Estimating Deterministics in Univariate Time Series

Inauguraldissertation
zur Erlangung des akademischen Grades
eines Doktors der Wirtschaftswissenschaften
der Universität Mannheim

vorgelegt von

Christopher Paul Walsh

Mai 2014

Abteilungssprecher:	Prof. Dr. Eckhard Janeba
Referent:	Prof. Dr. Enno Mammen
Korreferent:	Prof. Dr. Carsten Trenkler

Tag der Verteidigung:	26. Mai 2014

Acknowledgements

First of all, I wish to thank my advisors Carsten Trenkler and Enno Mammen. They were always there when I needed them and gave me enough freedom to potter on with my work. Thanks also goes to Kyusang Yu, who was literally always around to answer more or less qualified questions.

Special thanks goes to my former colleague Michael Vogt without whom the journey would have been so much more difficult and so much less amusing.

I would also like to thank Uta Pigorsch and Carsten Trenkler for giving me the opportunity to teach the next generation of econometricians, thereby learning something myself.

Further thanks goes to the fellow members of my CDSE cohort who endured all the course work with me and provided for interesting and fun conversations over the years we were together.

I am also very grateful to all the friends I have made in my time here in Mannheim, especially for all the moral support they provided. In total they, especially Lena, have been a vital component in making my time in Mannheim one of the most enjoyable periods of my life so far.

Last but by no way least, thanks goes to my mum and dad, who have always trusted and backed me. I am sure Lena and I will be very grateful if they continue to do so.

Mannheim, 29th April 2014

Christopher Walsh

Contents

Chapter 1

Introduction

This thesis deals with regression approaches to analyse certain nonlinear and non-stationary univariate discrete time series models. A univariate discrete time series consists of regular observations of a one dimensional dynamic process. Let $Y_t \in \mathbb{R}$ denote the observation of a particular dynamic process at time point t. The process is defined for equally spaced time points. Typically, it is assumed that the process has and always will exist. Thus the process can be described by the infinte sequence of random variables $\{Y_t\}_{t\in\mathbb{Z}}$[1].

We assume that we only observe the process at T equally spaced consecutive time points, say $t = 1, \ldots, T$. Thus the observations $\{Y_t\}_{t=1,\ldots,T}$ correspond to T regular measurements of a dynamic process. This type of data is collected in many scientic fields. These range from economics, finance, engineering, hydrology, atmospheric sciences, acoustics and seismology to mention just a few. Hence, there is a vast amount of different models for such data, spread across these fields (see e.g. Fuller (1996), Hamilton (1994), Mills and Markellos (2008) , Machiwal et al. (2012), Mudelsee (2010) and Quinn and Hannan (2001)).

The main objective in time series modelling is to either capture the dynamics of the underlying process or to replicate certain aspects of its stochastic behaviour such as certain moments or distribution functions. The models are estimated

[1] Alternatively the process is assumed to have a particular beginning in which case $\mathbb{Z}$ is replaced by $\mathbb{N}$ and the random variable Y_0 is denoted the intial value of the process.

based on the sample $\{Y_t\}_{t=1,\dots,T}$. Given the estimated model, further analysis can be conducted such as forecasting, signal extraction, turning point anaylsis or determining dynamic causal effects.[2]

Before introducing our models and their related results in Chapters 2, 3 and 4, we will use the rest of this introductory chapter to recap the concept of "stationarity" due to its importance in modelling time series processes. Following this we will see that real data series do not always seem to be stationary. This is illustrated by two data series that we will apply our models to in the subsequent chapters. Finally, we will review a common decomposition method to incorporate certain nonstationarities. We will also mention some modelling approaches that can be seen to fall within this framework, the focus being on models related to ours.

1.1 Stationarity

Most introductory texts on time series analysis deal with models for so called stationary processes. Stationarity essentially means that the stochastic mechanism of the process satisfies some sort of time invariancy. Before restating the most common types of stationary processes considered in the literature, we will give a brief example to illustrate why the concept of stationarity is so helpful.

The dynamics of the process $\{Y_t\}_{t\in\mathbb{Z}}$ are governed by the collection of all finite dimensional joint distribution functions of $\{Y_t\}_{t\in\mathbb{Z}}$. Suppose we were interested in looking at the dependence structure over one time period only. This would be contained in the collection of joint distributions $\{F_{Y_{t-1},Y_t}\}_{t\in\mathbb{Z}}$. In general, $\{F_{Y_{s-1},Y_s}\} = \{F_{Y_{t-1},Y_t}\}$ for $s \neq t$ so there is not really all that much we can say about the dependence structure over one period even if we observed the entire process $\{Y_t\}_{t\in\mathbb{Z}}$. The situation changes dramatically if we knew that the joint distributions $\{F_{Y_{t-1},Y_t}\}_{t\in\mathbb{Z}}$ were time invariant, i.e. $\{F_{Y_{s-1},Y_s}\} = F_{Y_0,Y_1}$ for $\forall t \in \mathbb{Z}$. In this case F_{Y_0,Y_1} would contain all one needs to know about the dependence structure over one time period and it could be arbitrarily well estimated for a large enough

[2]Sometimes no explicit model is used. An example is forecating using exponential smoothing. See Gijbels et al. (1999) for a comparison of exponential smoothing with kernel regression.

sample. This discussion extends to any other finite dimensional joint distribution of $\{Y_t\}_{t\in\mathbb{Z}}$. Stationarity essentially assumes some similar kind of time invariancy.

Two types of stationarity assumptions are most commonly invoked. Firstly, stationarity may refer to *strict stationarity*, which entails that the joint distribution of all finite dimensional tuples of the process $\{Y_t\}_{t\in\mathbb{Z}}$ are time invariant. Thus for any finite set of time indices $(t_1, \dots, t_n)$, we have

$$F_{Y_{t_1},\dots,Y_{t_n}} = F_{Y_{t_1+t},\dots,Y_{t_n+t}} \quad \forall t \in \mathbb{Z}. \tag{1.1}$$

Secondly, stationarity may refer to *weak* or *covariance stationarity* which means that the first two unconditional moments of the process coordinates $\{Y_t\}_{t\in\mathbb{Z}}$ are time invariant, i.e.

$$\mathbb{E}[Y_t] = \mu, \quad \forall t \in \mathbb{Z} \tag{1.2}$$

$$\mathbb{E}[Y_{t+h}Y_t] = c(h) \quad \forall t, h \in \mathbb{Z} \tag{1.3}$$

Hence, under weak stationarity the observations $\{Y_t\}_{t=1,\dots,T}$ form a sequence of random variables, whose means are constant and whose covariances between two observations h periods apart depend only on h.[3] The name weak stationarity is explained upon realising that a strictly stationary series is also weakly stationary as long as we allow for second moments to exist.

The striking thing is, that when we look at actual data many time series do not seem to satisfy either of these stationarity conditions. Two examples are provided in Figures 1.1 and 1.2. The example in Figure 1.1 is taken from the atmospheric sciences. The plot gives the minimum monthly surface temperature in the Antarctic from September 1957 to December 2004. It is clearly visible that there are some regular seasonal movements and a time varying mean, clearly violating (1.2)

[3] If the process is Gaussian, then the entire behaviour of the time series $\{Y_t\}_{t\in\mathbb{Z}}$ is modelled by the first two moments and the two types of stationarity conincide.

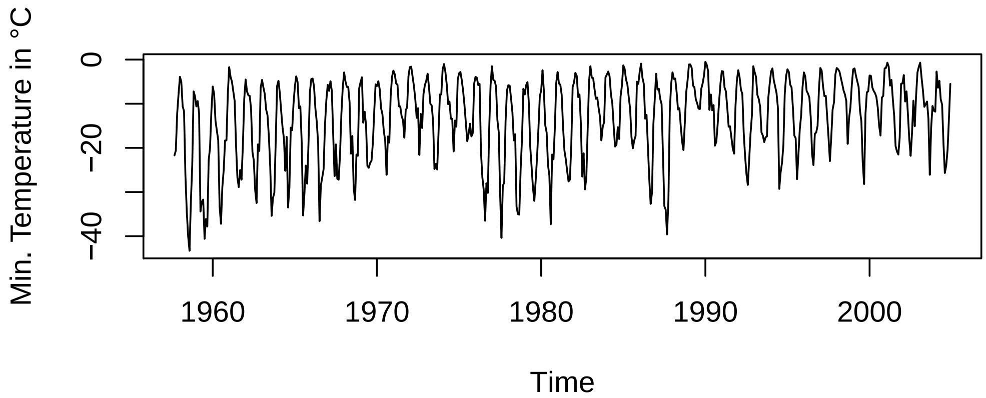


Figure 1.1: Monthly minimum near-surface temperatures (in °C) at Faraday station.

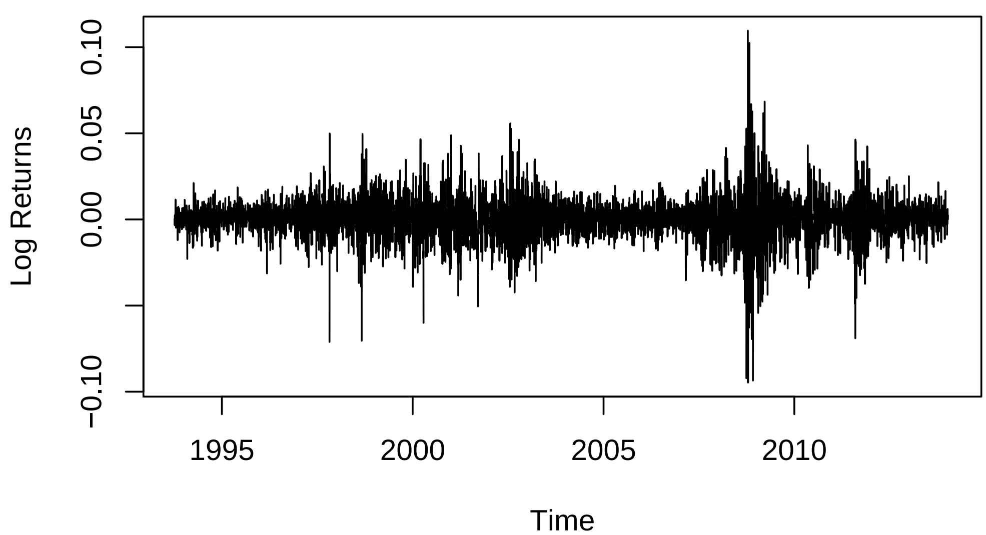


Figure 1.2: Plot of S&P 500 log returns from 10^{th} April 1993 until 2^{nd} February 2014

In Figure 1.2 on the other hand we see an example from the financial world. The plot shows the daily log returns of the S&P 500 index from 10^{th} April 1993 until 2^{nd} February 2014. In comparison to the previous example in Figure 1.1 there is no clear discernible seasonal pattern or a clearly visible time varying mean. However, one can see that the second moment of the daily log returns is time dependent. The variation in the log returns is fairly low at first. From about 1997 until 2004 it is higher before falling back to a more moderate level. The increase in the recent crisis is visible and is followed by a more stable lower level from about 2012 onwards. Both of the time series depicted in Figure 1.1 and Figure 1.2 will be used as illustrations of our modelling framework in Chapters 2 and 3 later on.

1.2 Modelling nonstationarity

To be able to construct models for the processes in Figure 1.1 or Figure 1.2 one needs to relax the stationarity assumption. Looking at the stationarity conditions (1.1) - (1.3) we see that a straightforward way to incorporate nonstationarity is to either allow for time varying first or second moments, i.e. a relaxation of the weak stationarity conditions or to allow for the joint distribution function of the observations to depend on time, i.e. a relaxation of the condition for strict stationarity. These are essentially the three deviations from stationarity mentioned in Chapter 9 of Fuller (1996).

For the time being we will postpone violations of the type seen in Figure 1.2 until we introduce our modelling framework for such data in Chapter 3 and concentrate on the frequently observed violation of the stationarity assumption arising from regularly occuring seasonal patterns or the existence of some trending behaviour in the series under study.

In terms of allowing for a time varying mean, Fuller (1996) goes on to establish what he calls the "*traditional model for economic time series*", which amounts to a decomposition of the observed time series $\{Y_t\}_{t\in\mathbb{Z}}$ into the sum of a "trend" (T_t), a "seasonal" (S_t) and an "irregular" or noise component (Z_t), i.e.[4]

$$Y_t = T_t + S_t + Z_t \quad \forall t \in \mathbb{Z}. \tag{1.4}$$

This decomposition or some form of it can be found in many other sources, e.g. Bosq (1998), Brockwell and Davis (2002), Heiler (2001) or Wildi (2006). The decomposition is also popular in economic data agencies (Eurostat (2009), Destatis (2004)) due to many researchers wanting so-called seasonally adjusted data, which removes an estimate for the seasonal component from the original data.

Without imposing any further structure on the components in (1.4), they are of course not identified. It should be clear that the components are only identified up to an additive constant. The identification issue goes beyond this. In fact, the irregular component may pick up some of the seasonal component or parts of the seasonal component my be shifted into the trend component. Thus the reasercher must in some way add additional structure to the modelling framework of (1.4). Sometimes there may be some form of guidance in setting these restrictions. Examples would include Mudelsee (2010) arguing for AR(1) structures in the irregular component of climate time series models or Hamilton (1994) for the use of the linear deterministic time trends when modelling GNP data. In general though the researcher is free to set the restrictions as he sees most appropriate for the problem at hand. This has led to a diverse amount of modelling approaches that can all be viewed as a version of the decomposition in (1.4) a point that is eloquently made by Fuller (1996) on p.475:

"While the model [in (1.4)] *is an old one indeed, a precise definition of the components has not evolved. This is not necessarily to be viewed as a weakness of the representation. In fact, the terms acquire meaning only when a procedure is used to estimate them, and the meaning is determined by the procedure."*

The dependence on the estimation method should always be borne in mind as different methods can lead to conflicting results as pointed out by Canova (1998) when using the estimated irregular component to model business cycles.

[4] A related model is the multiplicative model: $Y_t = T_t S_t Z_t$, an example of which will be given in Chapter 3.

In the remainder of this section we provide some examples of the modelling approaches encountered in the literature that can be cast in the form of (1.4). The overview is necessarily limited and we restrict the focus to models that are close to ours.[5] We will also try to categorize the approaches in terms of the assumptions they impose upon the components in the decomposition. This categorization will help position our models in the current literature. In all the remaining chapters we will treat the irregular component (Z_t) as a zero mean stationary process. Thus the non-irregular part captures the mean of Y_t, which we shall write as

$$\mu_t = T_t + S_t \quad \forall t \in \mathbb{Z}$$

One broad delineation of the models considered in the literature can then be made with regards to the modelling of the mean μ_t.

The first broad group of models treats the mean as being purely a function of time and thus deterministic. The modelling of the mean may then be done using parametric functions (see Fuller (1996), Hamilton (1994) or Hendry (1995)). Alternatively, one may use nonparametric models that refrain from specifying a parametric functional form. Examples include Truong (1991), Altman (1993), Härdle et al. (1997), Hall and Keilegom (2003), and Shao and Yang (2011).

The second broad group considers the mean μ_t to be stochastic. One way is to model the individual components in the mean μ_t as stochastic processes, the so called unobservable component models or structural time series models (see Harvey (1989), (1993)). It is also possible to model the mean μ_t stochastically by specfying it to be a function of other stationary processes. Note, that this is generally used to model stationary processes. In a parametric framework one can use distributed lag models (see Hendry (1995)). Using nonparametric functions of stationary processes to model the trend has been considered in Truong and Stone (1994), Schick (1994) and Lin et al. (1999). There has been some work to extend this approach to allow for nonparamtric functions of certain non-stationary processes. For a summary on such approaches see Tjøstheim (2012).

[5] For a further overview with a focus on the identifying assumptions on the individual compents see Wildi (2006).

The final method we will consider are so-called unit root processes (see Hamilton (1994)). This can also be included in the framework of (1.4) by setting $\mu_t = Y_{t-1}$. In this case interest lies in modelling the differenced series $\Delta Y_t = Y_t - Y_{t-1} = Z_t$. A popular approach is by considering this difference to follow an autoregressive moving average (ARMA) process, introduced by Box and Jenkins (1970). Extensions modelling higher order differences by ARMA models leads to the family of autoregressive integrated moving average (ARIMA) models.

It should be borne in mind that the modelling approaches mentioned above are not necessarily distinict from one another. For example, Harvey (1985) shows that a certain unobserved component model may be cast as an ARIMA model with additional restrictions or can be reduced to a model with a deterministic linear time trend.

To finish the discussion on models that fit (1.4), it should be noted that most of theses approaches rely on retrieveing the stationary irregular component from the original series by a suitable transform of the original series. But as the transform depends on the chosen model there can be no unique stationary rendering transform. A point made more pointedly by Harvey (1985) when referring to the use of ARIMA models:

"Indeed, the remarkable thing about differenced economics time series is not that they are sometimes nonstationary, but rather that they are occasionally stationary."

Lastly, we will mention one additional approach to dealing with nonstationarities that cannot be cast in the form of (1.4). However, our model in Chapter 3 can be seen as falling within this type of model class. The idea is to approximate nonstationary processes by stationary processes locally around each time point. The resulting processes are termed locally stationary. The concept and much of the theoretical work was done in in a series of papers by Dahlhaus (1996b), (1996a), (1997). An overview on the present state of modelling locally stationary processes is given in Dahlhaus (2012).

1.3 Summary

In this section we will provide a summary of the models considered in Chapters 2, 3 and 4. We will also show how these models relate to the decomposition in (1.4).

