of the corresponding shocks. However, since changing the signs of the columns is effectively the same as changing the signs of the corresponding shocks, and the structural shock has a distribution that is symmetric about zero, this condition is not restrictive. The assumption that the last d_1 shocks are identified is not restrictive either, as one may always reorder the structural shocks accordingly. See Chapter 4 for examples on identifying shocks with this proposition. Finally, note that Assumption 5.1 is not required for identification, when only some of the shocks are to be identified. In that case, it is replaced with the weaker Condition (1) (that is always satisfied under Assumption 5.1).

If Condition (1) is strengthened to Assumption 5.1, the model is statistically identified. Consequently, the constraints imposed in Condition (2) become testable. If Assumption 5.1 is not satisfied, the testing problem is nonstandard and the conventional asymptotic distributions of likelihood ratio and Wald test statistics are unreliable (see the related discussion in Chapter 4, Section 4.3.3).

5.5 Estimation

The parameters of the G-StMVAR model can be estimated by the method of maximum likelihood (ML). Even the exact log-likelihood function is available, as I have established the stationary distribution of the process in Theorem 5.2. Suppose the observed time series is $y_{-p+1}, ..., y_0, y_1, ..., y_T$ and that the initial values are stationary. Then, the log-likelihood function of the G-StMVAR model takes the form

$$L(\boldsymbol{\theta}) = \log\left(\sum_{m=1}^{M} \alpha_m d_{m,dp}(\boldsymbol{y}_0; \mathbf{1}_p \otimes \mu_m, \Sigma_{m,p}, \nu_m)\right) + \sum_{m=1}^{M} l_t(\boldsymbol{\theta}), \tag{5.5.1}$$

where $d_{m,dp}(\cdot; \mathbf{1}_p \otimes \mu_m, \Sigma_{m,p}, \nu_m)$ is defined in (5.3.4) and

$$l_t(\boldsymbol{\theta}) = \log\left(\sum_{m=1}^{M_1} \alpha_{m,t} n_d(y_t; \mu_{m,t}, \Omega_m) + \sum_{m=M_1+1}^{M} \alpha_{m,t} t_d(y_t; \mu_{m,t}, \Omega_{m,t}, \nu_m + dp)\right). \quad (5.5.2)$$

If stationarity of the initial values seems unreasonable, one can condition on the initial values and base the estimation on the conditional log-likelihood function, which is obtained by dropping the first term on the right side of (5.5.1).

If there are two regimes in the model ($M = 2$), the structural G-StMVAR model is obtained from the estimated reduced form model by decomposing the covariance matrices $\Omega_1, ..., \Omega_M$ as in (5.4.5). If $M \geq 3$ or overidentifying constraints are imposed on B_t through W, the model can be reparametrized with W and Λ_m ($m = 2, ..., M$) instead of $\Omega_1, ..., \Omega_M$, and the log-likelihood function can be maximized subject to the new set of parameters and constraints. In this case, the decomposition (5.4.5) is plugged in to the log-likelihood function and $vech(\Omega_1), ..., vech(\Omega_M)$ are replaced with $vec(W)$ and $\boldsymbol{\lambda}_2, ..., \boldsymbol{\lambda}_M$ in the parameter vector $\boldsymbol{\theta}$, where $\boldsymbol{\lambda}_m = (\lambda_{m1}, ..., \lambda_{md})$. Instead of constraining $vech(\Omega_1), ..., vech(\Omega_M)$ so that $\Omega_1, ..., \Omega_M$ are positive definite, the constraints $\lambda_{mi} > 0$ are imposed for all $m = 2, ..., M$ and $i = 1, ..., d$.

In the rest of this section, we assume that estimation is based on the conditional log-likelihood function $L_T^{(c)}(\boldsymbol{\theta}) = T^{-1} \sum_{m=1}^{M} l_t(\boldsymbol{\theta})$, i.e., the ML estimator $\hat{\boldsymbol{\theta}}_T$ maximizes $L_T^{(c)}(\boldsymbol{\theta})$. I have scaled the conditional log-likelihood function with the sample size T so that the notation is consistent with the literature cited.

Establishing the asymptotic properties of the ML estimator requires that it is uniquely identified. In order to achieve unique identification, the parameters need to be constrained so that the mixture components cannot be 'relabelled' to produce the same model with a different parameter vector. The required assumption is

$$\begin{aligned}&\alpha_1 > \cdots > \alpha_{M_1} > 0,\ \alpha_{M_1+1} > \cdots > \alpha_M > 0, \text{ and } \boldsymbol{\vartheta}_i = \boldsymbol{\vartheta}_j \text{ only if any of the conditions}\\ &(1)\ 1 \leq i = j \leq M,\ (2)\ i \leq M_1 < j,\ (3)\ i, j > M_1 \text{ and } \nu_i \neq \nu_j, \text{ is satisfied.}\end{aligned} \quad (5.5.3)$$

In the case of the structural G-StMVAR model, identification also requires that Assumption 5.1 is satisfied (see Section 5.4).[3] Then, identification of the structural model follows from the identification of the reduced form model.

The constraints imposed on the parameter space are summarized in the following assumption.

Assumption 5.2. *The true parameter value $\boldsymbol{\theta}_0$ is an interior point of $\boldsymbol{\Theta}$, which is a compact subset of $\{\boldsymbol{\theta} = (\boldsymbol{\vartheta}_1, ..., \boldsymbol{\vartheta}_M, \alpha_1, ..., \alpha_{M-1}, \boldsymbol{\nu}) \in \mathbb{R}^{M(d+d^2p+d(d+1)/2)} \times (0,1)^{M-1} \times (2,\infty)^{M_2}$: $\boldsymbol{A}_{m,p} \in \mathbb{S}^{d \times dp}, \Omega_m$ is positive definite, for all $m = 1, ..., M$, and (5.5.3) holds$\}$.*

[3] With the appropriate zero constraints on W, this condition can be relaxed, however (see the related discussion in Chapter 4).

Asymptotic properties of the ML estimator under the conventional high-level conditions are stated in the following theorem. Denote $\mathcal{I}(\boldsymbol{\theta}) = E\left[\frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right]$ and $\mathcal{J}(\boldsymbol{\theta}) = E\left[\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\right]$.

Theorem 5.3. *Suppose that y_t are generated by the stationary and ergodic G-StMVAR process of Theorem 5.2 and that Assumption 5.2 holds. Then, $\hat{\boldsymbol{\theta}}_T$ is strongly consistent, i.e., $\hat{\boldsymbol{\theta}}_T \to \boldsymbol{\theta}_0$ almost surely. Suppose further that (i) $T^{1/2}\frac{\partial}{\partial \boldsymbol{\theta}_0} L_T^{(c)}(\boldsymbol{\theta}_0) \xrightarrow{d} N(0, \mathcal{I}(\boldsymbol{\theta}_0))$ with $\mathcal{I}(\boldsymbol{\theta}_0)$ finite and positive definite, (ii) $\mathcal{J}(\boldsymbol{\theta}_0) = -\mathcal{I}(\boldsymbol{\theta}_0)$, and (iii) $E[\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0} |\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}|] < \infty$ for some $\boldsymbol{\Theta}_0$, compact convex set contained in the interior of $\boldsymbol{\Theta}$ that has $\boldsymbol{\theta}_0$ as an interior point. Then, $T^{1/2}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \xrightarrow{d} N(0, -\mathcal{J}(\boldsymbol{\theta}_0)^{-1})$.*

Given consistency, conditions (i)-(iii) of Theorem 5.3 are standard for establishing asymptotic normality of the ML estimator, but their verification can be tedious. If one is willing to assume the validity of these conditions, the ML estimator has the conventional limiting distribution, implying that the approximate standard errors for the estimates are obtained as usual. Furthermore, the standard likelihood based tests are applicable as long as the number of mixture components is correctly specified. This condition is important, because if the number of Gaussian or Student's t type mixture components is chosen too large, some of the parameters are not identified causing the result of Theorem 5.3 to break down. This particularly happens when one tests for the number of regimes, as under the null some of the regimes are removed from the model.[4] Likewise, when testing whether a regime is a Gaussian VAR against the alternative that it is a Student's t VAR, under the null, $\nu_m = \infty$ for the Student's t regime m to be tested, which violates Assumption 5.2.

Finding the ML estimate amounts to maximizing the log-likelihood function defined in (5.5.1) and (5.5.2) over a high dimensional parameter space satisfying the constraints summarized in Assumption 5.2. Due to the complexity of the log-likelihood function, numerical optimization methods are required. The maximization problem can be challenging in practice. This is particularly due to the mixing weights' complex dependence on the preceding observations, which induces a large number of modes to the surface of the log-likelihood function, and large areas to the parameter space, where it is flat in multiple directions. Also, the popular EM algorithm (Redner and Walker, 1984) is virtually useless here, as at each maximization step one faces a new optimization problem that is not much simpler than the original one. Following Meitz, Preve, and Saikkonen (2018), Meitz *et al.* (forthcoming), and Chapters 2, 3, and 4, I therefore employ a two-phase estimation procedure in which a genetic algorithm is used to find starting values for a gradient based method. The R package **gmvarkit** (Virolainen, 2018) that accompanies this essay employs a modified genetic algorithm that works similarly to the one described in the univariate context in Chapter 2, Section 2.3.1, and Appendix 2.A.

[4] Meitz and Saikkonen (2021) have, however, recently developed such tests for mixture autoregressive models with Gaussian conditional densities. Developing a test for the number of a regimes in the G-StMVAR model is a major task and beyond the scope of this paper.

5.6 Building a G-StMVAR model

Building a G-StMVAR model amounts to finding a suitable autoregressive order p, the number of Gaussian regimes M_1, and the number of Student's t regimes M_2. I propose a model selection strategy that takes advantage of the observation that the G-StMVAR model is a limiting case of the StMVAR model (in which all the mixture components are linear Student's t VARs).

It is easy to check that the linear Gaussian vector autoregression defined in Section 5.2 is a limiting case of the linear Student's t vector autoregression when the degrees of freedom parameter tends to infinity. As the mixing weights (5.3.5) are weighted ratios of the component process stationary densities, it then follows that a G-StMVAR(p, M_1, M_2) model is obtained as a limiting case of the StMVAR(p, M) model (or equivalently the G-StMVAR($p, 0, M$) model) with the degrees of freedom parameters of the first M_1 regimes tending to infinity. Since a StMVAR(p, M) model that is fitted to data generated by a G-StMVAR(p, M_1, M_2) process is, therefore, asymptotically expected to get large estimates for the degrees of freedom parameters of the first M_1 regimes, I propose starting model selection by finding a suitable StMVAR model. If the StMVAR model contains overly large degrees of freedom parameter estimates, one should switch the corresponding regimes to Gaussian VARs by estimating the appropriate G-StMVAR model.

For a strategy to find a suitable StMVAR model, I follow Kalliovirta *et al.* (2015), and suggest first considering the linear version of the model, that is, a StMVAR model with one mixture component. Partial autocorrelation functions, information criteria, and (quantile) residual diagnostics may be made use of as usual for selecting the appropriate autoregressive order p. If the linear model is found inadequate, mixture versions of the model can be examined. One should, however, be conservative with the choice of M, because if the number of regimes is chosen too large, some of the parameters are not identified. Adding new regimes to the model also vastly increases the number of parameters, and moreover, due to the increased complexity, it might be difficult to obtain the ML estimate in practice if there are many regimes in the model.

Overly large degrees of freedom parameters are redundant in the model, but their weak identification also causes numerical problems. Specifically, they induce a numerically nearly singular Hessian matrix of the log-likelihood function when evaluated at the estimate, which makes the approximate standard errors and the quantile residual diagnostic tests of Kalliovirta and Saikkonen (2010) often unavailable. Since removal of overly large degrees of freedom parameters by switching to the appropriate G-StMVAR model has little effect on the model's fit, the switch is advisable whenever overly large degrees of freedom parameter estimates are obtained.

5.7 Empirical application

As an empirical application, I study asymmetries in the expected effects of the monetary policy shock in the Euro area. Asymmetric effects of the Euro area monetary policy shock have been studied, among others, by Peersman and Smets (2002) and Dolado and María-Dolores (2006),

who found that monetary policy shock has larger effects on production during recessions than expansions. Pellegrino (2018) found real effects of the monetary policy shock weaker during uncertain times than tranquil times, whereas Burgard *et al.* (2019) found the effects of contractionary monetary policy shocks stronger but less enduring during "crisis" than during "normal times".

I consider a monthly Euro area data covering the period from January 1999 to December 2021 (276 observations) and consisting of four variables: industrial production index (IPI), harmonized consumer price index (HCPI), Brent crude oil price (Europe, OIL), and an interest rate variable (RATE). My policy variable is the interest rate variable, which is the Euro overnight index average (EONIA) from January 1999 to October 2008 and the Wu and Xia (2016) shadow rate from November 2008 to December 2021. The Wu and Xia (2016) shadow rate is a shadow interest rate that is not bounded by the zero-lower-bound and also quantifies unconventional monetary policy measures.[5] Overall, my empirical application closely resembles that in Chapter 4, where asymmetries in the expected effects of the U.S. monetary policy shock were studied in a structural GMVAR model.

I detrend the IPI by first separating its cyclical component from the trend with the linear projection filter proposed by Hamilton (2018) and then considering the cyclical component.[6] I thereby implicitly assume that the monetary policy shock does not have permanent effects on real industrial production. Hereafter, I often refer to the IPI's deviation from the trend as the output gap. The logs of HCPI and the oil price are detrended by taking first differences, whereas the interest rate variable is assumed stationary. For numerical reasons, the cyclical component of the IPI and the log-difference of HCPI are multiplied by 100 and the log-difference of OIL by 10. The series are presented in Figure 5.1, where the shaded areas indicate the periods of Euro area recessions defined by the OECD.[7]

For selecting the order my G-StMVAR model, I started by estimating one-regime StMVAR models with autoregressive orders $p = 1, ..., 12$ and found that AIC was minimized by the order $p = 1$. Then, I estimated a two-regime StMVAR model with $p = 1$ but found this model somewhat inadequate. Therefore, I increased the autoregressive order to $p = 2$, which increased the AIC. The overall adequacy of the StMVAR$(2, 2)$ model, i.e., G-StMVAR($p = 2$, $M_1 = 0$, $M_2 = 2$) model, was found reasonable, so I employ it for the further analysis. Because the model does not contain large degrees of freedom parameter estimates, I do not consider incorporating

[5] The IPI, HCPI, and EONIA were obtained from the European Central Bank Statistical Data Warehouse; the Brent crude oil prices were retrieved from the Federal Reserve Bank of St. Louis database; and the Wu and Xia (2016) shadow rate was obtained from the first author's website.

[6] Denoting the univariate, non-stationary time series as y_t, the filter defines its transient component at the time $t+h$ ($h > 0$) as the ordinary least squares residuals from regressing y_{t+h} on a constant and $y_t, ..., y_{t-s+1}$. When s is chosen larger than the order of integration, the residual process is stationary. I used the parameter values $h = 24$ and $s = 12$, as suggested by Hamilton (2018) for monthly data.

[7] OECD Composite Leading Indicators, "Composite Leading Indicators: Reference Turning Points and Component Series", `http://www.oecd.org/std/leading-indicators/oecdcompositeleadingindicatorsreferenceturningpointsandcomponentseries.htm` (2.2.2022). At the time of accessing the series, the publicly available data ended at August 2021. I assume that the rest of the year 2021 did not contain recessions.

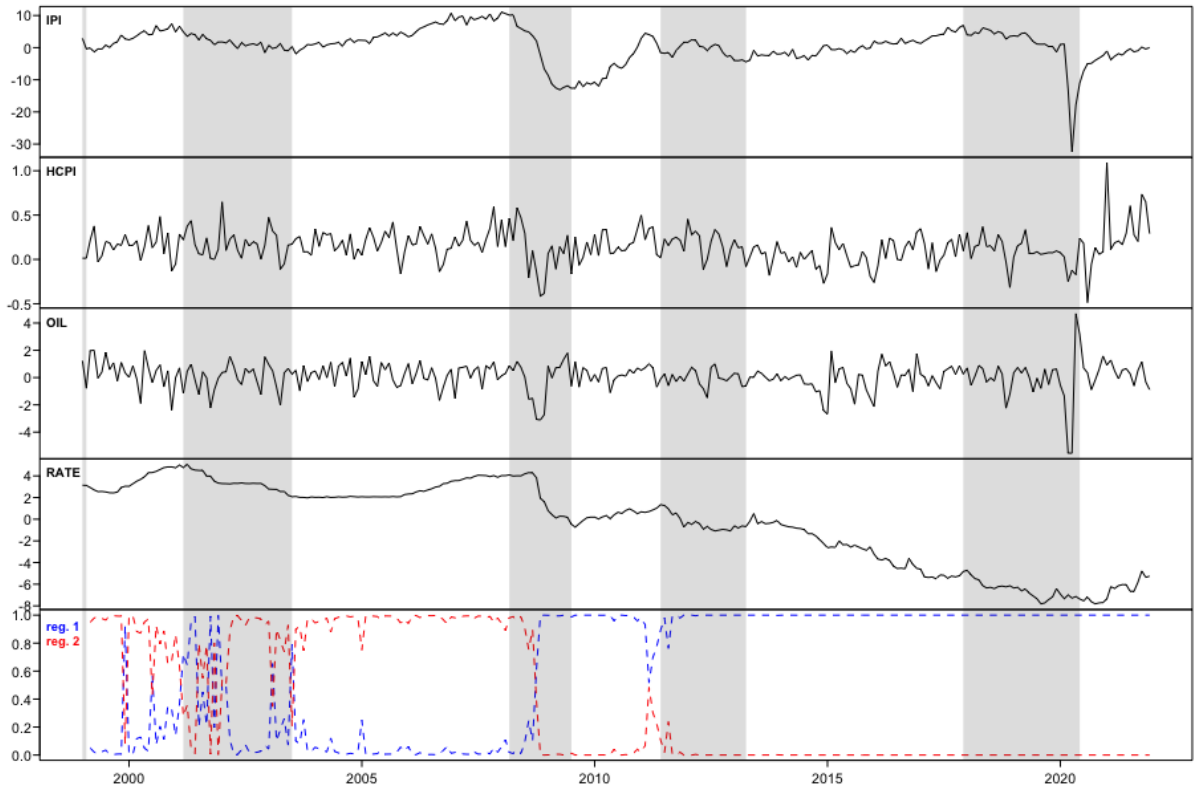

Figure 5.1: Monthly Euro area series covering the period from January 1999 to December 2021. The top panel presents the cyclical component of the industrial production index (IPI), which I separated from the trend using the linear projection filter proposed by Hamilton (2018). The second and third panels the log-differences of the harmonized consumer price index (HCPI) and Brent crude oil prices (Europe, OIL) multiplied by hundred and ten, respectively. The fourth panel presents an interest rate variable which is the EONIA from January 1999 to October 2008 and the Wu and Xia (2016) shadow rate from November 2008 to December 2021. The bottom panel shows the estimated mixing weights of the fitted StMVAR$(2, 2)$ model. The shaded areas indicate the periods of OECD based Euroa area recessions.

