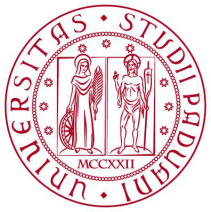


UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Scienze Statistiche
SCUOLA DI DOTTORATO DI RICERCA IN SCIENZE STATISTICHE
CICLO XXVII

ADVANCES IN NON LINEAR MODELS FOR TIME SERIES:
METHODS AND APPLICATIONS TO ECONOMIC AND FINANCIAL DATA

Direttore della Scuola: Ch.mo Prof. MONICA CHIOGNA

Supervisore: Ch.mo Prof. SILVANO BORDIGNON

Co-supervisore: Ch.mo Prof. SIEM JAN KOOPMAN

Dottorando: MARCO BAZZI

31 Gennaio 2015

To my beloved parents

Acknowledgements

At the end of a long journey, it is quite natural to look back to those who have made the trip smooth. First and foremost, Silvano Bordignon. I was an anonymous undergraduate student with a vague idea about time series when I knocked on his door for the first time. After eight years, three dissertations and uncountable afternoons spent in his office, I am deeply thankful to him for having believed in me so strongly.

I had the opportunity to spend two years in Amsterdam, as a visiting student at the Department of Econometrics and Operation Research at the Vrije Universiteit. I feel honored that Siem Jan Koopman accepted to welcome me in his group for such a long period. He taught me what it means being a researcher. I enjoyed the stimulating environment offered by the Department and this thesis benefited from the outstanding competences exhibited by him, Andre Lucas and Francisco Blasques. Siem Jan with his tireless enthusiasm, Andre with his insightful glance at empirical implications and Chico with his sound theoretical knowledge turned out to be the perfect team to work with.

I would like to mention and thank Guido Masarotto. He taught me how beautiful a well written code can be. The director of the doctoral school, Monica Chiogna, and her predecessor, Alessandra Salvan, were able to put students in the best position to conclude successfully the Ph.D Program. Some of us, like me, abroad for a long time but ever proud of our peaceful department. I thank Esther Ruiz and Sebastien Laurent for the helpful comments that they provided when they accepted to read my thesis allowing me to get the *Doctor Europaeus* mention.

These three years would have not been so enjoyable without my colleagues: Maeregu, Vera, Gloria, Valentina, Gianluca, Leonardo and Paola. Whether or not we are the best Ph.D cycle from an academic perspective is still an open question but I am quite sure that few cycles could remember so many brunches, dinner, drinking nights and babies like us. I am also thankful to the colleagues that I met in Amsterdam. Against my Italian habit, our early lunches at 12.00 have proved an extremely valuable time.

Last but not least, I should thank many friends. I hope that they will forgive me if I do not mention them one by one. A special thought to those who gave me a bed whenever I needed it and to those who helped me to discover how much *gezellig* Amsterdam is.

Abstract

When linear models fail to explain the dynamic behavior of economic and financial time series, the researcher has to turn his attention to the nonlinear world. This work has been devoted to develop novel methodological proposals that may be useful in explaining the evolution over time of economic indicators and financial instruments.

In Chapter 2, the well established Markov regime switching framework is extended letting the transition probabilities vary over time according to an observation-driven updating mechanism. An extensive simulation study shows the ability of our new model to track several dynamic patterns in transition probabilities. In the illustration to U.S. Industrial Production growth rate, we show that the model can capture the dynamic features of regime transition probabilities for means and variances.

In Chapter 3, we adapt the new methodology in order to model the electricity spot prices. The Markov regime switching has been extensively used in literature to deal with the spikes that affect the evolution over time of this commodity prices. The non-homogenous occurrence of jumps may be successfully explained by an hidden Markov chain with time-varying transition probabilities that can be also influenced by exogenous variables. The information related to forecasted reserve margin and forecast demand can be easily included in our proposal to improve the model fit as well as to describe the occurrence of spikes.

In Chapter 4, we propose a novel semi-nonparametric model to describe accurately the volatility of financial returns. The finite sample properties are investigated under both correct and incorrect model specification. The latter case suggests that our model is able to recover the functional form of volatility as long as the sample size increases. In empirical relevant settings, features like the asymmetric effect of negative and positive current shocks on future volatility, known as *leverage* effect, as well as the role played by the market condition in influencing the volatility evolution might be captured by our proposal.

Abstract

Quando i modelli lineari falliscono di spiegare il comportamento dinamico delle serie storiche finanziarie ed economiche, l'attenzione del ricercatore deve rivolgersi al mondo dei modelli non lineari. L'intento principale di questo lavoro è quello di sviluppare nuove metodologie che possano essere utili per spiegare l'evoluzione nel tempo sia di indicatori economici sia di strumenti finanziari.

Nel secondo capitolo, i modelli *Markov regime switching*, ampiamente discussi in letteratura, sono estesi per permettere alle probabilità di transizione di evolvere nel tempo in accordo con un meccanismo di aggiornamento guidato dalle osservazioni. Attraverso un approfondito studio di simulazione, siamo abili di mostrare la capacità del nuovo modello di replicare diversi andamenti dinamici nelle probabilità di transizione. Un'analisi empirica è condotta sul tasso di crescita della produzione industriale statunitense.

Nel terzo capitolo, la nuova metodologia è adattata per modellare efficacemente i prezzi del mercato elettrico. I modelli a cambio di regime sono stati usati ampiamente in questo contesto perchè sono in grado di cogliere i picchi che influenzano l'evoluzione temporale dei prezzi energetici. Dal momento che la presenza di questi picchi non è omogenea nel tempo, una catena di Markov latente e con probabilità di transizione dinamiche può essere utilizzata con successo. In particolare, variabili esogene possono essere impiegate per arricchire la dinamica delle probabilità di transizione. Ad esempio, l'informazione relativa al margine di riserva previsto e quella relativa alla domanda prevista possono essere incluse nel nostro modello migliorandone l'abilità descrittiva.

Nel quarto capitolo, proponiamo un nuovo modello semi-nonparametrico per descrivere accuratamente la volatilità dei rendimenti finanziari. Attraverso uno studio di simulazione, abbiamo investigato le proprietà dello stimatore da noi proposto in campioni finiti, sia quando il modello è specificato correttamente sia quando non lo è. In entrambi i casi, la forma funzionale della volatilità è stimata consistentemente anche quando il modello non è correttamente specificato. Da una prospettiva empirica, caratteristiche dei rendimenti finanziari quali l'effetto asimmetrico di shock negativi e positivi sulla futura volatilità e l'influenza delle condizione del mercato nell'evoluzione della volatilità possono essere colte dal nostro modello.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

The dynamic behavior of many economic and financial time series can be described by statistical models letting a subset of parameters changes over time. Cox (1981) have classified time series models with time-varying parameters into parameter-driven models and observation-driven models. In the first case, the driving mechanism for the parameter of interest does not depend on observations but has its own source of uncertainty. Apart from few cases like linear Gaussian state space models (Harvey, 1989) and Markov regime switching models (Hamilton, 1989), the likelihood function does not have a closed-form and its evaluation relies on computationally intensive techniques discussed, for example, in Durbin and Koopman (2012). The Stochastic Volatility model belongs to this class, see Shephard and Andersen (2009) for a detailed overview. In the latter case, the parameter is a deterministic function of lagged dependent variable and exogenous variables. Given the current information, the one step-ahead value of the parameter is perfectly predictable. Typically, the likelihood function is available in closed-form making the estimation procedure relatively easy. The Autoregressive Conditional Heteroskedasticity (ARCH) model of Engle (1982) and the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model Bollerslev (1986) are members of this class broadly used in financial applications.

Recently, a new observation-driven class has been introduced by Creal et al. (2011, 2013) and Harvey (2013), the Generalized Autoregressive Score (GAS) dynamics. In this class, the parameter is driven by the scores of the predictive likelihood function. Due to the generality of this class, score driven models include many well-known time series models in economics and finance and their flexibility results in a forecasting performance competitive with respect to correctly specified non linear non-Gaussian

state space models, as shown in Koopman et al. (2015).

There are situations where linear models are too restrictive and fail to explain the evolution of variables over time. For example, many economic indicators can be dramatically affected by extraordinary episodes such as financial panic, wars and major changes in policies. In these cases, it is reasonable to identify several subsamples, or regimes, for the observations letting the probability law change in each regime. In literature, there are many proposals to handle this nonlinearity as the Threshold Autoregressive (Tong, 1983), the Smooth Transition Autoregressive (Chan and Tong, 1986) and the Markov regime switching models (Hamilton, 1989). In particular, Markov regime switching models became widely popular among the econometric community. They constitute a very flexible class of non linear time series models able to capture many features of empirical time series. Furthermore, they overcome unrealistic limitations of ARMA models (Box and Jenkins, 1970) like normality of the predictive density as well as the marginal density, linearity of conditional expectation in the past observation and homoscedasticity of conditional variance.

A further distinction can be made between parametric and nonparametric models. As we observe only a finite number of data, it is impossible to identify the underlying process without imposing constraints. In a parametric approach, the probability law is chosen within a specified family with elements index by a finite-dimensional parameter. However, the parameter space can be thought as a subset of an infinite dimensional space. This generalization leads to a so called semi-nonparametric model (Chen, 2007). A comprehensive introduction and explanation of such methods for time series analysis is offered by Fan and Yao (2003).

1.2 Main contributions of the thesis

In the standard Markov regime switching proposed by Hamilton (1989), the statistical properties of the latent process which determines the switch across different regimes are summarized by a transition probability matrix with static transition probabilities. Several authors proposed to let the transition probabilities vary over time by using lagged values of the dependent variable (Diebold et al., 1994; Filardo, 1994), state duration (Durland and McCurdy, 1994) or exogenous variables (Perez-Quiros and Timmermann, 2000). In Chapter 2, along the lines of Filardo (1994) and Diebold et al. (1994), we assume an autoregressive dynamics for the transition variables. The dependent variable enters the equation non linearly through the scores of the predictive density, following the GAS updating mechanism proposed by Creal et al. (2011, 2013)

and Harvey (2013). In an extensive simulation study, we compare the goodness-of-fit ability as well as the one-step ahead forecast precision showing that our model outperforms the counterparts. Finally, we illustrate the new methodology by an empirical application to U.S. Industrial Production growth rate.

Markov regime switching models have been widely used to describe the dynamics of electricity spot prices. They are affected by extreme outliers that cannot be explained by linear models. Hence, it is particularly convenient to assume a switching mechanism which alternates quiet periods with jumps. Several specifications of Markov switching models have been proposed in the context of electricity commodities and a complete review can be found in Weron (2006). The homogenous latent Markov Chain implies a constant occurrence of the spikes which is found to be unrealistic in many empirical works. To overcome this limitation of the Hamilton model, the transition probabilities have been allowed to depend on exogenous variable like reserve margin and load (Mount et al., 2006) or temperatures (Huisman, 2008). Alternatively, Janczura and Weron (2010) proposed to estimate periodic transition probabilities directly on price data.

In Chapter 3 we propose to adapt the methodology introduced in Chapter 2 to model electricity spot prices. Through an empirical analysis of UK Automated Power Exchange (APX) spot prices, we find that the GAS methodology can be used to explain price dynamics successfully especially when exogenous information is not available. When the forecasted reserve margin and the forecasted demand are included, we propose a new formulation, namely GAS-exogenous (GASX), which outperforms the most common alternatives.

Starting from the seminal works of Engle (1982) and Bollerslev (1986), an increasing number of models has been proposed to explore the temporal dependency of the conditional second moment in financial time series. Additionally, the empirical evidence that negative and positive shocks affect the volatility asymmetrically, originally documented by Black (1976), has encouraged more sophisticated formulations of functional form of the conditional volatility. The Exponential GARCH model of Nelson (1991), the Quadratic ARCH proposed independently by Engle and Ng (1993) and Sentana (1995), the Threshold GARCH of Zaköian (1994) have been developed to overcome some weakness of the Standard GARCH model remaining in a parametric setting, see Rodríguez and Ruiz (2012) for an empirical comparison. Another line of research has been devoted to semi- and non-parametric formulations. Engle and Gonzalez-Rivera (1991) proposed to leave unspecified the functional form of the error terms and to estimate it non-parametrically by using a splines or a kernel estimator

(Linton, 1993). Alternatively, the relationship between the current volatility and the past returns has been considered as a smooth but unknown function by Pagan and Schwert (1990), Pagan and Hong (1991) and Linton and Mammen (2005) resulting in a non-parametric ARCH approach. A generalization has been offered by Audrino and Bühlmann (2001, 2009) who allowed volatility to be affected by its lagged values, too. In the latter paper, they estimated the unknown functional form by using multivariate B-splines. Despite the greater flexibility that all those techniques offer, the stochastic properties of underlying process and the asymptotic properties of the estimator are difficult to establish.

In Chapter 4, we propose a new semi-nonparametric methodology based on the transformed polynomials function introduced by Blasques (2014). The finite sample properties are studied by an intensive Monte Carlo study. The new formulation leads to a flexibility that is shown to be relevant in empirical settings. Analyzing several individual stock returns, it is found that volatility reacts differently depending on market conditions.

Chapter 2

Time-varying transition probabilities for Markov regime switching models

Abstract

We propose a new Markov switching model with time-varying transitions probabilities. The novelty of our model is that the transition probabilities evolve over time by means of an observation driven model. The innovation of the time-varying probability is generated by the score of the predictive likelihood function. We show how the model dynamics can be readily interpreted. We investigate the performance of the model in a Monte Carlo study and show that the model is successful in estimating a range of different dynamic patterns for unobserved regime switching probabilities. We also illustrate the new methodology in an empirical setting by studying the dynamic mean and variance behaviour of U.S. Industrial Production growth. We find empirical evidence of changes in the regime switching probabilities, with higher persistence for high volatility regimes in the earlier part of the sample, and higher persistence for low volatility regimes in the later part of the sample.

Some key words: Hidden Markov Models; observation driven models; time-varying parameter.

This chapter has been written during my visiting period at Department of Econometrics and Operation Research, at VU University, Amsterdam. I am deeply thankful to Silvano Bordignon for the original discussion and to Siem Jan Koopman, Andre Lucas and Francisco Blasques for their contribution in developing and writing this final version.

2.1 Introduction

Markov regime switching models have been widely applied in economics and finance. Since the seminal application of Hamilton (1989) to U.S. real Gross National Product growth and the well-known NBER business cycle classification, the model has been adopted in numerous other applications. Examples are switches in the level of a time series, switches in the (autoregressive) dynamics of vector time series, switches in volatilities, and switches in the correlation or dependence structure between time series; see Hamilton and Raj (2002) for a partial survey. The key attractive feature of Markov switching models is that the conditional distribution of a time series depends on an underlying latent state or regime, which can take only a finite number of values. The discrete state evolves through time as a discrete Markov chain and we can summarize its statistical properties by a transition probability matrix.

Diebold et al. (1994) and Filardo (1994) argue that the assumption of a constant transition probability matrix for a Markov switching model is too restrictive for many empirical settings. They extend the basic Markov switching model to allow for transition probabilities to vary over time using observable covariates, including strictly exogenous explanatory variables and lagged values of the dependent variable. Although this approach can be useful and effective, it is not always clear what variables or which functional specification we should use for describing the dynamics in the transition probabilities.

Our main focus is to develop a new dynamic approach for the analysis of time variation in transition probabilities of Markov switching models. We let the transition probabilities to vary over time as specific transformations of lagged dependent observations. Hence we adopt an observation driven approach to time-varying parameter models; see Cox (1981) for a detailed discussion. Observation driven models have the key advantage that the likelihood is typically available in closed form by means of the prediction error decomposition. Our main challenge is to specify a suitable functional form to link past observations to future transition probabilities. For this purpose, we use the scores of the predictive likelihood function. Such score driven dynamics have been introduced by Creal et al. (2011, 2013) and Harvey (2013). Score driven models encompass many well-known time series models in economics and finance, including the ARCH model of Engle (1982), the generalized ARCH (GARCH) model of Bollerslev (1986), the exponential GARCH (EGARCH) model of Nelson (1991), the autoregressive conditional duration (ACD) model of Engle and Russell (1998), and many more. Various new and successful applications of score models have appeared in the recent literature. For example, Creal et al. (2011) and Lucas et al. (2014) study

dynamic volatilities and correlations under fat-tails and possible skewness; Harvey and Luati (2014) introduce new models for dynamic changes in levels under fat tails; Creal et al. (2014) investigate score-based mixed measurement dynamic factor models; Oh and Patton (2013) and De Lira Salvatierra and Patton (2013) investigate factor copulas based on score dynamics; and Koopman et al. (2015) show that score driven time series models have a similar forecasting performance as correctly specified nonlinear non-Gaussian state space models over a range of model specifications.

We show that the score function in our Markov switching model has a highly intuitive form. The score combines all relevant innovative information from the separate models associated with the latent states. The updates of the time-varying parameters are therefore based on the probabilities of the states, given all information up to time $t-1$. In our simulation experiments, the new model performs well and succeeds in capturing a range of time-varying patterns for the unobserved transition probabilities. In addition to their intuitive appeal, score driven models also possess information theoretic optimality properties. Blasques et al. (2015) demonstrate that observation driven models with score dynamics improve upon the local Kullback-Leibler divergence of the statistical model from the true unknown data distribution.

We apply our model to study the monthly evolution of U.S. Industrial Production growth from January 1919 to October 2013. This long span of time series observations is known to be characterized by different growth and volatility regimes. Therefore, the assumption of constant transition probabilities between the different regimes may be too restrictive. We uncover three regimes for the mean and two regimes for the variance over the sample period considered. The corresponding transition probabilities turn out to be time-varying. In particular, the high volatility regime appears to be much more persistent in the earlier part of the sample compared to the later part. The converse holds for the low volatility regime. Such changes in the dynamics of the time series are captured in a straightforward way within our model. Moreover, the fit of the new model outperforms the fit of several competing models.

As a final contribution, it is worthwhile mentioning that our model also presents an interesting mix of parameter driven (Markov switching) dynamics with observation driven score dynamics for the corresponding (transition probability) parameters. In particular, it is interesting to see that score driven models can still be adopted when an additional filtering step (for the unobserved discrete states) is required to compute the score of the resulting conditional observation density. This feature of the new dynamic switching model is interesting in its own right. Similar developments for a linear Gaussian state space model have been reported by Creal et al. (2013) and Delle

Monache and Petrella (2014)

The remainder of the paper is organized as follows. In Section 2.2 we briefly discuss the main set-up of the Markov switching model and its residual diagnostics. In Section 2.3 we introduce the new Markov switching model with time-varying transition probabilities based on the score of the predictive likelihood function. In Section 2.4 we discuss some of the statistical properties of the model. In Section 2.5 we report the results of a Monte Carlo study. In Section 2.6 we present the results of our empirical study into the dynamic salient features of U.S. Industrial Production growth. Section 2.7 concludes.

2.2 Markov switching models

Markov switching models are well-known and widely used in applied econometric studies. We refer to the textbook of Frühwirth-Schnatter (2006) for an extensive introduction and discussion. The treatment below establishes the notation and discusses some basic notions of Markov switching models.

Let $\{y_t,\ t = 1, \ldots, T\}$ denote a univariate time series of T observations. We consider the time series $\{y_t,\ t = 1, \ldots, T\}$ as a subset of a stochastic process $\{y_t\}$. The probability distribution of the stochastic process y_t depends on the realizations of a hidden discrete stochastic process z_t. The stochastic process y_t is directly observable, whereas z_t is a latent random variable that is observable only indirectly through its effect on the realizations of y_t. The hidden process $\{z_t\}$ is assumed to be an irreducible and aperiodic Markov chain with finite state space $\{0, \ldots, K-1\}$. Its stochastic properties are sufficiently described by the $K \times K$ transition matrix, Π, where π_{ij} is the $(i+1, j+1)$ element of Π and is equal to the transition probability from state i to state j. All elements of Π are nonnegative and the elements of each row sum to 1, that is

$$\pi_{ij} = \mathbb{P}[z_t = j | z_{t-1} = i], \quad \sum_{j=0}^{K-1} \pi_{ij} = 1, \quad \pi_{ij} \geq 0, \qquad \forall i, j \in \{0, \ldots, K-1\}. \quad (2.1)$$

Let $p(\cdot\,|\theta_i, \psi)$ be a parametric conditional density indexed by parameters $\theta_i \in \Theta$ and $\psi \in \Psi$, where θ_i is a regime dependent parameter and ψ is not regime-specific. We assume that the random variables $y_1, \ldots, y_T$ are conditionally independent given $z_1, \ldots, z_T$, with densities

$$y_t |\, (z_t = i) \sim p(\cdot\,|\theta_i, \psi). \quad (2.2)$$

For the joint stochastic process $\{z_t, y_t\}$, the conditional density of y_t is

$$p(y_t|\psi, I_{t-1}) = \sum_{i=0}^{K-1} p(y_t|\theta_i, \psi)\ \mathbb{P}(z_t = i|\psi, I_{t-1}), \tag{2.3}$$

where $I_{t-1} = \{y_{t-1}, y_{t-2}, \ldots\}$ is the observed information available at time $t-1$. All parameters ψ and $\theta_0, \ldots, \theta_{K-1}$ are unknown and need to be estimated.

The conditional mean of y_t given z_t and I_{t-1} may contain lags of y_t itself. Francq and Roussignol (1998) and Francq and Zakoïan (2001) derive the conditions for the existence of an ergodic and stationary solution for the general class of Markov switching ARMA models. In particular, they show that global stationarity of y_t does not require the stationarity conditions within each regime separately.

As an example, consider the case $K = 2$ for a continuous variable y_t with conditional density

$$p(\,\cdot\,|z_t) = \mathrm{N}\Big(\ (1 - z_t)\mu_0 + z_t\mu_1\ ,\ \sigma^2\ \Big), \tag{2.4}$$

where μ_0 and μ_1 are static regime-dependent means, and σ^2 is the common variance. The latent two-state process $\{z_t\}$ is driven by the transition probability matrix Π

$$\Pi = \begin{pmatrix} \pi_{00} & 1 - \pi_{00} \\ 1 - \pi_{11} & \pi_{11} \end{pmatrix}, \tag{2.5}$$

where the transition probabilities satisfy $0 < \pi_{00},\ \pi_{11} < 1$. We have $\theta_i = \mu_i$ for $i = 0, 1$, and $\psi = (\sigma^2, \pi_{00}, \pi_{11})'$.

To evaluate equation (2.3), we require the quantities $\mathbb{P}(z_t = i|\psi, I_{t-1})$ for all t. We can compute these efficiently using the recursive filtering approach of Hamilton (1989). Assuming we have an expression for the filtered probability $\mathbb{P}(z_{t-1} = i|\psi, I_{t-1})$, we can obtain the predictive probabilities $\mathbb{P}(z_t = i|\psi, I_{t-1})$ as

$$\mathbb{P}(z_t = i|\psi, I_{t-1}) = \sum_{k=0}^{K-1} \pi_{ki} \cdot \mathbb{P}(z_{t-1} = k|\psi, I_{t-1}). \tag{2.6}$$

Hence, the conditional density of y_t given I_{t-1} is given by

$$p(y_t|\psi, I_{t-1}) = \sum_{i=0}^{K-1} \sum_{k=0}^{K-1} p(y_t|\theta_i, \psi) \cdot \pi_{ki} \cdot \mathbb{P}(z_{t-1} = k|\psi, I_{t-1}). \tag{2.7}$$

We can rewrite this expression more compactly in matrix notation. Define ξ_{t-1} as the

$K-$dimensional vector containing the filtered probabilities $\mathbb{P}(z_{t-1} = i|\psi, I_{t-1})$ at time $t-1$ and let η_t be the $K-$dimensional vector collecting the densities $p(y_t|\theta_i, \psi)$ at time t for $i = 0, \ldots, K-1$. It follows that (2.7) reduces to

$$p(y_t|\psi, I_{t-1}) = \xi'_{t-1}\Pi\eta_t. \tag{2.8}$$

The filtered probabilities ξ_t can be updated by the Hamilton recursion

$$\xi_t = \frac{\left(\Pi' \ \xi_{t-1}\right) \odot \eta_t}{\xi'_{t-1}\Pi\eta_t}, \tag{2.9}$$

where $\odot$ denotes the Hadamard element by element product. The filter needs to be started from an appropriate set of initial probabilities $\mathbb{P}(z_0 = i|\psi, I_0)$. The smoothed estimates of the regime probabilities $\mathbb{P}(z_t = i|\psi, I_T)$ can be obtained from the algorithm of Kim (1994). The Hamilton filter in (2.9) is implemented for the evaluation of the the log-likelihood function which is numerically maximized with respect to the parameter vector $(\theta'_0, \ldots, \theta'_{K-1}, \psi')'$ using a quasi-Newton optimization algorithm. To avoid local maxima, we consider different starting values for the numerical optimization.

Diagnostic checking in Markov regime switching models is somewhat more complicated when compared to other time series models because the true residuals depend on the latent variable z_t. Hence the residuals are unobserved. A standard solution is the use of generalized residuals which have been introduced by Gourieroux et al. (1987) in the context of latent variable models. They have been used in the context of Markov regime switching models by Turner et al. (1989), Gray (1996), Maheu and McCurdy (2000), and Kim et al. (2004). Given the filtered regime probabilities $\mathbb{P}(z_t = i|\psi, I_{t-1})$, for $i = 0, \ldots, K-1$, let e_t be the standardized generalized residual

$$e_t = \sum_{i=0}^{K-1} \frac{y_t - \mu_i}{\sigma_i}\mathbb{P}(z_t = i|\psi, I_{t-1}), \qquad t = 1, \ldots, T. \tag{2.10}$$

with conditional mean μ_i and variance σ_i^2 of y_t in regime i. In the context of switching models, Smith (2008) defines the Rosenblatt residual $\tilde{e}_t$, based on Rosenblatt (1952), as

$$\tilde{e}_t = \Phi^{-1}\left(\sum_{i=0}^{K-1} \mathbb{P}(z_t = i|\psi, I_{t-1})\Phi\left(\sigma_i^{-1}\left(y_t - \mu_i\right)\right)\right), \tag{2.11}$$

where Φ denotes the cumulative distribution function of a standard normal with the corresponding inverse function Φ^{-1}. If y_t is generated by the distribution implied by

the Markov switching model, then the Rosenblatt residual $\tilde{e}_t$ is standard normally distributed. Furthermore, Smith (2008) shows in an extensive Monte Carlo study that Ljung-Box tests based on the Rosenblatt transformation have good finite-sample properties for the diagnostic checking of serial correlation in the context of Markov regime switching models.