1.3.1 Chapter 2 - based on joint work with M.Vogt

The model in Chapter 2 will consist of specifying a seasonal component S_t with a known period. Furthermore, the trend component T_t will consist of nonparametric functions of time and other variables thus representing a mixture between the deterministic and stochastic case. Additional structure is imposed by requiring the components to be additive. This is done to circumvent the well known curse of dimensionality. Finally, we will model the stationary irregular component Z_t by an autoregressive process of order p, AR(p). Thus, we will be considering the model

$$Y_{t,T} = m_\theta(t) + m_0(\frac{t}{T}) + \sum_{j=1}^{d} m_j(X_t^j) + \varepsilon_t \text{ for } t = 1, \dots, T, \tag{1.5}$$

with $\{(X_t^1, \dots, X_t^d, \varepsilon_t)\}$ a $(d+1)$-dimensional stationary process. This model fits into the framework in (1.4) with:

1. $S_t = m_\theta(t)$, which is a periodic function with known period θ, i.e. $m_\theta(t) = m_\theta(t + k\theta)$ for all $k \in \mathbb{N}$.
2. $T_t = m_0(\frac{t}{T}) + \sum_{j=1}^{d} m_j(X_t^j)$, which is the sum of the smooth function of rescaled time m_0 and the sum of component functions m_j that depend smoothly on the regressors X_t^j.
3. $Z_t = \varepsilon_t$, which is a stationary AR(p) process.

Model (1.5) can be seen as a nonparametric extension of a distributed lag model that additionally allows for a nonparemtric trend function. Special mention should be made here to the fact that due to the inclusion of the nonparametric trend function m_0, the model cannot be viewed as an extension of the augmented distributed lag model, i.e. one cannot include lagged dependent variables in the set of regressor variables. We will suggest estimators for the components of the model in (1.5). We will also provide theoretical results on the asymptotic behaviour of our estimators and apply our model to the data plotted in Figure 1.1.

1.3.2 Chapter 3 - based on joint work with M.Vogt

In Chapter 3 we will apply a model similar to the one in Chapter 2 to a particular transform of the dependent data. The model will be used to analyse the data plotted in Figure 1.2. Thus, the emphasis will be on modelling the volatility. Specifically, we will model the process as being zero mean, but having a scaling function that is modelled as a product of individual components. The model will be given by

$$Y_{t,T} = \tau_0(\frac{t}{T})\prod_{j=1}^{d}\tau_j(X_t^j)\varepsilon_t \text{ for } t = 1,\dots,T \tag{1.6}$$

with ε_t a GARCH(1,1) process, a member of the fairly general yet parsimonious family of nonlinear processes introduced by Bollerslev (1986). By squaring and taking logarithms this model can also be written as in (1.4). Using the aforementioned transform, we obtain

$$\log Y_{t,T}^2 = \log\tau_0^2(\frac{t}{T}) + \sum_{j=1}^{d}\log\tau_j^2(X_t^j) + \log\varepsilon_t^2 \text{ for } t = 1,\dots,T \tag{1.7}$$

This model fits into the framework in (1.4) for the transformed dependent variable $\log Y_{t,T}^2$ with:

1. $S_t = 0$.

2. $T_t = \log\tau_0^2(\frac{t}{T}) + \sum_{j=1}^{d}\log\tau_j^2(X_t^j)$, which is the sum of a smooth function of rescaled time and the sum of component functions that depend smoothly on the regressors X_t^j.

3. $Z_t = \log\varepsilon_t^2$ a stationary process.

Estimators for the components of the model in (1.6) are suggested. Theoretical results on the asymptotic behaviour of our estimators are provided as well as a way to interpret them. Finally, the model will be applied to the data plotted in Figure 1.2.

1.3.3 Chapter 4

In Chapter 4 we will give a model that will serve as a generalization of the structure in (1.4) for the deterministic case with a known seasonal period θ. Using a nonparmateric formulation such a model can be written as

$$Y_{t,T} = m_\theta(t) + m_0(\frac{t}{T}) + \varepsilon_t \text{ for } t = 1, \dots, T, \tag{1.8}$$

with $\{\varepsilon_t\}$ some stationary process, m_θ a periodic function with known period θ and m_0 a smooth function of rescaled time. The intriguing thing about this model is the dual use of the oberservation time point t. Not only does it give us the time point, but it also carries the information on the season the observation was made in.

The model in (1.8) uses this information to model the season and the trend component as a sum. However, a priori, there seems to be no reasaon why such an approach should by adopted. In particular, it may be that there is some change in the season over time, which cannot be captured with the above model due to the constancy of each season over the entire obseration period.

Instead of specifying the seasonal component to be overlaid on the trend component we will think of the process as having an underlying season-trend function. Such a model will be given by

$$Y_{t,T} = m(\frac{t}{T}, s_t) + \varepsilon_t \text{ for } t = 1, \dots, T,$$

with $\{\varepsilon_t\}$ some stationary process, s_t denoting the season $Y_{t,T}$ was made in and $m(\cdot, \cdot) : [0, 1] \times \{0, \dots, \theta - 1\} \to \mathbb{R}$ the season-trend function. We will see that the season-trend function will be interpretable as a regression function with a categorical and a continuous regressor. We will estimate the model using a German temperature series and compare it to the additive specification in (1.8).

Chapter 2

An additive Model

In this chapter, we study a nonparametric additive regression model suitable for a wide range of time series applications. Our model includes a periodic component, a deterministic time trend, various component functions of stochastic explanatory variables, and an AR(p) error process that accounts for serial correlation in the regression error. We propose an estimation procedure for the nonparametric component functions and the parameters of the error process based on smooth backfitting and quasi-maximum likelihood methods. Our theory establishes convergence rates as well as asymptotic normality of our estimators. Moreover, we are able to derive an oracle type result for the estimators of the AR parameters: Under fairly mild conditions, the limiting distribution of our parameter estimators is the same as when the nonparametric component functions are known. Finally, we illustrate our estimation procedure by applying it to a sample of temperature and ozone data collected on the Antarctic Peninsula.

2.1 Introduction

In many time series applications, the data at hand exhibit seasonal fluctuations as well as a trending behaviour. A common way to incorporate these features is to assume that the data generating process can be written as the sum of a seasonal part, a deterministic time trend and a stationary stochastic process. In most cases, the structure of these three components is largely unknown. Hence, in order to estimate them, it is important to have flexible semi- and nonparametric methods at hand.

Let $\{Y_{t,T}, t = 1, \ldots, T\}$ be the time series under investigation. A general semiparametric framework which decomposes $Y_{t,T}$ into a seasonal, a trend and a stationary stochastic component is given by the regression model

$$Y_{t,T} = m_\theta(t) + m_0\Big(\frac{t}{T}\Big) + m(X_t) + \varepsilon_t \quad \text{for } t = 1, \ldots, T \tag{2.1}$$

with $\mathbb{E}[\varepsilon_t|X_t] = 0$. Here, m_θ is a periodic function with a known period θ and m_0 is a deterministic time trend. The stochastic component consists of the residual ε_t and of the term $m(X_t)$ which captures the influence of the d-dimensional stationary covariate vector $X_t = (X_t^1, \ldots, X_t^d)$. We do not impose any parametric restrictions on the component functions m_θ, m_0 and m. Moreover, we allow for correlation in the error terms ε_t which are modelled by a stationary AR(p) process. Note that, as usual in nonparametric regression, the time argument of the trend function m_0 is rescaled to the unit interval.

Two special cases of model (2.1) have been considered in the literature. The fixed design setting $Y_{t,T} = m_0(\frac{t}{T}) + \varepsilon_t$ has been analyzed for example in Truong (1991), Altman (1993), Hall and Keilegom (2003), and Shao and Yang (2011) who provide a variety of methods to estimate the nonparametric trend function m_0 and the AR parameters of the error term. Interestingly, they establish an oracle type result for the parameter estimators. In particular, they show that the limiting distribution of the estimators is unaffected by the need to estimate the nonparametric function m_0. A second special case of model (2.1) is the setting $Y_t = m(X_t) + \varepsilon_t$. The problem of estimating the AR parameters in this setup has been studied under the restriction that $\{X_t\}$ is independent of the error process $\{\varepsilon_t\}$.Truong and Stone (1994), Schick (1994) and Lin et al. (1999) show that under this restriction an oracle type result holds analogous to that in the fixed design setting.

In this chapter, we study estimation of the parametric and nonparametric components in the general model (2.1). We allow X_t and ε_t to be dependent, thus dispensing with the very restrictive assumption that the covariate process is independent of the errors. In order to circumvent the well-known curse of dimensionality we assume the function m to be additive with component functions m_j for $j = 1, \ldots, d$, thus yielding

$$Y_{t,T} = m_\theta(t) + m_0\Big(\frac{t}{T}\Big) + \sum_{j=1}^{d} m_j(X_t^j) + \varepsilon_t \quad \text{for } t = 1, \ldots, T. \tag{2.2}$$

A full description of model (2.2) together with a discussion of its components is given in Section 2.2.

Our estimation procedure is introduced in Section 2.3. The nonparametric components m_θ and $m_0, \ldots, m_d$ are estimated by extending the smooth backfitting approach of Mammen et al. (1999), who derived its asymptotic properties in a strictly stationary setup. Due to the inclusion of the periodic and the deterministic trend components our model dynamics are no longer stationary. In Subsections 2.3.1 and 2.3.2, we describe how to incorporate this type of nonstationarity into the smooth backfitting procedure. Given our estimates $\widetilde{m}_\theta$ and $\widetilde{m}_0, \ldots, \widetilde{m}_d$ of the functions m_θ and $m_0, \ldots, m_d$, we can construct approximate expressions $\widetilde{\varepsilon}_t$ of ε_t. Using these, the parameters of the AR(p) error process are estimated via a quasi-maximum likelihood based method, the details of which are given in Subsection 2.3.3.

Section 2.4 contains our results on the asymptotic properties of our estimators. In Subsections 2.4.2 and 2.4.3, we provide the convergence rates of the nonparametric estimators $\widetilde{m}_\theta$ and $\widetilde{m}_0, \ldots, \widetilde{m}_d$ as well as their Gaussian limit distribution. The asymptotic behaviour of the parameter estimators of the AR(p) error process is studied in Subsection 2.4.4. There, we show that the parameter estimators are asymptotically normal. Deriving the limit distribution of the parameter estimators is by far the most difficult part of the theory developed in this chapter. To do so, we need to establish a higher-order stochastic expansion of the first derivative of the likelihood function. This requires substantially different and much more intricate techniques than in the analysis of the special cases previously discussed in the literature.

As will be seen, the asymptotic distribution of our parameter estimators generally differs from that of the oracle estimators constructed under the assumption that the additive component functions are known. Thus, the additional uncertainty which stems from estimating the component functions becomes visible in the asymptotic distribution of our parameter estimators. Under fairly mild conditions on the dependence structure between the covariates X_t and the errors ε_t, however, the limiting distribution will turn out to coincide with that of the oracle estimators. The key restriction on the dependence structure is that $\mathbb{E}[\varepsilon_t|X_{t+k}] = 0$ for all $k = -p, \ldots, p$. This can be thought of as p-lag past and future exogeneity. Most importantly, it is much weaker than imposing independence between $\{X_t\}$ and $\{\varepsilon_t\}$. Thus, our theory generalizes the oracle type results found in the simpler settings discussed above.

Our estimation procedure is illustrated with a real data example in Section 2.5. We apply it to a sample of monthly minimum temperature and ozone data from the Faraday/Vernadsky research station on the Antarctic Peninsula. These data were first analyzed in Hughes et al. (2007) who use a parametric regression model with AR errors. Hence, our analysis can be regarded as a semiparametric extension to their study.

2.2 Model

Before we introduce our estimation procedure, we have a closer look at model (2.2) and comment on some of its features. We observe a sample of variables $\{Y_{t,T}, X_t\}$ for $t = 1, \ldots, T$, where $Y_{t,T}$ is real-valued and $X_t = (X_t^1, \ldots, X_t^d)$ is a strictly stationary $\mathbb{R}^d$-valued random vector. As already noted in the introduction, the data are assumed to follow the process

$$Y_{t,T} = m_\theta(t) + m_0\Big(\frac{t}{T}\Big) + \sum_{j=1}^{d} m_j(X_t^j) + \varepsilon_t \quad \text{for } t = 1, \ldots, T \tag{2.3}$$

with $\mathbb{E}[\varepsilon_t|X_t] = 0$, where m_θ is a periodic component with some known integer-valued period θ, m_0 is a deterministic trend, and the m_j are nonparametric functions of the regressors X_t^j for $j = 1, \ldots, d$. Moreover, $\{\varepsilon_t\}$ is a stationary AR(p)

process of the form

$$\varepsilon_t = \sum_{i=1}^{p} \phi_i^* \varepsilon_{t-i} + \eta_t \quad \text{for all } t \in \mathbb{Z},$$

where $\phi^* = (\phi_1^*, \dots, \phi_p^*)$ is the vector of parameters and the residuals η_t are assumed to be martingale differences.

The additive functions in model (2.3) are only identified up to an additive constant. To identify them, we assume that

$$\int_0^1 m_0(x_0)dx_0 = 0 \quad \text{and} \quad \int m_j(x_j)p_j(x_j)dx_j = 0 \quad \text{for } j = 1, \dots, d, \tag{2.4}$$

where p_j is the marginal density of X_t^j. The covariates X_t^j are supposed to take values in a bounded interval which without loss of generality is taken to be $[0, 1]$ for each $j = 1, \dots, d$. Throughout this chapter, x_0 is used to denote a point in rescaled time. Moreover, we write $x = (x_0, x_{-0})$ with $x_{-0} = (x_1, \dots, x_d)$.

To be able to do reasonable asymptotics, we let the trend function m_0 in model (2.3) depend on rescaled time $\frac{t}{T}$ rather than on real time t. If we defined m_0 in terms of real time, we would not get additional information on the structure of m_0 locally around a fixed time point t as the sample size increases. Within the framework of rescaled time, in contrast, the function m_0 is observed on a finer and finer grid of rescaled time points on the unit interval as T grows. Thus, we obtain more and more information on the local structure of m_0 around each point in rescaled time. This is the reason why we can make reasonable asymptotic considerations within this framework.

Unlike m_0, we let the periodic component m_θ in model (2.3) be a function of real time t. This allows us to exploit its periodic character when doing asymptotics: Assume we want to estimate m_θ at a time point $t_\theta \in \{1, \dots, \theta\}$. As m_θ is periodic, it has the same value at $t_\theta, t_\theta + \theta, t_\theta + 2\theta, t_\theta + 3\theta$, and so on. Hence, if m_θ depends on real time t, the number of time points in our sample at which m_θ has the value $m_\theta(t_\theta)$ increases as the sample size grows. This gives us more and more information about the value $m_\theta(t_\theta)$ and thus allows us to do asymptotics.

2.3 Estimation Procedure

We now describe how the various components of model (2.3) are estimated. Our procedure consists of three steps. In the first step, the periodic model component m_θ is estimated. The estimation of the nonparametric functions $m_0, \dots, m_d$ is addressed in the second step. Finally, we use the estimates of the additive component functions to construct estimators of the AR parameters.

2.3.1 Estimation of m_θ

For any time point $t = 1, \dots, T$, let $t_\theta = t - \lfloor \frac{t}{\theta} \rfloor \theta$ with $\lfloor x \rfloor$ denoting the largest integer, smaller than or equal to x. Our estimate of the periodic component m_θ is defined as

$$\widetilde{m}_\theta(t) = \frac{1}{K_{t_\theta,T}} \sum_{k=1}^{K_{t_\theta,T}} Y_{t_\theta+(k-1)\theta,T} \quad \text{for } t = 1, \dots, T, \tag{2.5}$$

where $K_{t_\theta,T} = 1 + \lfloor \frac{T-t_\theta}{\theta} \rfloor$ is the number of observations that satisfy $t = t_\theta + k\theta$ for some $k \in \mathbb{N}$. The estimate has a very simple structure: It is the empirical mean of observations that are separated by a multiple of θ periods.[1] Later on, we will show that $\widetilde{m}_\theta$ is asymptotically normal. Note that this result is robust to the presence of the deterministic trend function m_0. In particular, we will see that the effect of the unknown time trend m_0 on the estimate $\widetilde{m}_\theta$ can be asymptotically neglected.

2.3.2 Estimation of $m_0, \dots, m_d$

We next introduce the estimates of the functions $m_0, \dots, m_d$. For the time being let us assume that the periodic component m_θ is known. Later on, m_θ will be replaced by its estimate $\widetilde{m}_\theta$. Given that m_θ is known, $Z_{t,T} = Y_{t,T} - m_\theta(t)$ is

[1] We are estimating a periodic sequence instead of a periodic function as we only observe the periodic component at equidistant time points and do not want to make any additional functional form assumptions.

observable. This allows us to rewrite model (2.3) as

$$Z_{t,T} = m_0\Big(\frac{t}{T}\Big) + \sum_{j=1}^{d} m_j(X_t^j) + \varepsilon_t. \tag{2.6}$$

In order to estimate the functions $m_0, \dots, m_d$ in (2.6), we extend the smooth backfitting approach of Mammen et al. (1999). The asymptotic properties of this approach are well understood in a strictly stationary setup. Our setting, however, involves a deterministic time trend component which makes the model dynamics nonstationary. In what follows, we describe how to extend the smooth backfitting procedure to allow for the nonstationarities present in our setting.