Gaussian mixture components by switching to a G-StMVAR model with $M_1 > 0$. Details on model selection and the adequacy of the selected model are provided in Appendix 5.C.1.

The estimated mixing weights of the StMVAR$(2, 2)$ model are presented in the bottom panel of Figure 5.1. The first regime (blue) mainly prevails after the Financial crisis in 2008, but it obtains large mixing weights also before and during the early 2000's recession. The second regime (red) dominates when the first one does not, that is, mainly before the Financial crisis. After the Financial crisis, its mixing weights stay close to zero, excluding a short period before the early 2010's recession, however. Since the prevailing regime starts switching sharply from the second to the first in October 2008, my model is consistent with the evidence that the ECB changed

its reaction function after the bankruptcy of Lehman Brothers in September 2008 (Gerlach and Lewis, 2014).[8]

Based on unconditional means (and marginal variances) of the regimes (presented in Table 5.2 in Appendix 5.C.2), the post-Financial crisis regime is characterized by negative (but volatile) output gap as well as lower inflation, oil price inflation, and interest rate variable than the pre-Financial crisis regime, which is characterized by positive output gap. The post-Financial crisis regime also exhibits higher kurtosis and overall volatility than the pre-Financial crisis regime. Details on the characteristics of the regimes are discussed in Appendix 5.C.2. For the ease of communication, I will refer to the first regime as the low-growth post-Financial crisis regime and the second regime as the high-growth pre-Financial crisis regime without explicitly reminding that the classification is not a strict one: both regimes obtain large mixing weights before and after the Financial crisis, while both regimes also prevail during expansions and recessions.

5.7.1 Identification of a monetary policy shock

Decomposing the covariance matrices of the reduced form StMVAR$(2, 2)$ model as in (5.4.5) gives the following estimates for the structural parameters:

$$\hat{W} = \begin{bmatrix} 0.77\ (0.512) & -0.94\ (0.618) & \mathbf{1.82}\ (0.771) & -0.13\ (0.271) \\ -0.12\ (0.069) & 0.12\ (0.077) & \mathbf{0.17}\ (0.079) & 0.03\ (0.026) \\ \mathbf{-1.05}\ (0.447) & -0.39\ (0.482) & 0.56\ (0.306) & 0.21\ (0.143) \\ 0.01\ (0.016) & -0.02\ (0.022) & 0.02\ (0.030) & \mathbf{0.50}\ (0.199) \end{bmatrix}, \quad \hat{\lambda}_2 = \begin{bmatrix} 0.50\ (0.397) \\ 0.36\ (0.291) \\ 0.19\ (0.155) \\ 0.03\ (0.021) \end{bmatrix}, \tag{5.7.1}$$

where the ordering of the variables is $y_t = (\text{IPI}_t, \text{HCPI}_t, \text{OIL}_t, \text{RATE}_t)$, the estimates $\hat{\lambda}_{2i}$ are in decreasing order (which fixes an arbitrary ordering for the columns of $\hat{W}$), and approximate standard errors are given in parentheses next to the estimates. The estimates that deviate from zero by more than two times their approximate standard error are bolded. I assume that the λ_{2i} are all distinct, i.e., Assumption 5.1.

The estimates and approximate standard errors in (5.7.1) show that the fourth shock is the only shock that moves the interest variable (also statistically) significantly at impact, while it is also the only shock that moves production to the opposite direction. Therefore, I deem it as the monetary policy shock. The fourth shock, however, appears to move inflation and oil price inflation to the same direction as the interest rate variable, which is contrary to many of the standard economic theories stating that an increase in the nominal interest rate should decrease inflation by decreasing aggregate demand (e.g., Galí, 2015, and the references therein).

The monetary policy shock is identified with Proposition 5.1 (Proposition 4.2 in Chapter 4) by placing such constraints on W (or equivalently the B-matrix) that it is unambiguously distin-

[8] Gerlach and Lewis (2014) found that the ECB was cutting the interest rates faster at the time of the crisis, and that the ECB started a policy shift back in the late 2010. According my StMVAR model, however, the dominating regime never switches back to the pre-Financial crisis regime (in my sample period), although the second regime obtains mixing weights clearly larger than zero in the late 2010 and several relatively large mixing weights in the early 2011.

guished from the other shocks. I assume that the monetary policy shock moves the interest rate and production in opposite directions and that it does not move inflation nor oil price inflation at impact. The zero constraints on inflation and oil price inflation obtained the p-values 0.22 and 0.15 in a Wald test individually and the p-value 0.34 jointly, so they are not rejected. These zero constraints are useful for distinguishing the monetary policy shock from the other shocks, but they also dampen the arguably implausible instantaneous increase in prices in response to a contractionary monetary policy shock.[9]

I distinguish the monetary policy shock from the other shocks by assuming that first shock moves oil price inflation at impact and that the second and third shocks move inflation at impact. This is not economically restrictive, as the responses can be very small. The above-described identification produced the following estimates:

$$\hat{W} = \begin{bmatrix} 0.88\ (0.692) & -1.16\ (0.812) & 2.00\ (1.164) & -0.40\ (0.351) \\ -0.13\ (0.091) & 0.14\ (0.099) & 0.20\ (0.122) & 0 \\ -1.20\ (0.686) & -0.42\ (0.561) & 0.65\ (0.435) & 0 \\ 0.01\ (0.021) & -0.03\ (0.029) & 0.06\ (0.045) & 0.56\ (0.313) \end{bmatrix}, \quad \hat{\lambda}_2 = \begin{bmatrix} 0.37\ (0.423) \\ 0.27\ (0.303) \\ 0.14\ (0.161) \\ 0.02\ (0.027) \end{bmatrix}, \tag{5.7.2}$$

where the identified monetary policy shock is ordered last.[10]

5.7.2 Impulse response analysis

Following Chapter 4 (and others), I employ the generalized impulse response function (GIRF) (Koop, Pesaran, and Potter, 1996) for estimating the expected of effects of the monetary policy shock. The GIRF is defined as

$$\text{GIRF}(h, \delta_j, \mathcal{F}_{t-1}) = \text{E}[y_{t+h}|\delta_j, \mathcal{F}_{t-1}] - \text{E}[y_{t+h}|\mathcal{F}_{t-1}], \tag{5.7.3}$$

where $h \in \{0, 1, 2, ...\}$ is the horizon and $\mathcal{F}_{t-1} = \sigma\{y_{t-j}, j > 0\}$ as before. The first term on the right side is the expected realization of the process at time $t + h$ conditionally on a structural shock of size $\delta_j \in \mathbb{R}$ in the jth element of e_t at time t and the previous observations. The second term on the right side is the expected realization of the process conditionally on the previous observations only. The GIRF thus expresses the expected difference in the future outcomes when the structural shock of size δ_j in the jth element of e_t hits the system at time t as opposed to all shocks being random. Since the regimes of my StMVAR model have economic

[9] My results are, hence, contrary to Castelnuovo (2016) who argued that muted response of inflation in the Euro area could be caused by misspecified zero constraints in the impact matrix. My model suggests that the zero constraints instead dampen the price puzzle (see Sims, 1992).

[10] The parameters λ_{2i}, $i = 1, ..., 4$, were assumed distinct without a formal justification, which led to statistical identification of the model. However, by Proposition 4.3 in Chapter 4, the monetary policy shock is still identified if $\lambda_{2i} = \lambda_{2j}$ for any $i, j = 1, 2, 3$ and additionally $\lambda_{2i} = \lambda_{24}$ for any one of $i = 1, 2, 3$. But the approximate standard errors and the Wald test results are valid only if λ_{2i} are all distinct. In particular, the approximate standard errors cannot be used to infer about the (in)equality of λ_{2i}, $i = 1, ..., 4$, without considerably complex examinations, as they are (asymptotically) valid only if the λ_{2i} are different to each other in the first place (see Lütkepohl, Meitz, Netšunajev, and Saikkonen, 2021, for a related discussion).

interpretations, it is interesting to also study the effects of the monetary policy shock to the mixing weights $\alpha_{m,t}$. The related GIRFs are obtained by replacing y_{t+h} with $\alpha_{m,t+h}$ on the right side of (5.7.3).

The G-StMVAR model has a p-step Markov property, so the GIRF can be calculated conditionally on the (σ-algebra generated by the) p previous observations $\boldsymbol{y}_{t-1} = (y_{t-1}, ..., y_{t-p})$. I make use of this property by generating histories $\boldsymbol{y}_{t-1} = (y_{t-1}, ..., y_{t-p})$ from the stationary distribution of each regime separately, and thereby obtain GIRFs conditional on the starting values being from this regime. The GIRFs and confidence intervals that reflect uncertainty about the initial value within the given regime are estimated using the Monte Carlo algorithm presented in Appendix 4.B in Chapter 4, where the point estimate is the mean over the Monte Carlo replications and the confidence intervals are obtained from the empirical quantiles.

The StMVAR model accommodates asymmetries in the GIRFs with respect the initial state of the economy as well as to the sign and size of the shock. I study these three types of asymmetries by generating starting values from each regime separately and then estimating GIRFs for positive (contractionary) and negative (expansionary) one-standard-error (small) and two-standard-error (large) shocks. After estimating the GIRFs, I scale them so that they correspond to a 25 basis point instantaneous increase of the interest rate variable, making any asymmetries easy to detect.

Figure 5.2 presents the GIRFs $h = 0, 1, ..., 96$ months ahead estimated for the identified monetary policy shock.[11] The GIRFs of inflation and oil price inflation are accumulated to (scaled) log-levels. From the top to bottom, the responses of IPI, HCPI, oil price, interest rate, and the first regime's mixing weights are depicted in each row, respectively. The first [third] column shows the responses to small contractionary (blue solid line) and expansionary (red dashed line) shocks with the initial values generated from the stationary distribution of the low-growth post-Financial crisis [high-growth pre-Financial crisis] regime. The second [fourth] column shows the responses to large contractionary and expansionary shocks with the initial values generated from the low-growth post-Financial crisis [high-growth pre-Financial crisis] regime. The shaded areas are the 90% confidence intervals that reflect the uncertainty about the initial value within the given regime. Responses of the second regime's mixing weights are not depicted because they are the negative of those of the first regime.

In the low-growth post-Financial crisis regime (the first and second columns of Figure 5.2), the IPI decreases (increases) strongly at impact in response to a contractionary (expansionary) monetary policy shock. On average, the peak response is in the first period, and then the average response starts to slowly decay towards zero.[12] The confidence intervals show that with some of the starting values the peak effect occurs later, however. The effects seem to die out faster for an expansionary than a contractionary shock.

[11] I used $R_1 = R_2 = 2500$ in the Monte Carlo algorithm (Chapter 4, Appendix 4.B). That is, for each regime as well as sign and size of the shock, I generated 2500 initial values, and for each of those initial values the GIRF is estimated based on 2500 Monte Carlo repetitions.

[12] By zero, I mean the expected observation if all the shocks were random. Accordingly, by positive, I mean expected observations larger than that and by negative expected observations smaller than that.

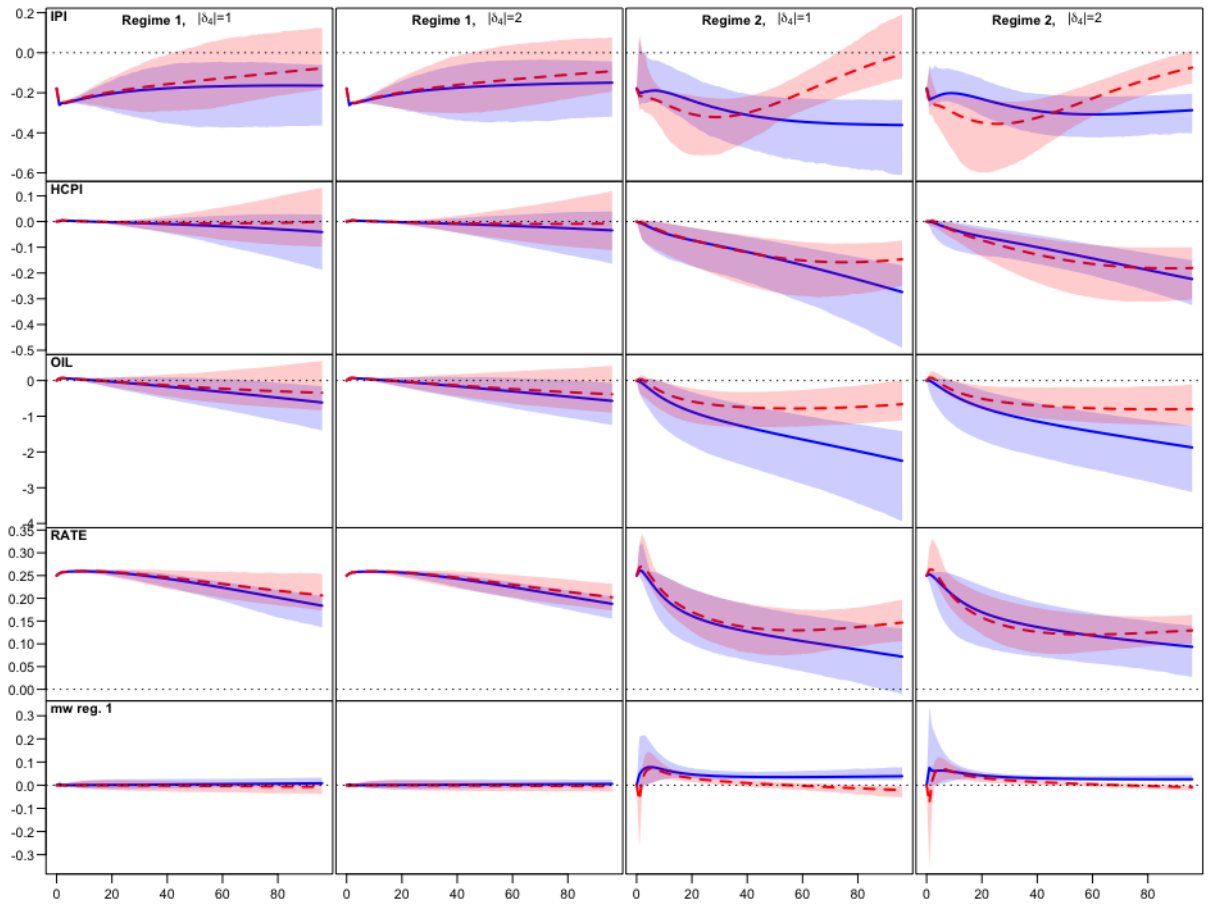

Figure 5.2: Generalized impulse response functions $h = 0, 1, ..., 96$ months ahead estimated for the monetary policy shock identified in Section 5.7.1 using $R_1 = R_2 = 2500$ in the Monte Carlo algorithm presented in Chapter 4, Appendix 4.B. From top to bottom, the responses of production, HCPI, oil price, the interest rate variable, and the first regime's mixing weights are depicted in each row, respectively. The GIRFs of the HCPI and oil price are accumulated to (scaled) log-levels. The first [third] column shows the responses to one-standard-error contractionary (blue solid line) and expansionary (red dashed line) shocks with the initial values generated from the stationary distribution of the low-growth post-Financial crisis [high-growth pre-Financial crisis] regime. The second [fourth] column shows the responses to two-standard error contractionary and expansionary shocks with the initial values generated from the low-growth post-Financial crisis [high-growth pre-Financial crisis] regime. After estimation, all GIRFs were scaled so that the instantaneous movement of the interest rate variable is 25 basis points. The shaded areas are the 90% confidence intervals that reflect uncertainty about the initial state within the given regime. Responses of the second regime's mixing weights are not depicted because they are the negative of those of the first regime.

In response to a contractionary shock, the price level stays roughly at zero for several years but decreases slowly and persistently. In response to an expansionary shock, the price level barely moves on average in the horizon of eight years, but confidence bounds show that with some of the starting values it decreases and from some increases.[13] The oil price seems to increase (decrease) slightly for roughly fifteen months before it decreases (increases) persistently. The confidence bounds, however, show that with some starting values the expansionary shock may decrease the oil price. The interest rate variable stays high (low) very persistently, and the GIRFs seem quite symmetric with respect to the size of the shock.