2.3 Time-varying transition probabilities

In the previous section we considered the transition probability matrix Π to be constant over time. Diebold et al. (1994) and Filardo (1994) argue for having time-varying transition probabilities Π_t. They propose to let the elements of Π_t be functions of past values of the dependent variable y_t and of exogenous variables. The Hamilton filter and Kim smoother can easily be generalized to handle such cases of time-varying Π_t. A key challenge is to specify an appropriate and parsimonious function that links the lagged dependent variables to future transition probabilities. For the specification of the dynamics of Π_t, we adopt the generalized autoregressive score dynamics of Creal et al. (2013); similar dynamic score models have been proposed by Creal et al. (2011) and Harvey (2013). We provide the details of the score driven model for time-varying transition probabilities in the Markov regime switching model. The new dynamic model is parsimonious and the updating mechanism is highly intuitive. Each probability update is based on the weighting of the likelihood information $p(\cdot\,|\theta_i, \psi)$ in (2.2) for each separate regime i.

2.3.1 Dynamics driven by score of predictive likelihood

The parameter vector ψ contains both the transition probabilities as well as other parameters capturing the shape of the conditional distributions $p(y_t|\psi, I_{t-1})$. With a slight abuse of notation, we split ψ into a dynamic parameter f_t that we use to capture the dynamic transition probabilities, and a new static parameter ψ^* that gathers all remaining static parameters in the model, as well as some new static parameters that govern the transition dynamics of f_t. For example, in the two-state example of Section 2.2 we may choose $f_t = (f_{00,t}, f_{11,t})'$ with $f_{00,t} = \text{logit}(\pi_{00,t})$ and $f_{11,t} = \text{logit}(\pi_{11,t})$, where $\text{logit}(\pi_{00,t}) = \log(\pi_{00,t}) - \log(1 - \pi_{00,t})$, and $\log(\cdot)$ refers to the natural logarithm. At the same time, we set $\psi^* = (\sigma^2, \omega, A, B)$, where ω, A, and B are defined below in equation (2.12). For the remainder of this paper, we denote the conditional observation density by $p(y_t|f_t, \psi^*, I_{t-1})$.

In the framework of Creal et al. (2013), the dynamic processes for the parameters are driven by information contained in the score of the conditional observation density $p(y_t|f_t, \psi^*, I_{t-1})$ with respect to f_t. The main challenge in the context of Markov switching models is that the conditional observation density is itself a mixture of densities using the latent mixing variable z_t. Therefore, the shape of our conditional observation density as given by equation (2.3) is somewhat involved.

The updating equation for the time-varying parameter f_t based on the score of the predictive density is given by

$$f_{t+1} = \omega + As_t + Bf_t, \qquad s_t = S_t \cdot \nabla_t, \qquad \nabla_t = \frac{\partial}{\partial f_t} \log p(y_t|f_t, \psi^*, I_{t-1}), \tag{2.12}$$

where ω is a vector of constants, A and B are coefficient matrices, and s_t is the scaled score of the predictive observation density with respect to f_t using the scaling matrix S_t. The updating equation (2.12) can be viewed as a steepest ascent or Newton step for f_t using the log conditional density at time t as its criterion function. An interesting choice for S_t, as recognized by Creal et al. (2013), is the square root matrix of the inverse Fisher information matrix. This particular choice of S_t accounts for the curvature of ∇_t as a function of f_t. Also, for this choice of S_t and under correct model specification, the scaled score function s_t has a unit variance; see also Section 2.4.

2.3.2 Time-varying transition probabilities: the case of 2 states

We first consider the two-state Markov regime switching model, $K = 2$. We let the transition probabilities $\pi_{00,t}$ and $\pi_{11,t}$ vary over time while the two remaining probabilities are set to $\pi_{01,t} = 1 - \pi_{00,t}$ and $\pi_{10,t} = 1 - \pi_{11,t}$ as in (2.5). We specify the transition probabilities for the two regimes as

$$\pi_{ii,t} = \delta_{ii} + (1 - 2\delta_{ii}) \exp(-f_{ii,t})/(1 + \exp(-f_{ii,t})), \qquad i = 0, 1,$$

where $f_{00,t}$ and $f_{11,t}$ are the only two elements in the time-varying parameter vector f_t, and where the two parameters $0 \leq \delta_{ii} \leq 0.5$, for $i = 0, 1$, can be set by the econometrician to limit the range over which $\pi_{ii,t}$ can vary. In the application in Section 2.6, we set $\delta_{ii} = 0$, for $i = 0, 1$, such that $\pi_{ii,t}$ can take any value in the interval $(0, 1)$.

We prefer to work with a parsimonious model specification and therefore we typically have diagonal matrices for A and B in (2.12). The updating equation for the time-varying parameter f_t is given by equation (2.12) where the scaling is set to

$S_t = \mathcal{I}_{t-1}^{-0.5}$ where $\mathcal{I}_{t-1}$ is the 2×2 Fisher information matrix corresponding to the 2×1 score vector ∇_t defined in (2.12). The score vector for the conditional density in (2.7) takes the form

$$\nabla_t = \frac{p(y_t|\theta_0, \psi^*) - p(y_t|\theta_1, \psi^*)}{p(y_t|\psi^*, I_{t-1})} g\Big(f_t, \psi^*, I_{t-1}\Big), \tag{2.13}$$

$$g\Big(f_t, \psi^*, I_{t-1}\Big) = \begin{pmatrix} \mathbb{P}[z_{t-1} = 0|\psi^*, I_{t-1}] \cdot (1 - 2\delta_{00})\pi_{00,t}\,(1 - \pi_{00,t}) \\ -\mathbb{P}[z_{t-1} = 1|\psi^*, I_{t-1}] \cdot (1 - 2\delta_{11})\pi_{11,t}\,(1 - \pi_{11,t}) \end{pmatrix}. \tag{2.14}$$

This expression has a highly intuitive form. The first factor in (2.13) is the difference in the likelihood of y_t given $z_t = 0$ versus $z_t = 1$. The difference is scaled by the total likelihood of the observation given all the static parameters. If the likelihood of y_t given $z_t = 0$ is relatively large compared to that for $z_t = 1$, we expect $f_{00,t}$ to rise and $f_{11,t}$ to decrease. This is precisely what happens in equations (2.13) and (2.14). The magnitudes of the steps are determined by the conditional probabilities of being in regime $z_{t-1} = 0$ or $z_{t-1} = 1$, respectively, at time $t - 1$. The remaining factors $(1 - 2\delta_{ii})\pi_{ii,t}(1 - \pi_{ii,t})$, for $i = 0, 1$, are due to the logit parameterization. In particular, if we are almost certain of being in regime $z_{t-1} = 0$ at time $t - 1$, that is $\mathbb{P}[z_{t-1} = 0|\psi^*, I_{t-1}] \approx 1$, then we take a large step with $f_{00,t}$ but we do not move $f_{11,t}$ by much. Obviously, if we are almost certain of being in regime $z_{t-1} = 0$, y_t can only learn us something about $\pi_{00,t}$. We do not learn much about $\pi_{11,t}$ in this case. The converse holds if we are almost certain of being in regime $z_{t-1} = 1$ at time $t - 1$, in which case we can only learn about $f_{11,t} = \text{logit}(\pi_{11,t})$. The weights for the filtered probabilities in the vector $g(f_t, \psi^*, I_{t-1})$ in (2.13) takes account of this.

The conditional Fisher information matrix based on (2.13) is singular by design. The vector $g(f_t, \psi^*, I_{t-1})$ on the right-hand side of (2.13) is I_{t-1}-measurable and hence the expectation of its outer product remains of rank 1. Therefore, we scale the score by a square root Moore-Penrose pseudo-inverse[2] of the conditional Fisher information matrix. We have

$$s_t = G_t \frac{\left[p(y_t|\theta_0, \psi^*) - p(y_t|\theta_1, \psi^*)\right] / p(y_t|\psi^*, I_{t-1})}{\sqrt{\int_{-\infty}^{\infty} \left[p(y_t|\theta_0, \psi^*) - p(y_t|\theta_1, \psi^*)\right]^2 / p(y_t|\psi^*, I_{t-1})\, \mathrm{d}y_t}}, \tag{2.15}$$

with $G_t = g\Big(f_t, \psi^*, I_{t-1}\Big) \,/\, \Big\|g\Big(f_t, \psi^*, I_{t-1}\Big)\Big\|$, and where the integral has no closed form

[2] If $x \in \mathbb{R}^n$ is a vector, then the Moore-Penrose pseudo-inverse of xx' is given by $\|x\|^{-4}xx'$, and its square root by $\|x\|^{-3}xx'$, as $\|x\|^{-3}xx'\|x\|^{-3}xx' = \|x\|^{-4}xx'$. As $g(f_t, \psi^*, I_{t-1})$ is I_{t-1}-measurable, scaling the score by the square root Moore-Penrose pseudo-inverse of the conditional Fisher information matrix yelds an expression proportional to $\|g(f_t, \psi^*, I_{t-1})\|^{-3} g(f_t, \psi^*, I_{t-1}) g(f_t, \psi^*, I_{t-1})' g(f_t, \psi^*, I_{t-1}) = \|g(f_t, \psi^*, I_{t-1})\|^{-1} g(f_t, \psi^*, I_{t-1})$.

in general and is computed numerically, for example using Gauss-Hermite quadrature methods. An alternative to the analytic Moore-Penrose pseudo-inverse is a numerical pseudo inverse; for example, we could use the Tikhonov regularized matrix inverse as given by $\mathcal{I}_{t-1}^{*1/2} = \left(\lambda \mathrm{I} + (1-\lambda)\mathcal{I}_{t-1}\right)^{-1/2}$, with unit matrix I and fixed scalar $0 < \lambda < 1$. For $\lambda \to 0$ the Tikhonov inverse collapses to the Moore-Penrose pseudo-inverse.

2.3.3 Time-varying transition probabilities: the case of K states

We can easily generalize the two-regime model to K regimes. To enforce that all transition probabilities are non-negative and sum to one (row-wise), we use the multinomial logit specification. Given a set of values for $0 \leq \delta_{ij} \leq 0.5$, we set

$$\pi_{ij,t} = \delta_{ij} + (1 - 2\delta_{ij}) \exp(f_{ij,t}) \left[1 + \sum_{j=1}^{K-1} \exp(f_{ij,t})\right]^{-1}, \tag{2.16}$$
$$\pi_{i,K-1,t} = 1 - \sum_{j=1}^{K-1} \pi_{ij,t}(\delta_{ij}),$$

for $i = 0, \ldots, K-1$ and $j = 0, \ldots, K-2$. The time-varying parameters $f_{ij,t}$, corresponding to the time-varying transition probabilities $\pi_{ij,t}$, are collected in the $K(K-1) \times 1$ vector f_t. The vector f_t is subject to the updating equation (2.12). The ingredients for the scaled score vector in the updating equation (2.12) are given by

$$\begin{array}{rcl} \nabla_t & = & J_t' \nabla_t^{\Pi}, \\ \mathcal{I}_{t-1} & = & \mathbb{E}[J' \nabla_t^{\Pi} \nabla_t^{\Pi\prime} J_t], \end{array} \qquad \nabla_t^{\Pi} = \frac{\partial \log p(y_t|\psi^*, I_{t-1})}{\partial \mathrm{vec}(\Pi)'} = \frac{\eta_t \otimes \xi_{t-1}}{p(y_t|\psi^*, I_{t-1})},$$

where $\otimes$ is the Kronecker product and the elements of $J_t = \partial \mathrm{vec}(\Pi_t)/\partial f_t'$ are given by

$$\frac{\partial \pi_{ij,t}}{\partial f_{i'j',t}} = \begin{cases} (1-2\delta_{ij})\pi_{ij,t}(1-\pi_{ij,t}), & \text{for } i = i' \wedge j = j', \\ -(1-2\delta_{ij})\pi_{ij,t}\ \pi_{ij',t}, & \text{for } i = i' \wedge j \neq j', \\ 0, & \text{otherwise,} \end{cases}$$

for $i, i' = 0, \ldots, K-1$ and $j, j' = 0, \ldots, K-2$.

2.4 Statistical properties

Next we study the stochastic properties of the estimated dynamic transition probabilities in our score driven Markov switching model. In particular, we analyze the

behavior of the estimated time-varying parameter as a function of past observations $y_1, \ldots, y_{t-1}$, parameter vector ψ^*, and initial point $\bar{f}_1$. We write the process as $\{\tilde{f}_t\}$ with $\tilde{f}_t := \tilde{f}_t(\psi^*, \bar{f}_1)$, for $t = 1, \ldots, T$. We follow Blasques et al. (2014b), who use the stationarity and ergodicity (SE) conditions formulated by Bougerol (1993) and Straumann and Mikosch (2006) for general stochastic recurrence equations. Define $\tilde{X}_t = (\tilde{f}_t', \tilde{\xi}_t')'$ as the stacked vector of filtered time-varying parameters $\tilde{f}_t$ and filtered probabilities $\tilde{\xi}_t$ as defined in (2.9). We define $\tilde{\xi}_t = \tilde{\xi}_t(\psi^*, \bar{\xi}_1)$ for some initial point $\bar{\xi}_1$. Our stochastic recurrence equation for the filtered process $\{\tilde{X}_t\}$ now takes the form $\tilde{X}_{t+1} = H(\tilde{X}_t, y_t; \psi^*)$, where

$$\tilde{X}_{t+1} = \begin{bmatrix} \tilde{\xi}_{t+1} \\ \tilde{f}_{t+1} \end{bmatrix} = H(\tilde{X}_t, y_t; \psi^*) := \begin{bmatrix} \alpha(\tilde{\xi}_t, \tilde{f}_t, y_t; \psi^*) \\ \omega + A\, s(\tilde{\xi}_t, \tilde{f}_t, y_t; \psi^*) + B\, \tilde{f}_t \end{bmatrix},$$

where $s(\tilde{\xi}_t, \tilde{f}_t, y_t; \psi^*)$ is the scaled score defined in (2.15) and $\alpha(\tilde{\xi}_t, \tilde{f}_t, y_t; \psi^*)$ is the fraction defined in (2.9) for the recursion $\tilde{\xi}_{t+1} = \alpha(\tilde{f}_t, \tilde{\xi}_t, y_t; \psi^*) := \left(\Pi' \ \tilde{\xi}_{t-1}\right) \odot \eta_t \ / \ \left(\tilde{\xi}_{t-1}' \Pi \eta_t\right)$. The following proposition states sufficient conditions for the filtered process $\{\tilde{X}_t(\psi^*, \bar{X}_1)\}$ with initialization at $\bar{X}_1 := (\bar{\xi}_1', \bar{f}_1')'$ to converge almost surely and exponentially fast (e.a.s) to a unique limit SE process $\{\tilde{X}_t(\psi^*)\}$.[3]

PROPOSITION 1. *Let $\{y_t\}$ be SE, with δ_{ij} in (2.16) satisfying $\delta_{ij} > 0$ for all pairs (i, j), and assume that for every $\psi^* \in \Psi^*$*

(i) $\mathbb{E} \log^+ \left\| H\left(\bar{\xi}_1, \bar{f}_1, y_1, \psi^*\right) \right\| < \infty$;

(ii) $\mathbb{E} \ln \sup_{(f,\xi)} \left\| \dot{H}\left(f, \xi, y_1, \psi^*\right) \right\| < 0$;

where $\dot{H}(X, y_1; \psi^*) = \partial H(X, y_1; \psi^*)/\partial X$ denotes the Jacobian function of H w.r.t. X. Then $\{\tilde{X}_t(\psi^*, \bar{X}_1)\}$ converges e.a.s. to a unique SE process $\{\tilde{X}_t(\psi^*)\}$, for every $\psi^* \in \Psi^*$, that is $\|\tilde{X}_t(\psi^*, \bar{X}_1) - \tilde{X}(\psi^*)\| \overset{e.a.s.}{\to} 0$ as $t \to \infty$ $\forall \psi^* \in \Psi^*$.

Proof. The assumption that $\{y_t\}$ is SE implies that $\{\eta_t\}$ in (2.9) is SE. Together with the continuity of H (and the resulting measurability w.r.t. the Borel σ-algebra), it follows that $\{H(\cdot, \cdot, y_t, \psi^*)\}$ is SE for every ψ^* by Krengel (1985, Proposition 4.3). Condition C1 in Bougerol (1993, Theorem 3.1) is immediately implied by assumption *(i)* for every $\psi^* \in \Psi^*$. Condition C2 in Bougerol (1993, Theorem 3.1) is implied, for every $\psi^* \in \Psi^*$, by condition *(ii)*. As a result, for every $\psi^* \in \Psi^*$, $\{\tilde{X}_t(\psi^*, f_1)\}$ converges almost surely to an SE process $\{\tilde{X}(\psi^*)\}$. Uniqueness and e.a.s. convergence is obtained by Straumann and Mikosch (2006, Theorem 2.8). □

[3] We say that a random sequence $\{\tilde{X}_t\}$ converges e.a.s. to another random sequence $\{\tilde{X}_t^*\}$ if there is a constant $c > 1$ such that $c^t \|\tilde{X}_t - \tilde{X}_t^*\| \overset{a.s.}{\to} 0$; see Straumann and Mikosch (2006) for further details.

Proposition 1 effectively defines those combinations of δ, A and B for which we can ensure that the filtered sequence $\{\tilde{X}_t(\psi^*, f_1)\}$ converges e.a.s. to an SE limit for a given SE data sequence $\{y_t\}$. We emphasize that a finite ω is required for condition *(i)* to hold since ω enters the H function. However, ω plays no role in the contraction condition *(ii)* as it does not influence the Jacobian $\dot{H}$. Numerical experiments (not reported here) suggest that stability is achieved by a wide range of combinations of the parameters δ, A and B, where δ is a vector containing all δ_{ij}. In particular, the set of stable combinations (A,B) becomes larger for higher values of δ. This mechanism is intuitive because the entries of the vector δ bound the elements of the vector ξ and Π away from zero and one. As a result, $\delta > 0$ ensures that the denominator in (2.9) is bounded away from zero and hence the sequence $\{\xi_t\}$ becomes more stable.

Proposition 1 is essential in characterizing the stochastic properties of the filtered time-varying parameters. It does not only allow us to have further insights into the nature of the filtered estimates f_t in the Monte Carlo study of Section 2.5, but it also enables us to interpret the parameter estimates of the score driven time-varying parameter model. The SE nature of the filtered sequence is also an important ingredient in obtaining proofs of consistency and asymptotic normality of the maximum likelihood estimator that rely on an application of laws of large numbers and central limit theorems; see Blasques et al. (2014a).

2.5 Monte Carlo study

2.5.1 Design of the simulation study

To investigate the performance of our estimation procedure for the Markov regime switching model with time-varying transition probabilities, we consider a Monte Carlo study for the two regime model (2.4). The two regimes consist of two normal distributions with common variance $\sigma^2 = 0.5$ and means $\mu_0 = -1$ and $\mu_1 = 1$. We set $\delta_{00} = \delta_{11} = 0$ and consider 6 different forms of time variation for the transition probabilities $\pi_{00,t}$ and $\pi_{11,t}$. The patterns are summarized in Table 2.1 and range from a constant set of transition probabilities, via slow and fast continuously changing transition probabilities, an incidental structural break in the middle of the sample, to a logistic dynamic transition probabilities driven by the lagged variable y_{t-1} in the spirit of Diebold et al. (1994) and Filardo (1994). We investigate the robustness of our estimation procedure in these different settings. It is clear from the outset that for all of the data generation processes considered in this study, our regime switching

Table 2.1: Simulation patterns for $\pi_{00,t}$ and $\pi_{11,t}$

	Model	$\pi_{00,t}$	$\pi_{11,t}$
1.	Constant	0.95	0.85
2.	Slow Sine	$0.5 + 0.45\cos(4\pi t/T)$	$0.5 - 0.45\cos(4\pi t/T)$
3.	Sine	$0.5 + 0.45\cos(8\pi t/T)$	$0.5 - 0.45\cos(8\pi t/T)$
4.	Fast Sine	$0.5 + 0.45\cos(20\pi t/T)$	$0.5 - 0.45\cos(20\pi t/T)$
5.	Break	$0.2\ 1_{\{T<T/2\}} + 0.8\ 1_{\{t\geq T/2\}}$	$0.8\ 1_{\{T<T/2\}} + 0.2\ 1_{\{t\geq T/2\}}$
6.	Autoregressive	$(1+\exp(-1.2y_{t-1}))^{-1}$	$(1+\exp(0.3y_{t-1}))^{-1}$

model with time-varying parameters is misspecified.

In our Monte Carlo study we consider three different sample sizes: $T = 250, 500, 1000$. The number of Monte Carlo replications is set to 1000. A burn-in period of $B = 100$ observations is simulated for all sample sizes to reduce the influence of a starting point. Our main model of interest is the score driven (SD) model with $K = 2$ states, as described in detail in Sections 2.3.1 and 2.3.2. For each data generating process and sample size, we estimate the static parameters ψ^* using the method of maximum likelihood. Given the estimated values for the static parameters, we compute the filtered parameters $\hat{\pi}_{00,t}$ and $\hat{\pi}_{11,t}$ using the updating equations in (2.12). We compare the results to those for the Markov switching model with time invariant transition probabilities (Const.) and to those for the Markov switching model with time-varying transition probabilities (TVP) as in specification 6 of Table 2.1 and similar to the model used by Diebold et al. (1994) and Filardo (1994).

2.5.2 The simulation results

Table 2.2 presents Monte Carlo averages and standard errors of the maximized log-likelihood, the corrected Akaike Information Criterion (AICc), and the Bayesian Information Criterion (BIC) of Schwarz (1978). The AICc is the original AIC of Akaike (1973), but with a stronger finite sample penalty as proposed by Hurvich and Tsai (1991).

We learn from Table 2.2 that the average maximized log-likelihood values are uniformly higher for the models with time-varying transition probabilities (TVP and SD) compared to the model with constant transition probabilities (Const.). This result is not surprising since the time-varying models are a dynamic generalization

of the static model and have more parameters. We notice that the score driven (SD) model outperforms the TVP model in terms of average log-likelihood values. The only exception is given by the Autoregressive DGP with a large sample size of $T = 1000$. More convincing than the results for average likelihood values are the fact that the time varying parameter models produce overall substantially smaller AICc and BIC values than the constant model. The only exception is the data generating process with constant transition probabilities. In all other cases, our SD model with time-varying transition probabilities performs substantially better than the static model. However, this finding does not hold for the TVP specification; the SD model only outperforms the TVP model, used as the data generating process, when the same autoregressive dynamics are adopted.

In Table 2.3, we present Monte Carlo averages of measures of forecast precision for the different models. Following Perez-Quiros and Timmermann (2001), the first two conditional moments for our two-regime Markov-switching model are given by

$$\mu_t = \mathbb{E}[y_t|\mathcal{I}_{t-1}] = \mu_0\mathbb{P}[x_t = 0|\mathcal{I}_{t-1}] + \mu_1\mathbb{P}[x_t = 1|\mathcal{I}_{t-1}],$$
$$\sigma_t^2 = \mathbb{V}[(y_t - \mu_t)^2|\mathcal{I}_{t-1}] = \sigma^2 + \mathbb{P}[x_t = 0|\mathcal{I}_{t-1}](\mu_0 - \mu_t)^2 + \mathbb{P}[x_t = 1|\mathcal{I}_{t-1}](\mu_1 - \mu_t)^2.$$

We use the two moments to compute the one-step-ahead Mean Absolute Forecast Error (MAFE) and the Mean Squared Forecast Error (MSFE) statistics

$$\text{MAFE} = \frac{1}{T}\sum_{t=1}^{T}|y_t - \mu_t|, \qquad \text{MSFE} = \frac{1}{T}\sum_{t=1}^{T}(y_t - \mu_t)^2,$$

together with their standardized counterparts

$$\text{MASFE} = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{y_t - \mu_t}{\sigma_t}\right|, \qquad \text{MSSFE} = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{y_t - \mu_t}{\sigma_t}\right)^2.$$

We find that for all data generating processes with time-varying transition probabilities, the new score driven model has again the best performance. Only for the data generating process with constant transition probabilities does the constant model perform best in terms of the standardized criteria. However, the differences are small in this case. Interestingly, in terms of forecasting performance the score driven model outperfoms the TVP model for the unstandardized forecast error criteria, even if the autoregressive specification is the data generating process. Only for large sample sizes and in terms of the standardized criteria does the TVP specification outperform. Hence even when the model is mis-specified, the score driven specification appears to

Table 2.2: Simulation results I: log-likelihood and goodness-of-fit statistics

We have simulated 1000 time series from each data generation process (DGP) listed in Table 2.1 and for sample sizes $T = 250, 500, 1000$. The static parameters are estimated by the method of maximum likelihood, for the Markov regime switching model with static probabilities (Const.), with autoregressive time-varying transition probabilities (TVP) and with score driven (SD) updating time-varying transition probabilities. In the latter case, the underlying time-varying parameters are updated using equation (2.12). We report the sample averages and standard error for the 1000 simulated series of the maximized log-likelihood value (LogLik), the corrected Akaike Information Criterion (AICc), the Bayesian Information Criterion (BIC).