To do so, we first introduce the auxiliary estimates

$$\widehat{q}(x) = \frac{1}{T}\sum_{t=1}^{T} K_h\Big(x_0, \frac{t}{T}\Big) \prod_{k=1}^{d} K_h(x_k, X_t^k)$$
$$\widehat{m}(x) = \frac{1}{T}\sum_{t=1}^{T} K_h\Big(x_0, \frac{t}{T}\Big) \prod_{k=1}^{d} K_h(x_k, X_t^k) Z_{t,T} \big/ \widehat{q}(x).$$

$\widehat{q}(x)$ is a kernel estimate of the density $q(x) := I(x_0 \in [0,1])p(x_{-0})$ with p being the joint density of the regressors $X_t = (X_t^1, \dots, X_t^d)$. Moreover, $\widehat{m}(x)$ is a $(d+1)$-dimensional Nadaraya-Watson estimate of the regression function $m(x) = m_0(x_0) + \ldots + m_d(x_d)$. In these definitions,

$$K_h(v,w) = \frac{K_h(v-w)}{\int_0^1 K_h(s-w)ds}$$

is a modified kernel weight, where $K_h(v) = \frac{1}{h}K(\frac{v}{h})$ and the kernel function $K(\cdot)$ integrates to one. These weights have the property that $\int_0^1 K_h(v,w)dv = 1$ for all w, which is needed to derive the asymptotic results for the backfitting estimates.

Given the smoothers $\widehat{q}$ and $\widehat{m}$, we define the smooth backfitting estimates $\widetilde{m}_0, \dots, \widetilde{m}_d$ as the minimizers of the criterion

$$\int_{[0,1]^{d+1}} \big(\widehat{m}(x) - g_0(x_0) - \ldots - g_d(x_d)\big)^2 \widehat{q}(x)dx, \tag{2.7}$$

where the minimization runs over all additive functions $g(x) = g_0(x_0)+\cdots+g_d(x_d)$ whose components satisfy $\int_0^1 g_j(x_j)\widehat{p}_j(x_j)dx_j = 0$ for $j = 0,\ldots,d$. Here, $\widehat{p}_j$ is a kernel estimator of p_j for $j = 0,\ldots,d$ and we define $p_0(x_0) = I(x_0 \in [0,1])$. Explicit expressions for these estimators are given below in (2.9) and (2.12).

According to the definition in (2.7), the backfitting estimate $\widetilde{m} = \widetilde{m}_0 + \ldots + \widetilde{m}_d$ is an L^2-projection of the $(d+1)$-dimensional Nadaraya-Watson smoother $\widehat{m}$ onto the space of additive functions with respect to the density $\widehat{q}$. In particular, note that $\widehat{q}$ estimates the product of a uniform density over $[0,1]$ and the density p of the regressors X_t. This shows that rescaled time is treated in a similar way to an additional stochastic regressor which is uniformly distributed over $[0,1]$ and independent of the variables X_t. The heuristic idea behind this is the following: Firstly, as the variables X_t are strictly stationary, their distribution is time-invariant. In this sense their stochastic behaviour is independent of rescaled time $\frac{t}{T}$. Thus rescaled time behaves similarly to an additional stochastic variable that is independent of X_t. Secondly, as the points $\frac{t}{T}$ are evenly spaced over the unit interval, a variable with a uniform distribution closely replicates the pattern of rescaled time.

By differentiation, we can show that the solution to the projection problem (2.7) is characterized by the system of integral equations

$$\widetilde{m}_j(x_j) = \widehat{m}_j(x_j) - \sum_{k\neq j} \int_0^1 \widetilde{m}_k(x_k)\frac{\widehat{p}_{k,j}(x_k,x_j)}{\widehat{p}_j(x_j)}\, dx_k - \widetilde{m}_c \tag{2.8}$$

with $\int_0^1 \widetilde{m}_j(x_j)\widehat{p}_j(x_j)dx_j = 0$ for $j = 0,\ldots,d$. As we do not observe the variables $Z_{t,T} = Y_{t,T} - m_\theta(t)$, we define the kernel estimates in (2.8) in terms of the approximations $\widetilde{Z}_{t,T} = Y_{t,T} - \widetilde{m}_\theta(t)$. In particular, we let

$$\widehat{p}_j(x_j) = \frac{1}{T}\sum_{t=1}^T K_h(x_j, X_t^j) \tag{2.9}$$

$$\widehat{p}_{j,k}(x_j,x_k) = \frac{1}{T}\sum_{t=1}^T K_h(x_j, X_t^j)K_h(x_k, X_t^k) \tag{2.10}$$

$$\widehat{m}_j(x_j) = \frac{1}{T}\sum_{t=1}^T K_h(x_j, X_t^j)\widetilde{Z}_{t,T}/\widehat{p}_j(x_j) \tag{2.11}$$

for $j,k = 1,\dots,d$ with $j \neq k$, where $\widehat{p}_j$ is the one-dimensional kernel density estimator of the marginal density p_j of X_t^j, $\widehat{p}_{j,k}$ is the two-dimensional kernel density estimate of the joint density $p_{j,k}$ of (X_t^j, X_t^k), and $\widehat{m}_j$ is a one-dimensional Nadaraya-Watson smoother. Moreover,

$$\widehat{p}_0(x_0) = \frac{1}{T}\sum_{t=1}^{T} K_h\Big(x_0, \frac{t}{T}\Big) \tag{2.12}$$

$$\widehat{p}_{0,k}(x_0, x_k) = \frac{1}{T}\sum_{t=1}^{T} K_h\Big(x_0, \frac{t}{T}\Big) K_h(x_k, X_t^k) \tag{2.13}$$

$$\widehat{m}_0(x_0) = \frac{1}{T}\sum_{t=1}^{T} K_h\Big(x_0, \frac{t}{T}\Big) \widetilde{Z}_{t,T}/\widehat{p}_0(x_0) \tag{2.14}$$

for $k = 1,\dots,d$ and $\widetilde{m}_c = \frac{1}{T}\sum_{t=1}^{T} \widetilde{Z}_{t,T}$. Note that it would be more natural to define $\widehat{p}_0(x_0) = I(x_0 \in [0,1])$, as we already know the "true density" of rescaled time. However, for technical reasons, we set $\widehat{p}_0(x_0) = \frac{1}{T}\sum_{t=1}^{T} K_h(x_0, \frac{t}{T})$. This creates a behaviour of the estimate $\widehat{p}_0$ in the boundary region of the support $[0,1]$ analogous to that of $\widehat{p}_j$ at the boundary.[2]

In our theoretical analysis, we work with the smooth backfitting estimators characterized as the solution to the system of integral equations (2.8). Note however that in general, the system of equations (2.8) cannot be solved analytically. Nevertheless, the solution can be approximated by an iterative projection algorithm which converges for arbitrary starting values; see Mammen et al. (1999), who establish the asymptotic properties of this algorithm under very general high order conditions. Our technical arguments will show that these high order conditions are satisfied in our framework.

[2]Alternatively, we could define $\widehat{p}_0(x_0) = \int_0^1 K_h(x_0, v)dv$. (Note that $\int_0^1 K_h(x_0, v)dv = 1$ for $x_0 \in [2C_1 h, 1 - 2C_1 h]$, where $[-C_1, C_1]$ is the support of the kernel function K.) Moreover, we could set $\widehat{p}_{0,k}(x_0, x_k) = \widehat{p}_0(x_0)\widehat{p}_k(x_k)$, thereby exploiting the "independence" of rescaled time and the other regressors.

2.3.3 Estimation of the AR Parameters

To motivate the third step in our estimation procedure, we shall initially consider an infeasible estimator of the model parameters. Suppose that the functions $m_\theta, m_0, \dots, m_d$ were known. In this situation, the AR(p) error process ε_t would be observable, since

$$\varepsilon_t = Y_{t,T} - m_\theta(t) - m_0\Big(\frac{t}{T}\Big) - \sum_{j=1}^{d} m_j(X_t^j). \tag{2.15}$$

The parameters $\phi^* := (\phi_1^*, \dots, \phi_p^*)$ of the error process could thus be estimated by standard maximum likelihood methods. In particular, we could use a conditional maximum likelihood estimator of the form

$$\widehat{\phi} = \arg\max_{\phi\in\Phi} l_T(\phi), \tag{2.16}$$

where Φ is a compact parameter space and l_T is the conditional log-likelihood given by

$$l_T(\phi) = -\sum_{t=p+1}^{T} \big(\varepsilon_t - \varepsilon_t(\phi)\big)^2 \tag{2.17}$$

with $\varepsilon_t(\phi) = \sum_{i=1}^{p} \phi_i \varepsilon_{t-i}$. Note that $\widehat{\phi}$ has a closed form solution which is identical to the usual least squares estimate. We will, however, not work with this closed form solution in what follows. Instead we will formulate our proofs in terms of the likelihood function. This makes it easier to apply our arguments to other error structures such as ARCH processes. We give some comments on how to extend our approach in this direction in Section 2.6.

As the functions $m_\theta, m_0, \dots, m_d$ are not observed, we cannot use the standard approach from above directly. However, given the estimates $\widetilde{m}_\theta, \widetilde{m}_0, \dots, \widetilde{m}_d$ from the previous estimation steps, we can replace ε_t by the estimates

$$\widetilde{\varepsilon}_t = Y_{t,T} - \widetilde{m}_\theta(t) - \widetilde{m}_0\Big(\frac{t}{T}\Big) - \sum_{j=1}^{d} \widetilde{m}_j(X_t^j) \tag{2.18}$$

and use these as approximations to ε_t in the maximum likelihood estimation. The log-likelihood then becomes

$$\widetilde{l}_T(\phi) = -\sum_{t=p+1}^{T} \left(\widetilde{\varepsilon}_t - \widetilde{\varepsilon}_t(\phi)\right)^2 \tag{2.19}$$

with $\widetilde{\varepsilon}_t(\phi) = \sum_{i=1}^{p} \phi_i \widetilde{\varepsilon}_{t-i}$. Our estimator $\widetilde{\phi}$ of the true parameter values ϕ^* is now defined as

$$\widetilde{\phi} = \arg\max_{\phi \in \Phi} \widetilde{l}_T(\phi). \tag{2.20}$$

2.4 Asymptotics

In this section, we analyze the asymptotic properties of our estimators. The first subsection lists the assumptions required for our analysis. The following subsections describe the main asymptotic results, with each subsection dealing with a separate step of our estimation procedure.

2.4.1 Assumptions

To derive the asymptotic properties of the nonparametric estimators $\widetilde{m}_\theta, \widetilde{m}_0, \ldots, \widetilde{m}_d$, the following assumptions are needed.

(A1) *The process* $\{X_t, \varepsilon_t\}$ *is strictly stationary and strongly mixing with mixing coefficients* α *satisfying* $\alpha(k) \le a^k$ *for some* $0 < a < 1$.

(A2) *The variables* X_t *have compact support, say* $[0,1]^d$. *The density* p *of* X_t *and the densities* $p_{(0,l)}$ *of* (X_t, X_{t+l}), $l = 1, 2, \ldots,$ *are uniformly bounded. Furthermore,* p *is bounded away from zero on* $[0,1]^d$.

(A3) *The functions* m_0 *and* m_j $(j = 1, \ldots, d)$ *are twice differentiable with Lipschitz continuous second derivatives. The first partial derivatives of* p *exist and are continuous.*

(A4) *The kernel K is bounded, symmetric about zero and has compact support ($[-C_1, C_1]$, say). Moreover, it fulfills the Lipschitz condition that there exists a positive constant L with $|K(u) - K(v)| \leq L|u - v|$.*

(A5) *There exists a real constant C and a natural number l^* such that $\mathbb{E}[|\varepsilon_t|^\rho \big| X_t] \leq C$ for some $\rho > \frac{8}{3}$ and $\mathbb{E}[|\varepsilon_t \varepsilon_{t+l}| \big| X_t, X_{t+l}] \leq C$ for all $l \geq l^*$.*

(A6) *The bandwidth h satisfies either of the following:*

(a) $T^{\frac{1}{5}} h \to c_h$ *for some constant c_h.*

(b) $T^{\frac{1}{4}+\delta} h \to c_h$ *for some constant c_h and some small $\delta > 0$.*

The above assumptions are very similar to the conditions needed for smooth backfitting in the stationary case to be found e.g. in Mammen et al. (1999), Mammen and Park (2006) or Yu et al. (2011). It should also be mentioned that we do not necessarily require exponentially decaying mixing rates as assumed in (A1). These could alternatively be replaced by sufficiently high polynomial rates. We nevertheless make the stronger assumption (A1) to keep the notation and structure of the proofs as clear as possible.

In order to show that the estimators of the AR parameters are consistent and asymptotically normal, we additionally require the following assumptions.

(A7) *The parameter space Φ is a compact subset of $\{\phi \in \mathbb{R}^p \mid \phi(z) = 1 - \phi_1 z - \ldots - \phi_p z^p \neq 0$ for all complex z with $|z| \leq 1$ and $\phi_p \neq 0\}$. The true parameter vector $\phi^* = (\phi_1^*, \ldots, \phi_p^*)$ is an interior point of Φ.*

(A8) $\mathbb{E}[\varepsilon_t^{4+\delta}] < \infty$, *for some $\delta > 0$.*

(A9) *There exists a real constant C and a natural number l^* such that $\mathbb{E}[|\varepsilon_t| \big| X_{t+k}] \leq C$ and $\mathbb{E}[|\varepsilon_t \varepsilon_{t+l}| \big| X_{t+k}, X_{t+l}] \leq C$ for all l with $|l| \geq l^*$ and $k = -p, \ldots, p$.*

The compactness assumption in (A7) is standard. (A8) and (A9) are technical assumptions needed to show asymptotic normality.

2.4.2 Asymptotics for $\widetilde{m}_\theta$

We start by considering the asymptotic behaviour of the estimate $\widetilde{m}_\theta$. The next theorem shows that it is asymptotically normal.

Theorem 2.4.1. *Assume that $\mathbb{E}|\varepsilon_t|^\rho < \infty$ for some $\rho > 2$ and let* (A1) *be fulfilled. Then*

$$\sqrt{T}(\widetilde{m}_\theta(t) - m_\theta(t)) \xrightarrow{d} N(0, V_\theta)$$

for all $t = 1, \dots, T$, where

$$V_\theta = \theta \sum_{k=-\infty}^{\infty} \mathrm{Cov}(W_0, W_{k\theta})$$

with $W_t = Y_{t,T} - m_\theta(t) - m_0(\frac{t}{T}) = \sum_{j=1}^d m_j(X_t^j) + \varepsilon_t$.

As $\widetilde{m}_\theta$ and m_θ are periodic, this trivially implies that

$$\sup_{t=1,\dots,T} |\widetilde{m}_\theta(t) - m_\theta(t)| = \sup_{t=1,\dots,\theta} |\widetilde{m}_\theta(t) - m_\theta(t)| = O_p\Big(\frac{1}{\sqrt{T}}\Big).$$

The proof of Theorem 2.4.1 is straightforward: We have

$$\begin{aligned}\widetilde{m}_\theta(t) - m_\theta(t) &= \frac{1}{K_{t_\theta,T}} \sum_{k=1}^{K_{t_\theta,T}} m_0\Big(\frac{t_\theta + (k-1)\theta}{T}\Big) + \frac{1}{K_{t_\theta,T}} \sum_{k=1}^{K_{t_\theta,T}} W_{t_\theta+(k-1)\theta} \\ &=: (A) + (B).\end{aligned}$$

The term (A) approximates the integral $\int_0^1 m_0(u)du$. It is easily seen that the convergence rate is $O(\frac{1}{T})$. As $\int_0^1 m_0(u)du = 0$ by the normalization in (2.4), we obtain that (A) is of the order $O(\frac{1}{T})$ and can thus be neglected asymptotically. Noting that $\{W_t\}$ is mixing by (A1) and has mean zero by our normalization, we can now apply a central limit theorem for mixing variables to the term (B) to get the normality result of Theorem 2.4.1.

2.4.3 Asymptotics for $\widetilde{m}_0, \ldots, \widetilde{m}_d$

The main result of this subsection characterizes the limiting behaviour of the smooth backfitting estimates $\widetilde{m}_0, \ldots, \widetilde{m}_d$. It shows that the estimators converge uniformly to the true component functions at the one-dimensional nonparametric rates no matter how large the dimension d of the full regression function. Moreover, it characterizes the asymptotic distribution of the estimators.

Theorem 2.4.2. *Suppose that conditions* (A1) – (A5) *hold.*

(a) Assume that the bandwidth h satisfies (A6)(a) *or* (A6)(b). *Then, for $I_h = [2C_1h, 1-2C_1h]$ and $I_h^c = [0, 2C_1h) \cup (1-2C_1h, 1]$,*

$$\sup_{x_j \in I_h} \left|\widetilde{m}_j(x_j) - m_j(x_j)\right| = O_p\Big(\sqrt{\frac{\log T}{Th}}\Big) \tag{2.21}$$

$$\sup_{x_j \in I_h^c} \left|\widetilde{m}_j(x_j) - m_j(x_j)\right| = O_p(h) \tag{2.22}$$

for all $j = 0, \ldots, d$.