In the high-growth pre-Financial crisis regime (the third and fourth columns of Figure 5.2), the IPI decreases (increases) strongly at impact in response to a contractionary (expansionary) monetary policy shock. On average, the response then decreases (increases) and peaks roughly after two and a half years for an expansionary shock, but stays low very persistently without a particular peak effect for a contractionary shock. The IPI stays low very persistently, because the probability of entering the low-growth post-Financial crisis regime stays above zero very persistently, as the responses of first regime's mixing weights show (the bottom panels of the third and fourth columns in Figure 5.2).

The price level starts to steadily decrease (increase) after the impact period. The average price level somewhat stabilizes roughly after two years when the shock is expansionary, but keeps decreasing over the horizon of eight years when the shock is contractionary. The confidence bounds show that with some of the initial values, the price level decays towards zero after several years when the shock is small and expansionary. The oil price moves similarly to the consumer prices, whereas the interest rate variable stays high (low) persistently, more so if the shock is expansionary. Interestingly, an expansionary shock significantly increases the probability of the low-growth regime in the first period, but in the following periods it significantly increases the probability of the high-growth regime.

Overall, I find strong asymmetries with respect to the initial state of the economy and the sign of the shock, but the asymmetries are weak with respect to the size of the shock. The real effects are less enduring for an expansionary shock than for a contractionary shock. Particularly in the high-growth pre-Financial crisis regime, a contractionary shock persistently drives the economy towards the low-growth post-Financial crisis regime, which translates to a very persistent decrease in the output gap. The inflationary effects of the monetary policy shock are

[13]The observation of prices rising in response to a contractionary monetary policy shock (or decreasing in response to an expansionary monetary policy shock) is often referred to as the price puzzle. Sims (1992) suggested that the price puzzle may appear if the monetary policy maker uses more information in predicting the future inflation than the autoregressive system of the variables included in the model. Consequently, the identified monetary policy shock would also contain a component that incorporates some of the policy maker's endogenous response to the prediction of the future inflation, which may then make it appear as if the prices increase (or decrease) in response to the shock. Another explanation proposes that the prices increase due to the cost-push effect of the monetary policy shock: an increase in the nominal interest rate increases the marginal cost of production of the firms who operate on borrowed money, and thereby decreases the aggregate supply and increases the price level (e.g., Barth and Ramey, 2001, Ravenna and Walsh, 2006). Several authors have, however, argued that the cost-channel is not likely strong enough to cause a price puzzle (e.g., Castelnuovo, 2012, Kaufmann and Scharler, 2009, Rabanal, 2007).

stronger in the high-growth regime, while they are on average weak in the low-growth regime. In the low-growth regime, however, monetary policy is mainly measured with the Wu and Xia (2016) shadow rate instead of EONIA, which is mostly close to zero after the Financial crisis (in my sample period). Thus, the outcome might differ in the two regimes also due to the different measures of monetary policy.

5.8 Summary

I introduced a new mixture vector autoregressive model, which has attractive theoretical and practical properties. The G-StMVAR model accommodates both, conditionally homoskedastic Gaussian VARs and conditionally heteroskedastic Student's t VARs as its mixture components. The mixing weights are defined as weighted ratios of the component process stationary densities corresponding to p previous observations. Therefore, the greater the relative weighted likelihood of a regime is, the more likely the process is to generate an observation from it. This facilitates associating statistical characteristic and economic interpretations to the regimes. The specific formulation of the mixing weights also leads to attractive theoretical properties such as ergodicity and full knowledge of the stationary distribution of $p+1$ consecutive observations. Moreover, the maximum likelihood estimator of a stationary G-StMVAR model is strongly consistent, and therefore, it has the conventional limiting distribution under conventional high level conditions.

The G-StMVAR model is a multivariate version of the G-StMAR model of Virolainen (forthcoming). As special case, by assuming that all the mixture components are of the Student's t type, a multivariate version of the StMAR model of Meitz *et al.* (forthcoming) is obtained, which I call the StMVAR model. In addition to the reduced form model, I introduced a structural version of the G-StMVAR model with a time-varying impact matrix and statistically identified shocks. Referring to Chapter 4, I discussed the problem identifying the structural shocks and presented a general set of conditions for identifying any subset of the shocks. I then employed the structural model in the empirical application. The accompanying CRAN distributed R package **gmvarkit** (Virolainen, 2018) provides a comprehensive set of tools for maximum likelihood estimation and other numerical analysis of the introduced models.

The empirical application studied asymmetries in the expected effects of the monetary policy shock in the Euro area and considered a monthly data covering the period from January 1999 to December 2021. My StMVAR model identified two regimes: a low-growth regime and a high-growth regime. The low-growth regime is characterized by a negative (but volatile) output gap, and it mainly prevails after the collapse of Lehman Brothers in the Financial crisis but obtains large mixing weights also during and before the early 2000's recession. The high-growth regime is characterized by a positive output gap and it mainly dominates before the Financial crisis.

I found strong asymmetries with respect to the initial state of the economy and the sign of the shock, but asymmetries with respect to the size of the shock were weak. The real effects are less enduring for an expansionary shock than for a contractionary shock. Particularly in the high-growth pre-Financial crisis regime, a contractionary shock persistently drives the economy towards the low-growth post-Financial crisis regime, which translates to a very persistent de-

crease in the output gap. The inflationary effects of the monetary policy shock are stronger in the high-growth regime than in the low-growth regime, and in the latter the price level did not move much on average. In the low-growth regime, however, monetary policy is mainly measured with the Wu and Xia (2016) shadow rate instead of EONIA, which is mostly close to zero after the Financial crisis (in my sample period).

Bibliography

Aomoto K., Kita M. (2011). *Theory of Hypergeometric Functions*. 1st edition. Springer-Verlag, Tokyo.

Barth M. J., Ramey V. A. (2001). "The Cost Channel of Monetary Transmission." In Bernanke B. S., Rogoff K. (eds.), *NBER Macroeconomics Annual*, volume 16, pp. 199–239. MIT Press, Cambridge.

Burgard J., Neuenkirch M., Nöckel M. (2019). "State-Dependent Transmission of Monetary Policy in the Euro Area." *Journal of Money, Credit and Banking*, **51**(7), 2053–2070.

Castelnuovo E. (2012). "Testing the Structural Interpretation of the Price Puzzle with a Cost-Channel Model." *Oxford Bulletin of Economics and Statistics*, **74**(3), 425–452.

Castelnuovo E. (2016). "Monetary policy shocks and Cholesky VARs: an assessment for the Euro area." *Empirical Economics*, **50**, 383–414.

Ding P. (2016). "On the Conditional Distribution of the Multivariate t Distribution." *The American Statistician*, **70**(3), 293–295.

Dolado J. J., María-Dolores R. (2006). "State Asymmetries in the Effects of Monetary Policy Shocks on Output: Some New Evidence for the Euro-Area." In Milas C., Rothman P., van Dijk D. (eds.), *Nonlinear Time Series Analysis of Business Cycles*, pp. 311–331. Emerald Publishing Limited.

Galí J. (2015). *Monetary Policy, Inflation, and the Business Cycle*. 2nd edition. Princeton University Press, Princeton and Oxford.

Gerlach S., Lewis J. (2014). "ECB Reaction Functions and the Crisis of 2008." *International Journal of Central Banking*, **10**(1), 137–158.

Hamilton J. D. (2018). "WHY YOU SHOULD NEVER USE THE HODRICK-PRESCOTT FILTER." *The Review of Economics and Statistics*, **100**(5), 831–843.

Heracleous M. S. (2003). *Volatility Modeling Using the Student's t Distribution*. Ph.D. thesis, Virginia Tech.

Holzmann H., Munk A., Gneiting T. (2006). "Identifiability of finite mixtures of elliptical distributions." *Scandinavian Journal of Statistics*, **33**(4), 753–763.

Kalliovirta L. (2012). "Misspecification tests based on quantile residuals." *The Econometrics Journal*, **15**(2), 358–393.

Kalliovirta L., Meitz M., Saikkonen P. (2015). "A Gaussian Mixture Autoregressive Model for Univariate Time Series." *Journal of Time Series Analysis*, **36**(2), 247–266.

Kalliovirta L., Meitz M., Saikkonen P. (2016). "Gaussian mixture vector autoregression." *Journal of Econometrics*, **192**(2), 465–498.

Kalliovirta L., Saikkonen P. (2010). "Reliable Residuals for Multivariate Nonlinear Time Series Models." *Unpublished revision of HECER discussion paper No. 247*.

Kaufmann S., Scharler J. (2009). "Financial systems and the cost channel transmission of monetary policy shocks." *Economic Modelling*, **26**(1), 40–46.

Koop G., Pesaran M., Potter S. (1996). "Impulse response analysis in nonlinear multivariate models." *Journal of Econometrics*, **74**(1), 119–147.

Lanne M., Lütkepohl H. (2010). "Structural Vector Autoregressions With Nonnormal Residuals." *Journal of Business & Economic Statistics*, **28**(1), 159–168.

Lanne M., Lütkepohl H., Maciejowsla K. (2010). "Structural vector autoregressions with Markov switching." *Journal of Economic Dynamics and Control*, **34**(2), 121–131.

Lütkepohl H. (2005). *New Introduction to Multiple Time Series Analysis*. 1st edition. Springer, Berlin.

Lütkepohl H., Meitz M., Netšunajev A., Saikkonen P. (2021). "Testing identification via heteroskedasticity in structural vector autoregressive models." *The Econometrics Journal*, **24**(1), 1–22.

Meitz M., Preve D., Saikkonen P. (2018). ***StMAR Toolbox****: A MATLAB Toolbox for Student's t Mixture Autoregressive Models*.

Meitz M., Preve D., Saikkonen P. (forthcoming). "A mixture autoregressive model based on Student's t-distribution." *Communications in Statistics - Theory and Methods*.

Meitz M., Saikkonen P. (2021). "Testing for observation-dependent regime switching in mixture autoregressive models." *Journal of Econometrics*, **222**(1), 601–624.

Meyn S., Tweedie R. (2009). *Markov Chains and Stochastic Stability*. 2nd edition. Cambridge University Press, Cambridge.

Muirhead R. (1982). *Aspects of Multivariate Statistical Theory*. 1st edition. John Wiley & Sons, Hoboken, New Jersey.

Newey W., McFadden D. (1994). "Large sample estimation and hyphothesis testing." In Eagle R.F., MacFadden D.L. (eds.), *Handbook of Econometrics*, volume 4, chapter 36. Elsevier Science B.V.

Peersman G., Smets F. (2002). "Are the effects of monetary policy in the euro area greater in recessions than in booms?" In Mahadeva L., Sinclair P. (eds.), *Monetary Transmission in Diverse Economies*, pp. 28–48. Cambridge University Press.

Pellegrino G. (2018). "Uncertainty and the real effects of monetary policy shocks in the Euro area." *Economics Letters*, **162**, 177–181.

Poudyal N. (2012). *Confronting Theory with Data: the Case of DSGE Modeling*. Ph.D. thesis, Virginia Tech.

Rabanal P. (2007). "Does inflation increase after a monetary policy tightening? Answers based on an estimated DSGE model." *Journal of Economic Dynamics and Control*, **31**(3), 906–937.

Ranga Rao R. (1962). "Relations between Weak and Uniform Convergence of Measures with Applications." *The Annals of Mathematical Statistics*, **33**(2), 659–680.

Ravenna F., Walsh C. E. (2006). "Optimal monetary policy with the cost channel." *Journal of Monetary Economics*, **53**(2), 199–216.

Redner R., Walker H. (1984). "Mixture Densities, Maximum Likelihood and the Em Algorithm." *Society for Industrial and Applied Mathematics*, **26**(2), 195–239.

Sims A. (1992). "Interpreting the macroeconomic time series facts." *European economic review*, **36**(5), 975–1000.

Virolainen S. (2018). ***gmvarkit**: Estimate Gaussian and Student's t Mixture Vector Autoregressive Models*. R package version 2.0.3 available at CRAN: `https://CRAN.R-project.org/package=gmvarkit`.

Virolainen S. (2021). "Structural Gaussian Mixture vector autoregressive model with application to the asymmetric effects of monetary policy shocks." *Unpublished working paper, available as arXiv:2007.04713*.

Virolainen S. (forthcoming). "A mixture autoregressive model based on Gaussian and Student's t-distributions." *Studies in Nonlinear Dynamics & Econometrics*.

Wu J., Xia F. (2016). "Measuring the Macroeconomic Impact of Monetary Policy at the Zero Lower Bound." *Journal of Money, Credit and Banking*, **48**(2-3), 253–291.

Appendix 5.A Properties of multivariate Gaussian and Student's t distribution

Denote a d-dimensional real valued vector by y. It is well known that the density function of a d-dimensional Gaussian distribution with mean μ and covariance matrix Σ is

$$n_d(y;\mu,\Sigma) = (2\pi)^{-d/2}\det(\Sigma)^{-1/2}\exp\left\{-\frac{1}{2}(y-\mu)'\Sigma^{-1}(y-\mu)\right\}. \tag{5.A.1}$$

Similarly to Meitz *et al.* (forthcoming) but differing from the standard form, I parametrize the Student's t-distribution using its covariance matrix as a parameter together with the mean and the degrees of freedom. The density function of such a d-dimensional t-distribution with mean μ, covariance matrix Σ, and $\nu > 2$ degrees of freedom is (see, e.g., Appendix A in Meitz *et al.*, forthcoming)

$$t_d(y;\mu,\Sigma,\nu) = C_d(\nu)\det(\Sigma)^{-1/2}\left(1+\frac{(y-\mu)'\Sigma^{-1}(y-\mu)}{\nu-2}\right)^{-(d+\nu)/2}, \tag{5.A.2}$$

where

$$C_d(\nu) = \frac{\Gamma\left(\frac{d+\nu}{2}\right)}{\sqrt{\pi^d(\nu-2)^d}\Gamma\left(\frac{\nu}{2}\right)}, \tag{5.A.3}$$

and $\Gamma(\cdot)$ is the gamma function. I assume that the covariance matrix Σ is positive definite for both distributions.

Consider a partition $X = (X_1, X_2)$ of either Gaussian or t-distributed (with ν degrees of freedom) random vector X such that X_1 has dimension $(d_1 \times 1)$ and X_2 has dimension $(d_2 \times 1)$. Consider also a corresponding partition of the mean vector $\mu = (\mu_1, \mu_2)$ and the covariance matrix

$$\Sigma = \begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{12}' & \Sigma_{22}\end{bmatrix}, \tag{5.A.4}$$

where, for example, the dimension of Σ_{11} is $(d_1 \times d_1)$. In the Gaussian case, X_1 then has the marginal distribution $n_{d_1}(\mu_1, \Sigma_{11})$ and X_2 has the marginal distribution $n_{d_2}(\mu_2, \Sigma_{22})$. In the Student's t case, X_1 has the marginal distribution $t_{d_1}(\mu_1, \Sigma_{11}, \nu)$ and X_2 has the marginal distribution $t_{d_2}(\mu_2, \Sigma_{22}, \nu)$ (see, e.g., Ding, 2016, also in what follows).

When X has Gaussian distribution, the conditional distribution of the random vector X_1 given $X_2 = x_2$ is

$$X_1 \mid (X_2 = x_2) \sim n_{d_1}(\mu_{1|2}(x_2), \Sigma_{1|2}(x_2)), \tag{5.A.5}$$

where

$$\mu(x_2) \equiv \mu_{1|2}(x_2) = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2) \quad \text{and} \tag{5.A.6}$$

$$\Omega \equiv \Sigma_{1|2}(x_2) = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}'. \tag{5.A.7}$$

When X has t-distribution, the conditional distribution of the random vector X_1 given $X_2 = x_2$ is

$$X_1 \mid (X_2 = x_2) \sim t_{d_1}(\mu_{1|2}(x_2), \Sigma_{1|2}(x_2), \nu + d_2), \tag{5.A.8}$$

where

$$\mu(x_2) = \mu_{1|2}(x_2) = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2) \quad \text{and} \tag{5.A.9}$$

$$\Omega(x_2) \equiv \Sigma_{1|2}(x_2) = \frac{\nu - 2 + (x_2 - \mu_2)'\Sigma_{22}^{-1}(x_2 - \mu_2)}{\nu - 2 + d_2}(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}'). \tag{5.A.10}$$

In particular, we have

$$n_d(x; \mu, \Sigma) = n_{d_1}(x_1; \mu_{1|2}(x_2), \Sigma_{1|2}(x_2))n_{d_2}(x_2; \mu_2, \Sigma_{22}) \quad \text{and} \tag{5.A.11}$$

$$t_d(x; \mu, \Sigma, \nu) = t_{d_1}(x_1; \mu_{1|2}(x_2), \Sigma_{1|2}(x_2), \nu + d_2)t_{d_2}(x_2; \mu_2, \Sigma_{22}, \nu). \tag{5.A.12}$$

Appendix 5.B Proofs

5.B.1 Proof of Theorem 5.1

Corresponding to $\phi_0 \in \mathbb{R}^d$, $\boldsymbol{A}_p \in \mathbb{S}^{d\times dp}$, $\Omega \in \mathbb{R}^{d\times d}$ positive definite, and $\nu > 2$, define the notation μ, Σ_p, $\Sigma_1(h)$ $(h = 0, 1, ..., p)$, Σ_{1p}, and Σ_{p+1} as in (5.2.4). Note that, by construction and the assumption $\boldsymbol{A}_p \in \mathbb{S}^{d\times dp}$, Σ_p and Σ_{p+1} are symmetric positive definite block Toeplitz matrices with the $(d \times d)$ blocks $\Sigma_1(h)$, $h = 0, 1, ..., p$. Analogously to Meitz *et al.* (forthcoming), I prove (i) by constructing a dp-dimensional Markov chain $\boldsymbol{z}_t = (z_t, ..., z_{t-p+1})$ $(t = 1, 2, ...)$ with the desired properties. Then, I make use of the theory of Markov chains to establish its stationary distribution. To that end, an appropriate transition probability measure and an initial distribution needs to be specified. For the former, assume that the transition probability of $\boldsymbol{z}_t$ is determined by the density function $t_d(z_t; \mu(\boldsymbol{z}_{t-1}), \Omega(\boldsymbol{z}_{t-1}), \nu + dp)$, where $\mu(\boldsymbol{z}_{t-1})$ and $\Omega(\boldsymbol{z}_{t-1})$ are obtained from (5.A.9) and (5.A.10), respectively, by replacing $\boldsymbol{x}_2$ with $\boldsymbol{z}_{t-1}$. Because the distribution of the current observation depends only on the previous one, $\boldsymbol{z}_t$ is a Markov chain on $\mathbb{R}^{dp}$.