			LogLik			AICc			BIC		
DGP	T		Const.	TVP	SD	Const.	TVP	SD	Const.	TVP	SD
Constant	250	Mean	-316.913	-315.832	-311.724	644.072	646.127	642.198	652.244	657.449	656.600
		MC SE	0.437	0.439	0.449	0.875	0.879	0.898	0.875	0.879	0.898
	500	Mean	-636.358	-635.461	-632.067	1282.838	1285.150	1282.502	1294.601	1301.559	1303.525
		MC SE	0.628	0.629	0.641	1.256	1.258	1.283	1.256	1.258	1.283
	1000	Mean	-1274.725	-1273.928	-1271.112	2559.510	2561.969	2560.406	2574.799	2583.345	2587.853
		MC SE	0.892	0.894	0.958	1.784	1.787	1.915	1.784	1.787	1.915
Slow Sine	250	Mean	-375.769	-373.917	-361.709	761.783	762.297	742.169	769.955	773.619	756.571
		MC SE	0.341	0.339	0.336	0.681	0.679	0.672	0.681	0.679	0.672
	500	Mean	-746.710	-743.013	-713.123	1503.541	1500.253	1444.614	1515.303	1516.662	1465.638
		MC SE	0.536	0.520	0.507	1.072	1.039	1.013	1.072	1.039	1.013
	1000	Mean	-1510.835	-1503.562	-1429.336	3031.730	3021.238	2876.854	3047.019	3042.614	2904.302
		MC SE	0.708	0.730	0.701	1.416	1.459	1.402	1.416	1.459	1.402
Sine	250	Mean	-375.530	-373.888	-365.224	761.306	762.238	749.197	769.478	773.560	763.599
		MC SE	0.338	0.337	0.342	0.675	0.674	0.684	0.675	0.674	0.684
	500	Mean	-753.633	-750.461	-726.737	1517.387	1515.150	1471.841	1529.149	1531.559	1492.865
		MC SE	0.479	0.475	0.473	0.958	0.951	0.946	0.958	0.951	0.946
	1000	Mean	-1502.200	-1494.190	-1434.108	3014.460	3002.493	2886.397	3029.749	3023.869	2913.844
		MC SE	0.759	0.736	0.710	1.519	1.472	1.420	1.519	1.472	1.420
Fast Sine	250	Mean	-375.584	-374.174	-369.202	761.413	762.811	757.155	769.585	774.133	771.557
		MC SE	0.338	0.335	0.346	0.676	0.669	0.692	0.676	0.669	0.692
	500	Mean	-753.745	-751.370	-737.734	1517.611	1516.968	1493.836	1529.373	1533.377	1514.859
		MC SE	0.476	0.473	0.472	0.952	0.945	0.943	0.952	0.945	0.943
	1000	Mean	-1507.202	-1501.431	-1459.101	3024.464	3016.975	2936.384	3039.753	3038.351	2963.831
		MC SE	0.698	0.695	0.688	1.396	1.390	1.377	1.396	1.390	1.377
Break	250	Mean	-375.158	-373.361	-360.459	760.562	761.185	739.668	768.734	772.507	754.070
		MC SE	0.349	0.353	0.363	0.698	0.707	0.726	0.698	0.707	0.726
	500	Mean	-756.263	-751.225	-721.268	1522.648	1516.677	1460.903	1534.410	1533.087	1481.926
		MC SE	0.517	0.560	0.488	1.035	1.120	0.976	1.035	1.120	0.976
	1000	Mean	-1518.157	-1507.037	-1441.303	3046.375	3028.188	2900.787	3061.664	3049.564	2928.234
		MC SE	0.814	0.943	0.720	1.627	1.886	1.441	1.627	1.886	1.441
Autoregressive	250	Mean	-376.482	-373.334	-369.484	763.210	761.131	757.717	771.382	772.453	772.120
		MC SE	0.305	0.310	0.336	0.610	0.619	0.672	0.610	0.619	0.672
	500	Mean	-755.220	-749.792	-749.116	1520.562	1513.811	1516.599	1532.324	1530.220	1537.622
		MC SE	0.415	0.417	0.445	0.829	0.835	0.891	0.829	0.835	0.891
	1000	Mean	-1512.601	-1502.849	-1507.000	3035.262	3019.812	3032.182	3050.551	3041.188	3059.629
		MC SE	0.587	0.590	0.604	1.175	1.180	1.207	1.175	1.180	1.207

Table 2.3: Simulation results II: forecast precision measures

We have simulated 1000 time series from each data generation process (DGP) listed in Table 2.1 and for sample sizes $T = 250, 500, 1000$. The static parameters are estimated by the method of maximum likelihood, for the Markov regime switching model with static probabilities (Const.), with autoregressive time-varying transition probabilities (TVP) and with score driven (SD) time-varying transition probabilities. In the latter case, the underlying time-varying parameters are updated using equation (2.12). We report the sample averages and standard errors for the 1000 simulated series of mean absolute one-step ahead forecast error (MAFE), mean squared one-step ahead forecast error (MSFE), mean absolute standardized one-step ahead forecast error (MASFE) and mean squared standardized one-step ahead forecast error (MSSFE). The one-step ahead forecast errors are computed within the sample that is used for parameter estimation.

		MAFE			MSFE			MASFE			MSSFE		
DGP	T	Const.	TVP	SD	Const.	TVP	SD	Const.	TVP	SD	Const.	TVP	SD
	250	0.7017	0.6998	0.6885	0.8314	0.8261	0.7991	0.7741	0.7786	0.7844	1.0001	1.0085	1.0177
Constant	500	0.7037	0.7029	0.6987	0.8360	0.8337	0.8236	0.7738	0.7762	0.7795	1.0001	1.0051	1.0105
	1000	0.7044	0.7040	0.7025	0.8375	0.8365	0.8325	0.7736	0.7752	0.7764	1.0003	1.0039	1.0050
	250	0.9449	0.9319	0.8925	1.3259	1.3006	1.2091	0.8266	0.8229	0.8064	1.0136	1.0124	0.9875
Slow Sine	500	0.9225	0.9045	0.8671	1.2863	1.2516	1.1575	0.8205	0.8147	0.7965	1.0165	1.0136	0.9776
	1000	0.9503	0.9279	0.8699	1.3383	1.2963	1.1639	0.8283	0.8210	0.7989	1.0166	1.0135	0.9809
	250	0.9439	0.9329	0.9073	1.3247	1.3026	1.2407	0.8275	0.8244	0.8126	1.0172	1.0156	0.9962
Sine	500	0.9451	0.9283	0.8972	1.3285	1.2967	1.2177	0.8271	0.8221	0.8052	1.0170	1.0157	0.9827
	1000	0.9335	0.9082	0.8741	1.3084	1.2599	1.1708	0.8235	0.8146	0.7973	1.0175	1.0123	0.9761
	250	0.9458	0.9383	0.9217	1.3309	1.3161	1.2774	0.8280	0.8258	0.8183	1.0193	1.0182	1.0075
Fast Sine	500	0.9461	0.9358	0.9200	1.3308	1.3117	1.2678	0.8282	0.8256	0.8150	1.0194	1.0199	0.9968
	1000	0.9433	0.9247	0.9019	1.3269	1.2920	1.2277	0.8269	0.8213	0.8065	1.0194	1.0182	0.9843
	250	0.9482	0.9349	0.8765	1.3283	1.3036	1.1913	0.8270	0.8233	0.8068	1.0104	1.0106	1.0148
Break	500	0.9562	0.9315	0.8715	1.3441	1.2986	1.1824	0.8298	0.8220	0.8022	1.0121	1.0108	1.0084
	1000	0.9622	0.9334	0.8672	1.3570	1.3040	1.1765	0.8314	0.8224	0.7992	1.0131	1.0118	1.0057
	250	0.9607	0.9482	0.9302	1.3485	1.3223	1.2864	0.8264	0.8253	0.8168	0.9963	0.9993	0.9876
Autoregressive	500	0.9632	0.9528	0.9497	1.3532	1.3307	1.3260	0.8267	0.8261	0.8227	0.9959	0.9991	0.9934
	1000	0.9635	0.9544	0.9573	1.3539	1.3340	1.3415	0.8267	0.8266	0.8255	0.9962	0.9996	0.9966

be highly competitive to the correctly specified statistical model in terms of its forecasting performance; similar findings, but in the context of nonlinear non-Gaussian state space models, are also reported in Koopman et al. (2015).

Next we verify the precision of the filtered transition probability estimates. In a Monte Carlo study, the transition probabilities $\pi_{00,t}$ and $\pi_{11,t}$ are simulated as part of the data generation process. Hence we are able to compare true transition probabilities with their filtered estimates and compute the mean squared error (MSE) and the mean

absolute error (MAE) statistics.

The results in Table 2.4 provide strong evidence that the Markov regime switching model with time-varying transition probabilities is successful in producing accurate filtered estimates of the probabilities for all time-varying patterns. It is only for the data generating process with constant transition probabilities that the MSE and MAE statistics for the static model are smaller than those for the model with score driven dynamics. For this case, however, the absolute value of all statistics are an order of magnitude smaller than for all the other data generating processes. We also observe that for increasing sample sizes T, the MSE and MAE statistics mostly decrease for the score driven model, while this does not happen for the static model. To put these findings in some perspective, we emphasize that the sinusoid patterns have the same number of swings over the entire sample for different sample sizes. Therefore, the change in the transition probabilities gets smaller per unit of time as T increases. It follows that the updating equation (2.12) for the time-varying model can track the true transition probability more accurately as T increases. The inaccuracy of the estimates obtained from the static model, however, remains unaffected.

Table 2.4: Simulation results III: filtered transition probabilities

We have simulated 1000 time series from each data generation process (DGP) listed in Table 2.1 and for sample sizes $T = 250, 500, 1000$. The static parameters are estimated by the method of maximum likelihood, for the Markov regime switching model with static probabilities (Const.), with autoregressive time-varying transition probabilities (TVP) and with score driven (SD) time-varying transition probabilities. In the latter case, the underlying time-varying parameters are updated using equation (2.12). We report the sample averages for the 1000 simulated series of the mean squared error (MSE) and the mean absolute error (MAE) of the one-step ahead forecast of two transition probabilities π_{00} and π_{11}. The one-step ahead forecast errors are computed within the sample that is used for parameter estimation.

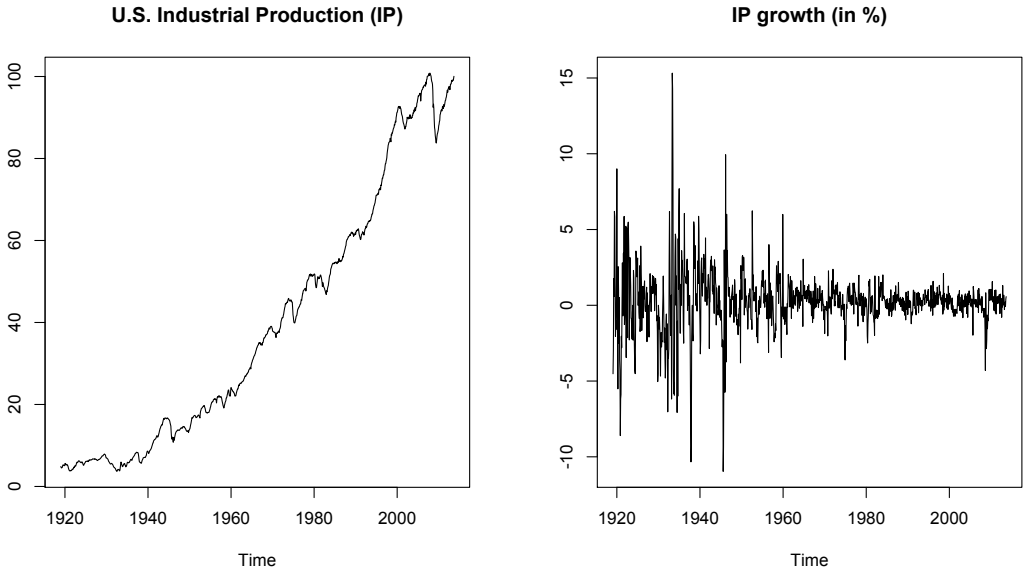
		MSE			MAE			MSE			MAE		
		π_{00}						π_{11}					
DGP	T	Const.	TVP	SD	Const.	TVP	SD	Const.	TVP	SD	Const.	TVP	SD
	250	0.0003	0.0140	0.0127	0.0146	0.0477	0.0494	0.0038	0.0796	0.0594	0.0447	0.1877	0.1655
Constant	500	0.0002	0.0047	0.0062	0.0102	0.0281	0.0321	0.0016	0.0440	0.0252	0.0297	0.1323	0.1065
	1000	0.0001	0.0015	0.0032	0.0072	0.0174	0.0211	0.0007	0.0284	0.0103	0.0204	0.0992	0.0688
	250	0.1723	0.2155	0.0572	0.3368	0.3729	0.1933	0.1862	0.2415	0.0408	0.3556	0.4015	0.1623
Slow Sine	500	0.2235	0.2993	0.0446	0.3928	0.4546	0.1759	0.1930	0.2476	0.0288	0.3486	0.3942	0.1343
	1000	0.1785	0.2756	0.0296	0.3540	0.4386	0.1425	0.1584	0.2305	0.0203	0.3190	0.3835	0.1120
	250	0.1867	0.2353	0.0742	0.3563	0.3949	0.2252	0.1742	0.2180	0.0633	0.3374	0.3733	0.2030
Sine	500	0.1755	0.2372	0.0576	0.3387	0.3910	0.1973	0.1872	0.2619	0.0408	0.3575	0.4208	0.1651
	1000	0.2013	0.2905	0.0446	0.3684	0.4435	0.1760	0.1898	0.2702	0.0280	0.3523	0.4203	0.1346
	250	0.1714	0.2136	0.1009	0.3371	0.3699	0.2611	0.1760	0.2221	0.0950	0.3444	0.3801	0.2538
Fast Sine	500	0.1767	0.2340	0.0780	0.3412	0.3884	0.2328	0.1809	0.2453	0.0674	0.3487	0.4018	0.2161
	1000	0.1808	0.2593	0.0619	0.3464	0.4119	0.2082	0.1832	0.2676	0.0488	0.3501	0.4207	0.1825
	250	0.1072	0.1301	0.0511	0.2203	0.2517	0.15836	0.1583	0.2347	0.0310	0.3650	0.4185	0.1339
Break	500	0.1357	0.1860	0.0250	0.2710	0.3240	0.1179	0.1656	0.2516	0.0230	0.3561	0.4251	0.1100
	1000	0.1426	0.2105	0.0162	0.2915	0.3543	0.0956	0.1571	0.2455	0.0168	0.3354	0.4089	0.0888
	250	0.1055	0.0452	0.1139	0.2710	0.1628	0.2769	0.0814	0.0413	0.0875	0.2633	0.1757	0.2620
Autoregressive	500	0.1027	0.0352	0.1043	0.2681	0.1520	0.2678	0.0796	0.0339	0.0797	0.2638	0.1605	0.2568
	1000	0.1017	0.0297	0.1011	0.2668	0.1460	0.2651	0.0785	0.0298	0.0764	0.2634	0.1513	0.2564

2.6 An empirical study of U.S. Industrial Production

Markov regime switching models are often used in empirical studies of macroeconomic time series. We therefore illustrate our new methodology for time-varying transition probabilities in an empirical study concerning a key variable for macroeconomic policy, U.S. Industrial Production (IP). The time series for IP is obtained from the Federal Reserve Bank of St. Louis economic database (FRED); we have monthly seasonally adjusted observations from January 1919 to October 2013, $T = 1137$. We analyze the

percentage growth of IP (log-differences $\times 100$) and consider the resulting series as our y_t variable. Figure 2.1 presents both the IP index and the IP percentage log differences y_t. Since the seminal paper of Hamilton (1989), modeling the different regimes in variables related to economic growth has proved important. Typically, different growth regimes can be distinguished, such as negative, slow, and high growth. Over long time spans as in our current application, additional challenges relate to the difference in the volatility of growth, as well as to the stability of the transition probabilities between the different regimes. For example, it may be challenging to empirically distinguish regimes of low volatility-high growth from regimes of high volatility-negative growth. Given the encouraging results from the simulation study, it is for this challenge where our new model with score driven dynamics can be useful.

Figure 2.1: U.S. Industrial Production (monthly, seasonally adjusted) and percentage growth rates (log-differences $\times 100$)



2.6.1 Three model specifications

After some preliminary analysis, we consider a model with *three* regimes for the mean ($m = 0, 1, 2$) and *two* regimes for the variance ($v = 0, 1$). The three regimes for the mean may represent recession, stability and growth periods in U.S. production. Each regime for the mean consists of a constant and p_m lagged dependent variables for y_t, with $m = 0, 1, 2$. The constant and the p_m autoregressive coefficients are collectively subject to the regime to which they belong. The two regimes for the variance may

simply distinguish periods of low and high volatility. To remain within a parsimonious modelling framework, we adopt the model specification of Doornik (2013) to structure the transition probability matrix. Doornik analyzes a quarterly time series of post-war U.S. gross domestic product growth by means of a Markov switching mean-variance component model.

Model I: static specification

Let $\{z_t^\mu\}$ and $\{z_t^\sigma\}$ denote the hidden processes that determine the mean and the variance for the density of y_t, respectively. We have

$$y_t \,|\, (z_t^\mu = m, z_t^\sigma = v, I_{t-1}) \sim \mathrm{N}\Big(\ \mu_{m,t}\ ,\ \sigma_v^2\ \Big), \qquad m = 0, 1, 2, \qquad v = 0, 1, \tag{2.17}$$

with the three mean equations

$$\mu_{m,t} = \phi_{0,m} + \phi_{1,m} y_{t-1} + \ldots + \phi_{p_m,m} y_{t-p_m}, \qquad m = 0, 1, 2, \tag{2.18}$$

where $\phi_{0,m}$ is an intercept, $\phi_{1,m}, \ldots, \phi_{p_m,m}$ are autoregressive coefficients, for $p_m \in \mathbb{N}^+$, and σ_1^2 and σ_2^2 are the two variances. The transition probabilities for the mean and variance are collected in the matrices Π^μ and Π^σ, respectively, which are given by

$$\Pi^\mu = \begin{pmatrix} \pi_{00}^\mu & \pi_{01}^\mu & 1 - \pi_{00}^\mu - \pi_{01}^\mu \\ \pi_{10}^\mu & \pi_{11}^\mu & 1 - \pi_{10}^\mu - \pi_{11}^\mu \\ \pi_{20}^\mu & \pi_{21}^\mu & 1 - \pi_{20}^\mu - \pi_{21}^\mu \end{pmatrix}, \qquad \Pi^\sigma = \begin{pmatrix} \pi_{00}^\sigma & 1 - \pi_{00}^\sigma \\ 1 - \pi_{11}^\sigma & \pi_{11}^\sigma \end{pmatrix}. \tag{2.19}$$

We follow Doornik (2013) in specifying the transition probability matrix for the $3 \times 2 = 6$ regimes as

$$\Pi = \Pi^\sigma \otimes \Pi^\mu,$$

where the 36 probabilities in Π are a function of 6 mean and 2 variance probabilities. Compared to a fully unrestricted model with 6 regimes, the current specification is much more parsimonious, while still allowing for a considerable degree of flexibility to capture the salient dynamics of the data. The conditional density of y_t given I_{t-1} can be expressed in terms of the filtered probabilities as in (2.8). We have

$$
\begin{aligned}
p(y_t|\psi, I_{t-1}) &= \xi_{t-1}^{'} \left(\Pi^{\sigma} \otimes \Pi^{\mu}\right) \eta_t \\
&= \begin{bmatrix} \mathbb{P}[z_{t-1}^{\mu}=0, z_{t-1}^{\sigma}=0|\psi, I_{t-1}] \\ \mathbb{P}[z_{t-1}^{\mu}=1, z_{t-1}^{\sigma}=0|\psi, I_{t-1}] \\ \vdots \\ \mathbb{P}[z_{t-1}^{\mu}=2, z_{t-1}^{\sigma}=1|\psi, I_{t-1}] \end{bmatrix}' \left(\Pi^{\sigma} \otimes \Pi^{\mu}\right) \begin{bmatrix} p(y_t; \mu_{0,t}, \sigma_0^2, I_{t-1}) \\ p(y_t; \mu_{1,t}, \sigma_0^2, I_{t-1}) \\ \vdots \\ p(y_t; \mu_{2,t}, \sigma_1^2, I_{t-1}) \end{bmatrix}.
\end{aligned} \tag{2.20}
$$

Model II: time-varying variance probabilities as function of $|y_{t-1}|$

As a benchmark, we also construct a model in the spirit of Diebold et al. (1994) and Filardo (1994). We specify the time-varying transition probabilities for the variance regimes in Π_t^{σ} as a logistic transformation of the lagged dependent variable, that is

$$
\pi_{vv,t}^{\sigma} = \frac{\exp(g_{vv,t})}{1+\exp(g_{vv,t})}, \qquad g_{vv,t} = c_{0,v} + c_{1,v}|y_{t-1}|, \qquad v = 0, 1, \tag{2.21}
$$

where $c_{0,v}$ is an intercept and $c_{1,v}$ is a fixed coefficient, for $v = 0, 1$. The four c coefficients are estimated by the method of maximum likelihood, jointly with the other coefficients.

Model III: time-varying variance probabilities as function of score

We considered several extensions of the current specification with time-varying transition probabilities. After a preliminary analysis, we found that for the current application only the introduction of time-varying transition probabilities for the variance regimes substantially improves the fit of the model. We therefore only present the results for this specification. Note that once the probabilities Π^{σ} are made time-varying, all regime transition probabilities $\Pi = \Pi^{\sigma} \otimes \Pi^{\mu}$ become time-varying.

Empirically, there appears no need to shrink the range of $\pi_{ii,t}$ ex ante, such that we can set $\delta_{ii} = 0$ for $i = 0, \ldots, 5$, where δ_{ii} was defined in Section 2.3. Using the framework of Section 2.3.2, we specify the time-varying matrix Π_t^{σ} as

$$
\pi_{vv,t}^{\sigma} = \frac{\exp(f_{vv,t})}{1+\exp(f_{vv,t})}, \qquad v = 0, 1, \qquad f_t = (f_{00,t}\ ,\ f_{11,t})',
$$

where f_t is updated over time as in (2.12).

The resulting conditional density for y_t is given by $p(y_t|\psi^*, I_{t-1}) = \xi_{t-1}^{'} \left(\Pi_t^{\sigma} \otimes \Pi^{\mu}\right) \eta_t$.

The regime probability structure of Doornik (2013) is more restricted than the one for the general K-regime Markov switching model in Section 2.3.3. We therefore have different expressions for the score vector and scaling matrix. The score vector is given by $\nabla_t = (\nabla_{00,t}\ ,\ \nabla_{11,t})'$, where

$$\nabla_{vv,t} = \frac{\pi^{\sigma}_{vv,t}(1-\pi^{\sigma}_{vv,t})}{p(y_t|\psi^*, I_{t-1})} \ \times\ \xi'_{t-1}\left(\frac{\partial \Pi^{\sigma}_t}{\partial \pi^{\sigma}_{vv,t}} \otimes \Pi^{\mu}\right)\eta_t, \qquad v = 0, 1,$$

with ξ_{t-1} and η_t defined implicitly in (2.20). To compute the conditional Fisher Information $\mathbb{E}[\nabla_t \nabla_t']$, where $\mathbb{E}[\,\cdot\,]$ is the expectations operator with respect to $p(y_t|\psi^*, I_{t-1})$, we evaluate four numerical integrals by a Gauss-Hermite method for every time t and for each value for the parameter vector ψ^*.

2.6.2 Parameter estimates, model fit and residual diagnostics

Table 2.5 presents the parameter estimates for the three model specifications. For all models we consider $p_m = 3$ lags in (2.18), with $m = 0, 1, 2$. Other values of p_m have been considered but do not substantially improve the fit. Parameter estimates are obtained by numerically maximizing the log-likelihood function with respect to the static parameter vector ψ or ψ^*. The associated standard errors are obtained by numerically inverting the Hessian matrix at the maximized log-likelihood value. The sets of estimated coefficients for the three mean regimes are very similar across the three model specifications. The introduction of time-varying variance transition probabilities only has a limited effect on the specification of the mean. The coefficient $\phi_{0,m}$ determines the interpretation of a regime: $m = 0$ corresponds to low IP growth, $m = 1$ represents a recession, and $m = 2$ identifies a high IP growth regime. The autoregressive coefficients $\phi_{1,m}, \ldots, \phi_{3,m}$ show that in "normal years" IP growth is persistent, while during recessions IP growth is subject to persistent cyclical dynamics. Periods of high IP growth are very short lived given the strongly negative autoregressive coefficients for $m = 2$. The estimated transition probabilities reveal the typical situation in regime switching models that once we are in a recession or low growth regime, it is most likely that we remain in this regime. It is only for the high growth regime that it is more likely to move to a low growth regime while the probability to stay, $\pi^{\mu}_{22} = 1 - \pi^{\mu}_{20} - \pi^{\mu}_{21}$, is estimated around 0.35.

For the variance regimes, the models clearly distinguish between a low (approximately 0.3) and a high (approximately 5.5 to 6.0) variance regime. The magnitudes of these variances are again comparable across the different models. For Model I, both

variance regimes are highly persistent with probabilities π^{σ}_{00} and π^{σ}_{11} both estimated close to 1. We also learn from Table 2.5 that the model fit improves upon introducing time-varying variance transition probabilities, both for Model II and Model III. The maximized log-likelihood values increase by 8 and 17 basis points at the cost of an additional 2 and 4 parameters for Model II and III, respectively. The corrected Akaike information criterion and the Bayesian information criterion clearly point to Model III as the best comprimise to model fit and a parsimonious model specification. The time-varying transition probabilities for the low volatility ($v = 0$) and high volatility ($v = 1$) regimes in Model III are highly persistent: the estimates for the diagonal elements of B are very close to unity. From Model II we may conclude that only the transition probability of a low volatility regime is time-varying: the estimate of $c_{1,1}$ is not significant. For Model III we find somewhat stronger evidence that both transition probabilities are time-varying: the estimates of both diagonal elements of A are significant at the usual level of 5%.

Finally, Table 2.5 presents diagnostic test statistics for the generalized and Rosenblatt residuals which we have discussed in Section 2.2. The p-values for the well-known Jarque-Bera χ^2 normality test and the Ljung-Box χ^2 serial correlation test, for the residuals and the squared residuals, indicate that all models are capable of describing the salient features in IP growth. There are some differences between the statistics for the generalized and Rosenblatt residuals, but the differences are small and do not affect the main conclusions. The Jarque-Bera test indicates that the IP growth time series is subject to a few outlying observations.