(b) Assume that the bandwidth h satisfies (A6)(a). *Then, for any $x_0, \ldots, x_d \in (0,1)$,*

$$T^{\frac{2}{5}} \begin{bmatrix} \widetilde{m}_0(x_0) - m_0(x_0) \\ \vdots \\ \widetilde{m}_d(x_d) - m_d(x_d) \end{bmatrix} \xrightarrow{d} N(B(x), V(x))$$

with the bias term $B(x) = [c_h^2(\beta_0(x_0) - \gamma_0), \ldots, c_h^2(\beta_d(x_d) - \gamma_d)]'$ and the covariance matrix $V(x) = \mathrm{diag}(v_0(x_0), \ldots, v_d(x_d))$. Here, $v_0(x_0) = c_h^{-1} c_K \sum_{l=-\infty}^{\infty} \gamma_\varepsilon(l)$ and $v_j(x_j) = c_h^{-1} c_K \sigma_j^2(x_j)/p_j(x_j)$ for $j = 1, \ldots, d$ with $\gamma_\varepsilon(l) = \mathrm{Cov}(\varepsilon_t, \varepsilon_{t+l})$, $\sigma_j^2(x_j) = \mathrm{Var}(\varepsilon_t | X_t^j = x_j)$ and the constants $c_h = \lim_{T\to\infty} T^{1/5} h$ and $c_K = \int K^2(u)du$. Furthermore, the functions β_j are the components of the $L^2(q)$-projection of the function β defined in Lemma A.3.3 of Appendix A.3 onto the space of additive functions. Finally, the constants γ_j can be characterized by the equation $\int_0^1 \alpha_{T,j}(x_j) \widehat{p}_j(x_j) dx_j = h^2 \gamma_j + o_p(h^2)$ for $j = 0, \ldots, d$, with $\alpha_{T,j}$ also given in Lemma A.3.3 of Appendix A.3.

As described in Subsection 2.3.2, rescaled time $\frac{t}{T}$ behaves similarly to an additional uniformly distributed regressor that is independent of the other regressors. This consideration allows us to derive the above result by extending the proving strategy of Mammen et al. (1999). The details are given in Appendix A.1.

2.4.4 Asymptotics for the AR Parameter Estimates

Lastly, we establish the asymptotic properties of our estimator $\widetilde{\phi}$ of the AR parameters ϕ^*. The technical details can be found in Appendix A.2. The first theorem shows that $\widetilde{\phi}$ is consistent.

Theorem 2.4.3. *Suppose that the bandwidth h satisfies* (A6)(a) *or* (A6)(b). *In addition, let assumptions* (A1) – (A5) *and* (A7) *be fulfilled. Then $\widetilde{\phi}$ is a consistent estimator of ϕ^*, i.e.* $\widetilde{\phi} \xrightarrow{P} \phi^*$.

The central result of our theory specifies the limiting distribution of $\widetilde{\phi}$.

Theorem 2.4.4. *Suppose that the bandwidth h satifies* (A6)(b) *and let assumptions* (A1) – (A5) *together with* (A7) – (A9) *be fulfilled. Then it holds that*

$$\sqrt{T}(\widetilde{\phi} - \phi^*) \xrightarrow{d} N(0, V^*)$$

with

$$V^* = \Gamma_p^{-1}(W + \Omega)\Gamma_p^{-1}.$$

Here, Γ_p is the autocovariance matrix of the AR(p) process $\{\varepsilon_t\}$, i.e. $\Gamma_p = (\gamma(i-j))_{i,j=1,\dots,p}$ with $\gamma(i-j) = \mathbb{E}[\varepsilon_0 \varepsilon_{i-j}]$. Moreover, $W = (\mathbb{E}[\eta_0^2 \varepsilon_{-i}\varepsilon_{-j}])_{i,j=1,\dots,p}$ and the matrix Ω is defined in equation (A.24) of Appendix A.2.

Consider for a moment the case in which the functions m_θ and $m_0, \dots, m_d$ are known. In this case, we can use the "oracle" estimator $\widehat{\phi}$ defined in (2.16) to estimate the AR parameters ϕ^*. Standard theory tells us that $\widehat{\phi}$ is asymptotically normal with asymptotic variance $\Gamma_p^{-1} W \Gamma_p^{-1}$. Theorem 2.4.4 thus shows that in general the limiting distribution of our estimator $\widetilde{\phi}$ differs from that of the oracle estimator. There is however a wide range of cases where $\widetilde{\phi}$ has the same asymptotic distribution as $\widehat{\phi}$. This oracle type result is stated in the following corollary.

Corollary 2.4.1. *Suppose that all the assumptions of Theorem 2.4.4 are fulfilled and that* $\mathbb{E}[\varepsilon_t | X_{t+k}] = 0$ *for all* $k = -p, \ldots, p$. *Then*

$$\sqrt{T}(\widetilde{\phi} - \phi^*) \xrightarrow{d} N(0, \Gamma_p^{-1} W \Gamma_p^{-1}).$$

Corollary 2.4.1 follows directly from the proof of Theorem 2.4.4: Inspecting the functions defined in Lemma A.2.1 and realizing that they are identically equal to zero under the assumptions of the corollary, the matrix Ω is immediately seen to be equal to zero as well. The corollary shows that the oracle result holds under fairly mild conditions on the dependence structure between X_t and ε_t, in particular under much weaker conditions than independence of the processes $\{X_t\}$ and $\{\varepsilon_t\}$. To give an example where the conditions of the corollary are satisfied but where the processes $\{X_t\}$ and $\{\varepsilon_t\}$ are not independent, consider the following: Let the errors be given by the AR(p) process $\varepsilon_t = \sum_{i=1}^{p} \phi_i^* \varepsilon_{t-i} + \eta_t$ with $\eta_t = \sigma(X_t)\xi_t$, where σ is a continuous volatility function and $\{\xi_t\}$ is a process of zero-mean i.i.d. variables that is independent of $\{X_t\}$. A simple argument shows that $\mathbb{E}[\varepsilon_t | \{X_t\}] = 0$ in this case, i.e. strict exogeneity holds. The assumption in the corollary, which can be thought of as p-lag past and future exogeneity also holds, whereas it is easily seen that the processes $\{X_t\}$ and $\{\varepsilon_t\}$ are not independent given that the function σ is non-constant.

Note that our theory also reestablishes the oracle result derived in the simpler setup without stochastic covariates, i.e. in the model

$$Y_{t,T} = m_\theta(t) + m_0\left(\frac{t}{T}\right) + \varepsilon_t \quad \text{for } t = 1, \ldots, T \tag{2.23}$$

with $\mathbb{E}[\varepsilon_t] = 0$. In this case, the periodic component can be estimated as described in Subsection 2.3.1. Moreover, we can use a Nadaraya-Watson smoother of the form (2.14) to approximate the trend component m_0. A vastly simplified version of the proof for Theorem 2.4.4 shows that the limiting distribution of the AR parameter estimates is identical to that of the oracle estimates in this setting. In particular, the stochastic higher-order expansion derived in Lemma A.2.1 is not required any more. The arguments of the much simpler Lemma A.2.2 are sufficient to derive the result. To understand the main technical reasons why the argument simplifies so substantially, we refer the reader to the remarks given after the proof of Lemma A.2.2 in Appendix A.2.

The normality results of Theorem 2.4.4 and Corollary 2.4.1 enable us to calculate confidence bands for the AR parameter estimators and to conduct inference based on these. To do so, we need a consistent estimator of the asymptotic variance of $\widetilde{\phi}$. Whereas such an estimator is easily obtained under the conditions of Corollary 2.4.1, it is not at all trivial to derive a consistent estimator of V^* in Theorem 2.4.4. This is due to the very complicated structure of the matrix Ω which involves functions obtained from a higher-order expansion of the stochastic part of the backfitting estimates (see Theorem A.1.1 in Appendix A.1). To circumvent these difficulties, one may try to set up a bootstrap approach to estimate confidence bands and to do testing. The normality result of Theorem 2.4.4 can be used as a starting point to derive consistency results for such a bootstrap procedure. Some suggestions how to bootstrap are given in Section 2.6.

2.5 Application

In this section we apply our estimation procedure to a set of monthly temperature and ozone data from the Faraday/Vernadsky research station on the Antarctic Peninsula.[3] A strong warming trend has been identified on the whole peninsula during the past 50 years. In particular, the monthly mean temperatures at Faraday station have considerably increased over this time (cf. Turner et al. (2002), Turner et al. (2005)). According to Hughes et al. (2007), the rise of the mean temperature is mostly due to an increase in the minimum temperature. They argue that to understand and quantify the warming on the peninsula an appropriate statistical model of the minimum temperature is called for. Following their lead we will focus on modelling the minimum temperature and consider stratosperic ozone as a potential explanatory variable. The data used in our analysis is plotted in Figure

[3] The data can be downloaded from the webpage of Suhasini Subba Rao `http://www.stat.tamu.edu/~suhasini/data.html`. Alternatively, it is available on request from the British Antarctic Survey, Cambridge.

2.1. The upper panel contains the monthly minimum near-surface temperatures at Faraday station from September 1957 to December 2004. The lower panel shows the monthly level of stratospheric ozone concentration measured in Dobson units over the same period. For more information on the data consult Hughes et al. (2007), where a detailed description of them can be found.

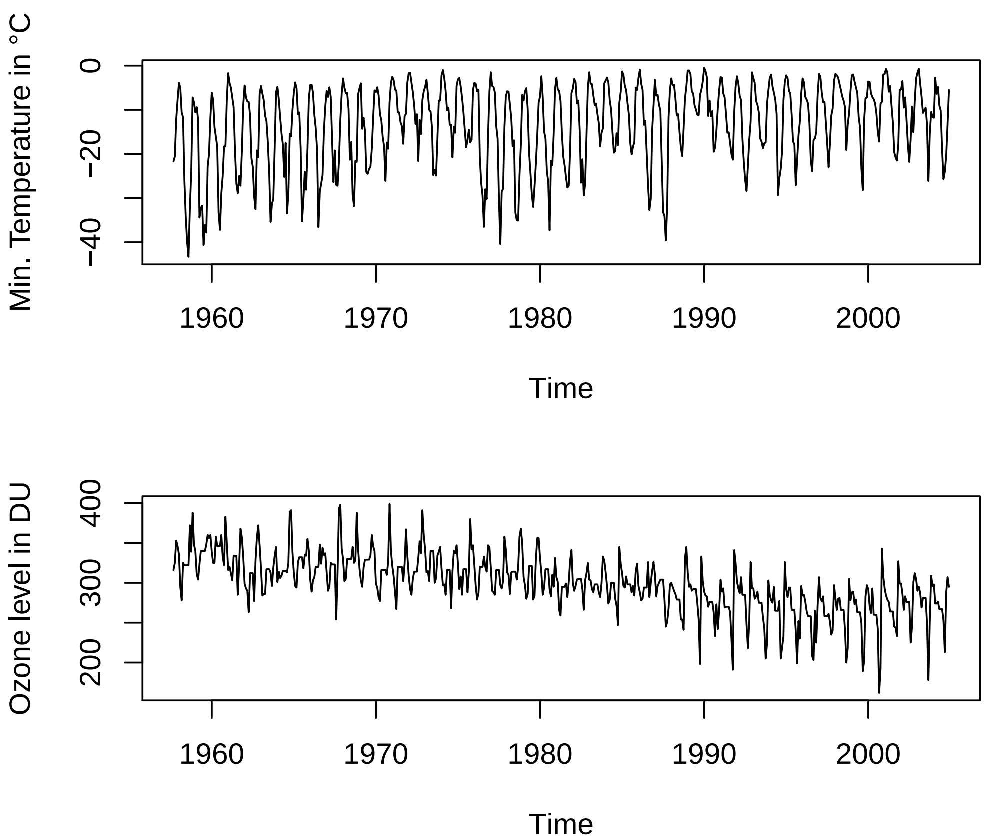


Figure 2.1: The upper panel shows the monthly minimum near-surface temperatures (in °C), the lower one the monthly stratospheric ozone concentrations (in Dobson units) at Faraday station.

Hughes et al. (2007) propose a parametric model with a linear time trend and a parametrically specified periodic component with a period of 12 months to fit the temperature and ozone data. Their baseline model is given by the equation

$$Y_t = a_0 + a_1 \sin\left(\frac{2\pi}{12}t\right) + a_2 \cos\left(\frac{2\pi}{12}t\right) + a_3 t + \varepsilon_t, \tag{2.24}$$

where Y_t denotes the minimum temperature and $a = (a_1, \ldots, a_3)$ is a vector of parameters. In addition, they consider the extended model

$$Y_t = a_0 + a_1 \sin\left(\frac{2\pi}{12}t\right) + a_2 \cos\left(\frac{2\pi}{12}t\right) + a_3 t + a_4 X_{t-1} + \varepsilon_t, \tag{2.25}$$

where the linear covariate X_{t-1} denotes the lagged detrended and deseasonalized ozone concentration. In their analysis, they find a strong linear upward trend in the minimum temperatures. Moreover, they observe considerable autocorrelation in the residuals and propose an AR model for ε_t. Using an order selection criterion, they find an AR(1) model to be most suitable, which also fits nicely with the preference for AR(1) errors when using discrete time series to model climate data as mentioned in Mudelsee (2010).

We now introduce a framework that can be regarded as a semiparametric extension of the parametric models (2.24) and (2.25). Our baseline model is given by

$$Y_{t,T} = m_\theta(t) + m_0\Big(\frac{t}{T}\Big) + \varepsilon_t \quad \text{for } t = 1, \ldots, T, \tag{2.26}$$

where $Y_{t,T}$ are minimum monthly temperatures, m_θ is a seasonal component and m_0 is a nonparametric time trend. We additionally consider an extended version of (2.26) having the form

$$Y_{t,T} = m_\theta(t) + m_0\Big(\frac{t}{T}\Big) + m_1(X_{t-1}) + \varepsilon_t \quad \text{for } t = 1, \ldots, T, \tag{2.27}$$

where as before, the variables X_{t-1} denote lagged monthly stratospheric ozone concentration levels that have been detrended and deseasonalized as in Hughes et al. (2007). The nonparametric functions in (2.26) and (2.27) are identified by the normalizations used in (2.4). Following Hughes et al. (2007) we assume the variables ε_t to have an AR(1) structure and allow for the minimum temperatures to have a 12-month cycle by setting $\theta = 12$.

Before giving our estimates we will provide the preferred fits of the models (2.24) and (2.25) given in Hughes et al. (2007) in order to compare our estimates to theirs. Their models are fitted using observations up until and including December 2003. For the model (2.24) their preferred fit is

$$Y_t = 6.25 \sin\left(\frac{2\pi}{12}t\right) + 6.95 \cos\left(\frac{2\pi}{12}t\right) + 0.0105t + \varepsilon_t, \tag{2.28}$$

with $\varepsilon_t = 0.566\varepsilon_{t-1} + \eta_t$ and η_t distributed as a converse GEV. Their preferred fit for the model in (2.25) is

$$Y_t = 6.61 \sin\left(\frac{2\pi}{12}t\right) + 7.22 \cos\left(\frac{2\pi}{12}t\right) + 0.0091t - 0.0267X_{t-1} + \varepsilon_t, \tag{2.29}$$

with $\varepsilon_t = 0.562\varepsilon_{t-1} + \eta_t$ and η_t distributed as a converse GEV.

We now turn to the estimation of our models (2.26) and (2.27). As in Hughes et al. (2007) we will estimate our models using the observed data up until December 2003. Using our three step procedure outlined in Section 2.3, we can estimate the additive component functions of (2.26) and (2.27) together with the AR parameter of the error term.

The estimate of the periodic component m_θ is given by the circles in Figure 2.2. The corresponding estimated 95% pointwise confidence bands are given by the dotted lines. Using the dashed line we have superimposed the estimated periodic function from the parametric model (2.29), whose values are given on the right y-axis. Two differences between our periodic component estimate and the parametric estimate given in (2.29) become apparent immediately. Firstly, our periodic component gives the lowest estimated monthly effect in the southern hemisphere winter month of August, whereas the lowest estimated monthly effect is in July and August, when using the parametric model. Secondly, in contrast to the parametric component our estimate is not symmetric: The fall in the minimum temperature from January to August is more gradual than the increase from August until January. Interestingly, the median monthly minimum temperature also follows this pattern as can be seen in the boxplot of the monthly minimum temperatures provided in Figure 1(b) of Hughes et al. (2007).

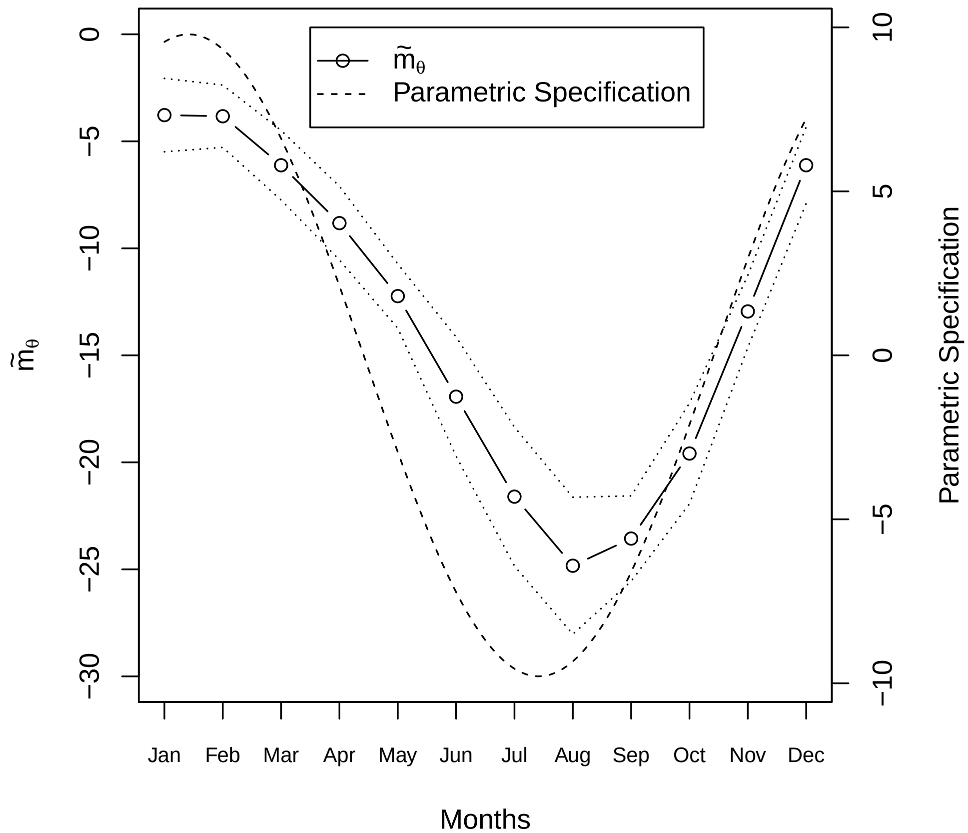


Figure 2.2: The circles represent the estimates of the seasonal component m_θ of models (2.26) and (2.27) along with the estimated 95% pointwise confidence bands (dotted lines). The dashed line is $6.61 \sin\left(\frac{2\pi}{12}t\right) + 7.22 \cos\left(\frac{2\pi}{12}t\right)$, the estimate of the seasonal component from the fitted parametric model in (2.29) obtained by Hughes et al. (2007).