Suppose the initial value $\boldsymbol{z}_0$ follows the t-distribution $t_{dp}(\mathbf{1}_p \otimes \mu, \Sigma_p, \nu)$. The properties of t-distribution (given in Appendix 5.A) then imply that if $\boldsymbol{z}_t^+ = (z_t, \boldsymbol{z}_{t-1})$, the density function of $\boldsymbol{z}_1^+$ is given by

$$t_{d(p+1)}(\boldsymbol{z}_1^+; \mathbf{1}_{p+1} \otimes \mu, \Sigma_{p+1}, \nu) = t_d(z_1; \mu(\boldsymbol{z}_0), \Omega(\boldsymbol{z}_0), \nu + dp)t_{dp}(\boldsymbol{z}_0; \mathbf{1}_p \otimes \mu, \Sigma_p, \nu). \tag{5.B.1}$$

Thus, $\boldsymbol{z}_1^+ \sim t_{d(p+1)}(\mathbf{1}_{p+1} \otimes \mu, \Sigma_{p+1}, \nu)$, and from the block Toeplitz structure of Σ_{p+1} it follows that the marginal distribution of $\boldsymbol{z}_1$ is the same as that of $\boldsymbol{z}_0$, i.e., $\boldsymbol{z}_1 \sim t_{dp}(\mathbf{1}_p \otimes \mu, \Sigma_p, \nu)$. Hence, as $\boldsymbol{z}_t$ is a Markov chain, it has a stationary distribution characterized by the density $t_{dp}(\mathbf{1}_p \otimes \mu, \Sigma_p, \nu)$ (Meyn and Tweedie, 2009, pp. 230-231), completing the proof of (i).

Denote by $\mathcal{F}_{t-1}^z$ the σ-algebra generated by the random vectors $\{z_s, s < t\}$. To prove (ii), note that due to the Markov property, $z_t|\mathcal{F}_{t-1}^z \sim t_d(\mu(\boldsymbol{z}_0), \Omega(\boldsymbol{z}_0), \nu + dp)$. Therefore, the conditional expectation and conditional variance of z_t given $\mathcal{F}_{t-1}^z$ can be written as

$$E[z_t|\mathcal{F}_{t-1}^z] = E[z_t|\boldsymbol{z}_{t-1}] = \mu + \Sigma_{1p}\Sigma_p^{-1}(\boldsymbol{z}_{t-1} - \mathbf{1}_p \otimes \mu) = \phi_0 + \boldsymbol{A}_p\boldsymbol{z}_{t-1}, \tag{5.B.2}$$

$$Var[z_t|\mathcal{F}_{t-1}^z] = Var[z_t|\boldsymbol{z}_{t-1}] = \frac{\nu - 2 + (\boldsymbol{z}_{t-1} - \mathbf{1}_p \otimes \mu)'\Sigma_p^{-1}(\boldsymbol{z}_{t-1} - \mathbf{1}_p \otimes \mu)}{\nu - 2 + dp}\Omega, \tag{5.B.3}$$

where $\Omega = \Sigma_1 - \Sigma_{1p}\Sigma_p^{-1}\Sigma_{1p}'$. I denote this conditional variance by $\Omega_t \equiv \Omega(\boldsymbol{z}_{t-1})$, which is positive definite due to the assumptions $\nu > 2$ and that Σ_p and Ω are both positive definite. Define the $(d \times 1)$ random vectors ε_t as

$$\varepsilon_t \equiv \Omega_t^{-1/2}(z_t - \phi_0 - \boldsymbol{A}_p\boldsymbol{z}_{t-1}), \tag{5.B.4}$$

where $\Omega_t^{-1/2}$ is a symmetric square root matrix of Ω_t^{-1}. Conditionally on $\mathcal{F}_{t-1}^z$, ε_t now follow the $t_d(0, I_d, \nu + dp)$ distribution, and therefore the 'VAR(p)-ARCH(p)' representation (5.2.9) is obtained. Because this conditional distribution does not depend on $\mathcal{F}_{t-1}^z$, it follows that the unconditional distribution of ε_t is also $t_d(0, I_d, \nu + dp)$. Hence, ε_t is independent of $\mathcal{F}_{t-1}^z$ (or of $\{z_s, s < t\}$), and as the random vectors $\{\varepsilon_s, s < t\}$ are functions of $\{z_s, s < t\}$, ε_t is also independent of $\{\varepsilon_s, s < t\}$. Thus, the proof of (ii) is completed by concluding that the random vectors ε_t are IID $t_d(0, I_d, \nu + dp)$ distributed.■

5.B.2 Proof of Theorem 5.2

The G-StMVAR process $\boldsymbol{y}_t$ is clearly a Markov chain on $\mathbb{R}^{dp}$. Let $\boldsymbol{y}_0 = (y_0, ..., y_{-p+1})$ be random vector whose distribution is characterized by the density $f(\boldsymbol{y}_0; \boldsymbol{\theta}) = \sum_{m=1}^{M_1} \alpha_m n_{dp}(\boldsymbol{y}_0; \mathbf{1}_p \otimes \mu_m, \Sigma_{m,p}) + \sum_{m=M_1+1}^{M} \alpha_m t_{dp}(\boldsymbol{y}_0; \mathbf{1}_p \otimes \mu_m, \Sigma_{m,p}, \nu_m)$. According to (5.2.3), (5.3.1), (5.3.5), and (5.B.1), the conditional density of y_1 given $\boldsymbol{y}_0$ is

$$\begin{aligned} f(y_1|\boldsymbol{y}_0; \boldsymbol{\theta}) = &\sum_{m=1}^{M_1} \frac{\alpha_m n_{dp}(\boldsymbol{y}_0; \mathbf{1}_p \otimes \mu_m, \Sigma_{m,p})}{f(\boldsymbol{y}_0; \boldsymbol{\theta})} n_d(y_1; \mu_{m,1}(\boldsymbol{y}_0), \Omega_{m,1}) \\ &+ \sum_{m=M_1+1}^{M} \frac{\alpha_m t_{dp}(\boldsymbol{y}_0; \mathbf{1}_p \otimes \mu_m, \Sigma_{m,p}, \nu_m)}{f(\boldsymbol{y}_0; \boldsymbol{\theta})} t_d(y_1; \mu_{m,1}(\boldsymbol{y}_0), \Omega_{m,1}(\boldsymbol{y}_0), \nu_m + dp) \end{aligned} \tag{5.B.5}$$

$$\begin{aligned} = &\sum_{m=1}^{M_1} \frac{\alpha_m}{f(\boldsymbol{y}_0; \boldsymbol{\theta})} n_{d(p+1)}((y_t, \boldsymbol{y}_0); \mathbf{1}_{p+1} \otimes \mu_m, \Sigma_{m,p+1}) \\ &+ \sum_{m=M_1+1}^{M} \frac{\alpha_m}{f(\boldsymbol{y}_0; \boldsymbol{\theta})} t_{d(p+1)}((y_t, \boldsymbol{y}_0); \mathbf{1}_{p+1} \otimes \mu_m, \Sigma_{m,p+1}, \nu_m). \end{aligned} \tag{5.B.6}$$

The random vector $(y_1, \boldsymbol{y}_0)$ therefore has the density

$$f(y_1, \boldsymbol{y}_0) = \sum_{m=1}^{M_1} \alpha_m n_{d(p+1)}((y_1, \boldsymbol{y}_0); \mathbf{1}_{p+1} \otimes \mu_m; \Sigma_{m,p+1}) + \sum_{m=M_1+1}^{M} \alpha_m t_{d(p+1)}((y_1, \boldsymbol{y}_0); \mathbf{1}_{p+1} \otimes \mu_m; \Sigma_{m,p+1}, \nu_m). \tag{5.B.7}$$

Integrating y_{-p+1} out, and using the properties of marginal distributions of a multivariate Gaussian and t-distributions (see Appendix 5.A) together with the block Toeplitz form of $\Sigma_{m,p+1}$, shows that the density of $\boldsymbol{y}_1$ is $f(\boldsymbol{y}_1; \boldsymbol{\theta}) = \sum_{m=1}^{M_1} \alpha_m n_{dp}(\boldsymbol{y}_1; \mathbf{1}_p \otimes \mu_m, \Sigma_{m,p}) + \sum_{m=M_1+1}^{M} \alpha_m t_{dp}(\boldsymbol{y}_1; \mathbf{1}_p \otimes \mu_m, \Sigma_{m,p}, \nu_m)$. Thus, $\boldsymbol{y}_0$ and $\boldsymbol{y}_1$ are identically distributed. As $\{\boldsymbol{y}_t\}_{t=1}^{\infty}$ is a (time-homogeneous) Markov chain, it follows that $\{\boldsymbol{y}_t\}_{t=1}^{\infty}$ has a stationary distribution, say $\pi_{\boldsymbol{y}}(\cdot)$, characterized by the density $f(\cdot; \boldsymbol{\theta}) = \sum_{m=1}^{M_1} \alpha_m n_{dp}(\cdot; \mathbf{1}_p \otimes \mu_m, \Sigma_{m,p}) + \sum_{m=M_1+1}^{M} \alpha_m t_{dp}(\cdot; \mathbf{1}_p \otimes \mu_m, \Sigma_{m,p}, \nu_m)$ (Meyn and Tweedie, 2009, pp. 230-231).

For ergodicity, let $P_{\boldsymbol{y}}(\boldsymbol{y}, \cdot) = \mathbb{P}(\boldsymbol{y}_p \in \cdot | \boldsymbol{y}_0 = \boldsymbol{y})$ signify the p-step transition probability measure of the process $\boldsymbol{y}_t$. Using the pth order Markov property of y_t, it is straightforward to check that $P_{\boldsymbol{y}}(\boldsymbol{y}, \cdot)$ has the density

$$f(\boldsymbol{y}_p | \boldsymbol{y}_0; \boldsymbol{\theta}) = \prod_{t=1}^{p} f(y_t | \boldsymbol{y}_{t-1}; \boldsymbol{\theta}) = \prod_{t=1}^{p} \left(\sum_{m=1}^{M_1} \alpha_m n_d(y_1; \mu_{m,t}(\boldsymbol{y}_{t-1}), \Omega_m) + \sum_{m=M_1+1}^{M} \alpha_m t_d(y_1; \mu_{m,t}(\boldsymbol{y}_{t-1}), \Omega_{m,t}(\boldsymbol{y}_{t-1}), \nu_m + dp) \right). \tag{5.B.8}$$

Clearly, $f(\boldsymbol{y}_p | \boldsymbol{y}_0; \boldsymbol{\theta}) > 0$ for all $\boldsymbol{y}_0 \in \mathbb{R}^{dp}$ and $\boldsymbol{y}_p \in \mathbb{R}^{dp}$, so it can be concluded that $\boldsymbol{y}_t$ is ergodic in the sense of Meyn and Tweedie (2009, Chapter 13) by using arguments identical to those used in the proof of Theorem 1 in Kalliovirta *et al.* (2015).■

5.B.3 Proof of Theorem 5.3

First note that $L_T^{(c)}(\boldsymbol{\theta})$ is continuous and that together with Assumption 5.2 it implies existence of a measurable maximizer $\hat{\boldsymbol{\theta}}_T$. To conclude that $\hat{\boldsymbol{\theta}}_T$ is strongly consistent, it needs to be shown that (see, e.g., Newey and McFadden, 1994, Theorem 2.1 and the discussion on page 2122)

(i) the uniform strong law of large numbers holds for the log-likelihood function; that is,

$$\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left| L_T^{(c)}(\boldsymbol{\theta}) - E[L_T^{(c)}(\boldsymbol{\theta})] \right| \to 0 \quad \text{almost surely as} \quad T \to \infty,$$

(ii) and that the limit of $L_T^{(c)}(\boldsymbol{\theta})$ is uniquely maximized at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

Proof of (i). By Theorem 5.2, the process $\boldsymbol{y}_{t-1} = (y_t, ..., y_{t-p+1})$, and hence also y_t, is stationary and ergodic, and $E[L_T^{(c)}(\boldsymbol{\theta})] = E[l_t(\boldsymbol{\theta})]$. To conclude (i), it therefore suffices to show that $E[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |l_t(\boldsymbol{\theta})|] < \infty$ (see Ranga Rao, 1962). I will do that by taking use of the compactness of the parameter space to derive finite lower and upper bounds for $l_t(\boldsymbol{\theta})$, which is given as

$$l_t(\boldsymbol{\theta}) = \log\left(\sum_{m=1}^{M_1} \alpha_{m,t} n_d(y_t; \mu_{m,t}, \Omega_m) + \sum_{m=M_1+1}^{M} \alpha_{m,t} t_d(y_t; \mu_{m,t}, \Omega_{m,t}, \nu_m + dp)\right). \quad (5.B.9)$$

Determinant of the positive definite conditional covariance matrix Ω_m is a continuous function of the parameters $vech(\Omega_m)$, and hence, compactness of the parameter space implies that the determinant is bounded from below by some constant that is strictly larger than zero and from above by some finite constant. Thus,

$$0 < c_1 \leq \det(\Omega_m)^{-1/2} \leq c_2 < \infty, \quad (5.B.10)$$

for some constants c_1 and c_2. Because Ω_m^{-1} is positive definite and the exponential function is bounded from above by one in the non-positive real axis, we obtain the upper bound

$$n_d(y_t; \mu_{m,t}, \Omega_m) = (2\pi)^{-d/2}\det(\Omega_m)^{-1/2} \exp\left\{-\frac{1}{2}(y_t - \mu_m)'\Omega_m^{-1}(y_t - \mu_m)\right\} \leq (2\pi)^{-d/2} c_2. \quad (5.B.11)$$

Next, I derive an upper bound for the t-distribution densities

$$\begin{aligned} t_d(y_t; \mu_{m,t}, \Omega_{m,t}, \nu_m + dp) =& \frac{\Gamma\left(\frac{\nu_m+(1+p)d}{2}\right)}{\sqrt{\pi^d(\nu_m + dp - 2)^d}\Gamma\left(\frac{\nu_m+dp}{2}\right)} \det(\Omega_{m,t})^{-1/2} \\ & \times \left(1 + \frac{(y_t - \mu_{m,t})'\Omega_{m,t}^{-1}(y_t - \mu_{m,t})}{\nu_m + dp - 2}\right)^{-(\nu_m+d(1+p))/2}. \end{aligned} \quad (5.B.12)$$

Since $\nu_m > 2$ and the parameter space is compact, $2 < c_3 \leq \nu_m \leq c_4 < \infty$ for some constants c_3 and c_4. Because the gamma function is continuous on the positive real axis, it then follows that

$$0 < c_5 \leq \frac{\Gamma\left(\frac{\nu_m+(1+p)d}{2}\right)}{\sqrt{\pi^d(\nu_m + dp - 2)^d}\Gamma\left(\frac{\nu_m+dp}{2}\right)} \leq c_6 \quad (5.B.13)$$

for some finite constants c_5 and c_6.