2.6.3 Signal extraction: regime transition probabilities

In Figure 2.2 we present the smoothed estimates of probabilities for mean and variance regimes and the filtered estimates of (time-varying) transition probabilities for all models. Model II appears unable to capture the dynamics in the transition probabilities. We have learned from Table 2.5 that the estimate of $c_{1,1}$ is not significant. This is also reflected in Figure 2.2: the filtered probability estimates for the high variance regime are almost constant over time. On the other hand, the filtered probability estimates for the low volatility regime are highly erratic. The filtered probabilities for Model III show an entirely different pattern. Both the low and high volatility transition probabilities evolve gradually over time. In particular, the persistence of the low volatility regime appears to have increased over time, with values around 0.7 in the early part of the sample, and values close to 1 in the second half of the sample. The converse holds for the high volatility regime. The persistence probability π^{σ}_{11} is

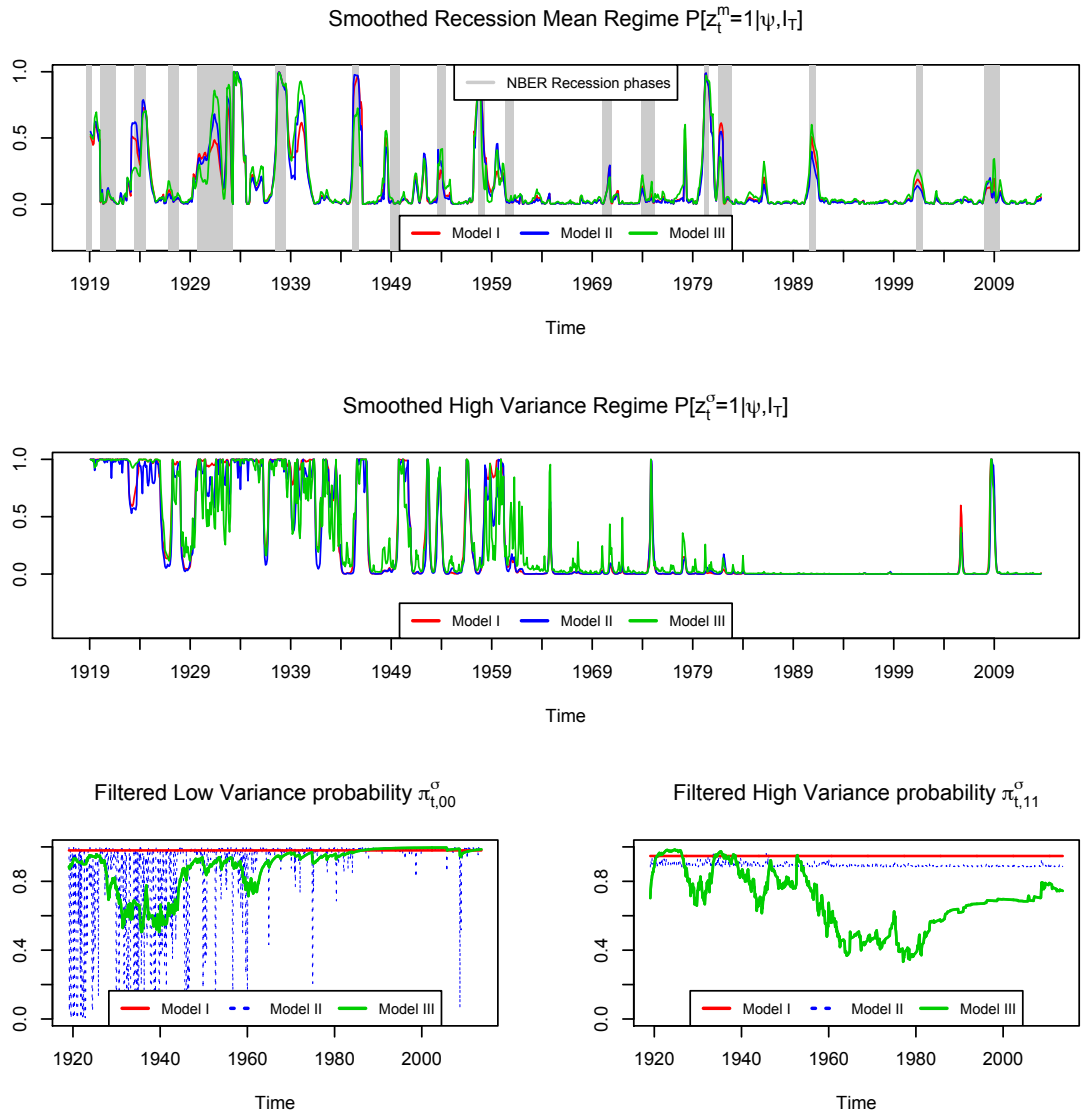
Table 2.5: U.S. Industrial Production: parameters, model fit and residual diagnostics

In the first two panels we report the maximum likelihood estimates with standard errors in parantheses below, for Models I, II and III. In the first panel the parameter estimates for the mean $\mu_{m,t}$ in (2.17) are reported for each regime $m = 0, 1, 2$: the intercept $\phi_{0,m}$, the autoregressive coefficients $\phi_{1,m}, \ldots, \phi_{3,m}$, and the transition probabilities π^{μ}_{mj}, for $j = 0, 1$, in Π^{μ} of (2.19). In the second panel the two regime variance estimates for σ^2_v are reported. The variance transition probability for Model I π^{σ}_{vv} is estimated directly while we have $\pi^{\sigma}_{vv} = \text{logit}^{-1}(x_v)$ for Model II ($x_v = c_{0,v}$) and for Model III ($x_v = \omega_v/(1 - B_{vv})$), for $v = 0, 1$, where ω_v and B_{vv} are the $(v+1)^{\text{th}}$ elements of vector ω and diagonal matrix B in (2.12), respectively. The time-varying variance probabilities are determined in Model II by $c_{1,v}$, and in Model III by A_{vv} and B_{vv} which are the $(v+1)^{\text{th}}$ diagonal elements of A and B in (2.12), respectively, for $v = 0, 1$. In the third panel we report model fit statistics: Fit(1) is maximized log-likelihood value; Fit(2) is AICc; Fit(3) is BIC, see Section 2.5.2. We further report the p-values of the residual diagnostic (RD) test statistics for the generalized (e_t) and Rosenblatt's residuals ($\tilde{e}_t$): RD(1) is Jarque-Bera normality $\chi^2(2)$ test; RD(2) is Ljung-Box serial correlation $\chi^2(6)$ test; RD(3) is as RD(2) for squared residuals.

	Model I (22 parameters)			Model II (24 parameters)			Model III (26 parameters)		
	$m=0$	$m=1$	$m=2$	$m=0$	$m=1$	$m=2$	$m=0$	$m=1$	$m=2$
$\phi_{0,m}$	0.076	-0.212	0.846	0.068	-0.277	0.837	0.043	-0.128	0.737
	(0.038)	(0.182)	(0.168)	(0.037)	(0.140)	(0.172)	(0.033)	(0.095)	(0.128)
$\phi_{1,m}$	0.316	1.121	-0.609	0.327	1.126	-0.621	0.351	1.087	-0.479
	(0.050)	(0.096)	(0.135)	(0.050)	(0.088)	(0.130)	(0.040)	(0.079)	(0.107)
$\phi_{2,m}$	0.212	-0.569	-0.395	0.220	-0.526	-0.484	0.234	-0.537	-0.221
	(0.050)	(0.146)	(0.123)	(0.040)	(0.139)	(0.086)	(0.037)	(0.108)	(0.086)
$\phi_{3,m}$	0.105	0.076	0.039	0.113	0.011	0.133	0.112	0.036	0.103
	(0.037)	(0.107)	(0.124)	(0.035)	(0.115)	(0.101)	(0.030)	(0.065)	(0.092)
π^{μ}_{m0}	0.909	0.111	0.577	0.909	0.113	0.592	0.864	0.145	0.576
	(0.041)	(0.103)	(0.140)	(0.032)	(0.072)	(0.142)	(0.036)	(0.070)	(0.175)
π^{μ}_{m1}	0.016	0.858	0.055	0.014	0.858	0.073	0.021	0.842	0.048
	(0.019)	(0.092)	(0.052)	(0.012)	(0.064)	(0.062)	(0.015)	(0.059)	(0.055)
	$v=0$	$v=1$		$v=0$	$v=1$		$v=0$	$v=1$	
σ^2_v	0.336	5.579		0.351	5.866		0.317	5.920	
	(0.025)	(0.629)		(0.026)	(0.691)		(0.023)	(0.541)	
π^{σ}_{vv}	0.980	0.947		0.996	0.883		0.886	0.702	
	(0.007)	(0.018)		(0.003)	(0.053)		(0.080)	(0.200)	
$c_{1,v}$				-1.899	0.108				
				(0.419)	(0.209)				
A_{vv}							0.132	0.148	
							(0.058)	(0.074)	
B_{vv}							0.998	0.989	
							(0.003)	(0.011)	
	$i=1$	$i=2$	$i=3$	$i=1$	$i=2$	$i=3$	$i=1$	$i=2$	$i=3$
Fit(i)	-1642	3330	3439	-1634	3317	3437	-1625	3302	3433
RD(i) e_t	0.065	0.772	0.556	0.032	0.968	0.659	0.019	0.792	0.924
RD(i) $\tilde{e}_t$	0.011	0.409	0.648	0.012	0.830	0.738	0.007	0.632	0.676

Figure 2.2: U.S. Industrial Production: smoothed and filtered transition probability

Smoothed probability estimates for the recession regime in the mean and for the high variance regime. Filtered transition probability estimates for the low and high variance regimes. In the first graph, the vertical gray areas indicate recessions according to the NBER business cycle classifications.



close to 1 up to the 1940s. After that, the probability decreases substantially to values around 0.5, and slowly rises again towards the end of the sample. The pattern for the filtered probabilities is consistent with the empirical pattern in the data in Figure 2.1. In the earlier part of the sample, high volatility levels are predominant. Towards the middle of the sample, large volatilities are incidental and short-lived. Interestingly, towards the end of the sample and particularly during the years of the financial crisis, the U.S. debt ceiling crisis, and the European sovereign debt crisis, higher volatility levels appear to cluster again.

The empirical patterns are corroborated by the parameter estimates in Table 2.5. In particular, the parameter estimates for the diagonal elements of B are both close to 1; this suggests that the dynamic transition probabilities evolve gradually over time. The estimates of both diagonal elements of A have the correct sign and lead to parameter changes that increase the local fit of the model in terms of log-likelihood.

Finally, we present the smoothed estimates of z_t in the top panels of Figure 2.2, together with the NBER business cycle classifications. We conclude that all models result in higher smoothed recession probabilities during NBER recessions. The model fit for a model with time-varying transition probabilities (Model II or III) is typically higher than that of the static Model I. From the smoothed probabilities for the high variance regime we see that most of the high variance regime period is located in the first half of the sample. The second episode of high variance is during the financial crisis, with the intermediate period having predominantly a low level of volatility. We notice that some, but not all, NBER recessions correspond to periods of high volatility. This supports the use of our current framework with separate regimes for the (conditional) means and for the variances.

2.7 Conclusion

We have introduced a new methodology for time-varying transition probabilities in Markov switching models. We have shown that an observation driven modeling framework based on the score of the predictive likelihood as in Creal et al. (2013) provides an effective tool to describe the dynamics of transition probabilities. The dynamics can easily be interpreted while the information embedded in the conditional observation densities is fully incorporated in the transition probability updates. We have formulated conditions for the estimated time-varying probabilities from our score driven model to converge to stationary and ergodic stochastic processes.

By means of an extensive Monte Carlo study, we have shown that our observation

driven model adequately tracks the dynamic patterns in transition probabilities, even if the model is misspecified. Both for structural breaks and slowly variations in regime transition probabilities, our model yields a large improvement in model fit compared to a model with constant transition probabilities only. The model also performs well with respect to nonlinear autoregressive parameterizations of the transition probabilities.

In our empirical study for U.S. Industrial Production growth, we have shown how the model can capture the dynamic features of regime transition probabilities for means and variances. Our score driven model outperforms both the Markov switching model with constant probabilities and with transition probabilities depending on a lagged dependent variable. In particular, the patterns filtered by our model can be interpreted and have economic relevance: a higher (lower) persistence for high (low) volatility regimes in the beginning (at the end) of our long time series span from January 1919 to October 2013. A key finding is that high volatility periods appear to re-occur and become more persistent again at the end of the sample, during the financial and sovereign debt crises. We conclude that the model can provide a useful benchmark in settings where transition probabilities in a regime switching model may vary over time.

Chapter 3

Markov switching model for electricity prices: an empirical comparison

Abstract

In empirical literature it is well known that Markov regime switching models represent an attractive choice to capture the complex dynamics of electricity spot prices characterized by the presence of jumps or spikes. A crucial decision that the researcher has to take is related to how many regimes to use. Through an empirical analysis of UK Automated Power Exchange (APX) electricity spot prices we found evidence in favour of a three regime specification. Due to the non-homogenous occurrence of spikes, the standard assumption that the transition probabilities are constant is too restrictive and several proposals have been made to overcome this limitation. We compare several specifications and we show that the mechanism driven by the scores of the predictive likelihood is a valid alternative to those models which exploit exogenous information like forecasted reserve margin and forecasted demand.

Some key words: Electricity spot prices, Spikes, Markov regime-switching, Time-varying transition probabilities, Score driven models.

I am deeply thankful to Silvano Bordignon for the numerous suggestions received.

3.1 Motivation

The electricity sector has been liberalized over the past three decades. In 1989 the UK Electricity Act determined the creation of the first organized electricity market in Europe. After the Central Electricity Generating Board was dismissed, a pool was constituted. However, the pool was heavily dominated by only two companies, National Power and Powergen. Attracted by the prospect of large profits new competitors put pressure on political authorities to enter to the market. The New Electricity Trading Arrangements, issued in March 2001 replaced the pool with a system of voluntary bilateral markets and power exchanges encouraging new competitors. In the same year, the UK Automated Power Exchange (APX) opened a spot market.

An increasing number of models have been developed in empirical literature for the financial instrument or other commodity markets. However, those models may be inadequate to deal with the uniqueness of electricity as a commodity. Indeed, electricity markets are characterized by extreme changes in spot prices known as jumps or spikes. Several nonlinear models have been proposed to include discontinuous components in realistic models of electricity price dynamics. Since the seminal work of Hamilton (1989), the Markov switching class of models became very popular. The key attractive feature of Markov switching models is that the conditional distribution of a time series depends on an underlying latent state or regime, which can take only a finite number of different values. The discrete state evolves through time as a discrete Markov chain and its statistical properties are summarized by a transition probability matrix. Ethier and Mount (1998) proposed a two state specification in which both regimes were governed by autoregressive processes of the first order with different or common variances. Huisman and Mahieu (2003) introduced a third regime, the jump reversal regime, that describes how prices move back to the baseline regime after the initial jump has occurred. Huisman and De Jong (2003) proposed a model with only two regimes, a stable mean reverting AR(1) regime and a spike regime. In order to cope with the heavy-tailed nature of spikes, Weron et al. (2004) and Janczura and Weron (2009, 2010) replaced the normal distribution of the spike regime with log-normal and Pareto distributions.

Another line of research has been devoted to overcome the unrealistic assumption of static transition probabilities. Huisman (2008) showed that spike intensity varies over time, influenced by seasonal fluctuations in climate conditions like temperature and the number of daylight hours. In econometric literature, Diebold et al. (1994) and Filardo (1994) allowed the transition probabilities dynamics to be driven by lagged values of the dependent variable or by explanatory variables (Gray, 1996). In the

same spirit, Mount et al. (2006) used fundamental variables as reserve margin forecast and load to model the transition probabilities. Janczura and Weron (2010) considered seasonal transition probability estimated directly on the price by a kernel density estimator.

In the previous chapter, we extended Diebold et al. (1994)'s and Filardo (1994)'s proposal by using the score of the predictive density in the updating mechanism according to the Generalized Autoregressive (GAS) models formulated by Creal et al. (2011, 2013) and Harvey (2013). In an extensive simulation study, we have emphasized the ability of their model to capture a range of dynamic patterns for the unobserved transition probabilities by linking non linearly past observations to future transition probabilities while in an empirical application to U.S. Industrial Production growth rate, we showed that the dynamic features of regime transition probabilities for means and variances can be well described by the novel formulation.

Main purpose of this chapter is to adapt the promising GAS approach to deal with the peculiarity of electricity price dynamic. Differently from what we did in the previous chapter, the normality assumption is not convenient for the regime-specific distributions and more sophisticated choices have been discussed and utilized for describing the extreme price regimes. Through an empirical comparison on UK APX spot prices we show its ability to accurately describe the non-linear price dynamics. A crucial decision the researcher has to take concerns the number of regimes used to describe the price dynamics. Similarly to Janczura and Weron (2010) we find that a three regime specification with a baseline regime alternated by two spikes regimes is the favorite choice both from a statistical and an economic perspective. The time-varying transition probabilities driven by the scores of the conditional densities are found to be a preferred choice over both the static and seasonal transition probabilities. When we include in our model the exogenous information contained in the forecasted demand and forecasted margin reserve, the fitting ability increases substantially. As second methodological result, we show that the GAS methodology can be easily extended to include explicitly the exogenous information in the transition probabilities dynamics. In our empirical application, we find that the novel proposal overcomes the alternatives.

In Section 3.2 the dataset is introduced and pre-filtering techniques are discussed to detect the cyclical and seasonal components from the raw prices. In Section 3.3 the econometrics methodology is presented. In Section 3.4, the results for both two and three regime specifications are presented and discussed. Finally, in Section 3.5, we conclude.

3.2 Dataset and pre-filtering

We analyze mean daily day-ahead spot prices from the UK Automated Power Exchange spot market. The data can be found on *www.bmreports.com*. The sample contains 2048 daily observations and covers the 5-year period April, 1, 2005 - November, 8, 2010. Mount et al. (2006) have shown that the load and the implicit reserve margin, defined as total offered capacity minus load, can be used as explanatory variables to describe the conditional mean and predict accurately the spikes in electricity prices. As the forecasted load is announced to suppliers one day-ahead, this information can be used to forecast prices for the next day. Also the forecasted reserve margin is known as it is calculated by subtracting the forecasted load from the day-ahead total offered capacity. The dataset analyzed in this paper contains the forecasted demand rather than the forecasted load. Although there exists some ambiguity in terminology, for the purpose of this paper we consider the two quantities substantially equivalent.

Spot prices and forecasted demand are far from being stationary. In order to deal with this characteristic, we consider the spot price (or the one day-ahead demand) X_t as a sum of three independent parts: a long-term component T_t, which represents the non periodic price levels, a weekly periodic component s_t which incorporates the seasonal fluctuations due to climate and consumption conditions and a stochastic component Y_t, $X_t = T_t + s_t + Y_t$. A multiplicative model and the logarithmic transformation applied to Y_t would be more appropriate if the researcher hypothesized interactions among the components. As in Weron (2009) and Janczura and Weron (2010), the trend component is estimated from raw data X_t using a wavelet filtering-smoothing technique. Then, the weekly component s_t is estimated on the de-trended series $X_t - T_t$ as the average week over all sample; we refer to Weron (2006) for alternative deseasonalization techniques. Lastly, the resulting stochastic component Y_t is shifted so that the minimum of the new process is the same as the minimum of X_t. In Figure 3.2, the original series is plotted together with the trend component. Notice that the reserve margin has not been treated as it can be assumed stationary. The Autocorrelation Function (ACF) for the estimated stochastic component Y_t shown in Figure 3.1, suggests that both the long term and the seasonal components have been successfully detected and removed by both the price and forecasted demand series. The Augmented Dickey-Fuller statistics, estimated to -9.521 for prices and to -7.692 for forecasted demand does not reject the null hypothesis of stationary at 0.01 significant level in both cases. In the remainder of this paper, we focus our attention in modeling the dynamic of the stochastic component of one-day-ahead price, Y_t.

Figure 3.1: UK APX Spot Prices: stochastic components

Autocorrelation function up to 100 lags for stochastic component of prices (on left) and of forecasted demand (on right).

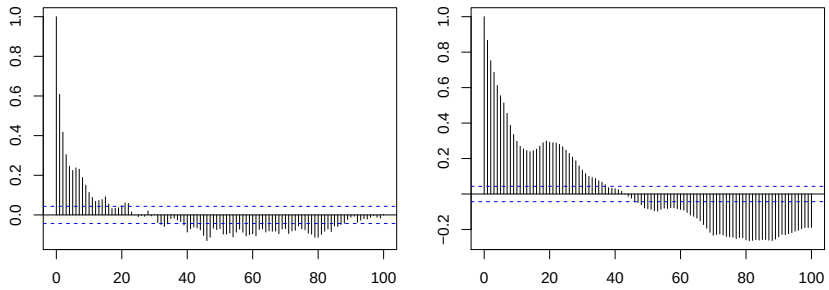
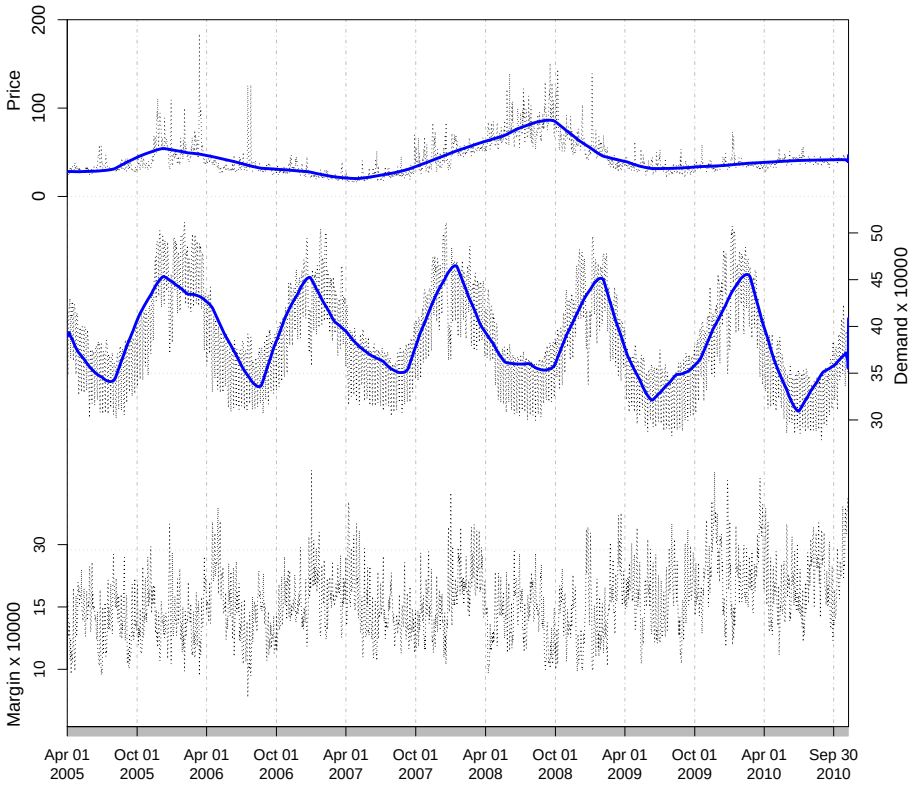


Figure 3.2: UK APX Spot Prices: original series, long-term and cyclical components

Prices, forecasted demand and forecasted reserve margin. Original series (dashed line) and trend/cyclical component estimated by wavelet filter (straight line).



3.3 Methodology

3.3.1 Markov-switching models

We invite the reader to refer to Subsection 2.2 for an introduction of the basics of Markov switching models. For the sake of clarity, we report the main notation. Let $\{y_t,\ t = 1, \cdots, T\}$ denote a univariate time series of T observations. The probability distribution of the stochastic process y_t depends on the realizations of a hidden discrete stochastic process z_t. The stochastic process y_t is directly observable, whereas z_t is a latent random variable that is observable only indirectly through its effect on the realizations of y_t. The hidden process $\{z_t\}_{t\in\mathbb{Z}}$ is assumed to be an irreducible and aperiodic Markov chain with finite state space $\{0, \cdots, K-1\}$. Its stochastic properties are sufficiently described by the $(K \times K)$ transition matrix, Π, where each element π_{ij} of Π is equal to the transition probability from state i to state j,

$$\pi_{ij} = \mathbb{P}[z_t = j | z_{t-1} = i], \qquad \forall i, j \in \{0, \cdots, K-1\}. \tag{3.1}$$

All elements of Π are nonnegative and the elements of each row sum to 1, i.e. $\pi_{ij} \geq 0$, $\forall i, j \in \{0, \cdots, K-1\}$ and $\sum_{j=0}^{K-1} \pi_{ij} = 1$, $\forall i = 0, \cdots, K-1$. Let $p(\cdot\,|\theta, \psi)$ be a parametric conditional density indexed by parameters $\theta \in \Theta$ and $\psi \in \Psi$. We assume that the random variables $y_1, \cdots, y_T$ are conditionally independent given $z_1, \cdots, z_T$, with densities $y_t|\,(z_t = i) \sim p(\cdot\,|\theta_i, \psi)$ with regime dependent parameter θ_i, and regime independent parameter ψ. The density of y_t conditional on the information available at time $t-1$, denoted by I_{t-1}, is

$$p(y_t|\psi, I_{t-1}) = \sum_{i=0}^{K-1} p(y_t|\theta_i, \psi)\ \mathbb{P}(z_t = i|\psi, I_{t-1}), \tag{3.2}$$

where both ψ and $\theta_0, \ldots, \theta_{K-1}$ need to be estimated. The conditional mean of y_t given z_t and I_{t-1} may contain lags of y_t itself. To evaluate (3.2), we require the quantities $\mathbb{P}(z_t = i|\psi, I_{t-1})$ for all t. We can compute these efficiently using the recursive filtering approach of Hamilton (1989). Define $\xi_{t-1|t-1}$ as the $K-$dimensional vector containing the filtered probabilities $\mathbb{P}(z_{t-1} = i|\psi, I_{t-1})$ at time $t-1$ and let η_t be the $K-$dimensional vector collecting the densities $p(y_t|\theta_i, \psi)$ at time t for $i = 0, \ldots, K-1$. Then the filtered probabilities $\xi_{t|t}$ are updated by the Hamilton recursion

$$\xi_{t|t} = \frac{\left(\Pi'\ \xi_{t-1|t-1}\right) \odot \eta_t}{\xi'_{t-1|t-1}\Pi\eta_t}, \qquad p(y_t|\psi, I_{t-1}) = \xi'_{t-1|t-1}\Pi\eta_t, \tag{3.3}$$

where $\odot$ denotes the Hadamard element by element product. The filter needs to be started from an appropriate set of initial filtered probabilities $\mathbb{P}(z_0 = i|\psi, I_0)$. If we are also interested in making inference about the smoothed regime probabilities $\mathbb{P}(z_t = i|\psi, I_T)$, then we can use the algorithm of Kim (1994) to compute these efficiently.