Figure 2.3 shows the smooth backfitting estimates of the additive functions m_0 and m_1 in model (2.27) along with the slope estimates obtained in the parametric model (2.29). As the Nadaraya-Watson estimate of m_0 in the simpler model (2.26) is very similar to the estimate in (2.27), we do not plot it separately. For the estimation of the functions m_0 and m_1, we have used an Epanechnikov kernel and bandwidths selected by a simple plug-in rule. To check the robustness of our results, we have additionally repeated our analysis for a wide range of different bandwidths. As the results are very similar, we only report the findings for the bandwidths chosen by the plug-in rule.

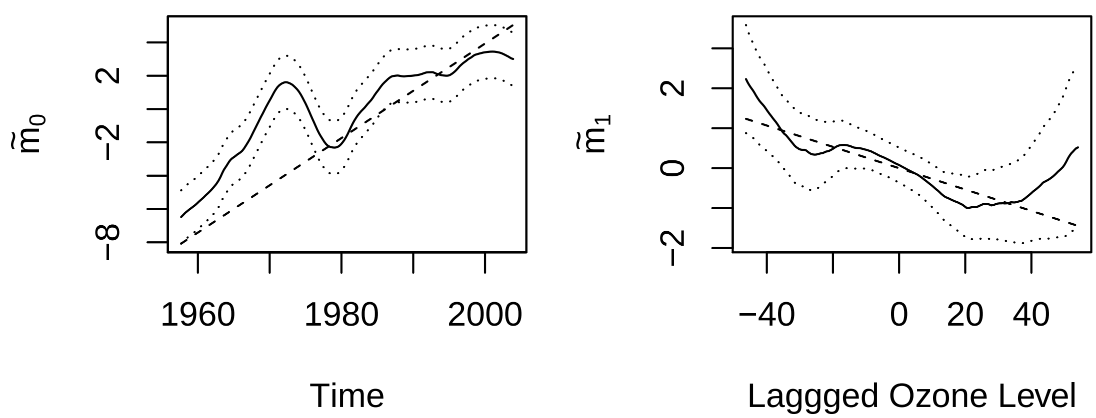


Figure 2.3: Estimation results for model (2.27). The solid lines are the smooth backfitting estimates $\widetilde{m}_0$ and $\widetilde{m}_1$, the dotted lines are pointwise 95% confidence bands. The dashed lines provide the slope estimates from the fitted parametric model in (2.29) obtained by Hughes et al. (2007).

From the shape of $\widetilde{m}_0$ together with the rather tight 95% confidence bands in the left hand panel of Figure 2.3, there seems to be a strongly nonlinear upward moving trend in the minimum monthly temperature. At first, the temperature increases quite sharply until about 1972. It then falls until roughly 1979. The subsequent equally steep increase until about 1986 is followed by a much flatter non-monotonic rise until 2004. Not only is the linear parametric trend in (2.29) not capable of capturing this nonlinear pattern, we can also see that it overestimates the overall trend increase in the monthly minimum temperature over the entire estimation period.

The estimate $\widetilde{m}_1$ in the right hand panel of Figure 2.3 suggests that the lagged ozone concentration level has a negative effect on the minimum temperature. Although the effect appears to be nonlinear again, the deviation from linearity does not seem to be as severe as for $\widetilde{m}_0$. Furthermore, the difference in the overall slope between our estimate and the parametric estimate provided in (2.29) is not so obvious as for the estimated trend components.

From the third step of our estimation procedure, we obtain estimated AR parameters of 0.57 and 0.58 for the models (2.26) and (2.27) respectively. These are essentially identical to the estimates obtained by Hughes et al. (2007) in the parametric models (2.28) and (2.29). Not only are the point estimates identical, but the parameter uncertainty is also fairly similar. Recall from the discussion in Subsection 2.4.4, that estimating confidence intervals for the parameter estimate in the extended model (2.27) is extremely involved if we are not willing to make the assumptions of Corollary 2.4.1. Thus, we shall be content with giving the estimated 95% confidence band for the simple model (2.26) here, which is given by $[0.49, 0.67]$. The corresponding estimated band for the simple parametric model (2.28) is $[0.51, 0.62]$. So, the estimated 95% confidence band for the parametric model (2.28) is slightly narrower than the one for our simple model (2.26). To summarize, it seems like the residual process displays significant positive persistence which is a common phenomenon for climate data (see Mudelsee (2010)).

Above we have compared the estimates of the different model components of our models (2.26) and (2.27) with their respective counterparts in the parametric models (2.28) and (2.29) of Hughes et al. (2007). As a final comparison we will repeat the forecasting exercise in Hughes et al. (2007), i.e. we compute rolling one-step ahead forecasts of the minimum temperature for the twelve months from January 2004 until December 2004. The results are presented in Figure 2.4.

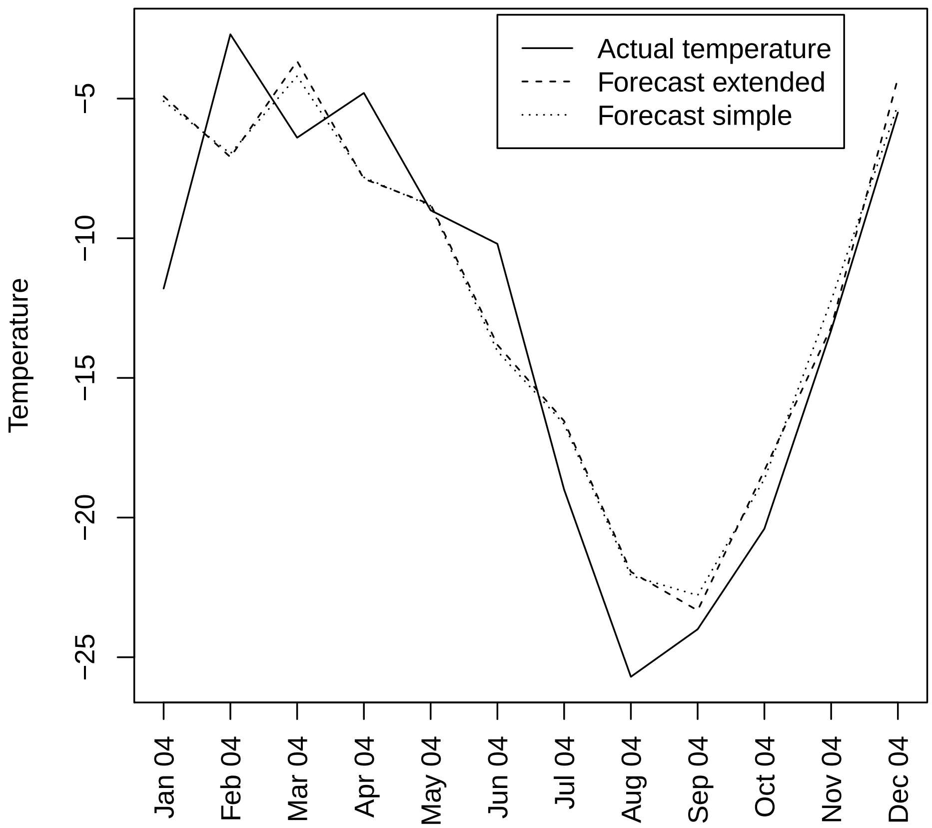


Figure 2.4: Forecasting results for the period from January 2004 to December 2004. The solid line shows the actual minimum temperatures in 2004, the dashed line gives the one-step ahead forecasts based on the extended model (2.27), and the dotted line depicts the corresponding forecasts based on the simple model (2.26).

To calculate the one-step ahead forecast for time point $t_0 + 1$, we estimate the model based on the observations at $t = 1, \ldots, t_0$. The estimated trend function $\widetilde{m}_0$ is extrapolated constantly into the future. The estimated mean squared error (MSE) of the forecasts based on model (2.27) is 10.27, whereas for the simple model (2.26) it amounts to 9.70. Note that these are somewhat lower than those in Hughes et al. (2007), who report an estimated MSE of 11.09 for model (2.28). Moreover, note that the estimated MSE for the simpler model (2.26) is slightly lower than the one for the extended model (2.27). This indicates that in terms of forecasting, we do not gain from including the lagged ozone level as an additional covariate. Thus, if the interest lies in forecasting then the simpler model (2.26) may be the better choice.

2.6 Concluding Remarks

We have studied a semiparametric regression framework whereby the time series under consideration is modelled as the sum of a periodic function, a deterministic time trend, an additive function of stationary covariates and an AR(p) residual. We have provided a method to estimate the various components of this model and have established the asymptotic properties of our estimators. In particular, we have shown that the estimators of the nonparametric component functions as well as those of the AR parameters are asymptotically normal. Importantly, in a wide range of cases the limiting distribution of the AR parameter estimators is the same as when the nonparametric component functions are known.

Our theory can be extended in several directions. As briefly mentioned in Subsection 2.3.3, our proving strategy may be applied to other error structures as well. An important example is the case in which we suspect the residuals to be heteroskedastic and model them via an ARCH(p) process. Going along the lines of the proofs for Theorems 2.4.3 and 2.4.4, the ARCH parameter estimators can be shown to be consistent and asymptotically normal. The only difference to the AR case is that the conditional likelihood has a more complicated form, making it more tedious to derive the expansion of the first derivative of the likelihood function in the normality proof.

Our proving strategy may also be applied to ARMA(p,q) and GARCH(p,q) residuals. This is most easily seen for a causal and invertible ARMA($1,1$) process

$\{\varepsilon_t\}$ which satisfies the equation

$$\varepsilon_t - \phi^* \varepsilon_{t-1} = \eta_t + \theta^* \eta_{t-1}$$

for some white noise residuals η_t. In this case, the conditional likelihood can be written as

$$l_T(\phi, \theta) = -\sum_{t=1}^{T} \big(\varepsilon_t - \varepsilon_t(\phi, \theta)\big)^2 \quad \text{with} \quad \varepsilon_t(\phi, \theta) = \sum_{k=1}^{t-1} (-\theta)^{k-1} (\phi + \theta) \varepsilon_{t-k},$$

which has a very similar structure to the likelihood function of the AR(p) case. The only notable difference is that the sum over k in the definition of $\varepsilon_t(\phi, \theta)$ now has $t-1$ elements rather than only a fixed number p. As the elements of the sum are weighted by the coefficients $(-\theta)^{k-1}(\phi + \theta)$ which decay exponentially fast to zero this does not cause any major problems in the proofs. In particular, we can truncate the sum at $\min\{t-1, C \log T\}$ for a sufficiently large C, the remainder being asymptotically negligible. After this truncation, the arguments of the AR(p) case apply more or less unchanged.

Moving to the higher order ARMA(p, q) setup, the structure of the likelihood function becomes much more complicated. It is thus convenient to base the estimation of the parameters on a criterion function which is a bit simpler to handle. In particular, consider a causal and invertible ARMA(p, q) process $\{\varepsilon_t\}$ of the form

$$\varepsilon_t - \sum_{i=1}^{p} \phi_i^* \varepsilon_{t-i} = \eta_t + \sum_{j=1}^{q} \theta_j^* \eta_{t-j}$$

and write $\phi^* = (\phi_1^*, \ldots, \phi_p^*)$ as well as $\theta^* = (\theta_1^*, \ldots, \theta_q^*)$. Due to the invertibility $1 + \sum_{j=1}^q \theta_j^* z^j \neq 0$ for all complex $|z| \leq 1$, there exist coefficients $\rho_k^* = \rho_k(\theta^*)$ with

$$\Big(1 + \sum_{j=1}^{q} \theta_j^* z^j\Big)^{-1} = \sum_{k=0}^{\infty} \rho_k^* z^k$$

for all $|z| \leq 1$. Using this, we obtain that

$$\sum_{k=0}^{\infty} \rho_k^* \Big(\varepsilon_{t-k} - \sum_{i=1}^{p} \phi_i^* \varepsilon_{t-k-i}\Big) = \eta_t.$$

Truncating the infinite sum on the left-hand side, we now define the expressions

$$\eta_t(\phi, \theta) = \sum_{k=0}^{t-p-1} \rho_k(\theta) \Big(\varepsilon_{t-k} - \sum_{i=1}^{p} \phi_i \varepsilon_{t-k-i} \Big)$$

and estimate the ARMA coefficients ϕ^* and θ^* by minimizing the least squares criterion

$$l_T(\phi, \theta) = \sum_{t=1}^{T} \eta_t(\phi, \theta)^2.$$

This criterion function again has a very similar structure to that of the AR(p) setup. In particular, setting $\rho_0(\theta) = 1$ and $\rho_k(\theta) = 0$ for $k > 0$ yields the conditional likelihood of the AR(p) case. As the coefficients $\rho_k(\theta)$ (as well as their derivatives with respect to θ) decay exponentially fast to zero, a truncation argument as in the ARMA$(1, 1)$ case allows us to adapt the proving strategy of Theorems 2.4.3 and 2.4.4 to the setup at hand.

Let us now turn to another important issue which concerns the limiting distribution of the AR parameter estimators. As discussed at the end of Section 2.4, the asymptotic variance of the estimators has a very complicated structure in general. This makes it extremely difficult to come up with a consistent estimator for the asymptotic variance. In many cases, it will thus not be possible to use the limiting distribution to compute confidence bands and critical values of test statistics. Bootstrap procedures may provide a way to circumvent this problem. In particular, it may be possible to extend standard bootstrap procedures for parametric AR processes as provided in Gonçalves and Kilian (2004) to the approximated AR variables $\widetilde{\varepsilon}_t$.

Our last remark is on the issue of data driven bandwidth selection in our framework. As shown e.g. in Altman (1990), Hart (1991), Herrmann et al. (1992) and Hart (1994), estimating the bandwidth in time direction in the fixed design setting $Y_{t,T} = m(\frac{t}{T}) + \varepsilon_t$ may become problematic when the errors are correlated. In particular, standard techniques like cross-validation perform very poorly in this case. In our setting, analogous difficulties are to be expected. A starting point to develop and analyze automatic bandwidth selection procedures in our framework may be the badnwidth selection techniques for smooth backfitting estimates in the stationary setup as discussed in Mammen and Park (2005).

Chapter 3

A Volatility Model

In this chapter, we study a semiparametric multiplicative volatility model, which splits up into a nonparametric part and a parametric GARCH component. The nonparametric part is modelled as a product of a deterministic time trend component and of further components that depend on stochastic regressors. We propose a two-step procedure to estimate the model. To estimate the nonparametric components, we extend the standard smooth backfitting procedure of Mammen et al. Mammen et al. (1999). The GARCH parameters are estimated in a second step via a quasi maximum likelihood approach. We show consistency and asymptotic normality of our estimators. Our results are obtained using mixing properties and local stationarity. Finally, we illustrate our method using financial data.

3.1 Introduction

Given the ever-changing economic and financial environment, it is quite plausible that many financial time series behave in a nonstationary way. Especially over longer horizons, structural changes may occur. Thus, the technical assumption of stationarity is likely to be violated in many cases. This issue has been pointed out by numerous authors in recent years. In particular, it has been claimed that many interesting stylized facts of financial return and volatility series can be neatly explained by employing nonstationary models (see e.g. Mikosch and Stărică (2000), (2003), (2004)).

One way to deal with nonstationarities in financial time series is the theory on locally stationary processes. The latter has been introduced in a series of papers by Dahlhaus (1996b), (1996a), (1997). Intuitively speaking, a process is locally stationary if over short periods of time (i.e. locally in time) it behaves approximately stationary, even though it is globally nonstationary. In recent years, many locally stationary models have been proposed in the financial time series context. Usually, these models are extensions of parametric time series models allowing for the parameters to change smoothly over time. An example is the class of ARCH processes with time-varying parameters introduced by Dahlhaus and Rao (2006).

A related locally stationary model which has been explored in a number of studies is given by the equation

$$Y_{t,T} = \tau\Big(\frac{t}{T}\Big)\varepsilon_t \quad \text{for } t = 1, \ldots, T, \tag{3.1}$$

where $Y_{t,T}$ are log-returns, τ is a smooth deterministic function of time and $\{\varepsilon_t\}$ is a standard stationary GARCH process with $\mathbb{E}[\varepsilon_t^2] = 1$. As usual in the literature on locally stationary models, the time-varying parameter τ does not depend on real time t, but on rescaled time $\frac{t}{T}$. We comment on this feature in more detail in Section 3.2. Model (3.1) has been considered for example in Feng (2004), where the τ-function is estimated nonparametrically. Engle and Rangel (2008) work with a closely related model, where the τ-component is modelled parametrically as a flexible exponential spline function. A multivariate generalization of model (3.1) is studied in Hafner and Linton (2010).