Using the bounds $2 < c_3 \leq \nu_m \leq c_4 < \infty$ and (5.B.10) together with the fact that $\Sigma_{m,p}^{-1}$ is

positive definite gives

$$\begin{aligned}\det(\Omega_{m,t})^{-1/2} &= \left(\frac{\nu_m - 2 + (\boldsymbol{y}_{t-1} - \mathbf{1}_p \otimes \mu_m)'\Sigma_{m,p}^{-1}(\boldsymbol{y}_{t-1} - \mathbf{1}_p \otimes \mu_m)}{\nu_m - 2 + dp}\right)^{-d/2} \det(\Omega_m)^{-1/2} \\ &\leq \left(\frac{c_3 - 2}{c_4 + dp - 2}\right)^{-d/2} c_2 < \infty.\end{aligned} \tag{5.B.14}$$

For a lower bound, note that $\Sigma_{m,p}^{-1}$ is a continuous function of the parameters and thereby its eigenvalues are as well. It then follows from the compactness of the parameter space that its largest eigenvalue, λ_1^{max}, is bounded from above by some finite constant, say c_7. The compactness of the parameter space also implies that there exist finite constant c_8 such that $\mu_{im} \leq c_8$ for all $i = 1, .., d$ (where μ_{im} is the ith element of μ_m). By using the orthonormal spectral decomposition of $\Sigma_{m,p}^{-1}$, we then obtain

$$\begin{aligned}(\boldsymbol{y}_{t-1} - \mathbf{1}_p \otimes \mu_m)'\Sigma_{m,p}^{-1}(\boldsymbol{y}_{t-1} - \mathbf{1}_p \otimes \mu_m) &\leq \lambda_1^{max}(\boldsymbol{y}_{t-1} - \mathbf{1}_p \otimes \mu_m)'(\boldsymbol{y}_{t-1} - \mathbf{1}_p \otimes \mu_m) \\ &\leq c_7(\boldsymbol{y}_{t-1}'\boldsymbol{y}_{t-1} - 2c_8\boldsymbol{y}_{t-1}'\mathbf{1}_{dp} + dpc_8^2).\end{aligned} \tag{5.B.15}$$

Thus,

$$\det(\Omega_{m,t})^{-1/2} \geq \left(\frac{c_4 - 2 + c_7(\boldsymbol{y}_{t-1}'\boldsymbol{y}_{t-1} - 2c_8\boldsymbol{y}_{t-1}'\mathbf{1}_{dp} + dpc_8^2)}{c_3 - 2 + dp}\right)^{-d/2} c_1. \tag{5.B.16}$$

As $-(\nu_m + (1+p)d)/2 < 0$ and $\Omega_{m,t}^{-1}$ is positive definite, we have that

$$\left(1 + \frac{(y_t - \mu_{m,t})'\Omega_{m,t}^{-1}(y_t - \mu_{m,t})}{\nu_m + dp - 2}\right)^{-(\nu_m+(1+p)d)/2} \leq 1. \tag{5.B.17}$$

Hence, $t_d(y_t; \mu_{m,t}, \Omega_{m,t}, \nu_m + dp) \leq \left(\frac{c_3-2}{c_4+dp-2}\right)^{-d/2} c_2c_6$. It then follows from the identity $\sum_{m=1}^M \alpha_{m,t} = 1$ that

$$l_t(\boldsymbol{\theta}) \leq \log\left(\max\left\{(2\pi)^{-d/2}c_2, \left(\frac{c_3 - 2}{c_4 + dp - 2}\right)^{-d/2} c_2c_6\right\}\right) < \infty. \tag{5.B.18}$$

That is, $l_t(\boldsymbol{\theta})$ is bounded from above by a finite constant.

Next, I proceed by bounding $l_t(\boldsymbol{\theta})$ from below. Since the eigenvalues of Ω_m^{-1} are continuous functions of the parameters bounded by compactness of the parameter space, the largest eigenvalue, λ_2^{max}, is bounded from above by some finite constant, say c_9. Taking use of the orthonormal spectral decomposition of Ω_m^{-1}, we then obtain

$$\begin{aligned}(y_t - \mu_{m,t})'\Omega_m^{-1}(y_t - \mu_{m,t}) &\leq \lambda_2^{max}(y_t - \boldsymbol{A}_{m,p}\boldsymbol{y}_{t-1})'(y_t - \boldsymbol{A}_{m,p}\boldsymbol{y}_{t-1}) \\ &\leq c_9(y_t'y_t - 2y_t'\boldsymbol{A}_{m,p}\boldsymbol{y}_{t-1} + \boldsymbol{y}_{t-1}'\boldsymbol{A}_{m,p}'\boldsymbol{A}_{m,p}\boldsymbol{y}_{t-1}).\end{aligned} \tag{5.B.19}$$

The compactness of the parameter space implies that

$$\boldsymbol{y}_{t-1}'\boldsymbol{A}_{m,p}'\boldsymbol{A}_{m,p}\boldsymbol{y}_{t-1} \leq c_{10}\sum_{i=1}^{dp}\sum_{j=1}^{dp}|\boldsymbol{y}_{j,t-1}\boldsymbol{y}_{i,t-1}| \tag{5.B.20}$$

for some finite constant c_{10}, where $\boldsymbol{y}_{i,t-1}$ is the ith element of $\boldsymbol{y}_{t-1}$. Denoting by $a_{m,i}(k,j)$ the kjth element of the autoregression matrix $A_{m,i}$ and y_{kt} the kth element of y_t, we have

$$y_t'\boldsymbol{A}_{m,p}\boldsymbol{y}_{t-1} = \sum_{k=1}^{d}\sum_{i=1}^{p}\sum_{j=1}^{d}a_{m,i}(k,j)y_{kt}y_{jt-i} \leq \sum_{k=1}^{d}\sum_{i=1}^{p}\sum_{j=1}^{d}c_{11}|y_{kt}y_{jt-i}|, \tag{5.B.21}$$

where c_{11} is a finite constant that bounds the absolute values of the autoregression coefficients from above (which exists due to compactness of the parameter space). Combining the above two bounds with (5.B.19) gives the upper bound

$$(y_t-\mu_{m,t})'\Omega_m^{-1}(y_t-\mu_{m,t}) \leq c_{12}\left(y_t'y_t + \sum_{i=1}^{dp}\sum_{j=1}^{dp}|\boldsymbol{y}_{j,t-1}\boldsymbol{y}_{i,t-1}| + \sum_{k=1}^{d}\sum_{i=1}^{p}\sum_{j=1}^{d}|y_{kt}y_{jt-i}|\right). \tag{5.B.22}$$

where c_{12} is a finite constant.

Using the fact that $\Sigma_{m,p}^{-1}$ is positive definite together with the bounds $2 < c_3 \leq \nu_m \leq c_4 < \infty$ shows that

$$\Omega_{m,t}^{-1} = \frac{\nu_m - 2 + dp}{\nu_m - 2 + (\boldsymbol{y}_{t-1} - \boldsymbol{1}_p \otimes \mu_m)'\Sigma_{m,p}^{-1}(\boldsymbol{y}_{t-1} - \boldsymbol{1}_p \otimes \mu_m)}\Omega_m^{-1} \leq \frac{c_4 - 2 + dp}{c_3 - 2}\Omega_m^{-1} \tag{5.B.23}$$

Using the above inequality together with $2 < c_3 \leq \nu_m$ and (5.B.22) then gives

$$\frac{(y_t-\mu_{m,t})'\Omega_{m,t}^{-1}(y_t-\mu_{m,t})}{v_m + pd - 2} \leq c_{13}\left(y_t'y_t + \sum_{i=1}^{dp}\sum_{j=1}^{dp}|\boldsymbol{y}_{j,t-1}\boldsymbol{y}_{i,t-1}| + \sum_{k=1}^{d}\sum_{i=1}^{p}\sum_{j=1}^{d}|y_{kt}y_{jt-i}|\right), \tag{5.B.24}$$

where $c_{13} = ((c_3-2)(c_3+pd-2))^{-1}(c_4-2+dp)c_{12}$ is a finite constant.

From $\sum_{m=1}^{M}\alpha_{m,t} = 1$, (5.B.10), (5.B.13), (5.B.16), (5.B.22), (5.B.24), and $\nu_m \leq c_4$, we

then obtain a lower bound for $l_t(\boldsymbol{\theta})$ as

$$\begin{aligned} l_t(\boldsymbol{\theta}) \geq \min \Bigg\{ & -\frac{d}{2}\log(2\pi) + \log(c_1) \\ & - \frac{1}{2}c_{12}\left(y_t'y_t + \sum_{i=1}^{dp}\sum_{j=1}^{dp} |\boldsymbol{y}_{j,t-1}\boldsymbol{y}_{i,t-1}| + \sum_{k=1}^{d}\sum_{i=1}^{p}\sum_{j=1}^{d} |y_{kt}y_{jt-i}| \right), \\ & c_{15} - \frac{d}{2}\log(c_4 - 2 + c_7(\boldsymbol{y}_{t-1}'\boldsymbol{y}_{t-1} - 2c_8\boldsymbol{y}_{t-1}'\mathbf{1}_{dp} + dpc_8^2)) \\ & -c_{14}\log\left(1 + c_{13}\left(y_t'y_t + \sum_{i=1}^{dp}\sum_{j=1}^{dp} |\boldsymbol{y}_{j,t-1}\boldsymbol{y}_{i,t-1}| + \sum_{k=1}^{d}\sum_{i=1}^{p}\sum_{j=1}^{d} |y_{kt}y_{jt-i}| \right)\right)\Bigg\}, \end{aligned} \tag{5.B.25}$$

where $c_{14} = (c_4 + (1+p)d)/2$ and $c_{15} = log(c_5) + \log(c_1) + \frac{d}{2}(c_3 - 2 + dp)$. Since y_t is stationary with finite second moments, it holds that

$$\begin{aligned} & E\left[y_t'y_t + \sum_{i=1}^{dp}\sum_{j=1}^{dp} |\boldsymbol{y}_{j,t-1}\boldsymbol{y}_{i,t-1}| + \sum_{k=1}^{d}\sum_{i=1}^{p}\sum_{j=1}^{d} |y_{kt}y_{jt-i}| \right] < \infty \text{ and} \\ & E[\boldsymbol{y}_{t-1}'\boldsymbol{y}_{t-1} - 2c_8\boldsymbol{y}_{t-1}'\mathbf{1}_{dp}] < \infty, \end{aligned} \tag{5.B.26}$$

and thereby we obtain from Jensen's inequality that also

$$\begin{aligned} & E\left[\log\left(1 + c_{13}\left(y_t'y_t + \sum_{i=1}^{dp}\sum_{j=1}^{dp} |\boldsymbol{y}_{j,t-1}\boldsymbol{y}_{i,t-1}| + \sum_{k=1}^{d}\sum_{i=1}^{p}\sum_{j=1}^{d} |y_{kt}y_{jt-i}| \right)\right)\right] < \infty \text{ and} \\ & E[\log(c_4 - 2 + c_7(\boldsymbol{y}_{t-1}'\boldsymbol{y}_{t-1} - 2c_8\boldsymbol{y}_{t-1}'\mathbf{1}_{dp} + dpc_8^2))] < \infty. \end{aligned} \tag{5.B.27}$$

The upper bound in Equation (5.B.18) together with (5.B.25), (5.B.26), and (5.B.27) shows that $E[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |l_t(\boldsymbol{\theta})|] < \infty$. ■

Proof of (ii). To prove that $E[l_t(\boldsymbol{\theta})]$ is uniquely maximized at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, it needs to be shown that $E[l_t(\boldsymbol{\theta})] \leq E[l_t(\boldsymbol{\theta}_0)]$, and that $E[l_t(\boldsymbol{\theta})] = E[l_t(\boldsymbol{\theta}_0)]$ implies

$$\begin{aligned} & \boldsymbol{\vartheta}_m = \boldsymbol{\vartheta}_{\tau(m),0} \text{ and } \alpha_m = \alpha_{\tau(m),0} \text{ when } m = 1, ..., M_1, \text{ and} \\ & (\boldsymbol{\vartheta}_m, \nu_m) = (\boldsymbol{\vartheta}_{\tau(m),0}, \nu_{\tau(m),0}) \text{ and } \alpha_m = \alpha_{\tau(m),0} \text{ when } m = M_1 + 1, ..., M, \end{aligned} \tag{5.B.28}$$

for some permutations $\{\tau_1(1), ..., \tau_1(M_1)\}$ and $\{\tau_2(M_1+1), ..., \tau_2(M)\}$. For notational clarity, I write $\mu_{m,t} = \mu(\boldsymbol{y}; \boldsymbol{\vartheta}_m)$, $\Omega_m = \Omega(\boldsymbol{\vartheta}_m)$, $\Omega_{m,t} = \Omega(\boldsymbol{y}; \boldsymbol{\vartheta}_m, \nu_m)$, and $\alpha_{m,t} = \alpha_m(\boldsymbol{y}; \boldsymbol{\theta})$, making clear their dependence on the parameter value.

The density of $(y_t, \boldsymbol{y}_{t-1})$ can be written as

$$f((y_t, \boldsymbol{y}_{t-1}); \boldsymbol{\theta}_0) = \sum_{n=1}^{M} \alpha_{n,0} d_{n,dp}(\boldsymbol{y}_{t-1}; \mathbf{1}_p \otimes \mu_{n,0}, \Sigma_{n,p,0}, \nu_{n,0}) \times \left(\sum_{m=1}^{M_1} \alpha_m(\boldsymbol{y}; \boldsymbol{\theta}_0) n_d(y_t; \mu(\boldsymbol{y}; \boldsymbol{\vartheta}_{m,0}), \Omega(\boldsymbol{\vartheta}_{m,0})) + \sum_{m=M_1+1}^{M} \alpha_m(\boldsymbol{y}; \boldsymbol{\theta}_0) t_d(y_t; \mu(\boldsymbol{y}; \boldsymbol{\vartheta}_{m,0}), \Omega(\boldsymbol{y}; \boldsymbol{\vartheta}_{m,0}, \nu_{m,0}), \nu_{m,0} + dp) \right), \tag{5.B.29}$$

where $d_{n,dp}(\cdot; \mathbf{1}_p \otimes \mu_{n,0}, \Sigma_{n,p,0}, \nu_{n,0})$ is defined in (5.3.4). By using this together with reasoning based on Kullback-Leibler divergence, arguments analogous to those in Kalliovirta *et al.* (2016, pp. 494-495) can be used to conclude that $E[l_t(\boldsymbol{\theta})] - E[l_t(\boldsymbol{\theta}_0)] \leq 0$, with equality if and only if for almost all $(y, \boldsymbol{y}) \in \mathbb{R}^{d(p+1)}$,

$$\begin{aligned} &\sum_{m=1}^{M_1} \alpha_m(\boldsymbol{y}; \boldsymbol{\theta}) n_d(y_t; \mu(\boldsymbol{y}; \boldsymbol{\vartheta}_m), \Omega(\boldsymbol{\vartheta}_m)) + \\ &\sum_{m=M_1+1}^{M} \alpha_m(\boldsymbol{y}; \boldsymbol{\theta}) t_d(y_t; \mu(\boldsymbol{y}; \boldsymbol{\vartheta}_m), \Omega(\boldsymbol{y}; \boldsymbol{\vartheta}_m, \nu_m), \nu_m + dp) \\ &= \sum_{m=1}^{M_1} \alpha_m(\boldsymbol{y}; \boldsymbol{\theta}_0) n_d(y_t; \mu(\boldsymbol{y}; \boldsymbol{\vartheta}_{m,0}), \Omega(\boldsymbol{\vartheta}_{m,0})) + \\ &\sum_{m=M_1+1}^{M} \alpha_m(\boldsymbol{y}; \boldsymbol{\theta}_0) t_d(y_t; \mu(\boldsymbol{y}; \boldsymbol{\vartheta}_{m,0}), \Omega(\boldsymbol{y}; \boldsymbol{\vartheta}_{m,0}, \nu_{m,0}), \nu_{m,0} + dp). \end{aligned} \tag{5.B.30}$$

For each fixed $\boldsymbol{y}$ at a time, the mixing weights, conditional means, and conditional covariances in (5.B.30) are constants, so the result on identification of finite mixtures of multivariate Gaussian and t-distributions in Holzmann, Munk, and Gneiting (2006, Example 1) can be applied (their parametrization of the t-distribution slightly differs from mine, but identification with their parametrization implies identification with my parametrization). For each fixed $\boldsymbol{y}$, there thus exists a permutations $\{\tau_1(1), ..., \tau_1(M_1)\}$ and $\{\tau_2(M_1+1), ..., \tau_2(M)\}$ (that may depend on $\boldsymbol{y}$) of the index sets $\{1, ..., M_1\}$ and $\{M_1+1, ..., M\}$ such that

$$\alpha_m(\boldsymbol{y}; \boldsymbol{\theta}) = \alpha_{\tau_1(m)}(\boldsymbol{y}; \boldsymbol{\theta}_0),\ \mu(\boldsymbol{y}; \boldsymbol{\vartheta}_m) = \mu(\boldsymbol{y}; \boldsymbol{\vartheta}_{\tau_1(m),0}),\ \text{and } \Omega(\boldsymbol{\vartheta}_m) = \Omega(\boldsymbol{\vartheta}_{\tau_1(m),0}), \tag{5.B.31}$$

for $m = 1, ..., M_1$ and almost all $y \in \mathbb{R}^d$, and

$$\begin{aligned} &\alpha_m(\boldsymbol{y}; \boldsymbol{\theta}) = \alpha_{\tau_2(m)}(\boldsymbol{y}; \boldsymbol{\theta}_0),\ \mu(\boldsymbol{y}; \boldsymbol{\vartheta}_m) = \mu(\boldsymbol{y}; \boldsymbol{\vartheta}_{\tau_2(m),0}), \Omega(\boldsymbol{y}; \boldsymbol{\vartheta}_m) = \Omega(\boldsymbol{y}; \boldsymbol{\vartheta}_{\tau_2(m),0}), \\ &\text{and } \nu_m = \nu_{\tau_2(m),0} \end{aligned} \tag{5.B.32}$$

for $m = M_1 + 1, ..., M$ and almost all $y \in \mathbb{R}^d$. Note that from (5.B.31) we readily obtain $vech(\Omega_m) = vech(\Omega_{\tau_1(m),0})$.