3.3.2 Choice of regime densities

A detailed overview of the Markov-switching models applied to the electricity spot prices appears in Weron (2006) and in Janczura and Weron (2010). The presence of spikes in the spot electricity prices suggests that there exists a non-linear mechanism switching between normal and high-price states or regimes that can be conveniently modeled as a latent process. Typically, the base regime is assumed to be driven by a mean-reverting diffusion process that is discretized as an autoregressive process of the first order,

$$y_{t,b} = \alpha + (1 - \beta)y_{t-1,b} + \sigma_b\varepsilon_t, \tag{3.4}$$

where the error terms, ε_t, follow a standard Gaussian distribution. As the regimes are often assumed to be independent from each other, then during a spike the base regime becomes latent. Janczura and Weron (2012) suggested to replace the latent variables $y_{t-1,b}$ in (3.4) with their expectations $\tilde{y}_{t-1,b} = \mathbb{E}[y_{t-1,b}|\psi_{t-1}]$ based on the whole information available at time $t-1$. The expected values $\tilde{y}_{t,b} = \mathbb{E}[y_{t,b}|\psi_t]$ can be computed using the following recursive formula

$$\mathbb{E}[y_{t,b}|\psi_t] = y_{t,b}\mathbb{P}[z_t = b|\psi_t] + \{\alpha + (1 - \beta)\mathbb{E}[y_{t-1,b}|\psi_{t-1}]\}\mathbb{P}[z_t \neq b|\psi_t]. \tag{3.5}$$

Following Mount et al. (2006) we can include in (3.4) exogenous variables as the forecasted demand and the forecasted reserve margin. As these variables may have a different scale, we standardized them by subtracting the sample mean and dividing by the sample standard deviation. The equation for the base regime becomes

$$y_{t,b} = \alpha + (1 - \beta)y_{t-1,b} + \gamma_{margin}\tilde{m}_t + \gamma_{demand}\tilde{d}_t + \sigma_b\varepsilon_t, \tag{3.6}$$

where $\tilde{m}_t$ denotes the (standardized) forecast reserve margin at time t, announced to suppliers one day ahead and $\tilde{d}_t$ denotes the (standardized) forecast demand.

For the spike and downward spike regimes, we follow Janczura and Weron (2009) and Janczura and Weron (2010), who used, respectively, a median-shifted log-normal

distribution and an inverted shifted log-normal distribution, i.e.,

$$\log\left(y_{t,s} - y(q_1)\right) \sim N(\mu_s, \sigma_s^2), \quad y_{t,s} > y(q_1), \tag{3.7}$$

$$\log\left(-y_{t,d} + y(q_2)\right) \sim N(\mu_d, \sigma_d^2), \quad y_{t,d} < y(q_2), \tag{3.8}$$

where $0 < q_1, q_2 < 1$ and $y(q_i)$ denotes the q_i-th percentile of the empirical cumulative distribution functions of the data. A typical choice is the median, i.e. $q_1 = q_2 = 0.5$ but in our empirical results, we obtain better results setting $q_1 = 0.25$ and $q_2 = 0.75$.

3.3.3 Time-varying transition probabilities specification

In (3.1) and (3.3), we assumed that the transition probabilities are constant over time. However, there is numerous empirical evidence that this assumption could be too restrictive. Diebold et al. (1994) and Filardo (1994) proposed to make the transition probabilities dynamic including past values of the dependent variable. Gray (1996) used past values of exogenous variables. For analysing energy prices Mount et al. (2006) modeled the transition probabilities as function of load and reserve margin and Huisman (2008) used temperature as a proxy for reserve margin.

Without introducing exogenous variables, Janczura and Weron (2010) assumed that the transition probabilities are periodic functions. Their estimation procedure relies on a modification of Expectation-Maximization (EM) algorithm (cfr. Hamilton, 1989). From the smoothed regime probabilities obtained with Kim's smoother, they estimated the transition probabilities independently for each of the four seasons. In order to make those probabilities smoothed, they applied a kernel density smoother with a Gaussian kernel. We introduce the following notation for the periodic time-varying transition probabilities

$$\pi_{ij,t} = \tilde{g}(\pi_{ij,Spring}, \pi_{ij,Summer}, \pi_{ij,Fall}, \pi_{ij,Winter}), \tag{3.9}$$

where $\pi_{ij,Spring}$ denotes the transition probability estimated during the Spring and $\tilde{g}(\cdot)$ the kernel smoother.

Mount et al. (2006) modeled dynamically the transition probability as function of (forecasted) demand and reserve margin. For the generic transition probabilities π_{ij},

(3.1) is replaced by

$$\pi_{ij,t} = \frac{\exp(f_{ij,t})}{1 + \sum_{j=1}^{K-1} \exp(f_{ij,t})}, \qquad \pi_{i,K-1,t} = 1 - \sum_{j=1}^{K-1} \pi_{ij,t}, \tag{3.10}$$

$$f_{ij,t} = \omega_{ij} + \delta_{ij}^{M} \tilde{m}_t + \delta_{ij}^{D} \tilde{d}_t, \tag{3.11}$$

where the multinomial logit transformation in (3.10) ensures all transition probabilities being non-negative and summing to one (row-wise) and in (3.11) the one day ahead forecasted values of the exogenous variables are included in the dynamics of the transition probability.

The novel approach proposed in the previous chapter specifies the dynamics of the transition probabilities according to the generalized autoregressive score (GAS) framework of Creal et al. (2013). After applying again the multinomial logit transformation in (3.10) to the transition probabilities, all the time-varying parameters f_t in (3.11) are collected into the vector $\mathbf{f}_t$. Then, the GAS updating mechanism is given by

$$\mathbf{f}_{t+1} = \omega + B\mathbf{f}_t + A\mathcal{I}_{t-1}^{-0.5}\nabla_t, \tag{3.12}$$

$$\nabla_t = \frac{\partial p(y_t|\psi, \mathbf{f}_t, I_{t-1})}{\partial \mathbf{f}_t}, \qquad \mathcal{I}_{t-1} = \mathbb{E}\left[\nabla_t \nabla_t'\right],$$

where ω is a vector of constants, A and B are diagonal matrices, and ∇_t is the score of the conditional observation density, $p(y_t|\psi, \mathbf{f}_t, I_{t-1})$, with respect to $\mathbf{f}_t$. With a slight abuse of notation, the vector ψ collects all the statistic parameter of regime-specific distributions as well as the parameters that drive the GAS updating mechanism and it needs to be estimated numerically by maximizing the log-likelihood function. $\mathcal{I}_{t-1}$ is the square root of the inverse Fisher Information matrix. The conditional observation density is computed by the Hamilton filter and given by (3.3). The GAS mechanism thus takes a steepest ascent or Newton type step in $\mathbf{f}_t$ using the time t log conditional density as its criterion function. The choice of the inverse square root of the Fisher Information matrix for scaling the score $\mathbf{f}_t$, as done in Creal et al. (2013), accounts for the curvature of ∇_t as a function of $\mathbf{f}_t$.

If the exogenous variables are included in the mean equation for the baseline regime and the GAS approach is chosen for modeling dynamically the transition probabilities, the update mechanism in (3.12) is still valid. Indeed, it can be easily shown that the score and the Fisher Information matrix of the conditional density are functions of the exogenous information. However we can also include directly in (3.12) the forecasted

reserve margin and demand in the following way

$$\left(\mathbf{f}_{t+1} - \Delta^M \tilde{m}_{t+1} - \Delta^D \tilde{d}_{t+1}\right) = \omega + B\left(\mathbf{f}_t - \Delta^M \tilde{m}_t - \Delta^D \tilde{d}_t\right) + A\mathcal{I}_{t-1}^{-0.5}\nabla_t, \qquad (3.13)$$

where Δ^M and Δ^D are diagonal matrix. In the rest of the paper, we call the GAS model with exogenous variables the GAS-eXogenous (GASX) model.

3.3.4 Estimation and diagnostic

Hamilton (1989) used the Expectation-Maximization (EM) algorithm, (cfr. Dempster, 1977) to maximize the expected log-likelihood with respect to the parameter vector, θ. The EM procedure is computationally efficient when the maximization step is available in closed-form. Following Janczura and Weron (2010) we use the EM algorithm to obtain the parameter estimates for the model with periodic time-varying transition probabilities introduced in (3.9). For all other specifications, we maximize directly the log-likelihood function computed by Hamilton filter using a quasi-Newton optimization algorithm. The Akaike Information Criterion (AIC) (see Akaike 1973, 1974) and the Bayesian Information Criterion (BIC) (see Schwarz, 1978) are chosen to measure the quality of a statistical model for a given set of data.

In order to assesses the correct specification of regime densities, we classified the observation y_t in regime i if the smoothed probability of being in regime i at time t, $\mathbb{P}[r_t = i]$, is bigger than the smoothed probability of being in any other regime j, for $\forall i, j \in 0, \cdots, K-1$ and $j \neq i$. We compare the empirical distribution of observations classified into regime i with the theoretical distribution through the Kolmogorov-Smirnov statistics. The unknown parameters that index the theoretical distribution are estimated by Maximum Likelihood Estimator. To verify the correct specification of the baseline regime, we require a further step as the Kolmogorov-Smirnov test can be used only on independent samples. After identifying these observation which are classified into the baseline regime, we estimated the standardized residuals $\hat{\varepsilon}_t$ from (3.4) or from (3.6). The Kolmogorov-Smirnov test can be performed on those standardized residuals to verify the normality assumption.

3.4 Empirical results

3.4.1 Results for Markov-switching models with 2 regimes

We start by modeling the stochastic component Y_t with the two-regime Markov Switching model with dynamics switching between the baseline regime in (3.4) and the spike regime in (3.7). Static transition probabilities are considered as a benchmark for all the time-varying specification introduced in Subsection 3.3.3. When the forecasted reserve margin and demand are taken into account, we replaced (3.4) by (3.6).

As shown in Table 3.1, the fit of the model with periodic transition probabilities is not statistically better than the benchmark model. The log-likelihood value improves roughly by 3 points but at the cost of six more parameters. This result is not surprising as we have removed the seasonal component of prices by using the pre-filtering techniques illustrated in Section 3.2. The other dynamic specifications show a statistically significant improvement in fitting as both AIC and the BIC are smaller than those of static model. We notice that the GAS specification outperforms the counterparts when no exogenous information is considered. Adding the forecasted reserve margin and demand into the baseline regime mean equation increases the log-likelihood values for both TVP and GASX models. The novel GASX specification results in a remarkably smaller values of the information criteria than those obtained with the TVP specification.

The Kolmogorov-Smirnov test is not able to give us an indication in favor of a particular transition probabilities specification. However, the statistics accepts the null hypothesis at level 0.10 of correct specification for the spike regime and does not accept it for the baseline regime. Also the QQ-plot of baseline regime residuals estimated by the GASX model, shown in Fig. 3.3 (a), indicates an heavy deviation from normality in the left tail. The comparison between theoretical and empirical distribution for the spike regimes, Fig. 3.3 (b), confirms substantially the correct specification indication provided by Kolmogorov-Smirnov statistics.

In Table 3.1 we report the parameter and standard error estimates for all the models. The parameter estimates are very similar among the different models. The baseline regime exhibits a strong mean-reverse dynamic, the estimates of β is 0.307 to 0.308 (SE 0.018 to 0.019) for models without exogenous information and 0.383 (SE 0.019) for models with exogenous information. To capture the low spikes, the baseline regime variance is estimated to be quite high, roughly 15 to 16, (SE 0.54 to 0.59). In both the TVP and GASX models the forecasted reserve margin does not improve the baseline regime mean specification while the forecasted demand does. The standard

deviation of the spike regime is estimated smaller than the mean indicating that the log-normal parametrization is able to accurately describe extreme prices.

We have already seen that when no extra-information is taken into account, the transition probabilities dynamics are well captured by the GAS specification in terms of goodness-of-fit. Looking at the updating equation parameters, we notice that the processes are highly persistent with B_{bb} estimated equal to 0.955 (SE 0.018) and B_{ss} estimated equal to 0.895 (SE 0.084). The scaled score is statistically different from zero for baseline regime transition probabilities whereas is not for the spike regime.

In the TVP model, the transition probabilities are driven by the exogenous variables. In the GASX model, the transition probabilities are driven by an autoregressive process as well as by the exogenous variables. For both the specifications, the forecasted reserve margin affects positively the probability of staying in baseline regime and negatively the probability of staying in spike regime. As expected, a reduction in forecasted margin level leads to an increasing probability that a spike occurs. Regarding the forecasted demand, the effect is not concordant. While in the TVP specification, the forecasted demand has a positive impact on both probabilities, in the GASX specification the forecasted demand is positively correlated with the probability of staying in spike regime and negatively with the probability of staying in baseline regime. This seems more coherent with the economic intuition: when a large demand is forecasted, it is more probable that a spike will occurs. The parameters related to the conditional scores, A_{bb} and A_{ss} are positive as expected and statistically different from zero. We noticed that the autoregressive dynamic is different in the GASX specification if compared with the GAS model. In particular, the estimates of B_{bb} reduces to 0.197 from the value 0.955 estimated by the model without exogenous information.

Table 3.1: UK APX Spot Prices: goodness-of-fit for models with 2 regimes

We have estimated the Markov switching model with two regimes on the stochastic component, y_t, of UK APX Spot Prices. For the first three models, the baseline state is given by the discretized mean-reversion diffusion process in (3.4) where the unknown value of $y_{t-1,b}$ is estimated by its conditional mean given by (3.5) and the spike state density is the shifted log-normal distribution in (3.7). For TVP and GASX models, (3.4) is replaced by (3.6). The transition probabilities follow the specification illustrated in Subsection 3.3.3. We report the maximized log-likelihood value (LogLik), the corrected Akaike Information Criterion (AICc), the Bayesian Information Criterion (BIC), the p-values of Kolmogorov-Smirnov tests on observation classified as spike (KS Spike) and on standardized residuals classified in baseline regime (KS Base), the estimated number of spikes (N Spikes).

	Without extra-info			With extra-info	
	Static	Periodic	GAS	TVP	GASX
logLik	-5856	-5853	-5824	-5682	-5653
N. par	7	13	11	13	17
AIC	11726	11732	11670	11391	11341
BIC	11765	11805	11731	11465	11436
KS Base	0.000	0.000	0.000	0.000	0.000
KS Spike	0.986	0.740	0.896	0.989	0.895
N. Spikes	366	358	355	358	346

Figure 3.3: UK APX Spot Prices: diagnostic for GASX model with 2 regimes

Theoretical normal quantities vs sample quantiles for standardized residuals of the baseline regime (a) and empirical and theoretical distribution for the spike regime (b).

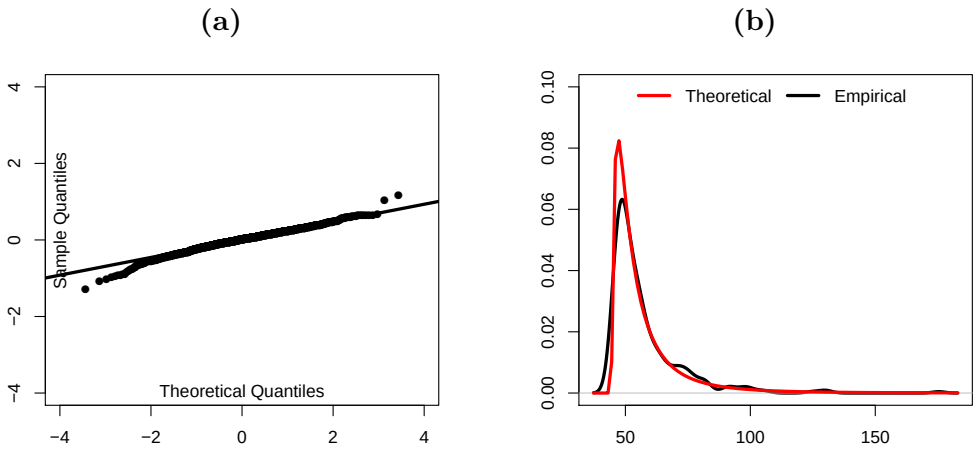


Table 3.2: UK APX Spot Prices: parameter estimates for models with 2 regimes

We have estimated the Markov switching model with two regimes on the stochastic component, y_t, of UK APX Spot Prices. For the first three models, the baseline state is given by the discretized mean-reversion diffusion process in (3.4) where the unknown value of $y_{t-1,b}$ is estimated by its conditional mean given by (3.5) and the spike state density is the shifted log-normal distribution in (3.7). For TVP and GASX models, (3.4) is replaced by (3.6). The transition probabilities follow the specification illustrated in Subsection 3.3.3. We report parameter estimates and numerical standard errors.

	Without exogenous information						With exogenous information			
	Static		Periodic		GAS		TVP		GASX	
α	11.823	(0.710)	11.842	(0.708)	11.838	(0.712)	14.707	(0.743)	14.762	(0.746)
β	0.307	(0.018)	0.308	(0.018)	0.307	(0.019)	0.383	(0.019)	0.383	(0.019)
σ^2_{base}	16.137	(0.586)	16.135	(0.583)	16.259	(0.592)	15.091	(0.541)	15.271	(0.539)
γ_{margin}							-0.093	(0.098)	-0.072	(0.098)
γ_{demand}							0.948	(0.105)	0.983	(0.105)
μ_{spike}	2.009	(0.033)	2.010	(0.066)	2.031	(0.067)	2.040	(0.066)	2.065	(0.062)
σ_{spike}	1.050	(0.047)	1.051	(0.047)	1.042	(0.047)	1.027	(0.045)	1.023	(0.043)
π_{bb}	0.923	(0.007)			0.920	(0.016)	0.959	(0.007)	0.909	(0.030)
π_{ss}	0.646	(0.027)			0.591	(0.043)	0.439	(0.040)	0.460	(0.018)
A_{bb}					0.384	(0.077)			1.539	(0.438)
A_{ss}					0.363	(0.217)			1.272	(0.275)
B_{bb}					0.955	(0.018)			0.197	(0.110)
B_{ss}					0.895	(0.084)			0.920	(0.033)
δ^{margin}_{bb}							1.597	(0.181)	1.752	(0.194)
δ^{margin}_{ss}							-0.694	(0.108)	-1.868	(0.305)
δ^{demand}_{bb}							0.407	(0.154)	-0.519	(0.140)
δ^{demand}_{ss}							7.615	(0.001)	0.537	(0.231)

3.4.2 Results for Markov-switching models with 3 regimes

In the previous subsection we noticed that the baseline regime was not correctly specified. The estimate of the variance parameter was quite high and the QQ-plot exhibited a deviation from the normality assumption in the left tail. To overcome those drawbacks we add a third regime to model the downward spike by using the inverted shifted log-normal distribution introduced in (3.8). In Table 3.3, we report the goodness-of fit results and in Table 3.4 the parameter and standard error estimates. The fitting performance is uniformly improved. The Kolmogorov-Smirnov statistics accepts the null hypothesis of correct specification for all the regimes. Also the empirical density functions of the two extreme regimes plotted against the theoretical density functions (Figure 3.4a and 3.4c) confirm that the distributions are well specified. The QQ-plot of the baseline regime standardized residual (Figure 3.4b) does not show a heavy left tail and deviations from normality, although not statistically significant, are equally distributed on both tails.

As already noticed for the Markov-switching models with 2 regimes, the GAS update mechanism for transition probabilities outperforms both the Static and the Periodic specification when no exogenous information is considered. When the forecasted reserve margin and the forecasted demand are included into the model, the GASX model outperforms the TVP model in terms of log-likelihood value that decreases about 86 points. This improvement is confirmed also by noticing that both information criteria take the smallest values in the novel specification.

Looking at Table 3.4, we see that for the baseline regime, the variance estimate is reduced to about 8 from the value equal to 16 estimated in the two-regime models while the estimate of β increases to 0.52-0.59 (it was 0.31 for models with two regimes). Also the mean estimate increases from 12 to 21/23. The magnitude of forecasted demand in the baseline mean equation reduces to about 0.7 from the value previously estimated equal to 0.9 while the reserve margin remains statistically not significant. The drastic reduction of the variance parameter estimate gives us another indication in favor of the three regime specification in which low prices are captured by the downward spike regime. The estimates related to the spike regime distribution remain substantially unchanged if compared with the two regime specifications.

In the GAS model, the transition probabilities dynamics are well specified. All the autoregressive parameter estimates suggest that the processes are highly persistent, furthermore, the parameters related to the scaled score are positive. In the TVP and GASX models, differently from the estimates obtained in the two regime specifications, the effects of the exogenous variables on transition probabilities have the same

sign. As expected, the forecasted reserve margin affects positively the probability of staying in the baseline regime and the probability of shifting from the spike to the baseline regime while affects negatively all the other transition probabilities. In particular, we highlight the large magnitude estimated for the parameter δ_{ds}, -2.999 for the TVP model and -2.977 for the GASX model. A reduction in the forecasted reserve margin increases largely the probability of switching from the downward spike regime to the spike regime. The forecasted demand influences positively all the transition probabilities. While in the two regime models, the parameter δ_{bb} was estimated positive in the TVP model and negative in the GASX model, now the estimates have the same sign. As we already pointed out, we cannot directly compare the autoregressive parameters estimated in the GAS model with these estimated in the GASX model. The dynamics are influenced by the exogenous information impact and result in two negative persistence parameter estimates, B_{sb} and B_{ss}. However, also in the GASX specification all the parameters related to the scaled score are positive as expected.

In Figure 3.5a-3.5c we show the expected duration of the three regimes for GAS and GASX models. For the regime i, the expected duration is given by $1/(1-\pi_{ii})$. Static transition probabilities imply constant expected duration that may not be reasonable for electricity prices. It is interesting to compare the time-varying expected duration with the spikes and downward spikes enlightened in Figure 3.5d. For the GASX model the exogenous information affects the expected duration dynamics over all the sample period. On the contrary, for the GAS model, when the probability of being in a particular regime is small, the related expected duration remains nearly equal to the previous level. When a particular regime is more likely to happen, also the expected duration reacts sensibly.

Figure 3.4: UK APX Spot Prices: diagnostic for GASX model with 3 regimes

Empirical and theoretical distribution for the spike regime (a) and the downward spike regime (c). Theoretical normal quantities vs sample quantiles for standardized residuals of the baseline regime (b).

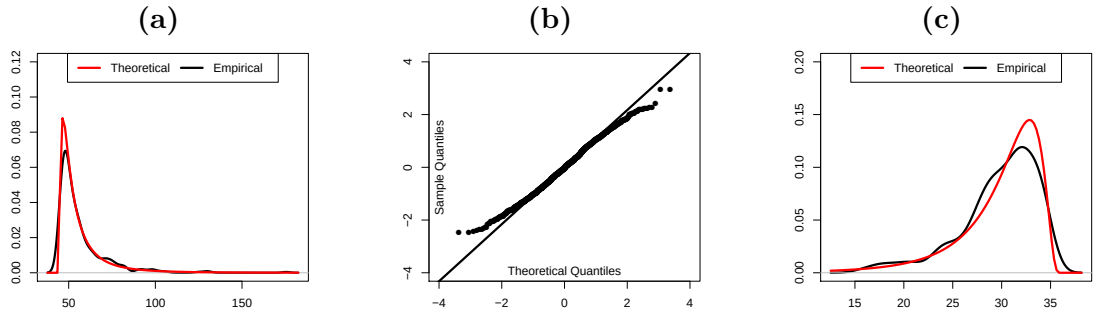


Table 3.3: UK APX Spot Prices: goodness-of-fit for models with 3 regimes

We have estimated the Markov switching model with three regimes on the stochastic component, y_t, of UK APX Spot Prices. For the first three models, the baseline state is given by the discretized mean-reversion diffusion process in (3.4) where the unknown value of $y_{t-1,b}$ is estimated by its conditional mean given by (3.5), the spike state density is the shifted log-normal distribution in (3.7), the downward spike regime density is the shifted inverse log-normal distribution in (3.8). For TVP and GASX models, (3.4) is replaced by (3.6). The transition probabilities follows the specification illustrated in Subsection 3.3.3. We report the maximized log-likelihood value (LogLik), the corrected Akaike Information Criterion (AICc), the Bayesian Information Criterion (BIC), the p-values of Kolmogorov-Smirnov tests on observation classified as spike/downward spike (KS Spike/ D. spike) and on standardized residuals classified in baseline regime (KS Base), the estimated number of spikes/downward spikes (N Spikes/ D. spikes).