Model (3.1) can be considered as a GARCH process with time-varying parameters, with certain restrictions imposed on the parameter functions. In particular, the unconditional volatility level $\mathbb{E}[Y_{t,T}^2]$ is given by the time-dependent function $\tau^2(\frac{t}{T})$, which is allowed to vary smoothly over time. In reality, the volatility level is unlikely to change deterministically over time. Instead it reflects and varies with changes in the economic and financial environment. Therefore, the τ-function should depend on certain economic and financial variables. In model (3.1), these dependencies are not modelled explicitly. Instead, rescaled time serves as a catch-all for omitted explanatory variables.

These considerations show that in a more realistic version of model (3.1), the τ-function should depend on economic and financial influences. However, there is clearly no way to come up with a model that incorporates all relevant variables. One way to deal with this is to use rescaled time as a proxy for the omitted variables. To formalize these ideas, we propose the model

$$Y_{t,T} = \tau\Big(\frac{t}{T}, X_t\Big)\varepsilon_t, \tag{3.2}$$

where $Y_{t,T}$ are log-returns, X_t is an $\mathbb{R}^d$-valued random vector of economic or financial covariates and τ is a smooth function of time and the variables X_t. As before, $\{\varepsilon_t\}$ is a standard GARCH process. To countervail the curse of dimensionality, we split up the τ-function into multiplicative components thus yielding the model

$$Y_{t,T} = \tau_0\Big(\frac{t}{T}\Big)\prod_{j=1}^{d}\tau_j(X_t^j)\varepsilon_t, \tag{3.3}$$

where τ_0 and τ_j for $j = 1, \dots, d$ are smooth functions of time and of the regressors X_t^j, respectively. As will be seen in Section 3.2, the multiplicative specification of the τ-function in (3.3) not only avoids the curse of dimensionality but also allows for a direct interpretation of the various components.

In the following sections, we give an in-depth theoretical treatment of model (3.3). The complete formulation of the model together with its assumptions is given in Section 3.2. In Section 3.3, we propose a two-step procedure to estimate both the nonparametric and the parametric components of the model. To estimate the nonparametric functions τ_j for $j = 0, \dots, d$, we extend the smooth backfitting procedure of Mammen et al. (1999) to our locally stationary stetting. Having estimates $\widetilde{\tau}_j$ of the functions τ_j, we can construct approximate expressions $\widetilde{\varepsilon}_t$ of the GARCH variables ε_t. This allows us to estimate the GARCH parameters of the model via approximate quasi-maximum likelihood methods in a second step. Consistency and asymptotic normality of our estimators are shown in Section 3.4.

The contribution in this chapter is twofold. From a technical point of view, we extend the asymptotic results for model (3.1) to a more general framework in which the τ-function depends both on rescaled time and stochastic regressors. This vastly complicates both steps of the asymptotic analysis and as a result, we cannot extend existing proving techniques as provided in Hafner and Linton (2010) in a straightforward manner. In particular, novel and intricate arguments are required to derive the asymptotic behaviour of the GARCH estimates obtained in the second estimation step. In terms of volatility modelling, we introduce a flexible framework which allows to capture both nonstationarities and influences from the economic and financial environment. As the component functions τ_j in our model are completely nonparametric, we are able to explore the form of the relationship between volatility and its potential sources. Therefore, our model allows us to extend existing parametric studies on the sources of volatility as conducted e.g. in Engle and Rangel (2008) and Engle et al. (2008).

To illustrate the usefulness of our model and to complement the technical analysis, we present an empirical example in Section 3.5. There, the model is applied to S&P 500 return data using various interest rate spreads as explanatory variables.

3.2 The Model

Suppose we observe a sample of daily log-returns $Y_{t,T}$ of a financial time series and a sequence of daily $\mathbb{R}^d$-valued random stationary covariate vectors $X_t = (X_t^1, \dots, X_t^d)$ for $t = 1, \dots, T$. We assume the log-return series follows the process

$$Y_{t,T} = \tau_0\Big(\frac{t}{T}\Big) \prod_{j=1}^{d} \tau_j(X_t^j)\varepsilon_t \quad \text{for } t = 1, \dots, T \tag{3.4}$$

with

$$\varepsilon_t = \sigma_t \eta_t$$

$$\sigma_t^2 = w_0 + a_0 \varepsilon_{t-1}^2 + b_0 \sigma_{t-1}^2.$$

Here, τ_0 and τ_j ($j = 1, \dots, d$) are smooth nonparametric functions of time and the stochastic regressors, respectively. Furthermore, $\{\varepsilon_t\}$ is a strictly stationary GARCH process, which is assumed to be independent of the covariate process

$\{X_t\}$. The residuals of the GARCH process are assumed to be i.i.d. with zero mean and unit variance. For simplicity, we restrict attention to the GARCH(1,1) specification.

In order to conduct meaningful asymptotics, we let the function τ_0 depend on rescaled time $\frac{t}{T}$ rather than on real time t. Thus, τ_0 is defined on $(0, 1]$ rather than on $\{1, \ldots, T\}$. In the remainder of this chapter, we denote rescaled time by $x_0 \in (0, 1]$. It relates to observed time $t \in \{0, \ldots, T\}$ through the mapping $t = [x_0 T]$, where $[x]$ denotes the smallest integer weakly larger than x. If we defined the function τ_0 in terms of observed time, we would not get additional information on the structure of τ_0 around a particular time point t as the sample size T increases. Within the framework of rescaled time, in contrast, the function τ_0 is observed on a finer and finer grid on the unit interval as T grows. Thus, we obtain more and more information on the local structure of τ_0 around each point x_0 in rescaled time. This is the reason why we can make meaningful asymptotic considerations within this framework. A detailed discussion of the concept of rescaled time can be found in Dahlhaus (1996a).

For a sufficiently smooth trend function τ_0, we have

$$\big|Y_{t,T} - Y_t(x_0)\big| \leq C\Big|\frac{t}{T} - x_0\Big|U_t, \tag{3.5}$$

where C is a constant independent of x_0, t and T, $Y_t(x_0) = \tau_0(x_0)\prod_{j=1}^{d}\tau_j(X_t^j)\varepsilon_t$, and $U_t = \prod_{j=1}^{d}\tau_j(X_t^j)\varepsilon_t$. Note that due to the stationarity of X_t and ε_t both $\{Y_t(x_0)\}$ and $\{U_t\}$ are strictly stationary processes. As $U_t = O_p(1)$, we obtain from (3.5) that

$$\big|Y_{t,T} - Y_t(x_0)\big| = O_p\Big(\Big|\frac{t}{T} - x_0\Big|\Big). \tag{3.6}$$

Therefore, if $\frac{t}{T}$ is close to x_0, then $Y_{t,T}$ is close to $Y_t(x_0)$ at least in a stochastic sense. Put differently, locally in time, the process $\{Y_{t,T}\}$ is close to the stationary process $\{Y_t(x_0)\}$. In this sense, the process $\{Y_{t,T}\}$ is locally stationary.

We close this section with a remark on the interpretation of the nonparametric components of model (3.4). First, note that the functions $\tau_0, \ldots, \tau_d$ and the GARCH residual ε_t are only identified up to a multiplicative constant in model (3.4). Thus we are free to rescale them in a suitable way. Given the independence between X_t and ε_t assumed later in (V3), normalizing the components such that $\mathbb{E}[\varepsilon_t^2] = 1$ yields

$$\mathbb{E}[Y_{t,T}^2 | X_t] = \tau_0^2\Big(\frac{t}{T}\Big) \prod_{j=1}^{d} \tau_j^2(X_t^j). \tag{3.7}$$

Thus, the product of the τ-components gives the volatility at time t conditional on the covariates X_t. If we additionally scale the model components to satisfy $\mathbb{E}[\prod_{j=1}^d \tau_j^2(X_t^j)] = 1$, we obtain that

$$\mathbb{E}[Y_{t,T}^2] = \tau_0^2\Big(\frac{t}{T}\Big), \tag{3.8}$$

i.e. the deterministic function of time $\tau_0^2(\frac{t}{T})$ gives the time-varying unconditional volatility level. In (3.7), $\tau_0^2(\frac{t}{T})$ thus specifies the unconditional volatility level and the product of the remaining components $\prod_{j=1}^d \tau_j^2(X_t^j)$ is the multiplicative factor by which the volatility conditional on X_t deviates from the unconditional level.

3.3 Estimation Procedure

We now turn to the two-step estimation procedure alluded to in the introduction to this chapter. In the first step, we provide estimates of the nonparametric functions $\tau_0, \ldots, \tau_d$. In the second step, we use these nonparametric estimates to obtain estimators of the GARCH parameters.

3.3.1 Estimation of the Nonparametric Model Components

In order to estimate the nonparametric functions $\tau_0, \ldots, \tau_d$, we first transform the multiplicative model (3.4) into an additive one and use the results from Chapter 2. Given the resulting estimators of the additive model we retrieve the estimimates of the components in the multiplicative model by applying the reverse transform. The transform we apply to (3.4) is to first square it and then take logarithms.

This yields

$$Z_{t,T} = m_0\Big(\frac{t}{T}\Big) + \sum_{j=1}^{d} m_j(X_t^j) + u_t, \tag{3.9}$$

where $Z_{t,T} := \log Y_{t,T}^2$, $m_j := \log \tau_j^2$ for $j = 0, \ldots, d$, and $u_t := \log \varepsilon_t^2$. This fits into the model structure considered in Chapter 2 without a periodic component, i.e. with $\theta \equiv 1$. Note that the functions $m_0, \ldots, m_d$ in (3.9) are only identified up to an additive constant. To identify them, we assume that

$$\int_0^1 m_0(x_0)dx_0 = 0 \quad \text{and} \quad \int_{\mathbb{R}} m_j(x_j)p_j(x_j)dx_j = 0 \quad \text{for } j = 1, \ldots, d,$$

where p_j is the marginal density of X_t^j. Furthermore, we normalize $\mathbb{E}[u_t] = 0$, which introduces a constant m_c to (3.9)[1], and we are left with

$$Z_{t,T} = m_c + m_0\Big(\frac{t}{T}\Big) + \sum_{j=1}^{d} m_j(X_t^j) + u_t, \tag{3.10}$$

In Section 3.4, we will give a set of sufficient conditions to ensure that the assumptions in Chapter 2 are fulfilled thus enabling us to obtain estimators of the nonparametric component functions $\widetilde{m}_0, \ldots, \widetilde{m}_d$. To get the estimators of the multiplicative components we apply the reverse transform to get

$$\widetilde{\tau}_j = \sqrt{\exp(\widetilde{m}_j)}$$

for $j = 0, \ldots, d$.

3.3.2 Estimation of the Parametric Model Components

To motivate the second step in our estimation procedure, we first consider an infeasible estimator of the model parameters. Suppose that the nonparametric components $\tau_0^2, ..., \tau_d^2$ were known. In this situation, the GARCH variables ε_t^2 would be observable, since

$$\varepsilon_t^2 = \frac{Y_{t,T}^2}{\tau_0^2(\frac{t}{T})\prod_{k=1}^{d}\tau_k^2(X_k^t)}. \tag{3.11}$$

[1]This constant was subsumed into the periodic component m_θ in Chapter 2.

The GARCH parameters $\phi_0 := (w_0, a_0, b_0)$ could thus be estimated by standard quasi maximum likelihood methods, where the quasi log-likelihood is given by

$$l_T(\phi) = -\sum_{t=1}^{T} \big(\log v_t^2(\phi) + \frac{\varepsilon_t^2}{v_t^2(\phi)} \big). \tag{3.12}$$

Here, $\phi = (w, a, b)$ and

$$v_t^2(\phi) = \begin{cases} \frac{w}{1-b} & \text{for } t = 1 \\ w + a\varepsilon_{t-1}^2 + b v_{t-1}^2(\phi) & \text{for } t = 2, \ldots, T \end{cases} \tag{3.13}$$

is the conditional volatility of the GARCH process with starting value $v_0^2(\phi) = w/(1-b)$.

As the functions $\tau_0^2, \ldots, \tau_d^2$ are not observed, we cannot apply this standard approach. However, given the estimates $\widetilde{\tau}_0^2, \ldots, \widetilde{\tau}_d^2$ from the first estimation step, we can replace ε_t^2 by the standardized residuals

$$\widetilde{\varepsilon}_t^2 = \frac{Y_{t,T}^2}{\widetilde{\tau}_0^2(\frac{t}{T}) \prod_{k=1}^d \widetilde{\tau}_k^2(X_k^t)} \tag{3.14}$$

and use these as approximations to ε_t^2 in the quasi maximum likelihood estimation. The quasi log-likelihood then becomes

$$\widetilde{l}_T(\phi) = -\sum_{t=1}^{T} \big(\log \widetilde{v}_t^2(\phi) + \frac{\widetilde{\varepsilon}_t^2}{\widetilde{v}_t^2(\phi)} \big), \tag{3.15}$$

where analogously to (3.13),

$$\widetilde{v}_t^2(\phi) = \begin{cases} \frac{w}{1-b} & \text{for } t = 1 \\ w + a\widetilde{\varepsilon}_{t-1}^2 + b \widetilde{v}_{t-1}^2(\phi) & \text{for } t = 2, \ldots, T \end{cases} \tag{3.16}$$

is the approximate conditional volatility. Our estimator $\widetilde{\phi}$ of the true parameter values ϕ_0 is now defined as

$$\widetilde{\phi} = \arg\max_{\phi \in \Phi} \widetilde{l}_T(\phi), \tag{3.17}$$

where the parameter space Φ is assumed to be compact. In comparison to this, the standard maximum likelihood estimator for the case in which the τ-components are known is defined as

$$\widehat{\phi} = \arg\max_{\phi\in\Phi} l_T(\phi). \tag{3.18}$$

3.4 Asymptotics

In Section 3.4.1 we treat the nonparametric estimates $\widetilde{\tau}_0, \ldots, \widetilde{\tau}_d$. Section 3.4.2 gives results on the asymptotic behaviour of the GARCH estimates $\widetilde{\phi}$. In order to establish the asymptotic properties of our nonparametric estimators we make the following assumptions on the model components.

(V1) *The process* $\{X_t, \varepsilon_t, \sigma_t\}$ *is strictly stationary and strongly mixing with mixing coefficients* α *satisfying* $\alpha(k) \le a^k$ *for some* $0 < a < 1$.

(V2) *The functions* τ_0 *and* τ_j $(j = 1, \ldots, d)$ *are twice (continuously) differentiable, strictly positive, and bounded away from zero with Lipschitz continuous second derivatives.*

(V3) *The variables* X_t *and* ε_t *are independent and the error process is normalized s.t.* $\mathbb{E}[\log \varepsilon_t^2] = 0$.

(V4) *The conditional volatility* σ_t^2 *is bounded away from zero and the GARCH residuals* η_t *have a density with respect to Lebesgue measure which is bounded in a neighbourhood of zero.*

(V5) *The variables* X_t *have compact support, say* $[0,1]^d$.

(V6) *The kernel* K *is bounded, has compact support (*$[-C_1, C_1]$*, say) and is symmetric about zero. Moreover, it fulfills the Lipschitz condition that there exists a positive constant* L *such that* $|K(u) - K(v)| \le L|u - v|$.

(V7) *The density* p *of* X_t *and the densities* $p_{(0,l)}$ *of* (X_t, X_{t+l}), $l = 1, 2, \ldots,$ *are uniformly bounded. Furthermore,* p *is bounded away from zero on* $[0,1]^d$. *The first partial derivatives of* p *exist and are continuous.*

(V8) *Let* $Z_t = Z_{t,T} - m_0(\frac{t}{T})$. *For some* $\theta > \frac{8}{3}$, $\mathbb{E}[|Z_t|^\theta] < \infty$.

(V9) *The conditional densities* $f_{X_t|Z_t}$ *of* X_t *given* Z_t *and* $f_{X_t,X_{t+l}|Z_t,Z_{t+l}}$ *of* (X_t, X_{t+l}) *given* (Z_t, Z_{t+l}), $l = 1, 2, \ldots,$ *exist and are bounded from above.*

(V10) *The bandwidth* h *satisfies either of the following:*

(a) $T^{\frac{1}{5}} h \to c_h$ *for some constant* c_h.

(b) $T^{\frac{1}{4}+\delta} h \to c_h$ *for some constant* c_h *and some small* $\delta > 0$.

As already mentioned in Section 3.2 assumption (V1) restricts the nonstationarity in the model to result from the time-varying component τ_0. The interpretation of $\tau_0^2(\cdot)$ as the unconditional volatility level is given by the independence of X_t and ε_t in (V3).[2] Note that in assumption (V3) we are stipulating the contemporaneous independence of the two processes not their full independence. Assumption (V4) validates the transform used in the first estimation step leading to the additive model (3.9). Assumption (V2) ensures the additive components in the transformed model (3.10) satisfy the appropriate degree of smoothness for the results from Chapter 2 to hold. The regression error u_t in the transformed model (3.10) is conditionally mean zero due to the independence of X_t and ε_t and the normalization of the error process in (V3). (V5) is only needed for the second estimation step. For the first step, we could allow the support of X_t to be unbounded and estimate the functions $\tau_0, \ldots, \tau_d$ uniformly over compact subsets of the support. However, for ease of notation, we assume (V5) throughout. The remaining conditions ensure that the respective assumptions in Chapter 2 are fulfilled.

Essentially the same remarks to those following the assumptions for the additive model in Chapter 2 can be made here. Again the assumptions needed to establish the asymptotic behaviour of the nonparametric functions are very similar to the conditions that can be found in Mammen et al. (1999) for the strictly stationary

[2] The independence condition could be replaced by the requirement that $\mathbb{E}[\varepsilon_t^2|X_t] = \mathbb{E}[\varepsilon_t^2]$ *a.s.* and $\mathbb{E}[\log \varepsilon_t^2|X_t] = 0$. However, these are so restrictive that not much is gained in moving away from independence apart from vastly complified arguments in the proofs and additional unverifiable low-level conditions.

case. It should also be mentioned again that we could do away with the assumption of exponentially decaying mixing rates in (V1) in favour of sufficiently high polynomial rates. The stronger assumption (V1) is again retained to keep the notation and structure of the proofs as clear as possible.