Arguments analogous to those in Kalliovirta *et al.* (2016, p. 495) can then be used to conclude from (5.B.31) and (5.B.32) that $\alpha_m = \alpha_{\tau_1(m),0}, \phi_{m,0} = \phi_{\tau_1(m),0,0}$, and $\boldsymbol{A}_{m,p} = \boldsymbol{A}_{\tau_1(m),p,0}$ for $m = 1, ..., M_1$, and $\alpha_m = \alpha_{\tau_2(m),0}, \phi_{m,0} = \phi_{\tau_2(m),0,0}$, and $\boldsymbol{A}_{m,p} = \boldsymbol{A}_{\tau_2(m),p,0}$ for $m = M_1 + 1, ..., M$. Given these identities and $\nu_m = \nu_{\tau_2(m),0}$, we obtain from $\Omega(\boldsymbol{y}; \boldsymbol{\vartheta}_m) = \Omega(\boldsymbol{y}; \boldsymbol{\vartheta}_{\tau_2(m),0})$ in (5.B.32) that

$$(\boldsymbol{y} - \mathbf{1}_p \otimes \mu_{\tau_2(m),0})' \Sigma_p(\boldsymbol{\vartheta}_m)^{-1} (\boldsymbol{y} - \mathbf{1}_p \otimes \mu_{\tau_2(m),0}) \Omega_m - (\boldsymbol{y} - \mathbf{1}_p \otimes \mu_{\tau_2(m),0})' \Sigma_p(\boldsymbol{\vartheta}_{\tau_2(m),0})^{-1} (\boldsymbol{y} - \mathbf{1}_p \otimes \mu_{\tau_2(m),0}) \Omega_{\tau_2(m),0} = (\nu_{\tau_2(m),0} - 2)(\Omega_{\tau_2(m),0} - \Omega_m). \tag{5.B.33}$$

The condition $\Omega(\boldsymbol{y}; \boldsymbol{\vartheta}_m) = \Omega(\boldsymbol{y}; \boldsymbol{\vartheta}_{\tau_2(m),0})$ implies that Ω_m is proportional to $\Omega_{\tau_2(m),0}$, say $\Omega_m = c(\boldsymbol{\vartheta}^+_{m,\tau_2(m)})\Omega_{\tau_2(m),0}$, where the strictly positive scalar $c(\boldsymbol{\vartheta}^+_{m,\tau_2(m)})$ may depend on the parameter $\boldsymbol{\vartheta}^+_{m,\tau_2(m)} \equiv (\boldsymbol{\vartheta}_m, \boldsymbol{\vartheta}_{\tau_2(m),0}, \nu_{\tau_2(m),0})$. It is then easy to see from the vectorized structure of $\Sigma_p(\cdot)$, given in (5.2.4), that $\Sigma_p(\boldsymbol{\vartheta}_m)^{-1} = c(\boldsymbol{\vartheta}^+_{m,\tau_2(m)})^{-1} \Sigma_p(\boldsymbol{\vartheta}_{\tau_2(m),0})^{-1}$. By using this together with the identity $\Omega_m = c(\boldsymbol{\vartheta}^+_{m,\tau_2(m)})\Omega_{\tau_2(m),0}$, the left side of (5.B.33) reduces to

$$\begin{aligned}
&(\boldsymbol{y} - \mathbf{1}_p \otimes \mu_{\tau_2(m),0})' (c(\boldsymbol{\vartheta}^+_{m,\tau_2(m)}) \Sigma_p(\boldsymbol{\vartheta}_m)^{-1} - \Sigma_p(\boldsymbol{\vartheta}_{\tau_2(m),0})^{-1}) (\boldsymbol{y} - \mathbf{1}_p \otimes \mu_{\tau_2(m),0}) \Omega_{\tau_2(m),0} \\
&= (\boldsymbol{y} - \mathbf{1}_p \otimes \mu_{\tau_2(m),0})' \left(\frac{c(\boldsymbol{\vartheta}^+_{m,\tau_2(m)})}{c(\boldsymbol{\vartheta}^+_{m,\tau_2(m)})} \Sigma_p(\boldsymbol{\vartheta}_{\tau_2(m),0})^{-1} - \Sigma_p(\boldsymbol{\vartheta}_{\tau_2(m),0})^{-1} \right) (\boldsymbol{y} - \mathbf{1}_p \otimes \mu_{\tau_2(m),0}) \\
&\times \Omega_{\tau_2(m),0} = 0.
\end{aligned} \tag{5.B.34}$$

Thereby (5.B.33) reduces to $(\nu_{\tau_2(m),0} - 2)(\Omega_{\tau_2(m),0} - \Omega_m) = 0$, which implies $\Omega_m = \Omega_{\tau_2(m),0}$, as $\nu_{\tau_2(m),0} > 2$. Since Condition (5.5.3) sets a unique ordering for the mixture components, it can be concluded that $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, completing the proof of consistency.

Given consistency and assumptions of the theorem, asymptotic normality of the ML estimator can be concluded using the standard arguments. The required steps can be found, for example, in Kalliovirta *et al.* (2016, proof of Theorem 3). I omit the details for brevity.■

Appendix 5.C Details on the empirical application

5.C.1 Model selection and adequacy of the selected model

I estimated the models based on the exact log-likelihood function.[14] The estimation and other numerical analysis is carried out with the CRAN distributed R package **gmvarkit** Virolainen

[14] I disregarded estimates that incorporated a near-singular error term covariance matrix or only a few observations from one of the regimes, and considered the largest local maximum of the log-likelihood function that incorporates clearly non-singular error term covariance matrices and a reasonable amount of observations from both regimes.

(2018) that accompanies this essay. The R package **gmvarkit** also contains the data studied in the empirical application to facilitate reproduction of my results. For evaluating adequacy of the estimated models, I employ quantile residuals diagnostics in the framework proposed by Kalliovirta and Saikkonen (2010). For a correctly specified G-StMVAR model, the multivariate quantile residuals are asymptotically standard normally distributed and can thereby be used for graphical diagnostics similarly to the conventional Pearson residuals (Kalliovirta and Saikkonen, 2010, Lemma 3).[15] A closed form expression for the quantile residual of the G-StMVAR model is derived in Appendix 5.D. For brevity, I show the diagnostic figures for the selected model only.

I started by estimating one-regime StMVAR models with autoregressive orders $p = 1, ..., 12$ and found that BIC, HQIC, and AIC were all minimized by the order $p = 1$. Graphical quantile residual diagnostics revealed that the StMVAR$(1, 1)$ model is somewhat inadequate particularly in capturing conditional heteroskedasticity of the series and movements of the interest rate variable, whose quantile residuals' time series displays a shift in volatility and marginal distribution significant excess kurtosis. Increasing the autoregressive order to 2 or 3 did not help much. Increasing p from 2 to 3 decreased the log-likelihood, suggesting that the order $p = 3$ might not be suitable for a StMVAR model. The log-likelihoods and values of the information criteria are presented in Table 5.1 for the discussed models.

Therefore, I estimated a two-regime StMVAR model with $p = 1$, i.e., a G-StMVAR($p = 1, M_1 = 0, M_2 = 2$) model. Compared to the one-regime models, particularly the time series plot and marginal distribution of the interest variable's quantile residuals became significantly more reasonable. However, I found that this model has several moderately sized correlation coefficients (CC) at small lags in the autocorrelation function (ACF) of its quantile residuals and squared quantile residuals. To improve the fitness, I increased the autoregressive order to $p = 2$, which decreased many of the moderately sized CCs but increased AIC (see Table 5.1). The AIC is, nevertheless, smaller than for any of the one-regime models (while the one-regime StMVAR $p = 1, 2$ models have smaller BIC and the $p = 1$ model also smaller HQIC). As is discussed next, I find the overall adequacy of this model reasonable.[16]

While my procedure is open for discussion, note that the disregarded estimates are useless for statistical inference, as the number of observations from the more rare regime is very small compared to the number of parameters. Such estimates arise due to the complex endogeneity of the mixing weights that makes the surface of the log-likelihood function extremely complex. Similar puzzle in the estimation is discussed in univariate context in more detail in Chapter 3.

[15] Kalliovirta and Saikkonen (2010) also propose formal diagnostic tests for testing normality, autocorrelation, and conditional heteroskedasticity of the quantile residuals. The tests take into account the uncertainty about the true parameter value and can be calculated based on the observed data or by employing a simulation procedure for better size properties. However, I found these tests quite forgiving without the simulation procedure, while with the simulation procedure using a sample of length 10000, all the tests rejected the adequacy my StMVAR model at all the conventional levels of significance. I therefore rather employ graphical diagnostics to examine to what extend the statistical properties of the quantile residuals plausibly resemble those on an IID standard normal process.

[16] It is possible that superior fitness of the two-regime models is due to the accommodation of regime-switching error covariance matrices or kurtosis and cannot be attributed to the time-varying AR matrices or intercepts. To test

Model	Log-lik	BIC	HQIC	AIC
StMVAR$(1,1)$	-2.486	5.604	5.360	5.197
StMVAR$(2,1)$	-2.447	5.851	5.482	5.235
StMVAR$(3,1)$	-2.471	6.224	5.729	5.398
StMVAR$(1,2)$	-2.211	5.705	5.210	4.878
StMVAR$(2,2)$	-2.138	6.210	5.464	4.964

Table 5.1: The log-likelihoods and values of the information criteria divided by the number of observations for the discussed StMVAR(p, M) models.

Figure 5.3 presents the ACF and crosscorrelation function (CCF) of the quantile residuals of the StMVAR$(2,2)$ model for the first 20 lags. As the figure shows, there is not much autocorrelation in the residuals, but CCs of almost 0.2 in absolute value stick out in the ACF of IPI's quantile residuals at the lag 10, in the CCF of OIL and HCPI at the lag 6, and in the CCF of HCPI and RATE at the lag 15. There are also a moderately sized CCs at small lags in the ACF of the IPI's quantile residuals at the lag 3 and in the RATE's quantile residuals at the lag 2. These CCs are not, however, very large, and as 316 CCs are presented, some of them are expected to be moderate for a correctly specified model. Therefore, my model appears to capture the autocorrelation structure of the series reasonably well, although some of the CCs are slightly larger than what one would expect for an IID process.[17]

whether this is the case, I estimated two constrained StMVAR$(2,2)$ models. In the first one, I constrained the AR matrices to identical in both regimes, whereas in the second one, I constrained AR matrices and intercepts to be identical in both regimes. Because these models are nested to the unconstrained StMVAR model, the constraints can be tested with a likelihood ratio test (assuming the validity of the unverified assumption made in Theorem 5.3). The former type constraints obtained the p-value 0.022 and the latter type constraints the p-value $2 \cdot 10^{-4}$. I then repeated the exercise for the StMVAR$(1,2)$ model, which minimized AIC, and obtained the p-values 0.003 and $2 \cdot 10^{-7}$ for the constraints, respectively. As the constraints are rejected at the 5% level of significance or less, it seems plausible that AR matrices and intercepts vary in time.

[17] It does not show up in the figure but there is a large negative CC at the lag 24 in autocorrelation function the IPI's quantile residuals. I suspect that this might be related to the employed detrending method (Hamilton, 2018), where the cyclical component of IPI_t at the time $t+h$ is defined as the OLS residual from regressing IPI_{t+h} on a constant and $IPI_t, ..., IPI_{t-s+1}$, where I used $h = 24$ and $s = 12$. So I experimented with the univariate log industrial production series and detrended it with the filter using $h = 5, 6, 7, 10, 12, 18, 19, 24$ and $s = 12$. First, I examined the partial autocorrelation functions (PACF) of the cyclical component and found that with each h, there is a large or moderate positive partial autocorrelation coefficient (PACC) at the lags $h+1$ and $2h+1$ with the latter one being smaller. Except that for $h = 24$ both of the PACCs were quite small, however, and for $h = 19$ the PACC at the lag $2h + 1$ was relatively small. Then, I fitted one-regime GMAR and StMAR models (Kalliovirta *et al.*, 2015, Meitz *et al.*, forthcoming) to the cyclical component using the autoregressive orders $p = 2, 3, 11, 12$ and examined the autocorrelation functions of the quantile residuals (which equal to the Pearson residuals in the linear Gaussian case, see Kalliovirta, 2012, for details). In each of the cases, there was a large negative CC in the quantile residuals' ACF at the lag h when $p \leq h$ but not when $p > h$ (in which case there was often a large negative CC at some lag larger than p). Hence, the large negative CC seems to be related the detrending method. Nonetheless, since accommodating the large negative CC with the autoregressive order $p = 25$ seems excessive even for a linear VAR (given my sample of 276 observations), I will only note its existence. The details of this

The ACF and CCF of the squared quantile residual are presented in Figure 5.4 for the first 20 lags. The figure shows that there is a moderately large CC at the first lag in the ACF of the IPI's squared quantile residuals and a slightly larger one (roughly 0.2) at the fourth lag in the ACF of RATE's squared quantile residuals. There is a particularly large CC at the lag 10 in the CCF of HCPI's and IPI's squared quantile residuals, and a somewhat large CC at the lag 16 in the CCF of IPI's and RATE's squared quantile residuals, at the lag 9 in the CCF of OIL's and IPI's squared quantile residuals, and at the lag 10 in the OIL's and HCPI's squared quantile residuals. Nonetheless, the model seems to capture the conditional heteroskedasticity of the series moderately well, as the inadequacies do not seem very severe, with the exception of the single large CC at the lag 10 in the CCF between squared quantile residuals of HCPI and IPI.

The marginal quantile residual time series are presented in the top panels of Figure 5.5. The time series seem reasonable, as the are no apparent shifts in the mean, volatility, or dynamics. The COVID-19 lockdown shows as a large negative (marginal) quantile residual of IPI, but I do not view this as an inadequacy, as the fast drop in the cycle is know to be caused by an exceptionally large exogenous shock, and a correctly specified model should thereby produce a large negative residual. Also, the high inflation rates during the COVID-19 crisis show as consecutive positive (marginal) quantile residuals of HCPI. The normal quantile-quantile plots (the bottom panels of Figure 5.5) show that the marginal distribution of the series appears to be captured relatively well. Overall, I find the adequacy of my model reasonable enough for further analysis.

investigation are not shown for brevity.

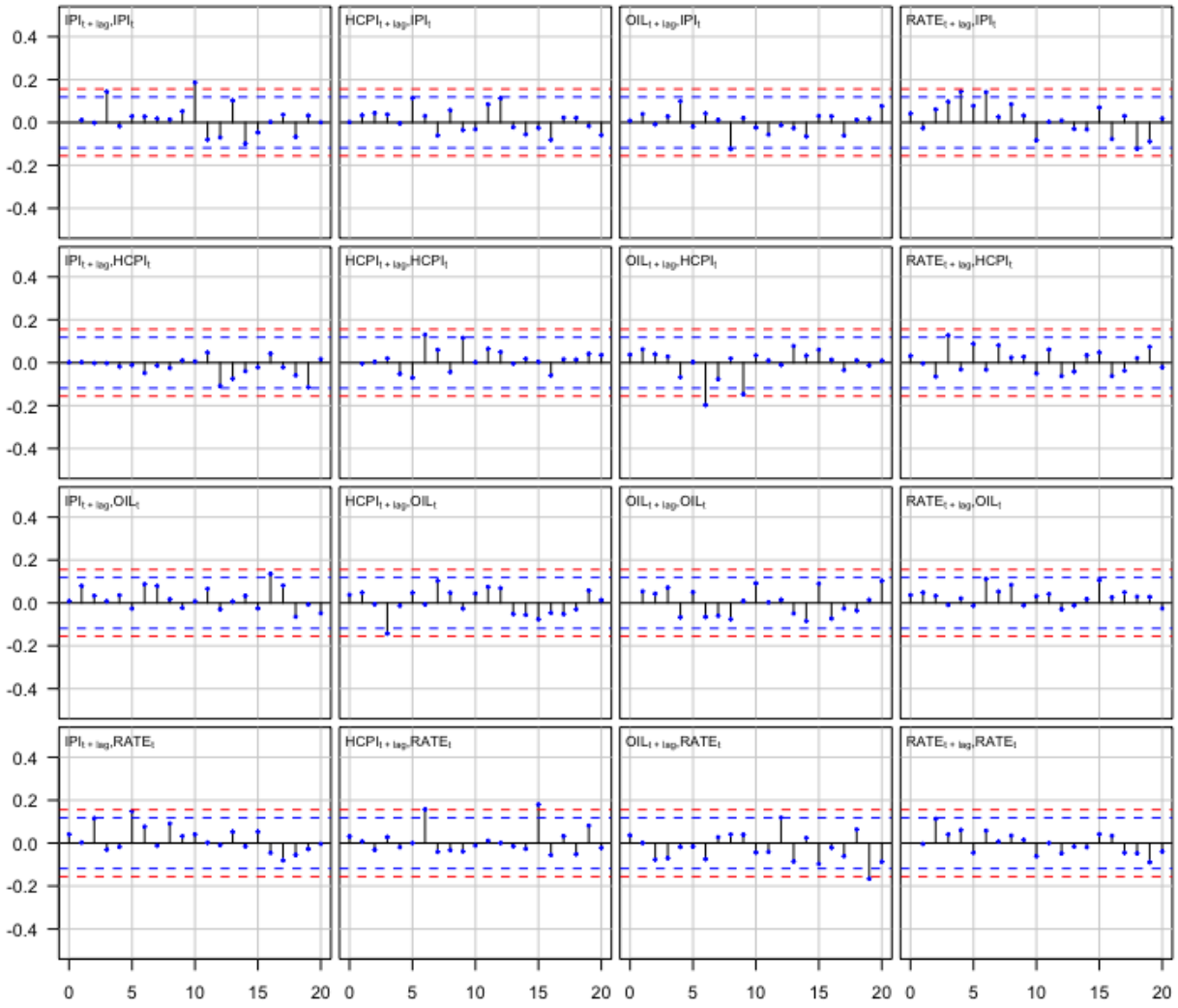

Figure 5.3: Auto- and crosscorrelation functions of the quantile residuals of the fitted StMVAR$(2, 2)$ model for the lags $0, 1, ..., 20$. The lag zero autocorrelation coefficients are omitted, as they are one by convention. The blue dashed lines are the 95% bounds $\pm 1.96/\sqrt{T}$ ($T = 274$ as the first $p = 2$ observations were used as the initial values) for autocorrelations of IID observations, whereas the red dashed lines are the corresponding 99% bounds $\pm 2.58/\sqrt{T}$. These bounds are presented to give an approximate perception on the magnitude of the correlation coefficients.