	Without extra-info			With extra-info	
	Static	Periodic	GAS	TVP	GASX
logLik	-5302	-5282	-5201	-5126	-5040
N. par	13	31	25	27	39
AIC	10631	10627	10452	10307	10159
BIC	10704	10801	10593	10459	10379
KS Base	0.256	0.162	0.114	0.457	0.595
KS Spike	0.768	0.787	0.372	0.731	0.705
KS D. Spike	0.239	0.197	0.125	0.225	0.181
N. Spikes	405	408	396	395	379
N. D. Spikes	322	326	330	322	335

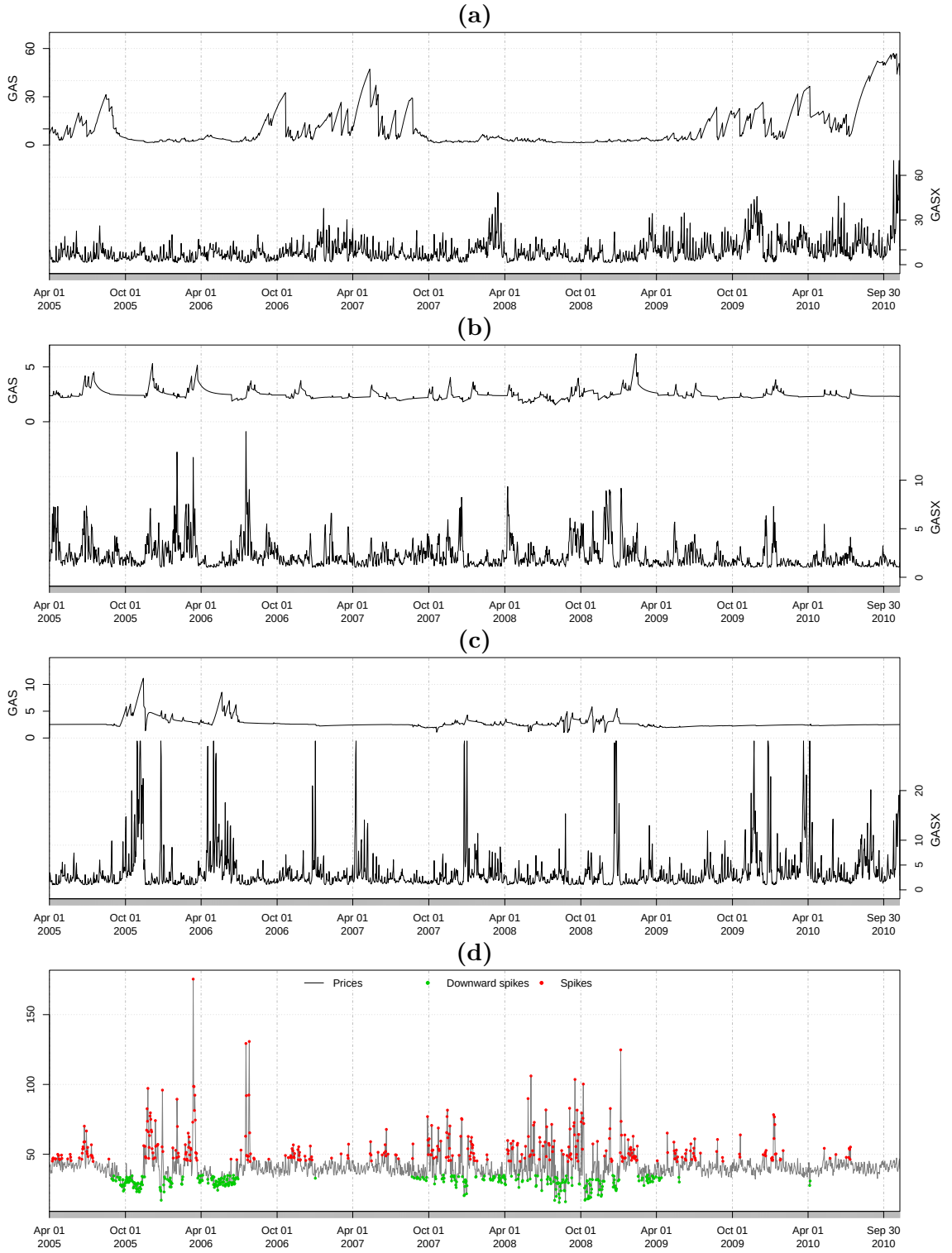
Table 3.4: UK APX Spot Prices: parameter estimates for models with 3 regimes

We have estimated the Markov switching model with three regimes on the stochastic component, y_t, of UK APX Spot Prices. For the first three models, the baseline state is given by the discretized mean-reversion diffusion process in (3.4) where the unknown value of $y_{t-1,b}$ is estimated by its conditional mean given by (3.5), the spike state density is the shifted log-normal distribution in (3.7), the downward spike regime density is the shifted inverse log-normal distribution in (3.8). For TVP and GASX models, (3.4) is replaced by (3.6). The transition probabilities follows the specification illustrated in Subsection 3.3.3. We report parameter estimates and numerical standard errors.

	Without exogenous information						With exogenous information			
	Static		Periodic		GAS		TVP		GASX	
α	20.992	(1.225)	21.115	(1.220)	21.283	(1.200)	23.576	(1.234)	23.468	(0.462)
β	0.528	(0.031)	0.531	(0.031)	0.534	(0.030)	0.593	(0.031)	0.588	(0.012)
σ^2_{base}	8.567	(0.393)	8.418	(0.375)	8.575	(0.394)	8.136	(0.366)	8.260	(0.253)
γ_{margin}							-0.027	(0.082)	-0.034	(0.082)
γ_{demand}							0.706	(0.096)	0.706	(0.091)
μ_{spike}	1.912	(0.063)	1.901	(0.062)	1.975	(0.065)	1.948	(0.064)	2.007	(0.061)
σ_{spike}	1.071	(0.044)	1.076	(0.044)	1.059	(0.045)	1.050	(0.044)	1.044	(0.020)
μ_{down}	1.669	(0.043)	1.653	(0.043)	1.635	(0.045)	1.669	(0.043)	1.631	(0.045)
σ_{down}	0.673	(0.032)	0.683	(0.032)	0.676	(0.034)	0.666	(0.032)	0.678	(0.033)
π_{bb}	0.848	(0.007)			0.859	(0.024)	0.886	(0.007)	0.329	(0.001)
π_{sb}	0.330	(0.010)			0.383	(0.013)	0.515	(0.019)	0.940	(0.009)
π_{db}	0.234	(0.026)			0.358	(0.082)	0.287	(0.034)	0.373	(0.001)
π_{bs}	0.102	(0.006)			0.091	(0.008)	0.065	(0.006)	0.347	(0.030)
π_{ss}	0.636	(0.010)			0.582	(0.013)	0.421	(0.019)	0.043	(0.006)
π_{ds}	0.025	(0.011)			0.032	(0.016)	0.011	(0.008)	0.252	(0.002)
A_{bb}					0.446	(0.086)			0.486	(0.084)
A_{sb}					0.077	(0.092)			0.010	(0.004)
A_{db}					0.252	(0.098)			0.308	(0.098)
A_{bs}					1.103	(0.279)			8.583	(1.935)
A_{ss}					0.461	(0.216)			4.386	(0.995)
A_{ds}					4.073	(2.637)			2.257	(0.775)
B_{bb}					0.994	(0.003)			0.992	(0.001)
B_{sb}					0.992	(0.008)			-0.979	(0.003)
B_{db}					0.987	(0.017)			0.983	(0.002)
B_{bs}					0.932	(0.025)			0.232	(0.061)
B_{ss}					0.951	(0.037)			-0.141	(0.166)
B_{ds}					0.743	(0.180)			0.902	(0.005)
δ^{margin}_{bb}							0.324	(0.157)	0.119	(0.171)
δ^{margin}_{sb}							0.069	(0.383)	0.097	(0.368)
δ^{margin}_{db}							-0.661	(0.233)	-0.896	(0.237)
δ^{margin}_{bs}							-1.045	(0.220)	-1.614	(0.263)
δ^{margin}_{ss}							-0.832	(0.389)	-1.167	(0.399)
δ^{margin}_{ds}							-2.999	(0.749)	-2.977	(0.529)
δ^{demand}_{bb}							0.703	(0.163)	0.725	(0.162)
δ^{demand}_{sb}							0.501	(0.350)	0.564	(0.188)
δ^{demand}_{db}							0.392	(0.172)	0.233	(0.323)
δ^{demand}_{bs}							1.137	(0.204)	1.072	(0.230)
δ^{demand}_{ss}							0.842	(0.355)	0.904	(0.396)
δ^{demand}_{ds}							1.151	(0.484)	1.574	(0.556)

Figure 3.5: UK APX Spot Prices: graphical results for GAS and GASX models with 3 regimes

Comparison between GAS and GASX model for expected duration of the baseline regime (a), spike regime (b) and downward spike (c) regime . Stochastic component y_t and identified downward spikes and spikes (d) from the TVP model.



3.5 Conclusions

Electricity prices represent an interesting challenge for econometricians. In order to explain the spiky behavior that characterizes their evolution over time, Markov-switching models are broadly used. Although many authors proposed to use a two-regime specification we found empirical evidence in favor of a more complex three-regime specification. A discretized mean-reverting diffusion process is able to capture the price evolution when the market is quiet while low and high prices are modeled by a log-normal and a shifted log-normal distributions, respectively.

The spikes do not occur homogeneously over time suggesting that static transition probabilities may be too restrictive in modeling electricity prices. To overcome this limitation of the standard Markov switching model, we compare several alternatives to model dynamically the transition probabilities. In particular, we assumed a seasonal structure or we used exogenous variables like forecasted reserve margin and demand to explain their dynamics. Again, we adopted the novel GAS updating mechanism which exploits the information contained in the score of the conditional density to drive the transition probabilities dynamics. We extended the GAS formulation to include auxiliary information into the dynamics of transition probabilities explicitly and we called the new specification GAS-eXogenous (GASX).

The alternatives have been compared by analyzing the UK APX spot prices and we have found a strong empirical evidence in favor of the score driven specifications. In particular, the GASX model outperforms the counterparts suggesting that the transition probabilities dynamics may be explained accurately by combining the autoregressive structure driven by the scores of the conditional densities and the exogenous information. Our findings are similar to those of the recent literature in pointing out how the forecasted reserve margin and forecast demand are able to improve the model fit as well as to describe the occurrence of spikes.

Chapter 4

Transformed polynomials for modeling conditional volatility

Abstract

We propose a flexible model for filtering time-varying conditional volatilities. In particular, we make use of novel transformed polynomial functions to update the unobserved time-varying conditional volatility parameter. The flexible updating equation is shown to approximate arbitrarily well any continuous function and to have known convergence rates of approximation on Hölder spaces. A Monte Carlo study explores the finite sample properties of the estimator. Finally, a number of applications shows the good performance of the model in empirically relevant settings.

Some key words: Conditional volatility, GARCH, Transformed polynomial function, Semi-nonparametric models.

This chapter has been written during my visiting period at Department of Econometrics and Operation Research, at VU University, Amsterdam. I am deeply thankful to Francisco Blasques for the original discussion and to Andre Lucas and Siem Jan Koopman for the numerous suggestions that improved this final version.

4.1 Introduction

Financial time-series are often characterized by clusters of volatility and heavy tails. The autoregressive heteroskedasticity (ARCH) model introduced in the seminal contributions of Engle (1982) and the generalized autoregressive heteroskedasticity (GARCH) model of Bollerslev (1986) have become the benchmark against which more sophisticated models are compared. The GARCH(1,1) models a sequence of returns $\{y_t\}_{t\in\mathbb{Z}}$ as

$$y_t = \sigma_t u_t, \quad \forall t \in \mathbb{Z},$$

where $\{u_t\}_{t\in\mathbb{Z}}$ is a sequence of normal independently distributed random variables, and σ_t evolves according to

$$\sigma_{t+1}^2 = \omega + \alpha y_t^2 + \beta \sigma_t^2, \quad \forall t \in \mathbb{Z}.$$

In a parametric setting, many efforts have been made to model the asymmetric response of volatility to positive and negative past returns. The Generalized Quadratic ARCH (GQARCH) model proposed independently by Engle and Ng (1993) and Sentana (1995), the Threshold GARCH (TGARCH) model of Zaköian (1994), the GJR-GARCH of Glosten et al. (1993), the Asymmetric Power ARCH (APARCH) of Ding et al. (1993) are the most popular alternatives, capable of addressing this phenomenon which is known as leverage effect and was originally documented by Black (1976). In a semi-parametric, nonparametric or semi-nonparametric setting, many models have been designed to overcome some of the restrictiveness of the parametric assumptions in Gaussian GARCH models. For example, Linton (1993) and Drost and Klaassen (1997) proposed kernel-based estimates of the density of the error term. A different stream of research focused on the functional form of the volatility function. These nonparametric models focus however on the ARCH volatility and do not nest the GARCH(1,1). Pagan and Schwert (1990), Pagan and Hong (1991) and Linton and Mammen (2005) considered nonparametric ARCH model where the conditional volatility is a smooth function of previous values of returns. The estimated news impact curves for S&P500 returns in Linton and Mammen (2005) turn out to be considerably asymmetric with a minimum at positive returns (rather than zero) and non-quadratic tails. More recently, Audrino and Bühlmann (2001,2009) proposed an estimation algorithm for the more general case where volatility is assumed to be a smooth but unknown function of past returns and past realizations of the volatility. However, they did not derive the stochastic properties of their model or establish asymptotic properties for their

estimator.

This chapter proposes a flexible model for filtering time-varying conditional volatilities. Following Audrino and Buhlmann (2009), we assume that the conditional variance depends on past values of returns and volatility through an unknown continuous function ϕ, i.e.

$$\sigma_{t+1}^2 = \phi(\sigma_t^2, y_t), \quad \forall t \in \mathbb{Z}.$$

We make use of novel transformed polynomial functions introduced by Blasques (2014) to update the unobserved time-varying conditional volatility parameter. In Section 4.2 we introduce the transformed polynomial functions to model the conditional volatility. In Section 4.3, we show through an extensive Monte Carlo simulation study the ability of our novel approach in approximating the true and unknown function both under correct specification and misspecification. In Section 4.4 we analyze the returns of ten different stocks which compose the S&P100 Index and we find that our method is robust. The comparison for a number of parametric and semi-nonparametric models is illustrated on IBM stock returns. Finally, in Section 4.5 we conclude.

4.2 Transformed polynomial functions

The transformed polynomial function has been introduced by Blasques (2014) to model the conditional expectation. Assume that the univariate sequence $\{y_t\}_{t\in\mathbb{Z}}$ is generated by

$$y_t = \phi(y_{t-1}) + \varepsilon_t, \quad t \in \mathbb{Z},$$

where $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a i.i.d. sequence. When the true function ϕ is assumed to be continuous in y, it is shown that it can be approximated arbitrarily well by the $k-$order transformed polynomial function

$$\tilde{p}_{k,\beta}(y, \boldsymbol{\theta}) = \theta_0 + \theta_1 y + \left(\sum_{i=2}^{k} \theta_i y^i\right) \cdot \phi_\beta(y), \tag{4.1}$$

where $\boldsymbol{\theta} = (\theta_0, \theta_1, \cdots, \theta_k) \in \mathbb{R}^{k+1}$ is the parameter vector and the transformation function is defined as $\phi_\beta(y) = \exp(\beta\ y^2)$ with β is a scalar satisfying $\beta < 0$. The transformation function plays a crucial role in bounding the tail behavior by a linear function.

If the researcher follows a parametric approach and adopts a given fixed order k, then the parametric ML estimator:

(i) converges to a pseudo-true parameter $\boldsymbol{\theta}_0^* \in \Theta_k$ that minimizes the Kullback-Leibler divergence between the true probability measure of the data and the model implied measure if k is not large enough for the model to be well specified;

(ii) converges to the true parameter $\boldsymbol{\theta}_0 \in \Theta_k$ if k is large enough for the model to be well specified.

Instead, in a semi-nonparametric (SNP) approach, the researcher lets the order k diverge to infinity with sample size T. As a result, the transformed polynomial function $\tilde{p}_{k,\beta}$ can be used in conjunction with a sieve maximum likelihood (ML) estimator to consistently estimate any true function ϕ that is continuous in y. As $\beta \to 0$, the consistency of the SNP estimator is guaranteed by the fact that the space of transformed polynomials $\tilde{\mathbb{P}}(\mathcal{Y})$ is dense on the space of continuos functions on $\mathcal{X}$, denoted by $\mathbb{C}(\mathcal{X})$ in sup-norm for every compact $\mathcal{Y} \subset \mathbb{R}$, see Proposition 1 in Blasques (2014).

In the analysis of financial time series, the researcher is interested in modeling the conditional variance dynamics rather than the conditional mean. The data generating process for the time-varying volatilities is assumed to take the form

$$y_t = \sigma_t u_t, \quad \forall t \in \mathbb{Z},$$

where $\{u_t\}_{t\in\mathbb{Z}}$ is a i.i.d. sequence with common density $u_t \sim f_u$. Typical choices for f_u are normal and Student's t-distributions. The volatility σ_t evolves according to

$$\sigma_{t+1}^2 = \phi(y_t, \sigma_t^2), \quad \forall t \in \mathbb{Z},$$

where ϕ is a continuous function in y_t and σ_t^2. In order to approximate the unknown function ϕ the transformed polynomial function in (4.1) can be generalized as a function of y_t and σ_t^2. This yields the filtering equation of the form

$$\tilde{\sigma}_{t+1}^2 = \tilde{p}^k(y_t, \tilde{\sigma}_t^2; \boldsymbol{\theta}), \quad \forall t \in \mathbb{N}.$$

where $\boldsymbol{\theta} \in \Theta_k \subseteq \mathbb{R}^{n(k)}$ is the vector of parameters that parametrizes the transformed polynomial $\tilde{p}^k(y_t, \tilde{\sigma}_t^2; \boldsymbol{\theta})$ of order k. The dimension of parameter vector, $n(k)$, is an increasing function of transformed polynomial order k. While in modeling the conditional expectation it is convenient that the transformation function recovers a linear behavior in the tails, the positivity of the conditional variance can be guaranteed allowing the tail behavior to be dominated by the quadratic term in y_t and the linear

term in $\tilde{\sigma}_t^2$. Thus, the transformed polynomial $\tilde{p}^k(y_t, \tilde{\sigma}_t^2; \boldsymbol{\theta})$ takes the form

$$\begin{aligned}\tilde{p}^k(y_t, \tilde{\sigma}_t^2; \boldsymbol{\theta}) =& \psi_{00} + \psi_{20}\ y_t^2 + \psi_{01}\ \tilde{\sigma}_t^2 \\ &+ \sum_{1 \leq i+j \leq k} \psi_{ij}\ y_t^i\ \tilde{\sigma}_t^{2j}\ \exp\left(\beta_i^y\ y_t^2 + \beta_j^{\sigma^2}\ \tilde{\sigma}_t^4\right)\end{aligned} \tag{4.2}$$

where $(i, j) \in \{\mathbb{N}_0 \times \mathbb{N}_0\} \setminus (2,0)\ (0,1)\}$. The vector $\boldsymbol{\theta}$ contains the unknown parameters: the intercept ψ_{00}, all polynomial coefficients $\psi_{i,j}$, as well as the coefficients of the exponential term β_i^y and $\beta_j^{\sigma^2}$. As in (4.1), the exponential term coefficients have to be not positive. It is quite intuitive to notice that in the tails the behavior of the transformed polynomial function is bounded by the parametric GARCH(1,1) while the true function $\phi(y_t, \sigma_t^2)$ is well approximated in those regions for which the data provide information.

Blasques (2014) shows that the transformed polynomials functions can (i) approximate arbitrarily well any continuous functions, (ii) perform better than polynomials in approximating contracting functions, and, (iii) ensure that the filter is invertible and that the parameter vector $\boldsymbol{\theta}$ is identified. As a data generating process (DGP), the transformed polynomials also ensure that the process is strictly stationary, ergodic, has fading memory and bounded unconditional moments. However, the results in Blasques (2014) may not apply directly to the conditional volatility model. Indeed the transformed polynomial functions in (4.2) are bivariate as they are function of y_t and $\tilde{\sigma}_t^2$. Furthermore, the inclusion of the filtered conditional volatility sequence may complicate the stochastic properties of the process. In this paper we investigate the finite sample properties of the MLE through a Monte Carlo study.

4.3 Finite sample properties

4.3.1 Maximum likelihood estimator under correct specification

In Section 4.2 we pointed out the difference between a parametric and a semi-nonparametric approach to the parameter estimation. In particular, when the order of the transformed polynomial, k, is large enough or the model is correctly specified, we expect that the MLE, $\hat{\boldsymbol{\theta}}(\sigma_1^2)$, converges to the true parameter $\boldsymbol{\theta}_0$. In this first experiment, we sample an artificial sequence of $\{y_t\}_{t \in \mathbb{Z}}$ from a standard normal distribution with conditional variance given by the transformed polynomial function introduced in (4.2). We set $k = 3$, and $\boldsymbol{\theta} = [\psi_{00},\ \psi_{10},\ \beta_1^y,\ \psi_{01},\ \beta_1^{\sigma^2},\ \psi_{20},\ \psi_{11},\ \psi_{30},\ \beta_3^y] = [0.5,$-

0.15, −0.08, 0.60, −0.01, 0.25, 0.35, −0.15, 0.05]. This choice for the parameters leads to a conditional variance which reacts asymmetrically to positive and negative past shocks. Furthermore, this leverage effect is more pronounced when the market is already excited. We simulate a very long sequence of observations setting $T = 10000$. A burn-in period equal to $BI = 500$ observations is simulated to reduce the effect of starting point, σ_1^2, in the estimates.

The results averaged over 5000 replications reported in Table 4.1 seem to suggest the consistency of parametric ML estimator. The empirical distribution of the estimates, shown in Figure 4.2, does not deviate from the normality assumption.

The News Impact Curve (NIC) function has been introduced by Engle and Ng (1993) to measure the impact of shocks y on the one-step ahead conditional variance. For the GARCH(1,1) model, it takes the form $NIC(y) = \psi_{00} + \psi_{20}y^2$. Given the flexibility of the TP-GARCH model which takes into account also the interaction effect between the shock y and the volatility σ^2, we define the News Impact Curve function as the estimated transformed polynomial function in (4.2) in which the volatility is fixed to an arbitrary value, $\bar{\sigma}^2$, $NIC(y, \bar{\sigma}^2) = \tilde{p}^k(y, \bar{\sigma}^2)$. If we look at Figure 4.1, the parametric ML estimator for $\boldsymbol{\theta}$ under correct specification seems to be able to estimate correctly also the News Impact Curve functions.

Table 4.1: Simulation results I: maximum likelihood estimator under correct specification

We have simulated 5000 time series of length 10000 from the DGP $y_t = \sigma_t^2 u_t$, $\forall t \in \mathbb{N}$, where $u_t \sim N(0,1)$ and $\sigma_{t+1}^2 = \tilde{p}^k(y_t, \sigma_t^2; \boldsymbol{\theta})$, see details in the text. Taking the first 500 observations as burn-in, we estimated the parameter vector, θ, by maximizing numerically the log-likelihood function. We report the sample averages and the Monte Carlo Standard Errors for the parameter vector estimates, the maximized log-likelihood value (LogLik), the averaged mean squared error of the one-step ahead forecast of σ_t (MSE) and its averaged mean absolute value (MAE).

	ψ_{00}	ψ_{10}	β_1^y	ψ_{01}	$\beta_1^{\sigma^2}$	ψ_{20}	ψ_{11}	ψ_{30}	β_3^y
True	0.50	-0.15	-0.08	0.60	-0.01	0.25	0.35	-0.15	-0.05
Mean	0.5014	-0.1682	-0.0859	0.6000	-0.0102	0.2494	0.3622	-0.1490	-0.0483
MC SE	0.0005	0.0021	0.0006	0.0002	0.0002	0.0002	0.0011	0.0003	0.0001

	logLik	MSE $\tilde{\sigma}^2$	MAE $\tilde{\sigma}^2$
Mean	-19047	0.0782	0.1204
MC SE	2	0.0013	0.0005

Figure 4.1: News Impact Curve functions under correct specification

True (red line) and 95% point-wise confidence intervals (blue region) over 5000 MC replications. Sample size T=10000, BI =500.

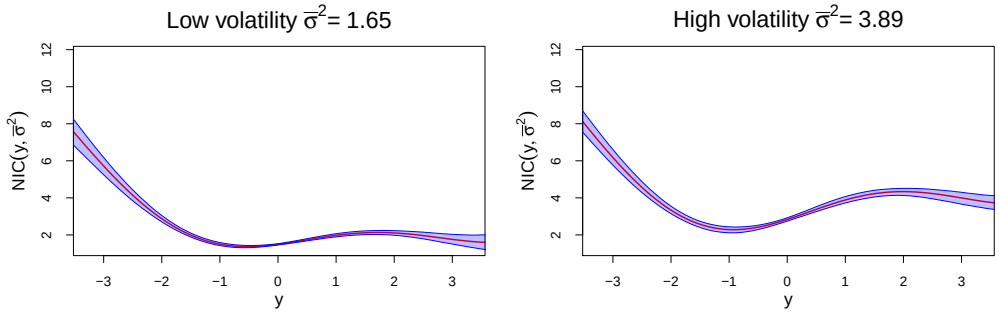
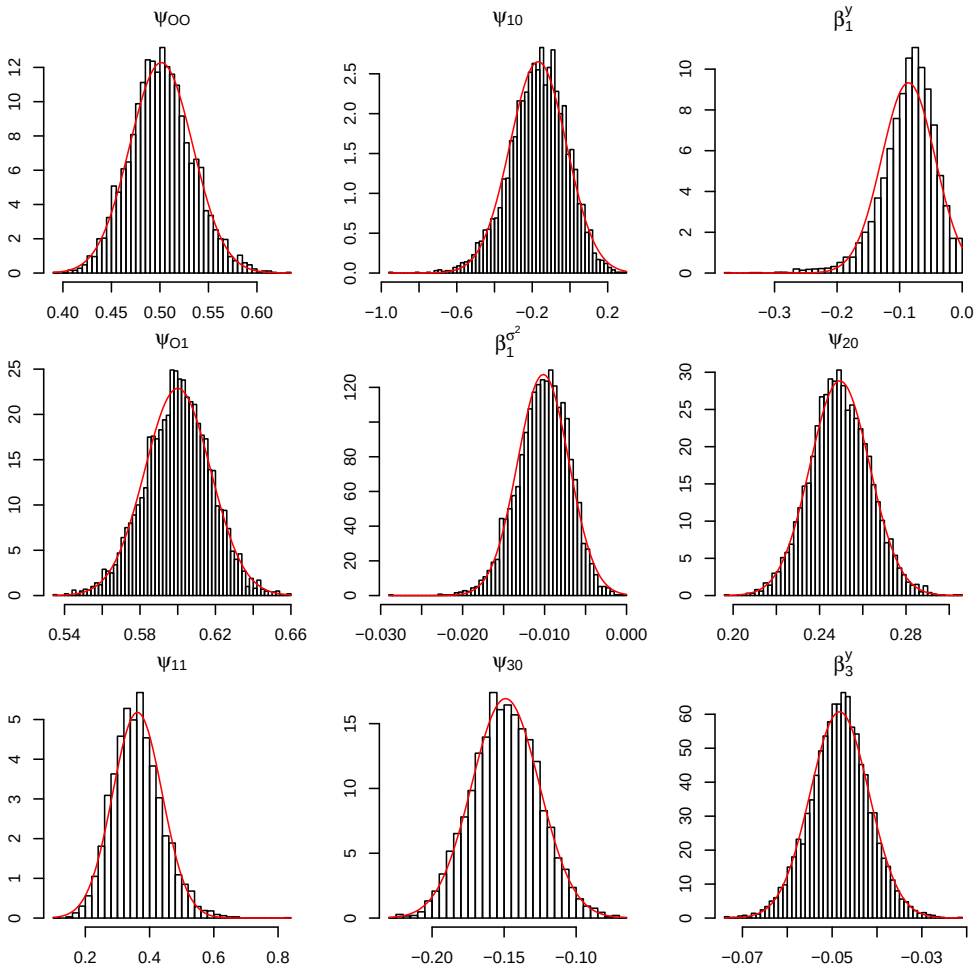


Figure 4.2: Parameter estimation under correct specification

Empirical density (histogram) and normal density (red line) over 5000 MC replications. Sample size T=10000, BI =500.