In order to derive the consistency and asymptotic normality of the GARCH parameter estimators in the second estimation step we will additionally need the following assumptions.

(V11) *The parameter space* Φ *is a compact subset of* $\{\phi \in \mathbb{R}^3 \mid \phi = (w, a, b)$ *with* $0 < \underline{\kappa} \le w, a \le \overline{\kappa} < \infty$ *and* $0 \le b < 1\}$ *with constants* $\underline{\kappa}$ *and* $\overline{\kappa}$. *The true parameter* $\phi_0 = (w_0, a_0, b_0)$ *is an interior point of* Φ *and* $a_0 + b_0 < 1$.

(V12) $\mathbb{E}[\varepsilon_t^{8+\delta}] < \infty$, *for some* $\delta > 0$.

(V11) is a standard assumption in the theory on GARCH models. Note it also implies that σ_t^2 is bounded away from zero, which was assumed in (V4). The moment condition in (V12) is needed to show asymptotic normality of the GARCH estimates.

3.4.1 Asymptotics for the Nonparametric Model Components

As we are mainly interested in the squared version of the $\widetilde{\tau}_0, \ldots, \widetilde{\tau}_d$ in our multiplicative model, we will restrict ourselves to reporting these. Their derivation is based on obtaining the asymptotic properties of the estimators $\widetilde{m}_0, \ldots, \widetilde{m}_d$ for the additive components in the transformed model (3.10) and then using the fact that due to the transform $\widetilde{\tau}_j^2 = \exp(\widetilde{m}_j)$ for $j = 0, \ldots, d$ these results easily carry over to the multiplicative components in (3.4).

Theorem 3.4.1. *Suppose that conditions* (V1) – (V9) *hold.*

(a) Assume that the bandwidth h *satisfies either* (V10)(a) *or* (V10)(b). *Then,*

for $I_h = [2C_1h, 1-2C_1h]$ *and* $I_h^c = [0, 2C_1h) \cup (1-2C_1h, 1]$,

$$\sup_{x_j \in I_h} \left| \widetilde{\tau}_j^2(x_j) - \tau_j^2(x_j) \right| = O_p\Big(\sqrt{\frac{\log T}{Th}}\Big) \tag{3.19}$$

$$\sup_{x_j \in I_h^c} \left| \widetilde{\tau}_j^2(x_j) - \tau_j^2(x_j) \right| = O_p(h) \tag{3.20}$$

for all $j = 0, \ldots, d$.

(b) *Assume that the bandwidth h satisfies* (V10)(a). *Then, for any* $x_0, \ldots, x_d \in (0,1)$,

$$T^{\frac{2}{5}} \begin{bmatrix} \widetilde{\tau}_0^2(x_0) - \tau_0^2(x_0) \\ \vdots \\ \widetilde{\tau}_d^2(x_d) - \tau_d^2(x_d) \end{bmatrix} \xrightarrow{d} N(B_\tau(x), V_\tau(x)),$$

with the bias term $B_{\tau^2}(x) = [\tau_0^2(x_0)c_h^2\beta_0(x_0), \ldots, \tau_d^2(x_d)c_h^2\beta_d(x_d)]'$ *and the covariance matrix* $V_{\tau^2}(x) = \mathrm{diag}(\tau_0^4(x_0)v_0(x_0), \ldots, \tau_d^4(x_d)v_d(x_d))$. *Here,* $v_0(x_0) = c_h^{-1}c_K \sum_{l=-\infty}^{\infty} \gamma_u(l)$ *and* $v_j(x_j) = c_h^{-1}c_K\sigma^2/p_j(x_j)$ *for* $j = 1, \ldots, d$ *with* $c_K = \int K^2(u)du$, $\gamma_u(l) = \mathrm{Cov}(u_t, u_{t+l})$ *and* $\sigma^2 = \mathrm{Var}(u_t)$ *for* $u_t = \log \varepsilon_t^2$. *Furthermore, the functions* $\beta_j(x_j)$ *are the components of the* $L_2(p)$*-projection of the function* β *defined in Lemma A.3.3 of Appendix A.3 onto the space of additive functions.*

As already remarked, the above follows from Theorem 2.4.2 established in Chapter 2 and the smoothness of the transform $\widetilde{\tau}_j^2 = \exp(\widetilde{m}_j)$ for $j = 0, \ldots, d$. Restatements of the expansions needed to show the equivalent of Theorem 2.4.2 for the transformed model (3.10) are provided in Appendix B.1. Recalll that the proofs exploit the fact that rescaled time behaves similarly to a uniformly distributed random variable that is independent of the other covariates.

3.4.2 Asymptotics for the Parametric Model Components

Given the estimators for $\tau_0^2, \ldots, \tau_d^2$ from the first step, the GARCH parameters ϕ_0 are estimated by $\widetilde{\phi}$ as outlined in Section 3.3.2. In this subsection, we look at consistency and asymptotic normality of $\widetilde{\phi}$. The following theorem establishes consistency.

Theorem 3.4.2. *Suppose that the bandwidth h satisfies* (V10)(a) *or* (V10)(b). *In addition, let assumptions* (V1) – (V9) *and* (V11) *be fulfilled. Then $\widetilde{\phi}$ is a consistent estimator of ϕ_0, i.e.*

$$\widetilde{\phi} \xrightarrow{P} \phi_0.$$

We next give a result on the limiting distribution of the GARCH estimates which shows that these are asymptotically normal.

Theorem 3.4.3. *Suppose that the bandwidth h satifies* (V10)(b) *and let assumptions* (V1) – (V9) *together with* (V11) – (V12) *be fulfilled. Then it holds that*

$$\sqrt{T}(\widetilde{\phi} - \phi_0) \xrightarrow{d} N(0, \Sigma).$$

Details on the covariance matrix Σ can be found in Appendix B.2 (see equation (B.14)).

The proof of asymptotic normality is the theoretically most challenging part in this chapter. The details are postponed to the Appendix B.2. For now we will be content with providing an outline. By the usual Taylor expansion argument, we arrive at

$$\sqrt{T}(\widetilde{\phi} - \phi_0) = -\Big(\frac{1}{T}\frac{\partial^2 \widetilde{l}_T(\bar{\phi})}{\partial\phi\partial\phi^T}\Big)^{-1}\frac{1}{\sqrt{T}}\frac{\partial \widetilde{l}_T(\phi_0)}{\partial\phi},$$

where $\bar{\phi}$ is an intermediate point between $\widetilde{\phi}$ and ϕ_0. As in the standard case, we can show that the second derivative on the right-hand side converges in probability to a deterministic matrix. The asymptotic distribution is thus determined by the term $\frac{1}{\sqrt{T}}\frac{\partial \widetilde{l}_T(\phi_0)}{\partial\phi}$, which we rewrite as

$$\frac{1}{\sqrt{T}}\frac{\partial \widetilde{l}_T(\phi_0)}{\partial\phi_i} = \underbrace{\frac{1}{\sqrt{T}}\frac{\partial l_T(\phi_0)}{\partial\phi_i}}_{=:A_1} + \underbrace{\Big(\frac{1}{\sqrt{T}}\frac{\partial \widetilde{l}_T(\phi_0)}{\partial\phi_i} - \frac{1}{\sqrt{T}}\frac{\partial l_T(\phi_0)}{\partial\phi_i}\Big)}_{=:A_2}.$$

We will show that this term is asymptotically normal. The main challenge to do so is to derive a stochastic expansion of the term A_2. This requires rather involved and nonstandard arguments which are presented in detail in Appendix B.2. In particular, we cannot just extend the arguments presented in Hafner and Linton (2010) to fit our setting. Once we have provided the expansion of A_2, we are in a position to apply a central limit theorem to the sum $A_1 + A_2$, which completes the proof. We will see that the term A_2 is itself asymptotically normal and thus contributes to the limit distribution of $\frac{1}{\sqrt{T}}\frac{\partial \widetilde{l}_T(\phi_0)}{\partial \phi}$. As a consequence, we obtain a larger asymptotic variance than in the standard case (where only the term A_1 occurs). This reflects the additional uncertainty that results from not knowing the functions $\tau_0, \ldots, \tau_d$.

3.5 Application

To illustrate our model, we apply it to a sample of daily financial data spanning the period from 10th April 1993 until 2nd February 2014. The estimated model is given by

$$Y_{t,T}^2 = \tau_0^2\left(\frac{t}{T}\right)\prod_{j=1}^{3}\tau_j^2(X_{t-1}^j)\varepsilon_t^2, \tag{3.21}$$

where $Y_{t,T}$ are S&P 500 log-returns and the covariates are three different lagged interest rate spreads all calculated from data provided in the H.15 release of the Federal Reserve.[3] One of the spreads we will use as a regressor is the difference between the yields on Moody's seasoned Baa and Aaa corporate bonds.[4] This can be thought of as a credit default spread as it in some way caputres the difference in the default risk of high graded and low graded corporate debt. Our second regressor is a measure of credit risk for highly rated corporate debt as provided by the difference between the yield on Moody's Aaa corporate bonds and the interest rate of 20 year constant maturity U.S. treasuries.[5] The final regressor we include

[3]The interest rate data are from the Federal Reserve Statistical Release H.15 available online at www.federalreserve.org/releases/h15/data.htm. The historical prices of the S&P 500 are from Yahoo! Finance available at finance.yahoo.com.

[4]The original source is Moody's Investor Services. More information can be found on the research pages of the St.Louis Fed at research.stlouisfed.org.

[5]The 20 year treausuries were used to get a close maturity match to the corporate bonds.

is the difference between the interest rate on 3 month eurodollar deposits and the interest rate for 3 month U.S. treasury bills.[6] This can be interpreted as a measure for the additional default risk faced by non-U.S. versus U.S. banks. It is also related to the TED spread, an indicator for the risk of bank default, which is defined as the difference between the 3 month LIBOR and the interest rate on 3 month U.S. treasuries.

The estimation results for the nonparametric model components are presented in Figures 3.1 and 3.3. The bandwidths for the function fits are chosen by a rule of thumb following the application in Yu et al. (2011).

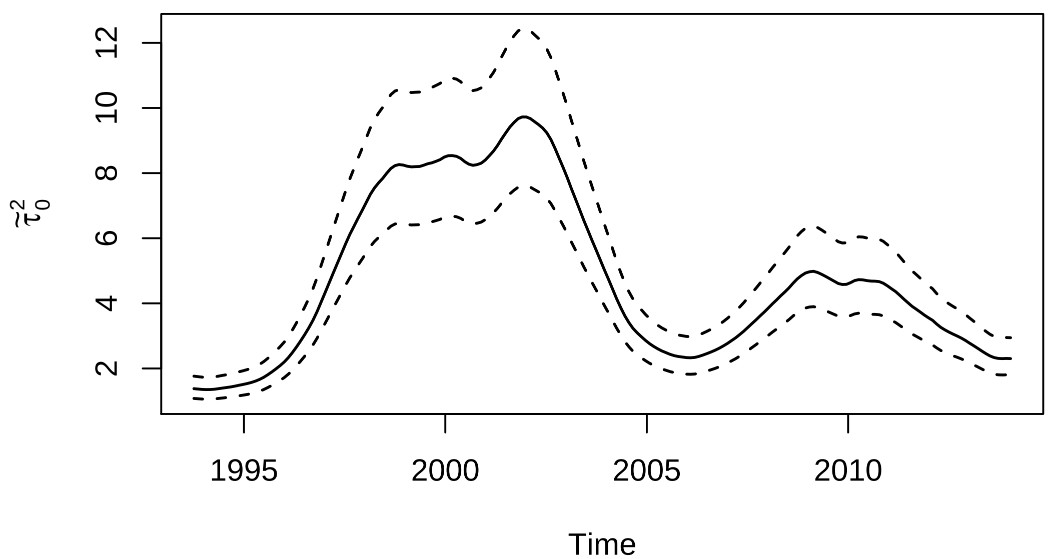


Figure 3.1: Plot of S&P 500 log returns from 10th April 1993 until 2nd February 2014

[6]The original source is Bloomberg and CRTB ICAP Fixed Income & Money Market Products.

The solid line in Figure 3.1 gives the estimate of $\widetilde{\tau}_0^2$. The dashed lines are the pointwise 95% confidence intervals. Due to the normalization of the other component estimates discussed later on, $\widetilde{\tau}_0^2$ only estimates the time varying unconditional volatility level in (3.8) up to a multiplicative constant. Comparing the estimate in Figure 3.1 with the log return series of the S&P 500 in Figure 3.2 we see that the estimate captures the periods of increased log return variance from 1997 until 2003 and from 2007 until 2012.

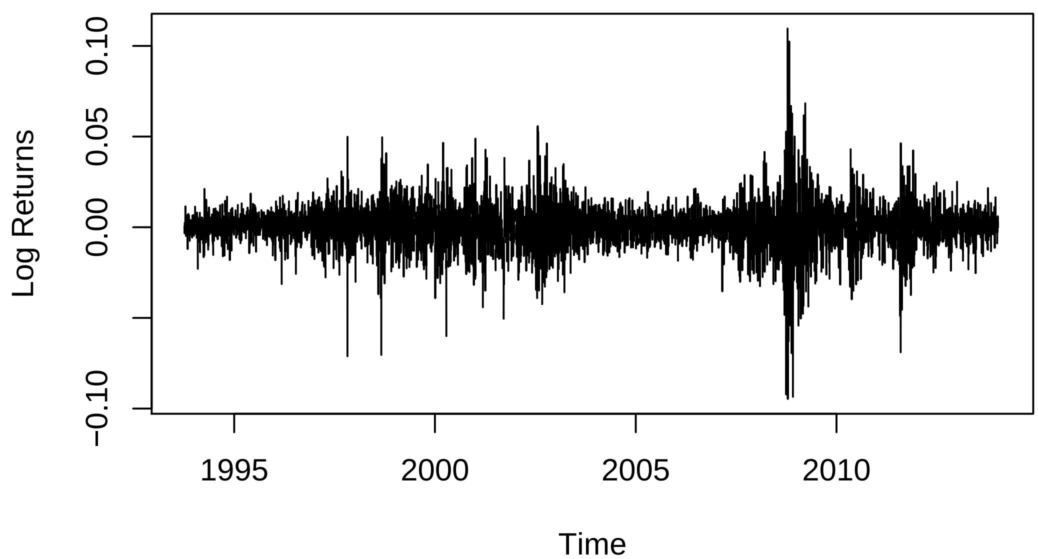


Figure 3.2: Plot of S&P 500 log returns from 10$^{\text{th}}$ April 1993 until 2$^{\text{nd}}$ February 2014

However, we can also see in Figure 3.2 that the second period was more severe than the first in terms of magnitude, which is not captured by our estimate. The main reason for this is that our regressors have more explanatory power in the recent crisis.

The estimated components $\widetilde{\tau}_j^2$ for $j = 1, 2, 3$ are given in Figure 3.3. The solid lines again represent the estimators $\widetilde{\tau}_j^2$ and the dashed lines are the pointwise 95% confidence intervals.

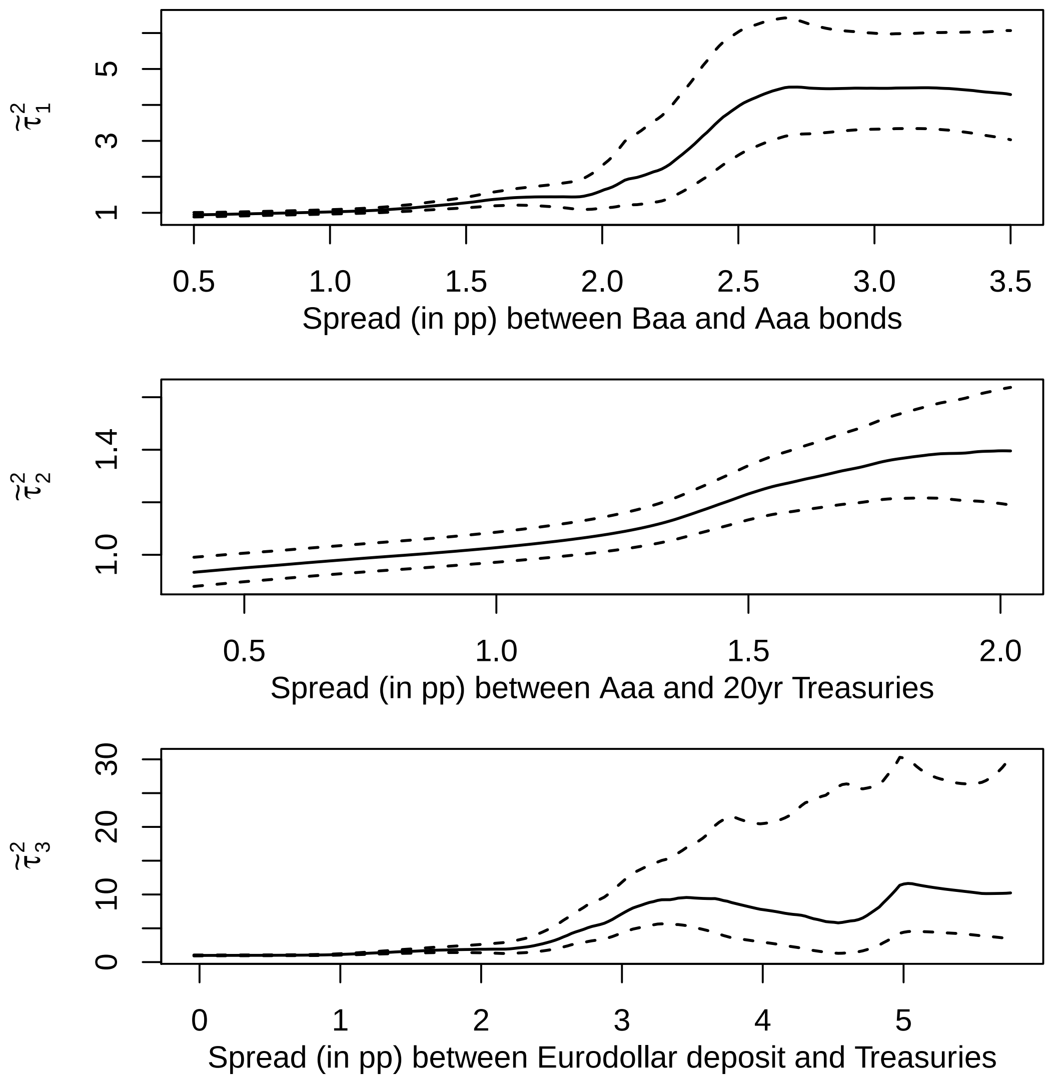


Figure 3.3: Estimates of τ_j^2 for $j = 1, 2, 3$. Spreads measured in percentage points.