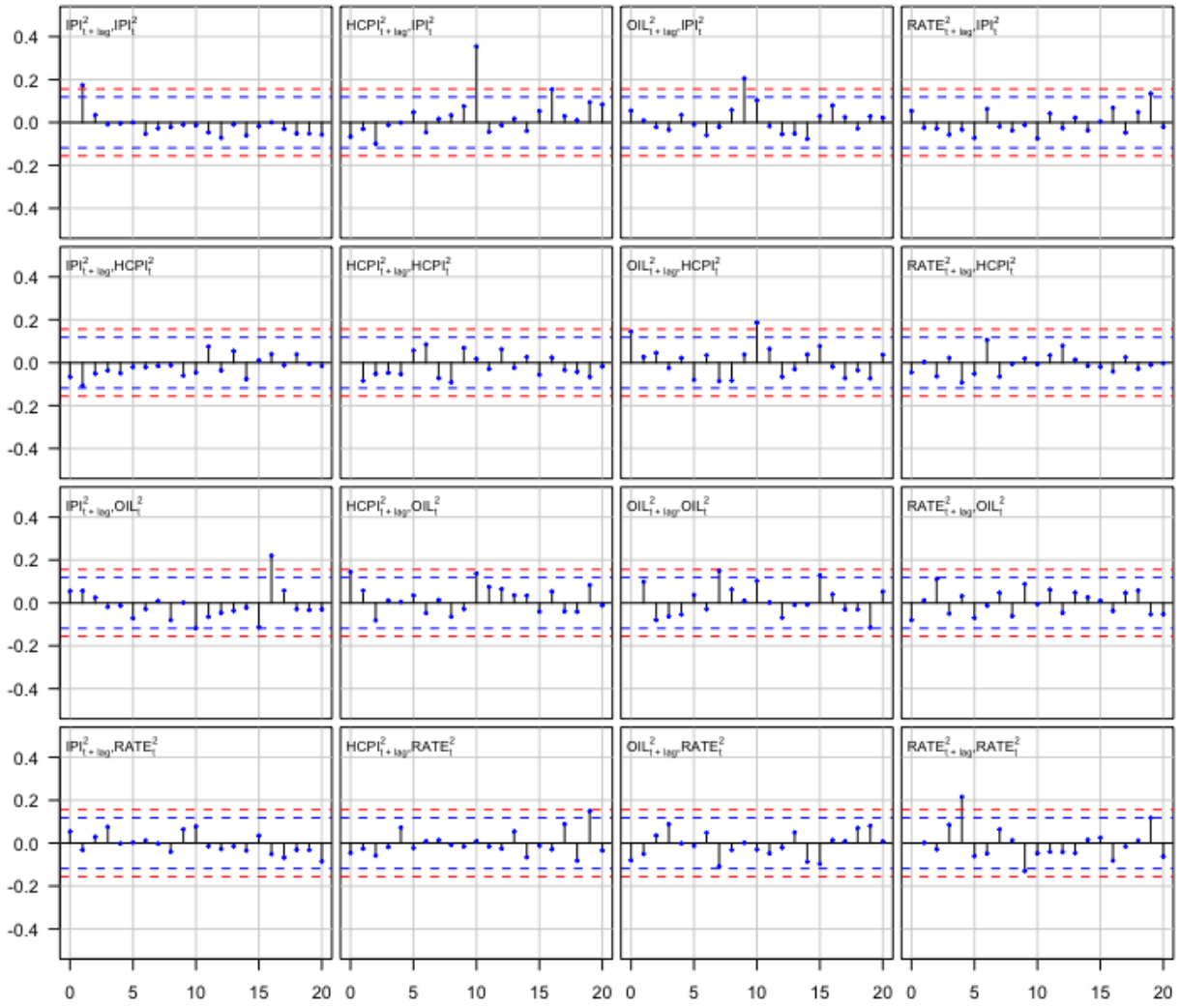

Figure 5.4: Auto- and crosscorrelation functions of the squared quantile residuals of the fitted StMVAR$(2, 2)$ model for the lags $0, 1, ..., 20$. The lag zero autocorrelation coefficients are omitted, as they are one by convention. The blue dashed lines are the 95% bounds $\pm 1.96/\sqrt{T}$ ($T = 274$ as the first $p = 2$ observations were used as the initial values) for autocorrelations of IID observations, whereas the red dashed lines are the corresponding 99% bounds $\pm 2.58/\sqrt{T}$. These bounds are presented to give an approximate perception on the magnitude of the correlation coefficients.

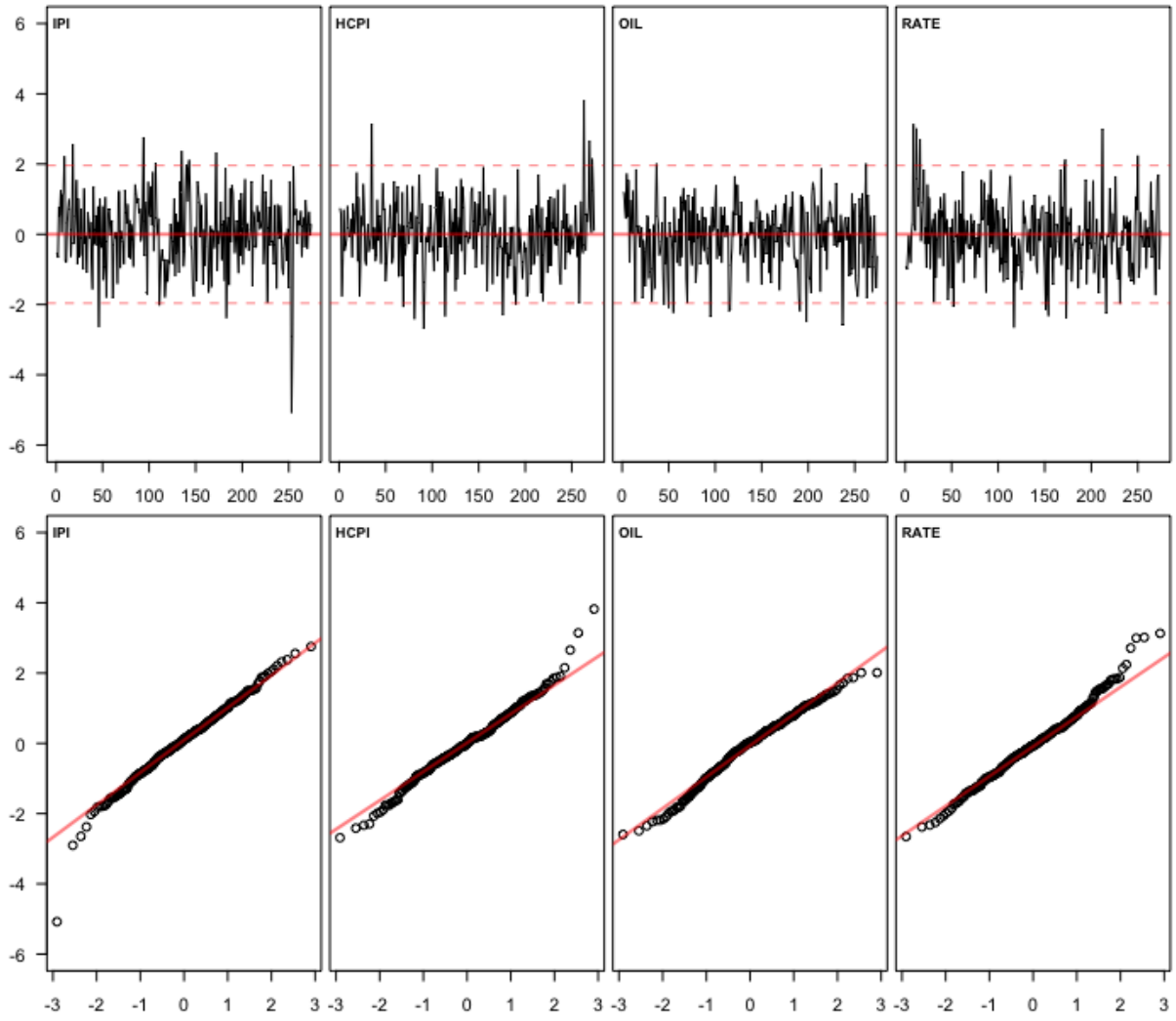

Figure 5.5: Quantile residual time series and normal quantile-quantile-plots of the fitted StMVAR$(2, 2)$ model.

5.C.2 Characteristics of the selected model

Table 5.2 presents the estimates for the mixing weight parameters, degrees of freedom parameters, and the estimated unconditional means and variances of each marginal series. The mixing weight parameters have the interpretation of being the unconditional probabilities for an observation being generated from each regime. For a correctly specified model, they should hence approximately reflect the proportions of observations generated from each regime. The first regime has a mixing weight parameter estimate 0.77, and it covers approximately 63% of the series (approximated as the mean of the estimated mixing weights), whereas the second regime has the implied mixing weight parameter estimate 0.23 and it covers approximately 37% of the

series. Therefore, they are somewhat disproportionate, but seem reasonable enough not to distort the generalized impulse response functions too much.

			IPI		HCPI		OIL		RATE	
	$\hat{\alpha}_m$	$\hat{\nu}_m$	$\hat{\mu}_{m,1}$	$\hat{\sigma}^2_{m,1}$	$\hat{\mu}_{m,2}$	$\hat{\sigma}^2_{m,2}$	$\hat{\mu}_{m,3}$	$\hat{\sigma}^2_{m,3}$	$\hat{\mu}_{m,4}$	$\hat{\sigma}^2_{m,4}$
Regime 1	0.77	3.40	-2.89	111.22	0.12	0.08	0.02	1.79	0.24	78.96
Regime 2	0.23	12.89	2.79	10.25	0.19	0.02	0.33	0.77	2.71	0.47

Table 5.2: Mixing weight parameter estimates ($\hat{\alpha}_m$), degrees of freedom parameter estimates ($\hat{\nu}_m$), and marginal stationary means ($\hat{\mu}_{m,i}$) and variances ($\hat{\sigma}^2_{m,i}$) of the component series implied by the fitted StMVAR$(2, 2)$ model for each of the regimes.

The degrees of freedom parameter estimates in Table 5.2 show that the first regime has fatter tailed distribution than the second one. The first regime also has negative and volatile long-run output gap, while the second regime has positive and less volatile long-run output gap. Long-run inflation, is low in the first regime (roughly 1.4% yearly), whereas it is moderate in the second regime (roughly 2.3% yearly). Also oil price inflation is relatively low in first regime, whereas it is high in the second one.[18] The interest rate variable has low mean in first regime, but the variance is high, which reflects wandering movements of the shadow rate (Wu and Xia, 2016) after the early 2010's recession. In the second regime, the interest rate variable has moderate mean and low variance. According to the unconditional variances of the observable variables in Table 5.2, the first regime appears overall more volatile than the second.

According to the estimated mixing weights presented in Figure 5.1, the first regime (blue) mainly prevails after the collapse of Lehman Brothers in the Financial crisis in September 2008. The first regime also obtains large mixing weights during and before the early 2000's recession, however. The second regime (red) dominates when the first one does not; that is, mainly before the Financial crisis, but excluding the aforementioned periods when the first regime obtains large mixing weights. After the Financial crisis, the second regime's mixing weights stay close to zero, excluding a short period before the early 2010's recession. Because the first regime is characterized by negative output gap and it mainly prevails after the Financial crisis, I refer to it as the low-growth post-Financial crisis regime. Accordingly, because the second regime is characterized by positive output gap and it mainly prevails before the Financial crisis, I refer to it as the high-growth pre-Financial crisis regime.

Appendix 5.D Quantile residual of the G-StMVAR model

This Appendix derives a closed form formula for the quantile residual of the G-StMVAR model that was employed in the model diagnostics in Appendix 5.C.1.

[18]The log-difference of oil price was multiplied by 10 and not 100 for numerical reasons, so the unconditional means should be multiplied 10 to obtain estimates for the (approximate) monthly long-run oil price inflation in percentage units.

5.D.1 The definition of quantile residual

Denote by y_t, $t = 1, 2, ...$, the time series of interest and $\mathcal{F}_{t-1}$ the σ-algebra generated by the random variables or vectors $\{y_{t-j}, j > 0\}$. Moreover, let $\boldsymbol{\theta}$ denote the relevant parameter vector. Kalliovirta (2012) defines univariate quantile residuals as

$$R_{t,\boldsymbol{\theta}} = \Phi^{-1}(F(y_t; \boldsymbol{\theta} \mid \mathcal{F}_{t-1})), \tag{5.D.1}$$

where $\Phi(\cdot)^{-1}$ is the standard normal quantile function and $F(\cdot \mid \mathcal{F}_{t-1})$ is the conditional distribution function of the considered model.

Kalliovirta and Saikkonen (2010) define multivariate quantile residuals analogously to the univariate ones but by taking into account the dependence of the component time series from each other. Denote $\mathcal{A}_{j-1} = \sigma(y_{1,t}, ..., y_{j-1,t})$ and by $f(\cdot|\sigma(\mathcal{F}_{t-1}, \mathcal{A}_{j-1})) = f_{j-1,t-1}(\cdot)$ the conditional density function conditional on the σ-algebra $\sigma(\mathcal{F}_{t-1}, \mathcal{A}_{j-1})$

The conditional density function of the random vector y_t can be expressed in a product form by conditioning to the components y_t in addition to the history $\mathcal{F}_{t-1}$ as

$$f(y_t; \theta|\mathcal{F}_{t-1}) = \prod_{j=1}^{d} f_{j-1,t-1}(y_{j,t}; \boldsymbol{\theta}), \tag{5.D.2}$$

where $y_{j,t}$ is the jth component of y_t and $f_{0,t-1}(y_{1,t}; \boldsymbol{\theta}) = f_{t-1}(y_{1,t}; \boldsymbol{\theta})$ is the marginal conditional density function of $y_{1,t}$ conditional on $\mathcal{F}_{t-1}$.

The conditional distribution functions corresponding to the density functions $f_{j-1,t-1}(\cdot; \boldsymbol{\theta})$ in (5.D.2) are of the form

$$F_{j-1,t-1}(y_{j,t}; \boldsymbol{\theta}) = \int_{-\infty}^{y_{j,t}} f_{j-1,t-1}(u; \boldsymbol{\theta}) du. \tag{5.D.3}$$

The multivariate quantile residuals are then defined as

$$R_{t,\boldsymbol{\theta}} = \begin{bmatrix} R_{1t,\boldsymbol{\theta}} \\ R_{2t,\boldsymbol{\theta}} \\ \vdots \\ R_{dt,\boldsymbol{\theta}} \end{bmatrix} = \begin{bmatrix} \Phi^{-1}(F_{0,t-1}(y_{1,t}; \boldsymbol{\theta})) \\ \Phi^{-1}(F_{1,t-1}(y_{2,t}; \boldsymbol{\theta})) \\ \vdots \\ \Phi^{-1}(F_{d-1,t-1}(y_{d,t}; \boldsymbol{\theta})) \end{bmatrix}, \tag{5.D.4}$$

and its empirical counterpart, $r_{t,\hat{\boldsymbol{\theta}}}$, is obtained by replacing the parameter $\boldsymbol{\theta}$ with its maximum likelihood (ML) estimate $\hat{\boldsymbol{\theta}}$.

5.D.2 Quantile residual of the G-StMVAR model

The conditional density function of the d-dimensional G-StMVAR process y_t conditional on $\mathcal{F}_{t-1}$ is

$$f_{t-1}(y_t; \boldsymbol{\theta}) = \sum_{m=1}^{M_1} \alpha_{m,t} n_d(y_t; \mu_{m,t}, \Omega_m) + \sum_{m=M_1+1}^{M} \alpha_{m,t} t_d(y_t; \mu_{m,t}, \Omega_{m,t}, \nu_m + dp), \tag{5.D.5}$$

where $n_d(\cdot; \mu_{m,t}, \Omega_m, \nu_m + dp)$ is the density function of d-dimensional normal distribution with mean $\mu_{m,t}$ and covariance matrix Ω_m; and $t_d(\cdot; \mu_{m,t}, \Omega_m, \nu_m + dp)$ is the density function of d-dimensional t-distribution with mean $\mu_{m,t}$, covariance matrix $\Omega_{m,t}$, and $\nu_m + dp$ degrees of freedom.