4.3.2 Maximum likelihood estimator under incorrect specification

In a semi-nonparametric approach, we let the order k of the transformed polynomial function in (4.2) diverge to infinity as the sample size increases. The sieve ML estimator is expected to consistently estimate any true function ϕ continuous in (y_t, σ_t^2). Similarly to the study in Blasques (2014), the aim of this second Monte Carlo simulation experiment is to recreate artificially this situation. We use the following functional form for the conditional variance,

$$\begin{aligned}\sigma_{t+1}^2 = \phi(y_t, \sigma_t^2) =& 0.5 + 0.25y_t^2 + 0.5\sigma_t^2 \\ &+ 5.25\left[1 + \tanh(-0.5y_t)\right]\left\{0.5 + 0.2\tanh\left[0.5(\sigma_t^2 - 9)\right]\right\}, \qquad (4.3)\end{aligned}$$

where $\tanh(\cdot)$ denotes the hyperbolic tangent. We remark that the function defined in (4.3) cannot be expressed as a finite number of polynomial terms. Furthermore, the conditional volatility exhibits a leverage effect, particularly pronounced when the market is excited. Then y_t, $t = 1, \cdots, T$ are sampled from a normal distribution with mean zero and variance σ_t^2. For three sample sizes, $T = 1000, 2500$ and 5000 we estimated a GARCH(1,1) model and a TP-GARCH model in which order k and/or number of parameter n increase along with sample size. A burn-in period equal to $BI = 500$ observations is simulated to reduce the effect of starting point, σ_1^2. The results are averaged over 1000 replications. In Table 4.2 we report the findings on goodness-of-fit. For all sample sizes, the TP-GARCH model outperforms the GARCH counterpart in fitting ability. This is not surprising as the GARCH model is nested in the TP-GARCH model. More interestingly, the superiority of the TP-GARCH model is confirmed by the corrected Akaike Information Criterion (AICc) and the Bayesian Information Criterion (BIC) of Schwarz (1978). The AICc is the original AIC of Akaike (1974), but with a stronger finite sample penalty as proposed by Hurvich and Tsai (1991). Some caution should be used in looking at BIC as all the estimated models are clearly misspecified. Both the Mean Squared Error (MSE) and the Mean Absolute Error (MAE) computed between the true σ^2 and the filtered estimate $\tilde{\sigma}^2$ decrease as the sample size increases but the filtered estimate obtained with the TP-GARCH model is always more precise than its Standard GARCH counterpart. A remarkable exception is given by the MSE measure for $T = 5000$ that has a bigger value than the one for $T = 2500$. We have chosen not to report the parameter vector estimates but the third order transformed polynomial function with seven extra-parameters used for the largest sample size suffers from a certain degree of uncertainty that increases

the variability in the filtered conditional variance, especially for extreme values of y and/or $\tilde{\sigma}^2$. On the other hand, the MAE measure, less sensitive to extreme values by construction, results in a smaller value for $T = 5000$ than for 2500. We leave further investigation to future research. If we decided to adopt a rule to fix the transformed function coefficients *a priori* reducing the number of parameters that have to be estimated, as done in the empirical application, the issue could be overcome.

In order to better understand the ability of the semi-nonparametric approach in estimating the true function $\phi(y_t, \sigma_t^2)$ we define another precision measure, namely the absolute norm of the News Impact Curve function, denoted by $\ell_1(\bar{\sigma}_t^2)$, for a fixed value of the conditional variance, $\bar{\sigma}_t^2$,

$$\ell_1(\bar{\sigma}^2) = \int_{\mathbb{R}} \left| \phi(y, \bar{\sigma}^2) - \tilde{p}^k(y, \bar{\sigma}^2) \right| \, dP_0(y), \tag{4.4}$$

where $P_0(y)$ denotes the true unconditional probability measure of y_t. A consistent estimator for (4.4) is given by

$$\tilde{\ell}_1(\bar{\sigma}^2) = \frac{1}{T} \sum_t \left| \phi(y_t, \bar{\sigma}^2) - \tilde{p}^k(y_t, \bar{\sigma}^2) \right|. \tag{4.5}$$

The TP-GARCH model clearly outperforms the parametric counterpart both for a low level of $\bar{\sigma}^2$ equal to the 20th percentile of the true conditional variance (ℓ_1^L in Table 4.2) and for a high level equal to the 80th percentile (ℓ_1^H in Table 4.2). As the sample size increases, the precision measure improves uniformly. For the TP-GARCH model this result seems able to confirm the consistency of the sieve ML estimator. The difference becomes more evident if we take a look at Figure 4.3 in which the 95% point-wise confidence intervals for the News Impact Curve functions are displayed together with the true function. While the confidence bands estimated by the TP-GARCH model always contain the true function, this is not true for those estimated by the GARCH model. Although the true model is clearly misspecified in both cases, one could argue that the approximation implied by the TP-GARCH model leads to a consistent estimator of $\phi(y_t, \sigma_t^2)$ whereas the GARCH model is unable to approximate the true function even if the sample size diverges.

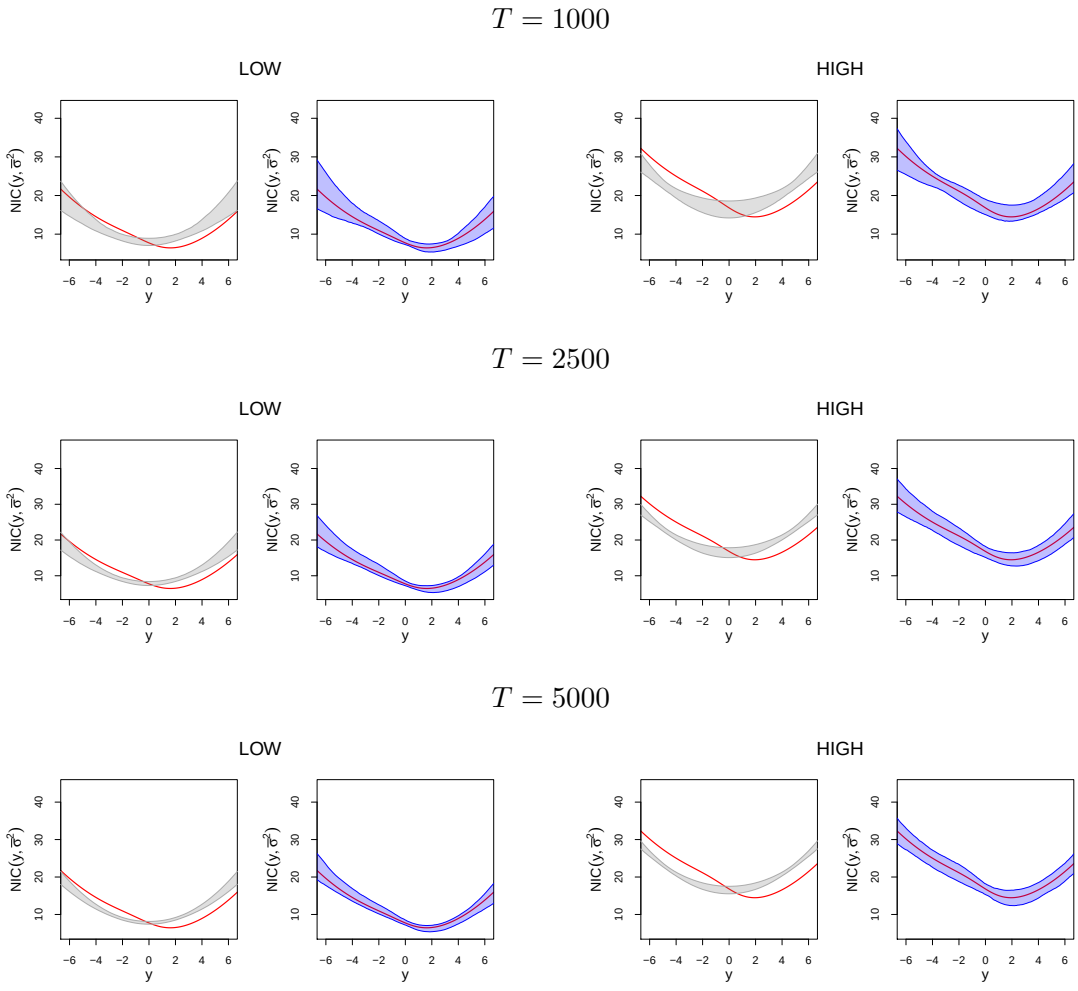
Table 4.2: Simulation results II: maximum likelihood estimator under incorrect specification

We have simulated 1000 time series from the model $y_t = \sigma_t^2 u_t$, $t = 1, \cdots, T$, where $u_t \sim N(0,1)$ and σ_t^2 is given by Eq. 4.3 for three sample sizes, $T = 1000, 2500$ and 5000. For each series, we estimated the GARCH (1,1) and the TP-GARCH model by maximizing numerically the log-likelihood function. In the latter case, we set a larger degree, k, of the transformed polynomial and/or a larger number of parameter, n, as T increases. We report the sample averages over the 1000 simulated series of the maximized log-likelihood value (LogLik), the AICc and the BIC, the mean squared error of the one-step-ahead forecast of σ^2 (MSE), its mean absolute error value (MAE), the estimate of the absolute norm of the NIC function for $\tilde{\sigma}^2$ equal to the 20th and 80th percentiles of the true σ^2, $\tilde{\ell}_1^L$ and $\tilde{\ell}_1^H$, defined in (4.5).

Sample Size	Model	k	n	logLik	AICc	BIC	MSE $\hat{\sigma}^2$	MAE $\hat{\sigma}^2$	$\tilde{\ell}_1^L$	$\tilde{\ell}_1^H$
1000	GARCH	2	3	-2764	5534	5549	26.80	3.67	2.29	3.32
	TP-GARCH	2	5	-2746	5503	5528	13.09	2.02	1.09	1.32
2500	GARCH	2	3	-6911	13828	13845	22.68	3.54	2.19	3.23
	TP-GARCH	2	9	-6860	13738	13790	6.01	1.39	0.84	1.12
5000	GARCH	2	3	-13827	27661	27681	21.36	3.50	2.15	3.20
	TP-GARCH	3	16	-13724	27480	27584	8.52	1.26	0.77	1.02

Figure 4.3: News Impact Curve functions under incorrect specification

True (red line) and 95% point-wise confidence intervals for the GARCH model (grey region) and for the TP-GARCH model (blue region) over 1000 MC replications. Sample size T=1000, 2500, 5000.



4.4 Applications

4.4.1 Comparative results for ten different stocks

In Section 4.2, we introduced the TP-GARCH model to overcome some restrictiveness of the GARCH model like the inability to deal with the leverage effect and the lack of interaction between volatility and returns in the functional form of the conditional variance. Here, we propose to compare empirically the novel method with other well-established counterparts. In the broad class of parametric candidates, we have chosen the GARCH(1,1) model for its simplicity and the GQARCH model, proposed by Sentana (1995), which includes the leverage effect in a very parsimonious way. The Spline-GARCH, described in Chen (2011), is chosen as semi-nonparametric alternatives because it nest the GARCH(1,1) model, it is relatively simple to estimate although it has strong theoretical properties. We analyzed the daily log-returns from ten different stocks exchanged in the USA stock market observed from January 1, 1993 to December 31, 2013, for a total of $T = 5477$ observations. We analyzed the stocks form the following firms: Apple, IBM, Coca Cola, Boeing, Mc Donald's, Microsoft, Exxon, Walt Disney, Citigroup and General Electrics.

In financial econometrics, it is widely known that returns are affected by leptokurtosis, i.e. heavy tails. For this reason, we assume both Gaussian and Student's t-density for the marginal distribution of the errors, u_t. The Student's t-distribution, which contains the normal one as special case, has been broadly used to deal with the leptokurtosis of the financial returns. More complex alternatives are left to future research.

As we already pointed out, all the competing models nest the GARCH (1,1) formulation, in which the conditional variance evolution can be expressed as

$$\sigma_t^2 = \psi_{00} + \psi_{20} y_{t-1}^2 + \psi_{01} \sigma_{t-1}^2. \tag{4.6}$$

In the GQARCH model, the asymmetric effect is modeled by introducing y_{t-1} in (4.6)

$$\sigma_t^2 = \psi_{00} + \psi_{20} y_{t-1}^2 + \psi_{01} \sigma_{t-1}^2 + \psi_{10} y_{t-1}. \tag{4.7}$$

Sentana (1995) has extensively studied the parameter vector restrictions that ensure the stationarity and positiveness of the volatility filtered sequence. In the Spline-

GARCH model, the conditional volatility is specified as follows

$$\sigma_t^2 = \psi_{00} + \psi_{01}\sigma_{t-1}^2 + \sum_{d=1}^{D} \psi_{d0} y_{t-1}^d + \sum_{k=1}^{K} \gamma_k (y_{t-1} - \xi_k)_+^D, \tag{4.8}$$

where D is the degree of the polynomial and ξ_k, $k = 1, \ldots, K$ are the internal knots, typically estimated from the empirical distribution of the data. Common choices for D and K are 3 and 2, respectively. Following Chen (2011) the GARCH(1,1) formulation is recovered by setting $\gamma_k = 0$ for all k and $\psi_{d0} = 0$ for all $d \neq 2$. By using the GARCH estimates as starting values, the parameters are jointly estimated by maximizing numerically the penalized log-likelihood, where the penalization is introduced to ensure the positivity of the conditional volatility.

A similar approach can be used for our formulation. After having obtained the GARCH estimates, the other parameters can be added and jointly estimated. However, here we adopt a rule to fix the β parameters rather than estimating them, differently from what we have done in our simulation studies. As we have already seen, the transformation function $\exp(\beta_i^y \; y_t^2 + \beta_j^{\sigma^2} \; \tilde{\sigma}_t^{2j})$ in (4.2) plays an important role in avoiding a divergent behavior in the tails. In the same spirit of Linton and Mammen (2005), who proposed to combine their semi-nonparametric estimator with a parametric model for the tails, we can choose values of β^y and β^{σ^2} so that the parametric GARCH functional form is recovered for values larger in absolute values than an arbitrary but reasonable threshold. More precisely, we suggest the following rule for the parameters related to the transform functions on the return y, β_i^y,

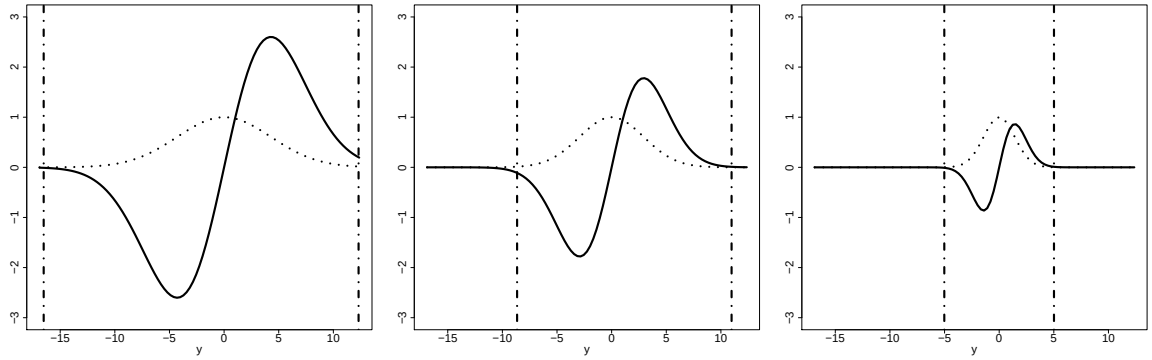
$$\beta_i^y = \frac{\log\left(e/|y_q|^i\right)}{y_q^2}, \quad \forall i = 1, 2, \cdots, \tag{4.9}$$

where e is a small constant, y_q is the qth percentile of the empirical cumulative distribution of the data. Analogously, the parameters related to the conditional variance, $\beta_j^{\sigma^2}$, can be fixed by a similar rule where y_q is replaced by $\tilde{\sigma}_q^2$. For estimating $\tilde{\sigma}_q^2$, we suggest to use the filtered conditional variance obtained by the GARCH(1,1) model. In Figure 4.4 we show the effects of choosing different y_q on β_1^y. Differently from the method proposed by Linton and Mammen (2005) the nature of the transform polynomial function guarantees the continuity of $\tilde{p}^k(y_t, \tilde{\sigma}_t^2; \boldsymbol{\theta})$ in $(y_t, \tilde{\sigma}_t^2)$.

For the TP-GARCH model, a quadratic and a cubic transformed polynomial functions are estimated. Then, in order to obtain a parsimonious representation, the parameters in the cubic expansion which are not statistically different from 0 are discarded, i.e. which a significance smaller than the 0.05.

Figure 4.4: Cutting rule for β_1^y

Cutting rule for three different choices of y_q: $\max(y_{0.0001}, y_{0.9999})$ (on left), $\max(y_{0.001}, y_{0.999})$ (in the middle),$\max(y_{0.01}, y_{0.99})$ (on right). Dashed lines: $\exp(-\beta_1^y y^2)$, straight lines: $y \exp(-\beta_1^y y^2)$. The vertical dashed lines corresponds to the empirical percentiles of y. $e = 0.01$.



In Table 4.3 we report the Bayesian Information Criterion (BIC) estimated for the different models. The BIC values have been chosen to compare the fitting ability of the competing models because we found that this criterion penalizes the models with a large parameter vector more strongly than other criteria like AICc. A overfitting issue could occur if a model with too many parameters was selected and this situation is particularly dangerous in a semi-nonparametric setting. Nevertheless some caution should be used as the Bayesian Information criterion is asymptotic valid only if the model is correctly specified. Summarizing what we learn from Table 4.3 we can say the following: (i) The Likelihood Ratio Test (not shown here) performed on the degree-of-freedom parameter estimated when the error term is assumed Student's t-distributed is statistically relevant for all stocks. Also the BIC values are estimated bigger when the error term is normal distributed. (ii) The Likelihood Ratio Test (not shown here) performed on the asymmetry parameter ϕ_{10} of the GQARCH model reveals that the leverage effect is statistically relevant for all stocks. Furthermore, the GQARCH model results in a smaller BIC than the GARCH model. (iii) The TP-GARCH model is highly competitive. In 12 cases, it is selected as the best model according to the information criterion. Even when the GQARCH model is preferred, the TP-GARCH model beats the other semi-nonparametric model, the Spline-GARCH.

In Figure 4.5 we compare the News Impact Curve functions estimated by GQARCH and TP-GARCH models with Student's t-disturbances. The greater flexibility offered by our model becomes clear if we look at the difference in the functional form of conditional variance given a low and an high value of volatility, $\bar{\sigma}^2$, respectively equal to the 20th and the 80th percentiles of the empirical cumulative distribution of the

filtered volatility, $\tilde{\sigma}_t^2$, $t = 1, \cdots, T$, estimated by the GARCH model. When the volatility level increases, the NIC function estimated by the GQARCH model is shifted up but the form remains unchanged. On the contrary, the different market condition, synthesized by the volatility level, affects the functional form of the News Impact Curve estimated by the TP-GARCH model. When the market is calm, i.e. low level of $\bar{\sigma}^2$, the impact of the news (returns) is mild. Consequently the TP-GARCH News Impact Curve function is estimated below the parametric counterpart for a large subset of the return range. Exceptions can be found for Boeing and Walt Disney returns. Exactly the opposite happens when the market is excited, i.e. high level of $\bar{\sigma}^2$. The TP-GARCH News Impact Curve function is estimated at an higher level, especially for negative shocks. The greater flexibility offered by our novel model can be appreciated by noticing another peculiarity. The leverage effect is particularly pronounced when the market is excited but it tends to be mitigated for extreme negative shocks. In the GARCH model as well as in the GQARCH formulation, the conditional volatility explodes for large shocks (in absolute value). This undesirable drawback seems to be substantially corrected by our proposal that reduces the effect of outliers on the filtered volatility estimation.

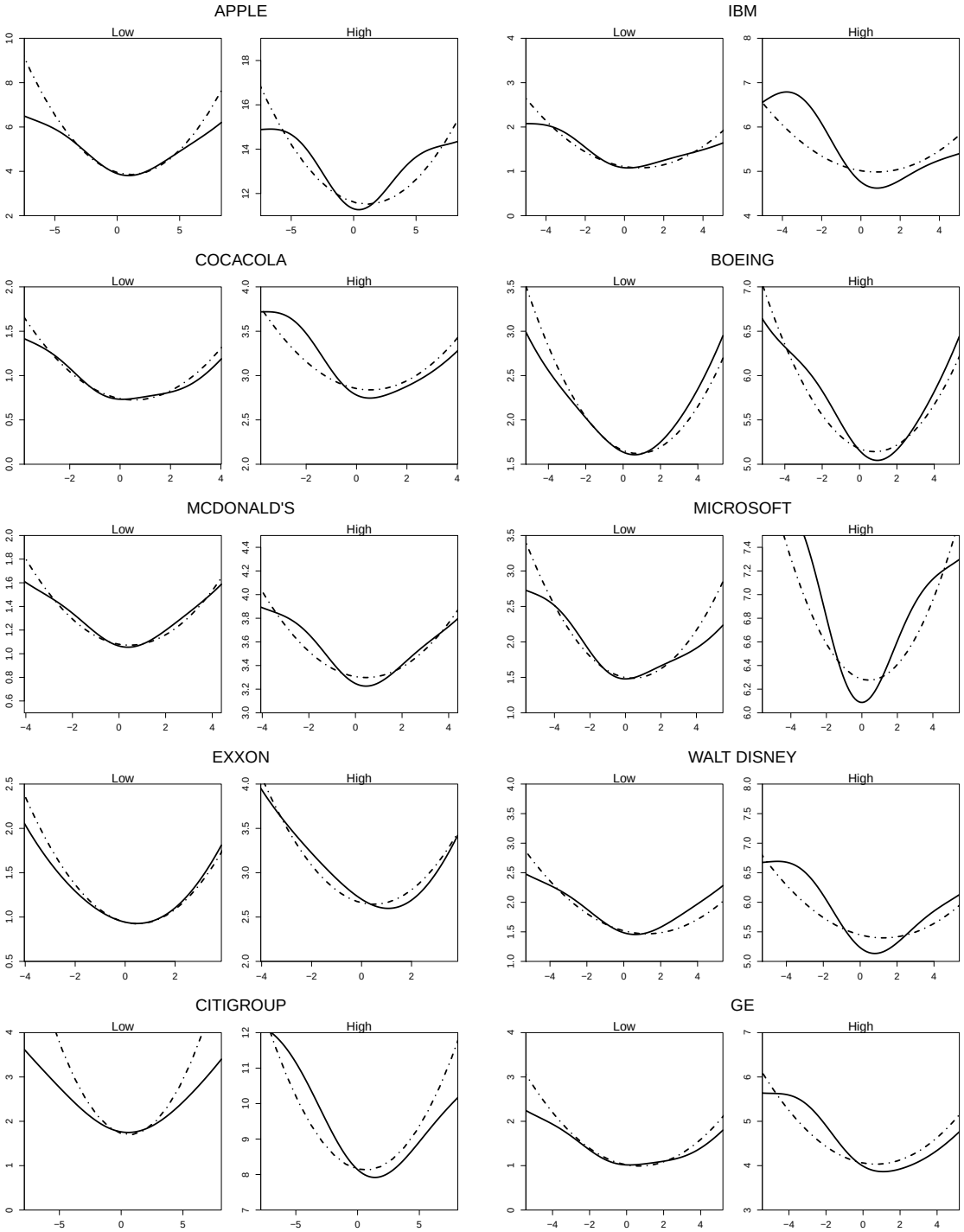
Table 4.3: Goodness-of-fit results for ten log-returns from S&P100 index

We analyzed the daily log-returns of ten stocks observed from January 4, 1993 to December 31, 2013, for a total of 5477 observations. We assume the model $y_t = \sigma_t^2 u_t$ as DGP, where u_t follows a Normal distribution (N) and a Student's t-distribution (ST). The functional form of σ_t^2 is given by (4.6) for the Standard GARCH (1,1) model, by (4.7) for the GQARCH model, by (4.8) for the Spline-GARCH model and by (4.2) for the TP-GARCH model. For the Spline-GARCH D has been set equal to 3 and K equal to 2. ξ_1 and ξ_2 are estimated by the 25th and 75th percentiles of data. For the TP-GARCH model we assume a second and a third order transformed polynomial. Then, we removed the parameter not significant at the 0.05 level from the cubic specification. The statistic parameter are estimated by maximizing numerically the log-likelihood function. We report the Bayesian Information Criterion (BIC).