Denote $y_t^{(k)} = (y_{1,t}, ..., y_{k,t})$ $(k \times 1)$, $k \leq d$, $\mu_{m,t}^{(k)} = (\mu_{1,m,,t}, ..., \mu_{k,m,t})$ $(k \times 1)$, $k \leq d$, and by $\Omega_{m,t}^{(k)}$ $(\Omega_m^{(k)})$ the upper left $(k \times k)$ block matrix of $\Omega_{m,t}$ (Ω_m). Then, the properties of the marginal distributions of multivariate Gaussian and t-distributions (see Appendix 5.A) show that conditional on $\mathcal{F}_{t-1}$, the random vectors $y_t^{(j)}$, $j = 1, .., d$, follow the distribution that is a mixture M_1 j-dimensional normal distributions (with means $\mu_{m,t}^{(j)}$ and covariance matrices $\Omega_m^{(j)}$) and $M_2 \equiv M - M_1$ j-dimensional t-distributions (with means $\mu_{m,t}^{(j)}$, covariance matrices $\Omega_{m,t}^{(j)}$, and $\nu_m + dp$ degrees of freedom). The mixing weights $\alpha_{m,t}$ are not affected, as they are $\mathcal{F}_{t-1}$-measurable. Therefore, the marginal density function of $y_t^{(j)}$ is

$$f_{t-1}(y_t^{(j)}; \boldsymbol{\theta}) = \sum_{m=1}^{M_1} \alpha_{m,t} n_j(y_t^{(j)}; \mu_{m,t}^{(j)}, \Omega_m^{(j)}) + \sum_{m=M_1+1}^{M} \alpha_{m,t} t_j(y_t^{(j)}; \mu_{m,t}^{(j)}, \Omega_{m,t}^{(j)}, \nu_m + dp), \quad \text{(5.D.6)}$$

The conditional density function $f_{0,t-1}(y_{1,t}; \boldsymbol{\theta})$ in (5.D.2) is obtained from (5.D.6) by choosing $j = 1$. For $j = 2, ..., d$, the conditional density functions $f_{j-1,t-1}(y_{j,t}; \boldsymbol{\theta})$ are obtained by substituting Equation (5.D.6) to the formula of conditional density function:

$$f_{j-1,t-1}\left(y_{j,t}; \boldsymbol{\theta}\right) = \frac{f_{t-1}(y_t^{(j)}; \boldsymbol{\theta})}{f_{t-1}(y_t^{(j-1)}; \boldsymbol{\theta})} = \frac{\sum_{m=1}^{M_1} \alpha_{m,t} n_j(y_t^{(j)}; \mu_{m,t}^{(j)}, \Omega_m^{(j)}) + \sum_{m=M_1+1}^{M} \alpha_{m,t} t_j(y_t^{(j)}; \mu_{m,t}^{(j)}, \Omega_{m,t}^{(j)}, \nu_m + dp)}{\sum_{n=1}^{M_1} \alpha_{n,t} n_{j-1}(y_t^{(j-1)}; \mu_{n,t}^{(j-1)}, \Omega_n^{(j-1)}) + \sum_{n=M_1+1}^{M} \alpha_{n,t} t_{j-1}(y_t^{(j-1)}; \mu_{n,t}^{(j-1)}, \Omega_n^{(j-1)}, \nu_m + dp)}. \quad \text{(5.D.7)}$$

It follows from the properties of the conditional distributions of multivariate normal distribution that the density of the j-dimensional normal distributions can be expressed as

$$n_j(y_t^{(j)}; \mu_{m,t}^{(j)}, \Omega_m^{(j)}) = n_1(y_{j,t}; \mu_{m,t,j|j-1}, \Omega_{m,j|j-1}) n_{j-1}(y_t^{(j-1)}; \mu_{m,t}^{(j-1)}, \Omega_m^{(j-1)}), \quad \text{(5.D.8)}$$

where $\mu_{m,t,j|j-1}$ and $\Omega_{m,j|j-1}$ are the conditional mean and covariance matrix of $y_{j,t}$ conditional on $\sigma(\mathcal{A}_{j-1}, \mathcal{F}_{t-1})$. Likewise, it follows from the properties of the conditional distributions of multivariate t-distribution that the density of the j-dimensional t-distributions can be expressed as

$$\begin{aligned} t_j(y_t^{(j)}; \mu_{m,t}^{(j)}, \Omega_{m,t}^{(j)}, \nu_m + dp) =& t_1(y_{j,t}; \mu_{m,t,j|j-1}, \Omega_{m,t,j|j-1}, \nu_m + dp + j - 1) \\ & \times t_{j-1}(y_t^{(j-1)}; \mu_{m,t}^{(j-1)}, \Omega_{m,t}^{(j-1)}, \nu_m + dp), \end{aligned} \quad \text{(5.D.9)}$$

where $\mu_{m,t,j|j-1}$ and $\Omega_{m,t,j|j-1}$ are the conditional mean and covariance matrix of $y_{j,t}$ conditional on $\sigma(\mathcal{A}_{j-1}, \mathcal{F}_{t-1})$.

Appendix

By denoting

$$\beta_{m,t,j} \equiv \frac{\alpha_{m,t} n_{j-1}(y_t^{(j-1)}; \mu_{m,t}^{(j-1)}, \Omega_m^{(j-1)})}{\sum_{n=1}^{M_1} \alpha_{n,t} n_{j-1}(y_t^{(j-1)}; \mu_{n,t}^{(j-1)}, \Omega_n^{(j-1)}) + \sum_{n=M_1+1}^{M} \alpha_{n,t} t_{j-1}(y_t^{(j-1)}; \mu_{n,t}^{(j-1)}, \Omega_{n,t}^{(j-1)}, \nu_n + dp)} \tag{5.D.10}$$

for $m = 1, .., M_1$, $j = 2, ..., d$, and

$$\beta_{m,t,j} \equiv \frac{\alpha_{m,t} t_{j-1}(y_t^{(j-1)}; \mu_{m,t}^{(j-1)}, \Omega_{m,t}^{(j-1)}, \nu_m + dp)}{\sum_{n=1}^{M_1} \alpha_{n,t} n_{j-1}(y_t^{(j-1)}; \mu_{n,t}^{(j-1)}, \Omega_n^{(j-1)}) + \sum_{n=M_1+1}^{M} \alpha_{n,t} t_{j-1}(y_t^{(j-1)}; \mu_{n,t}^{(j-1)}, \Omega_{n,t}^{(j-1)}, \nu_n + dp)} \tag{5.D.11}$$

for $m = M_1 + 1, ..., M$, $j = 2, ..., d$, and using the expressions (5.D.8) and (5.D.9), the conditional density function (5.D.7) can be expressed as

$$\begin{aligned} f_{j-1,t-1}(y_{j,t}; \boldsymbol{\theta}) = &\sum_{m=1}^{M_1} \beta_{m,t,j} n_1(y_{j,t}; \mu_{m,t,j|j-1}, \Omega_{m,j|j-1}) \\ &+ \sum_{m=M_1+1}^{M} \beta_{m,t,j} t_1(y_{j,t}; \mu_{m,t,j|j-1}, \Omega_{m,t,j|j-1}, \nu_m + dp + j - 1), \quad j = 2, .., d. \end{aligned} \tag{5.D.12}$$

For $m = 1, ..., M_1$, the conditional means $\mu_{m,t,j|j-1}$ and covariance matrices $\Omega_{m,j|j-1}$ are as in (5.A.6) and (5.A.7) when for each $j = 2, ..., d$ and $m = 1, ..., M$, we consider the partition $y_t^{(j)} = (y_t^{(j-1)}, y_{j,t})$, $\mu_{m,t}^{(j)} = (\mu_{m,t}^{(j-1)}, \mu_{j,m,t})$, and

$$\Omega_m^{(j)} = \begin{bmatrix} \Omega_m^{(j-1)} & \Omega_{(j-1),j,m} \\ \Omega'_{(j-1),j,m} & \Omega_m(j,j) \end{bmatrix}, \tag{5.D.13}$$

where $\Omega_m(j,j)$ is the jjth elementh of Ω_m and $\Omega_{(j-1),j,m}$ $((j-1) \times 1)$ consists of the rows $1, ..., j-1$ of the jth column of Ω_m. In particular, we have

$$\mu_{m,t,j|j-1} = \mu_{j,m,t} + \Omega'_{(j-1),j,m}(\Omega_m^{(j-1)})^{-1}(y_t^{(j-1)} - \mu_{m,t}^{(j-1)}), \tag{5.D.14}$$

$$\Omega_{m,j|j-1} = \Omega_m(j,j) - \Omega'_{(j-1),j,m}(\Omega_m^{(j-1)})^{-1}\Omega_{(j-1),j,m}. \tag{5.D.15}$$

For $m = M_1 + 1, .., M$, the conditional means $\mu_{m,t,j|j-1}$ and covariance matrices $\Omega_{m,t,j|j-1}$ are as in (5.A.9) and (5.A.10) when for each $j = 2, ..., d$ and $m = 1, ..., M$, we consider the partition $y_t^{(j)} = (y_t^{(j-1)}, y_{j,t})$, $\mu_{m,t}^{(j)} = (\mu_{m,t}^{(j-1)}, \mu_{j,m,t})$, and

$$\Omega_{m,t}^{(j)} = \begin{bmatrix} \Omega_{m,t}^{(j-1)} & \Omega_{(j-1),j,m,t} \\ \Omega'_{(j-1),j,m,t} & \Omega_{m,t}(j,j) \end{bmatrix}, \tag{5.D.16}$$

where $\Omega_{m,t}(j,j)$ is the jjth elementh of $\Omega_{m,t}$ and $\Omega_{(j-1),j,m,t}$ ($(j-1)\times 1$) consists of the rows $1,...,j-1$ of the jth column of $\Omega_{m,t}$. In particular, taking use of the relation $\Omega_{m,t}=\omega_{m,t}\Omega_m$ (where $\omega_{m,t}$ is scalar), we have

$$\begin{aligned}\mu_{m,t,j|j-1} &= \mu_{j,m,t}+\Omega'_{(j-1),j,m,t}(\Omega_{m,t}^{(j-1)})^{-1}(y_t^{(j-1)}-\mu_{m,t}^{(j-1)}) \\ &= \mu_{j,m,t}+\Omega'_{(j-1),j,m}(\Omega_{m}^{(j-1)})^{-1}(y_t^{(j-1)}-\mu_{m,t}^{(j-1)}),\end{aligned} \tag{5.D.17}$$

and

$$\begin{aligned}\Omega_{m,t,j|j-1} &= \frac{\nu_m+dp+(y_t^{(j-1)}-\mu_{m,t}^{(j-1)})'(\Omega_{m,t}^{(j-1)})^{-1}(y_t^{(j-1)}-\mu_{m,t}^{(j-1)})}{\nu_m+dp+j-3}\tilde{\Omega}_{m,t,j|j-1} \\ &= \frac{\nu_m+dp+\omega_{m,t}^{-1}(y_t^{(j-1)}-\mu_{m,t}^{(j-1)})'(\Omega_{m}^{(j-1)})^{-1}(y_t^{(j-1)}-\mu_{m,t}^{(j-1)})}{\nu_m+dp+j-3}\tilde{\Omega}_{m,t,j|j-1},\end{aligned} \tag{5.D.18}$$

where

$$\begin{aligned}\tilde{\Omega}_{m,t,j|j-1} &\equiv \Omega_{m,t}(j,j)-\Omega'_{(j-1),j,m,t}(\Omega_{m,t}^{(j-1)})^{-1}\Omega_{(j-1),j,m,t} \\ &= \omega_{m,t}(\Omega_m(j,j)-\Omega'_{(j-1),j,m}(\Omega_{m}^{(j-1)})^{-1}\Omega_{(j-1),j,m}).\end{aligned} \tag{5.D.19}$$

It then remains to find expressions for the conditional distribution functions $F_{j-1,t-1}(y_{j,t};\boldsymbol{\theta})$, $j=1,...,d$, in (5.D.4). For notational convenience, I write

$$\begin{aligned}f_{j-1,t-1}(y_{j,t}\boldsymbol{\theta}) = &\sum_{m=1}^{M_1}\beta_{m,t,j}n_1(y_{j,t};\mu_{m,t,j|j-1},\Omega_{m,j|j-1}) \\ &+\sum_{m=M_1+1}^{M}\beta_{m,t,j}t_1(y_{j,t};\mu_{m,t,j|j-1},\Omega_{m,t,j|j-1},\nu_m+dp+j-1)\end{aligned} \tag{5.D.20}$$

for all $j=1,...,d$ by defining $\beta_{m,t,1}\equiv\alpha_{m,t}$, $\mu_{m,t,1|0}\equiv\mu_{m,t}^{(1)}$, $\Omega_{m,1|0}\equiv\Omega_m^{(1)}$, and $\Omega_{m,t,1|0}\equiv\Omega_{m,t}^{(1)}$. For $j=2,...,d$, these quantities are defined in (5.D.10), (5.D.11), (5.D.14), (5.D.15), (5.D.17), and (5.D.18). Then,

$$\begin{aligned}F_{j-1,t-1}(y_{j,t};\boldsymbol{\theta}) = &\sum_{m=1}^{M_1}\beta_{m,t,j}\int_{-\infty}^{y_{j,t}}n_1(u;\mu_{m,t,j|j-1},\Omega_{m,j|j-1})du \\ &+\sum_{m=M_1+1}^{M}\beta_{m,t,j}\int_{-\infty}^{y_{j,t}}t_1(u;\mu_{m,t,j|j-1},\Omega_{m,t,j|j-1},\nu_m+dp+j-1)du,\end{aligned} \tag{5.D.21}$$

where I seek to solve the integrals inside the sums.

Regarding the first sum, for $m = 1, ..., M_1$, it is easy to see that the integrals can be expressed using the standard normal distribution function $\Phi(\cdot)$ as

$$\int_{-\infty}^{y_{j,t}} n_1(u; \mu_{m,t,j|j-1}, \Omega_{m,j|j-1})du = \Phi\left(\frac{y_{j,t} - \mu_{m,t,j|j-1}}{\sqrt{\Omega_{m,j|j-1}}}\right). \tag{5.D.22}$$

Next, consider the second sum, $m = M_1 + 1, ..., M$. By taking use of the symmetry of the t-distribution about its mean, we obtain

$$\begin{aligned}&\int_{-\infty}^{y_{j,t}} t_1(u; \mu_{m,t,j|j-1}, \Omega_{m,t,j|j-1}, \nu_m + dp + j - 1)du \\ &= \frac{1}{2} + \int_{\mu_{m,t,j|j-1}}^{y_{j,t}} t_1(u; \mu_{m,t,j|j-1}, \Omega_{m,t,j|j-1}, \nu_m + dp + j - 1)du.\end{aligned} \tag{5.D.23}$$

By applying the change of variables $\tilde{u}_{m,t,j} = u - \mu_{m,t,j|j-1}$ in the integral, the right side of (5.D.23) can be expressed as

$$\frac{1}{2} + \frac{\Gamma\left(\frac{\nu_m+dp+j}{2}\right)}{\sqrt{\pi(\nu_m + dp + j - 3)}\Gamma\left(\frac{\nu_m+dp+j-1}{2}\right)}\Omega_{m,t,j|j-1}^{-1/2}\int_0^{\tilde{y}_{m,t,j}}\left(1 + \frac{\tilde{u}_{m,t,j}^2}{a_{m,t,j}}\right)^{-b_{m,j}} d\tilde{u}_{m,t,j}, \tag{5.D.24}$$

where $\tilde{y}_{m,t,j} \equiv y_{j,t} - \mu_{m,t,j|j-1}$, $a_{m,t,j} \equiv (\nu_m + dp + j - 3)\Omega_{m,t,j|j-1}$, and $b_{m,j} \equiv (\nu_m + dp + j)/2$.

Then, by applying the change of variables $z_{m,t,j} = \tilde{u}_{m,t,j}^2/\tilde{y}_{m,t,j}$, the integral in (5.D.24) can be expressed as

$$\int_0^{\tilde{y}_{m,t,j}}\left(1 + \frac{\tilde{u}_{m,t,j}^2}{a_{m,t,j}}\right)^{-b_{m,j}} d\tilde{u}_{m,t,j} = \frac{1}{2}\int_0^{\tilde{y}_{m,t,j}}\left(\frac{\tilde{y}_{m,t,j}}{z_{m,t,j}}\right)^{1/2}\left(1 + \frac{z_{m,t,j}\tilde{y}_{m,t,j}}{a_{m,t,j}}\right)^{-b_{m,j}} dz_{m,t,j}. \tag{5.D.25}$$

By applying the third change of variables $x_{m,t,j} = z_{m,t,j}/\tilde{y}_{m,t,j}$ and using the properties of the gamma function, the right side of (5.D.25) can be expressed as

$$\frac{\tilde{y}_{m,t,j}}{2}\int_0^1 x_{m,t,j}^{-1/2}\left(1 - x_{m,t,j}\left(-\frac{\tilde{y}_{m,t,j}^2}{a_{m,t,j}}\right)\right)^{-b_{m,j}} dx_{m,t,j} = \tilde{y}_{m,t,j} \times {}_2\mathrm{F}_1\left(\frac{1}{2}, b_{m,j}, \frac{3}{2}; -\frac{\tilde{y}_{m,t,j}^2}{a_{m,t,j}}\right), \tag{5.D.26}$$

where the hypergeometric function is defined as (Aomoto and Kita, 2011, Section 1.3.1)

$${}_2F_1(a, b, c; x) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(c-a)}\int_0^1 s^{a-1}(1-s)^{c-a-1}(1-sx)^{-b}ds, \tag{5.D.27}$$

when $|x| < 1$, $a > 0$, and $c - a > 0$ (when $a, c \in \mathbb{R}$).

Using the above result, we have

$$\begin{aligned}&\int_{-\infty}^{y_{j,t}} t_1(u; \mu_{m,t,j|j-1}, \Omega_{m,t,j|j-1}, \nu_m + dp + j - 1)du \\ &= \frac{1}{2} + \frac{\Gamma\left(\frac{\nu_m+dp+j}{2}\right)}{\sqrt{\pi(\nu_m + dp + j - 3)}\Gamma\left(\frac{\nu_m+dp+j-1}{2}\right)}\Omega_{m,t,j|j-1}^{-1/2}\tilde{y}_{m,t,j} \times {}_2\mathrm{F}_1\left(\frac{1}{2}, b_{m,j}, \frac{3}{2}; -\frac{\tilde{y}_{m,t,j}^2}{a_{m,t,j}}\right)\end{aligned} \tag{5.D.28}$$

whenever $\left|-\frac{\tilde{y}_{m,t,j}^2}{a_{m,t,j}}\right| < 1$. That is, the closed form expression (5.D.28) exists when

$$|y_{j,t} - \mu_{m,t,j|j-1}| < \sqrt{(v_m + dp + j - 3)\Omega_{m,t,j,|j-1}}. \tag{5.D.29}$$

If this condition does not hold, the quantile residual can be obtained by numerically integrating the conditional density function $t_1(u; \mu_{m,t,j|j-1}, \Omega_{m,t,j|j-1}, \nu_m + dp + j - 1)$.