Stock	Error	GARCH	GQARCH	Spline-GARCH	TP-GARCH		
					Quad.	Cubic	Final
APPLE	N	26709	26605	26586	26604	26555	**26546**
	ST	25771	25758	25788	25767	25756	**25719**
IBM	N	20777	20694	20660	20646	20610	**20593**
	ST	20068	20040	20049	20037	20030	**20006**
COCACOLA	N	18001	**17931**	17947	17958	17971	17940
	ST	17639	**17616**	17634	17639	17640	17616
BOEING	N	21706	**21648**	21674	21662	21690	21662
	ST	21285	**21262**	21288	21274	21305	21271
MCDONALDS	N	19240	19220	19241	19221	19232	**19209**
	ST	18890	18888	18930	18897	18917	**18886**
MICROSOFT	N	22186	22190	22175	22195	22148	**22139**
	ST	21550	21549	21554	21557	21546	**21524**
EXXON	N	18601	**18574**	18603	18584	18606	18591
	ST	18439	**18424**	18463	18436	18461	18433
Walt Disney	N	21574	21513	21522	21509	21504	**21497**
	ST	21021	**20988**	21020	21013	21022	21002
Citigroup	N	23469	23435	23442	23429	23443	**23418**
	ST	23096	**23072**	23098	23077	23093	23078
GE	N	20034	19985	19999	19985	19975	**19951**
	ST	19758	19728	19734	19745	19746	**19721**

Figure 4.5: News Impact Curve functions for ten log-returns from S&P100 index

We display the estimated News Impact Curve functions for a low and a high values of the filtered conditional variance estimated by the GARCH model equal to the 20th and the 80th percentiles of the empirical cumulative distribution functions, respectively. The error term is assumed to follow the Student's t-distribution. The range of x-axis corresponds to 1st-99th percentiles of log-returns. GQARCH (dashed line) and TP-GARCH (straight line).



4.4.2 Case study: IBM log-return

While in the previous Subsection, we were able to show the robustness of the TP-GARCH model with respect to other parametric and semi-nonparametric alternatives by analyzing a large number of financial series; here, we try to offer the reader a deep insight of our model and for helping us, we decided to focus our attention to the IBM stock return, displayed in Figure 4.4 (a). In Table 4.4 we report the results on goodness-of-fit and on parameter estimates. As we have already pointed out, the Student's t-distribution assumption leads to uniformly better results. Not surprisingly, in the models with Student's t-distributed error term, the inverse of degree of freedom parameter, $1/\nu$, is estimated statistically far from 0. Consistently to the findings in financial econometrics literature, we found an empirical evidence that the financial returns exhibit a leptokurtosis that cannot well captured by the normal distribution. The Student's t-distribution seems to be more appropriate but other choices are left to future research. The GARCH(1,1) model, used as benchmark, results in the largest values of both AICc and BIC. The information criteria supports those models that are more flexible in modeling the conditional volatility. For example, the GQARCH model, adding just one more parameter, it is able to increase substantially the fitting ability. The leverage effect included by this model through the parameter ψ_{10} is estimated statistically relevant with estimates equal to -0.099 (SE 0.013) with normally distributed error terms and -0.073 (SE 0.013) when the normal density is replaced by the Student's t. The findings on the Spline-GARCH model have mixed interpretation. The improvement gained with respect to the GARCH model is supported by the information criteria for both the error term specification. However, the more parsimonious GQARCH model is favored when the Student's t-distribution is assumed for the errors.

The results are strongly more convincing for the TP-GARCH. Both information criteria choose the TP-GARCH model as the best compromise between model fitting ability and parsimony. In the TP-GARCH model with Student's t-error terms, the leverage effects is captured by the negative sign estimated for the parameter ψ_{30}. An interaction effect between volatility and returns is modeled by the parameter ψ_{12} while the statistically significant estimate of parameter ψ_{21} suggests a sort of "leverage-interaction" mixed effect. However, the large number of parameter estimated by the TP-GARCH model makes the interpretation of parameters difficult.

When we have commented the estimates of the News Impact Curve functions in the previous subsection, we argued that the TP-GARCH model was less sensitive to outliers than the competing models. The filtered conditional volatility shown in Figure

4.4 (b) confirms that intuition. In particular, during the so-called "dot-com" bubble crisis and, more recently, during the "subprime crisis" the TP-GARCH approach mitigates the effect of big shocks on the volatility estimation. The same conclusion can be reached by comparing the the News Impact Curve functions displayed in Figure 4.4 (c) for a low and a high level of volatility, $\tilde{\sigma}^2$. Although, at a first sight, the Spline-GARCH model looks appealing for the strong leverage phenomenon estimated, the volatility is not guaranteed to be positive at extreme values of y. Our proposal does not suffer from this undesirable drawback. Outside of those regions in which data give more information, the GARCH(1,1) functional form is recovered and the conditional variance function is ensured to be positive. For moderate values of y, the leverage effect is accurately described and results to be especially pronounced when the market is already excited, i.e., high level of $\tilde{\sigma}^2$.

Finally, some diagnostic analysis is proposed to assess the statistical quality of the TP-GARCH model with Student's t-distributed error term. The Rosenblatt's transformation can be applied to the standardized residuals defined as $\tilde{\varepsilon}_t = y_t/\tilde{\sigma}_t$ for $t = 1, \cdots, T$. The normalization obtained by the Rosenblatt's transformation enables us to use the standard diagnostic tools, shown Figure 4.6. The normal QQ-Plot does not exhibit significative deviations from the distribution assumption. The normalized residuals as well as their squares are serially uncorrelated suggesting that the TP-GARCH model is able to capture sufficiently accurately the conditional variance dynamics.

Figure 4.6: IBM log-return: diagnostics results

As the Student's t-distribution is assumed for the errors, the Rosenblatt's transformation is applied to the standardized residuals. QQ normal plot (left), Autocorrelation function of residuals (centre) and of squared residuals (right).

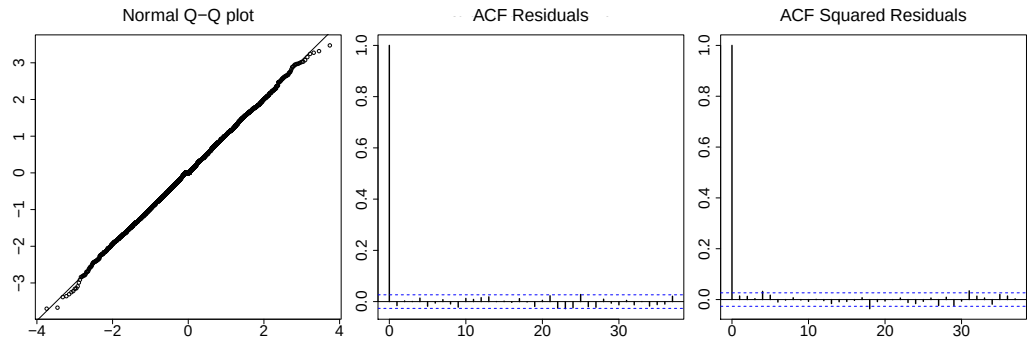
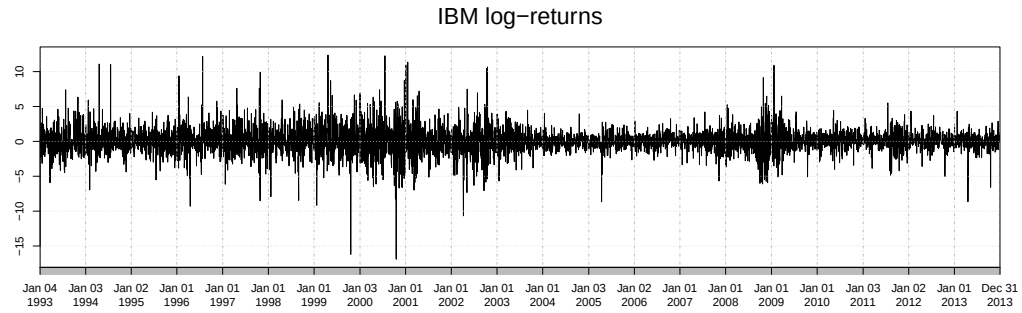


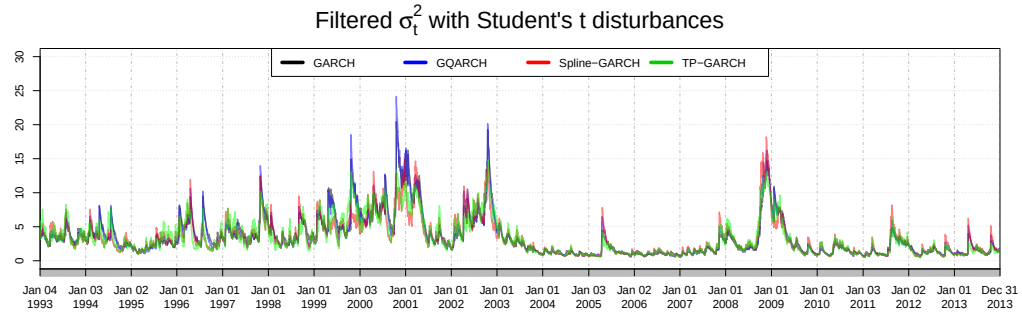
Figure 4.7: IBM log-return: filtered conditional variance and News Impact Curve functions

The Student's t-distribution is assumed for the errors. (a) Daily IBM log-returns January 1, 1993 to December 31, 2013 (b) Filtered conditional volatility for different models. (c) News Impact Curves: Low variance (LOW) corresponds to the 20th percentile of the filtered conditional variance estimated by TP-GARCH and high variance to the 80th percentile. Vertical dashed lines correspond to 0.1th and 99.9th percentiles of y_t.

(a)



(b)



(c)

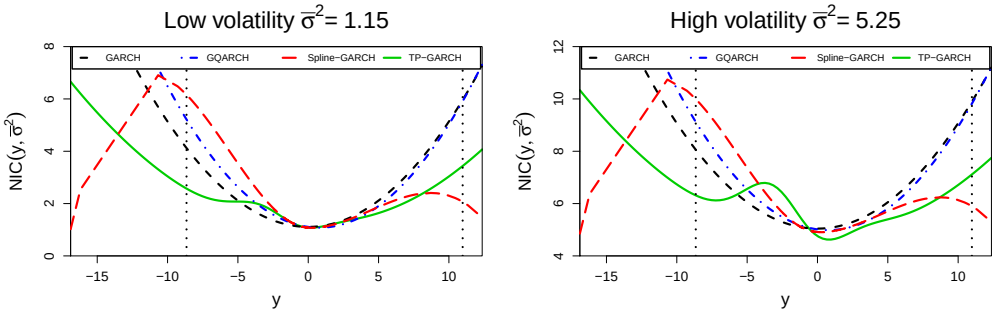


Table 4.4: IBM log-return: parameter estimates and goodness-of-fit

We analyzed the daily log-returns of IBM stock observed from January 4, 1993 to December 31, 2013, for a total of 5477 observations. We assumed as DGP the following model $y_t = \sigma_t^2 u_t$, where u_t follows a Normal distribution (a) and a Student's t-distribution (b). The functional form of σ_t^2 is given by (4.6) for the Standard GARCH (1,1) model, by Eq.(4.7) for the GQARCH model, by (4.8) for the Spline-GARCH model and by (4.2) for the TP-GARCH model. For the Spline-GARCH D has been setter equal to 3 and K equal to 2. ξ_1 and ξ_2 are estimated by the 25th and 75th percentiles of data. For the TP-GARCH model we use a third-degree transformed polynomial specification and then we removed the parameter not significant at the 0.05 level. The statistic parameter are estimated by maximizing numerically the log-likelihood function. We report the maximized log-likelihood value (LogLik), the corrected Akaike Information Criterion (AICc), the Bayesian Information Criterion (BIC) and the parameter estimates with numerical SE between parentheses.

(a) Normal disturbances

Model	Goodness-of-fit				Parameter Estimates				
	LogLik	N par.	AICc	BIC	ψ_{00}	ψ_{01}	ψ_{20}	ψ_{10}	
GARCH	-10375.75	3	20757.50	20777.32	0.030	0.926	0.068		
					(0.007)	(0.011)	(0.010)		
GQARCH	-10330.01	4	20668.02	20694.45	0.038	0.930	0.061	-0.099	
					(0.007)	(0.009)	(0.009)	(0.013)	
Spline-GARCH	-10291.53	9	20601.09	20660.54	0.038	0.918	0.149	-0.022	
					(0.008)	(0.009)	(0.026)	(0.031)	
TP-GARCH	-10262.48	8	20540.99	20593.83	0.030	0.955	0.017	0.068	
					(0.004)	(0.002)	(0.004)	(0.029)	
					ψ_{30}	γ_1	γ_2	ξ_1	ξ_2
Spline-GARCH					0.009	-0.022	0.006	-0.812	0.890
					(0.002)	(0.007)	(0.007)		
					ψ_{02}	ψ_{11}	ψ_{03}	ψ_{21}	
TP-GARCH					-0.033	-0.139	0.003	-0.071	
					(0.006)	(0.022)	(0.001)	(0.009)	

(b) Student's t-disturbances

Model	Goodness-of-fit				Parameter Estimates				
	LogLik	N par.	AICc	BIC	ψ_{00}	ψ_{01}	ψ_{20}	ψ_{10}	$1/\nu$
GARCH	-10017.26	4	20042.52	20068.95	0.010	0.958	0.040		0.228
					(0.003)	(0.002)	(0.003)		(0.014)
GQARCH	-9998.80	5	20007.62	20040.65	0.018	0.951	0.046	-0.073	0.224
					(0.004)	(0.004)	(0.005)	(0.013)	(0.014)
Spline-GARCH	-9981.49	10	19983.01	20049.06	0.023	0.933	0.139	0.001	0.219
					(0.009)	(0.011)	(0.034)	(0.037)	(0.014)
TP-GARCH	-9969.04	8	19954.11	20006.95	0.009	0.967	0.020		0.214
					(0.003)	(0.002)	(0.003)		(0.013)
					ψ_{30}	γ_1	γ_2	ξ_1	ξ_2
Spline-GARCH					0.008	-0.023	0.010	-0.812	0.890
					(0.002)	(0.009)	(0.008)		
					ψ_{02}	ψ_{30}	ψ_{12}	ψ_{21}	
TP-GARCH					-0.028	-0.016	-0.030	0.073	
					(0.007)	(0.008)	(0.007)	(0.015)	

4.5 Conclusion

We extended the transformed polynomial functions introduced by Blasques (2014) by proposing a novel semi-nonparametric approach to model the conditional volatility, the Transformed Polynomials GARCH (TP-GARCH). Our model is particularly relevant in those fields, like empirical finance, where the researcher is interested in modeling financial time series characterized by cluster volatility, leptokurtosis and leverage effect.

Through two Monte Carlo simulation studies we explored the finite sample properties both under correct and incorrect specification. In the first artificial study we simulated observations from a model in which the conditional variance evolves according to the TP-GARCH updating equation and the findings suggested that the parametric ML estimator of the parameter vector may be consistent and normally distributed. In the second experiment, the misspecified functional form of the conditional variance is approximated accurately well by the TP-GARCH model whereas the GARCH model fails. The sieve ML estimator seems to be consistent as the approximation become more accurate as long as the sample size and the order of transformed polynomial increase.

In an empirical analysis conducted on the log-returns of ten blue chips stocks traded in the USA market, the TP-GARCH formulation has been tested against a number of parametric and semi-parametric alternatives. The filtered volatility estimated by the TP-GARCH model has been shown to react more flexibly to different market conditions and, at the same time, to be less influenced by outliers.

The extension of theoretical framework developed by Blasques (2014) remains still a challenging open research question and efforts will be devoted to fill the gap in the next future.

Bibliography

Akaike, H. (1973). Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika 60*(2), 255–265.

Akaike, H. (1974). A new look at the statistical model identification. *Automatic Control, IEEE Transactions on 19*(6), 716–723.

Audrino, F. and P. Bühlmann (2001). Tree-structured generalized autoregressive conditional heteroscedastic models. *Journal of the Royal Statistical Society: Series B 63*(4), 727–744.

Audrino, F. and P. Buhlmann (2009). Splines for financial volatility. *Journal of the Royal Statistical Society: Series B 71*(3), 655–670.

Black, F. (1976). Studies of Stock Price Volatility Changes. In *Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economic Statistics Section*.

Blasques, F. (2014). Transformed polynomials for nonlinear autoregressive models of the conditional mean. *Journal of Time Series Analysis 35*(3), 218–238.

Blasques, F., S. J. Koopman, and A. Lucas (2014a). Maximum Likelihood Estimation for Generalized Autoregressive Score Models. *Discussion Paper Tinbergen Institute TI 14-029/III*.

Blasques, F., S. J. Koopman, and A. Lucas (2014b). Stationarity and Ergodicity of Univariate Generalized Autoregressive Score Processes. *Electronic Journal of Statistics 8*, 1088–1112.

Blasques, F., S. J. Koopman, and A. Lucas (2015). Information Theoretic Optimality of Observation Driven Time Series Models. *Biometrika*, forthcoming.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*(3), 307–327.

Bougerol, P. (1993). Kalman filtering with random coefficients and contractions. *SIAM Journal on Control and Optimization 31*(4), 942–959.

Box, G. E. and G. M. Jenkins (1970). *Time Series Analysis, Forecasting and Control*. San Francisco, Holden Day.

Chan, K. S. and H. Tong (1986). On estimating thresholds in autoregressive models. *Journal of Time Series Analysis 7*(3), 179–190.

Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. In *Handbook of Econometrics*, Volume 6, Part B, Chapter 76, pp. 5549 – 5632. James J. Heckman and Edward E. Leamer.

Chen, X. (2011). Penalized sieve estimation and inference of semi-nonparametric dynamic models: a selective review. Technical report, Cowles Foundation discussion paper.

Cox, D. R. (1981). Statistical Analysis of Time Series: Some Recent Developments. *Scandinavian Journal of Statistics 8*, 93–115.

Creal, D., S. J. Koopman, and A. Lucas (2011). A Dynamic Multivariate Heavy-Tailed Model for Time-Varying Volatilities and Correlations. *Journal of Business & Economic Statistics 29*(4), 552–563.

Creal, D., S. J. Koopman, and A. Lucas (2013). Generalized Autoregressive Score Models with Applications. *Journal of Applied Econometrics 28*(5), 777–795.

Creal, D., B. Schwaab, S. J. Koopman, and A. Lucas (2014). Observation Driven Mixed-Measurement Dynamic Factor Models with an Application to Credit Risk. *Review of Economics and Statistics 96*(5), 916–935.

De Lira Salvatierra, I. and A. J. Patton (2013). Dynamic copula models and high frequency data. *Duke University Discussion Paper*.

Delle Monache, D. and I. Petrella (2014). A Score driven approach for Gaussian State-Space models with Time-Varying Parameter. *Working Paper, Imperial College London*.

Dempster, A., N. Laird, and D. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B*, 1–38.

Diebold, F., J. Lee, and G. Weinbach (1994). Regime Switching with Time-Varying Transition Probabilities. In C. Hargreaves (Ed.), *Nonstationary Time Series Analysis and Cointegration*, pp. 283–302. Oxford University Press.

Ding, Z., C. W. Granger, and R. F. Engle (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance 1*(1), 83–106.

Doornik, J. (2013). A Markov-switching model with component structure for US GNP. *Economics Letters 118*(2), 265–268.

Drost, F. C. and C. A. Klaassen (1997). Efficient estimation in semiparametric GARCH models. *Journal of Econometrics 81*(1), 193–221.

Durbin, J. and S. J. Koopman (2012). *Time series analysis by state space methods*. Oxford University Press.

Durland, J. M. and T. H. McCurdy (1994). Duration-dependent transitions in a Markov model of US GNP growth. *Journal of Business & Economic Statistics 12*(3), 279–288.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica 50*(4), 987–1007.

Engle, R. F. and G. Gonzalez-Rivera (1991). Semiparametric ARCH models. *Journal of Business & Economic Statistics 9*(4), 345–359.

Engle, R. F. and V. K. Ng (1993). Measuring and testing the impact of news on volatility. *The Journal of Finance 48*(5), 1749–1778.

Engle, R. F. and J. R. Russell (1998). Autoregressive Conditional Duration: A New Model for irregularly Spaced Transaction Data. *Econometrica 66*(5), 1127–1162.

Ethier, R. and T. Mount (1998). Estimating the volatility of spot prices in restructured electricity markets and the implications for option values. Technical report, PSerc Working Paper.

Fan, J. and Q. Yao (2003). *Nonlinear time series*. New York, Springer-Verlag.

Filardo, A. J. (1994). Business-cycle phases and their transitional dynamics. *Journal of Business & Economic Statistics 12*(3), 299–308.

Francq, C. and M. Roussignol (1998). Ergodicity of autoregressive processes with Markov-switching and consistency of the maximum-likelihood estimator. *Statistics: A Journal of Theoretical and Applied Statistics 32*(2), 151–173.

Francq, C. and J.-M. Zakoïan (2001). Stationarity of multivariate markov–switching arma models. *Journal of Econometrics 102*(2), 339–364.

Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. Springer.

Glosten, L. R., R. Jagannathan, and D. E. Runkle (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance 48*(5), 1779–1801.

Gourieroux, C., A. Monfort, E. Renault, and A. Trognon (1987). Generalised residuals. *Journal of Econometrics 34*(1), 5–32.

Gray, S. F. (1996). Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics 42*(1), 27–62.

Hamilton, J. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica 57*(2), 357–384.

Hamilton, J. D. and B. Raj (2002). New directions in business cycle research and financial analysis. *Empirical Economics 27*(2), 149–162.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

Harvey, A. C. (2013). *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series*. Econometric Series Monographs. Cambridge University Press.

Harvey, A. C. and A. Luati (2014). Filtering with heavy tails. *Journal of the American Statistical Association 109*(507), 1112–1122.

Huisman, R. (2008). The influence of temperature on spike probability in day-ahead power prices. *Energy Economics 30*(5), 2697–2704.

Huisman, R. and C. De Jong (2003). Option pricing for power prices with spikes. *Energy Power Risk Management 7*(11), 12–16.

Huisman, R. and R. Mahieu (2003). Regime jumps in electricity prices. *Energy Economics 25*(5), 425 – 434.

Hurvich, C. M. and C.-L. Tsai (1991). Bias of the corrected AIC criterion for underfitted regression and time series models. *Biometrika 78*(3), 499–509.

Janczura, J. and R. Weron (2009). Regime-switching models for electricity spot prices: Introducing heteroskedastic base regime dynamics and shifted spike distributions. Technical report, IEEM Conference Proceedings (EEM'09).

Janczura, J. and R. Weron (2010). An empirical comparison of alternate regime-switching models for electricity spot prices. *Energy economics 32*(5), 1059–1073.

Janczura, J. and R. Weron (2012). Efficient estimation of markov regime-switching models: An application to electricity spot prices application to electricity spot prices. *AStA Advances in Statistical Analysis 96*(3), 385–407.

Kim, C. (1994). Dynamic linear models with Markov-switching. *Journal of Econometrics 60*(1), 1–22.

Kim, C.-J., J. C. Morley, and C. R. Nelson (2004). Is there a positive relationship between stock market volatility and the equity premium? *Journal of Money, Credit and Banking 36*, 339–360.

Koopman, S. J., A. Lucas, and M. Scharth (2015). Predicting time-varying parameters with parameter-driven and observation-driven models. *Review of Economics and Statistics*, forthcoming.

Krengel, U. (1985). *Ergodic theorems*. Berlin: De Gruyter studies in Mathematics.

Linton, O. (1993). Adaptive estimation in arch models. *Econometric Theory 9*(4), 539–569.

Linton, O. and E. Mammen (2005). Estimating Semiparametric ARCH (∞) Models by Kernel Smoothing Methods. *Econometrica 73*(3), 771–836.

Lucas, A., B. Schwaab, and X. Zhang (2014). Measuring credit risk in a large banking system: econometric modeling and empirics. *Journal of Business and Economic Statistics*, forthcoming.

Maheu, J. M. and T. H. McCurdy (2000). Identifying bull and bear markets in stock returns. *Journal of Business & Economic Statistics 18*(1), 100–112.

Mount, T. D., Y. Ning, and X. Cai (2006). Predicting price spikes in electricity markets using a regime-switching model with time-varying parameters. *Energy Economics 28*(1), 62–80.

Nelson, D. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica 59*(2), 347–370.

Oh, D. H. and A. J. Patton (2013). Time-varying systemic risk: Evidence from a dynamic copula model of CDS spreads. *Duke University Discussion Paper*.

Pagan, A. R. and Y. Hong (1991). Nonparametric estimation and the risk premium. In J. P. W. Barnett and G. Tauchen (Eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, pp. 51–75.

Pagan, A. R. and G. W. Schwert (1990). Alternative models for conditional stock volatility. *Journal of Econometrics 45*(1), 267–290.

Perez-Quiros, G. and A. Timmermann (2000). Firm size and cyclical variations in stock returns. *The Journal of Finance 55*(3), 1229–1262.

Perez-Quiros, G. and A. Timmermann (2001). Business cycle asymmetries in stock returns: Evidence from higher order moments and conditional densities. *Journal of Econometrics 103*(1), 259–306.

Rodríguez, M. J. and E. Ruiz (2012). Revisiting several popular GARCH models with leverage effect: Differences and similarities. *Journal of Financial Econometrics 10*(4), 637–668.

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *The Annals of Mathematical Statistics 23*(3), 470–472.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics 6*(2), 461–464.

Sentana, E. (1995). Quadratic ARCH Models. *Review of Economics Studies 62*, 639–661.

Shephard, N. and T. G. Andersen (2009). Stochastic Volatility: Origins and Overview. In *Handbook of Financial Time Series*. Springer: Berlin Heidelberg.

Smith, D. R. (2008). Evaluating Specification Tests for Markov-Switching Time-Series Models. *Journal of Time Series Analysis 29*(4), 629–652.

Straumann, D. and T. Mikosch (2006). Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: a stochastic recurrence equations approach. *The Annals of Statistics 34*(5), 2449–2495.

Tong, H. (1983). *Threshold models in non-linear time series analysis*. Springer-Verlag.

Turner, C. M., R. Startz, and C. R. Nelson (1989). A Markov model of heteroskedasticity, risk, and learning in the stock market. *Journal of Financial Economics 25*(1), 3–22.

Weron, R. (2006). *Modeling and forecasting electricity loads and prices: A Statistical Approach*. Wiley.

Weron, R. (2009). Heavy-tails and regime-switching in electricity prices. *Mathematical Methods of Operations Research 69*(3), 457–473.

Weron, R., M. Bierbrauer, and S. Trück (2004). Modeling electricity prices: jump diffusion and regime switching. *Physica A: Statistical Mechanics and its Applications 336*(1), 39–48.

Zaköian, J.-M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and control 18*(5), 931–955.