The estimates $\widetilde{\tau}_j^2$ have been normalized such that $\widetilde{\tau}_j^2(x_j^m) = 1$, where x_j^m is the median observed realization of the j-th covariate X_t^j over the modelling period. This means that the effect of the j-th covariate on volatility is normalized to 1 if it takes a "normal" (i.e. its median) value. As

$$\mathbb{E}[Y_{t,T}^2|X_t] = \tau_0^2\Big(\frac{t}{T}\Big)\prod_{j=1}^{3}\tau_j^2(X_{t-1}^j), \tag{3.22}$$

the normalization allows for the estimates $\widetilde{\tau}_j^2$ for $j = 1, 2, 3$ to be interpreted as the multiplicative effect of the covariate X_{t-1}^j on S&P 500 volatility. To illustrate this, let us compare volatility between two different settings: Hold all the covariates except the j-th fixed at some value x_{-j} and change the j-th regressor X_{t-1}^j from its median x_j^m to some value x_j. From (3.22), one can then see that the conditional volatility is changed by the factor $\tau_j^2(x_j)/\tau_j^2(x_j^m) = \tau_j^2(x_j)$ as $\tau_j^2(x_j^m)$ has been normalized to one. Consequently, the fits $\widetilde{\tau}_j^2(x_j)$ estimate the factor by which the volatility level gets increased or dampened, when the j-th covariate changes from a normal value (i.e. its median) to some other more extreme value.

We now look at the estimated component functions in Figure 3.3. First of all, the top panel shows the estimated multiplicative effect on volatility of the lagged difference in the corporate bond yields between Moody's Baa and Aaa rated bonds. The estimate is increasing and highly nonlinear. In particular, for low spreads of up until 2 percentage points the multiplicative effect is close to 1 and lower than 2. For values of the spread between 2 and 2.5 percentage points the effect increases linearly up until about 4.5. For all larger yield differences the effeect remains at that level. Notice, that for high spreads the neutral multiplicative factor of 1 is well outisde the 95% confidence bands.

The middle panel gives the estimated factor on volatility of the lagged difference between interest on Moody's Aaa graded corporate bonds and 20 year constant maturity U.S. treausuries. Although the effect is increasing again, it is much more linear. Furthermore, we can see that the estimated effects are much lower, ranging from about 0.9 until just below 1.4.

Finally, the bottom panel gives the estimated multiplicative factor for the difference between the interest rates on three month Eurodollar deposits and three month treasuries. The shape is similar to the first effect in the top panel. However, the range is much larger with the largest estimated effect being above 10. Note though, that the confidence bands are also much wider showing the imprecision in the estimate due to having observed few spreads larger than 3 percentage points.

We will finish the discussion of the nonparametric estimates by comparing the estimates of time varying unconditional volatility in our model and the simpler model without covariates (see (3.1)). In Figure 3.4 the solid line is a rescaled version of $\widetilde{\tau}_0^2$ that estimates the unconditional volatility level in our model. The dashed line is the estimated unconditional volatility obtained from the simpler model.

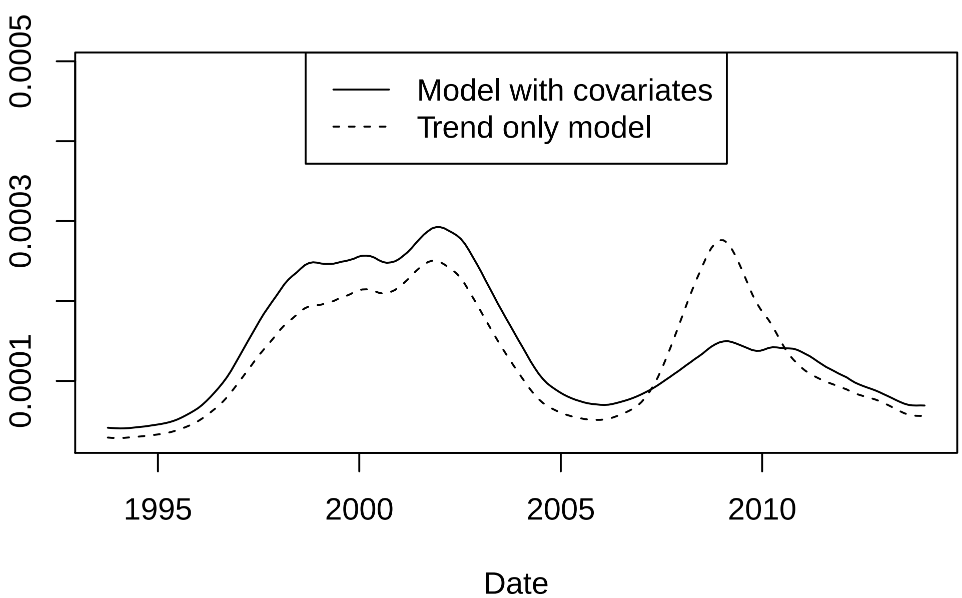


Figure 3.4: Time-varying unconditional volatilities for our model and the simpler model (3.1) without regressors.

Both curves in Figure 3.4 clearly show the volatility increase in the two recent crises periods. We can also see that the estimated unconditional volatility level in our model is much lower in the recent financial crisis than the estimate from the simpler model without covariates. As already mentioned, this suggests that our regressors explain a considerable part of volatility in the recent crisis. During the earlier crisis, however, the difference between the two curves is not so striking. Thus, the explanatory power of our covariates in this period seems to be much lower. This is quite plausible as our regressors are mainly from the U.S. financial sector and the turbulences between 1997 and 2003 were not primarily driven by events in this sector.

We finish our application with the estimation results for the parametric model components. In Table 3.1, we compare the GARCH estimates of our model with the ones obtained from the simpler model (3.1) and from a standard GARCH(1,1) model.

	$\widetilde{w}$	$\widetilde{a}$	$\widetilde{b}$	$\widetilde{a}+\widetilde{b}$	$\widetilde{HL}$
Standard GARCH(1,1)	0.000	0.085	0.908	0.992	90
Model with trend	0.035	0.078	0.885	0.963	19
Model with trend and covariates	0.047	0.073	0.878	0.951	15

Table 3.1: GARCH parameter estimates for GARCH(1,1) and for models (3.1) and (3.21),

The sum of the two estimated parameters $\widetilde{a} + \widetilde{b}$ reported in the penultimate column of Table 3.1 measures the persistence of shocks to volatility. One can see that this persistence measure decreases from 0.992 to 0.963 when accounting for time-varying unconditional volatility. This is in line with previous findings in the literature (compare e.g. Feng (2004)). Including our covariates in the model further decreases the estimated persistence to 0.951. Note that the reported decrease in persistence is quite dramatic even though it may seem rather small at first sight (compare the discussion in Lamoureux and Lastrapes (1990) and Mikosch and Stărică (2000) on this issue). To give some meaning to the numerical values of the persistence we will consider the half life of variance as in Lamoureux and Lastrapes (1990), which for a GARCH(1,1) model with parameters (ω, a, b) is defined by $HL = 1 - [ln(2)/ln(a + b)]$. The half life of volatility for the GARCH component gives the number of days it takes for a shock to the GARCH component to diminish to half its initial value. The last column of Table 3.1 provides the estimated half-lifes for the three competing models. Allowing for time varying unconditional volatility leads to a substantial decrease of the estimated half life from 90 trading days, which is more than four months to 19 trading days, which corresponds to roughly one month. Additionally including our regressors leads to a further decrease of the estimated half life to 15 trading days, which corresponds to 3 weeks.

To sum up, our results suggest that we can explain a good deal of S&P 500 volatility by our model. We have also seen that the regressors we included were more important in the recent financial crisis. Furthermore, it was seen that the persistence remaining in the GARCH component falls even further upon including our regressors. It should also be noted that we included additional regressors in the model (3.21).[7] Not only were some of the variables considered not available for the entire modelling period, but they also seemd to have little extra explanatory power: Using our procedure with all the considered regressors led to some of the estimated 95% confidence bands containing the horizontal line at one. This was taken as indication for no effect of the respective regressor. By successively removing such regressors we arrived at our model (3.21). Of course, for an in-depth analysis one would also need to validate the model. Such a model validation procedure would also help in choosing the covariates. One possible model selection method is described in Nielsen and Sperlich (2003). Finally, it would also be interesting to look at the forecasting performance of the model.

---

[7] These included lags of: an estimate for the slope of the yield curve given by the difference in interest rates on ten year and three month U.S. treasuries; the difference in corporate bond rates for financial and non-financial companies; the growth rate in the number of trades and the difference between the interest rate of ten year U.S. treasuries and their inflation protected variants.

3.6 Extensions

We use this section to discuss possible extensions and amendments to the model.

3.6.1 Estimation of the Covariance Matrix Σ

It is not at all trivial to construct a consistent estimate of the covariance matrix Σ introduced in Theorem 3.4.3. This is due to very complicated structure of Σ. In particular, the exact expression for Σ involves functions obtained from a higher order expansion of the stochastic part of the backfitting estimates (see Theorem B.1.1 in Appendix B.1). It is very complicated to calculate the exact form of these functions and even more challenging to give consistent estimates for them. The construction of a consistent estimate of Σ is thus a difficult theoretical problem.

3.6.2 Efficiency Gains

We next discuss how to gain efficiency in the estimation of both the nonparametric and parametric components of the model. For this purpose, we adapt the procedure in Hafner and Linton (2010).

First consider the nonparametric model components. If we knew the variables σ_t, we could divide the multiplicative model (3.4) by them to obtain

$$\frac{Y_{t,T}}{\sigma_t} = \tau_0\Big(\frac{t}{T}\Big) \prod_{j=1}^{d} \tau_j(X_t^j)\eta_t. \tag{3.23}$$

Squaring and taking the logarithm would then yield an additive regression model with error terms $v_t := \log \eta_t^2 - \mathbb{E}[\log \eta_t^2]$. These terms have a smaller variance than the errors $u_t = \log \varepsilon_t^2$ in the additive regression (3.10). In particular, $\text{Var}(v_t) = \text{Var}(\log \eta_t^2) \leq \text{Var}(\log \sigma_t^2) + \text{Var}(\log \eta_t^2) = \text{Var}(u_t)$. This suggests that at least for $j = 1, \ldots, d$, the infeasible smooth backfitting estimates based on equation (3.23) are more efficient in terms of asymptotic variance than our estimates.[8] Not

[8] Whether the infeasible estimate for $j = 0$ is more efficient depends on the autocorrelations of the errors u_t. Specifically, there are efficiency gains if and only if $\sum_{k=-\infty}^{\infty} \text{Cov}(u_0, u_k) > \text{Var}(v_t)$.

knowing the variables σ_t, we could use our estimation procedure to get initial estimates of them. Plugging these estimates into (3.23), it should be possible to obtain feasible smooth backfitting estimates with smaller asymptotic variance.

We now come to the parametric model components. Again, it should be possible to adapt the procedure described in Hafner and Linton (2010) to our setting in order to gain efficiency in the estimation of the parametric model parts. In the case of normally distributed GARCH residuals η_t, we may even be able to obtain estimates that reach the semiparametric efficiency bound. We omit the details and refer the interested reader to the description of the procedure in Hafner and Linton (2010).

3.6.3 Locally Stationary Covariates

It should also possible to allow for locally stationary regressors in model (3.4). In this case,

$$Y_{t,T} = \tau_0\Big(\frac{t}{T}\Big) \prod_{j=1}^{d} \tau_j(X^j_{t,T})\varepsilon_t \quad \text{for } t = 1, \ldots, T,$$

where ε_t is a strictly stationary GARCH residual as before, but where the covariates $X^j_{t,T}$ now form a locally stationary process for each $j = 1, \ldots, d$.

In this extended model, we face the following problem: If the regressors are locally stationary, their stochastic behaviour may change over time. As a consequence, rescaled time will not behave like an additional regressor any more that is *independent* of the other covariates, thus drastically complicating the asymptotic analysis.

If the stochastic behaviour of the regressors changes smoothly over time, we should nevertheless be able to get the smooth backfitting procedure to work. In particular, we conjecture that in this case we still obtain one-dimensional uniform nonparametric convergence rates. Moreover, if the covariates are assumed to be mixing, it should also be possible to prove asymptotic normality of the GARCH estimates.

3.7 Conclusion

We have proposed a new semiparametric volatility model, which generalizes the class of models $Y_{t,T} = \tau(\frac{t}{T})\varepsilon_t$, as for example considered in Feng (2004) and Engle and Rangel (2008). These models are able to account for nonstationarities in the volatility process. In addition, we are able to include covariates in a nonparametric way, hence allowing us to flexibly capture the effects of the financial and economic environment.

We have derived the asymptotic theory both for the nonparametric and the parametric part of the model. To estimate the nonparametric model components, we have extended the smooth backfitting approach of Mammen et al. (1999) to our nonstationary setting. Given the backfitting estimators, we were able to construct GARCH parameter estimates and to show that they are asymptotically normal. In particular, they converge at the fast parametric rate even though the nonparametric smoothers from the first step have slower nonparametric convergence rates. We concluded by illustrating the strengths of our model by applying it to financial data. In particular, our semiparametric approach allows us to estimate the form of the relationship between volatility and its potential sources. Therefore, we manage to go beyond existing parametric approaches such as in Engle and Rangel (2008) and Engle et al. (2008).

Chapter 4

Non-additive Season-trend Model

In this chapter we shall return to the "classical" decomposition given in (1.4)

$$Y_t = T_t + S_t + Z_t \quad \forall t \in \mathbb{Z}. \tag{4.1}$$

However, we will only consider the deterministic case, i.e. restricting the seasonal component S_t and the trend component T_t to be functions of time. Furthermore, it shall be assumed that the seasonal component has a known period θ, which will essentially be given by the frequency at which we observe the underlying process. Thus, for quarterly data, we would set $\theta = 4$, whereas for monthly observations we would choose $\theta = 12$. We will provide a model that refrains from additively decomposing the trend and seasonal components as in (4.1). The resulting season-trend function can be interpreted as a regression function with a categorical and a continuous covariate by rearranging the data. Based on this interpretation an estimator for the season-trend function will be suggested. Finally, an application of the model to a a German monthly temperature series illustrates the use of the model and compares its fit to one obtained from an additive model.

4.1 Introduction

In the deterministic case a nonparametric model for the setting in (4.1) was given in Chapter 1 by

$$Y_{t,T} = m_0(\frac{t}{T}) + m_\theta(t) + Z_t \quad \forall t \in 1, \ldots, T. \tag{4.2}$$

with m_0 a smooth deterministic trend and m_θ a periodic function with known period θ. In Chapter 2 we considered an extenstion to (4.2) by including further additive components that were functions of stationary regressors. Extensions to (4.2) necessitating the estimation of the period have also been considered. These include the recent contributions of Vogt and Linton (2014), who allow for the error process to be nonstationary and Sun et al. (2012), who do not include a trend and consider an i.i.d. error process. These papers also provide further references to models dealing with period estimation. Most notably, these include the classical parametric models and models similar to (4.2) with the sampling of the observations done at random and hence non-equidistant time points.

Both the aforementioned papers by Vogt and Linton (2014) and Sun et al. (2012), as indeed the model in Chapter 2, deal with the seasonal component by essentially using the dummy variable approach to obatin what they term a seasonal sequence.[1] The seasonal component is modelled by

$$S_t = m_\theta(t) = \sum_{l=1}^{\theta} \theta_l I(t = n\theta + l \text{ for some } n \in \mathbb{N}) \tag{4.3}$$

with θ_l giving the seasonal effect in season l. To avoid the lengthy notation in the indicator function we will introduce the modulo operator and write the above as

$$S_t = m_\theta(t) = \sum_{k=0}^{\theta-1} \theta_k I(t \bmod \theta = k). \tag{4.4}$$

[1]The term seasonal sequence is used as there are infinitely many functions of time that have a periodic behaviour at equidistant time points. We will not make this explicit nomenclatural distinction and refer to a periodic function even if we are only actually dealing with a periodic sequence.

The change of index from l to k is intentional to highlight that the interpretation of the seasonal effects has changed due to the use of the modulo operator. As the modulo operator gives us the remainder, the seasonal effect $\theta_1, \ldots, \theta_{11}$ are the same in (4.3) and (4.4). However, the seasonal effect θ_{12} in (4.3) now corresponds to θ_0 in (4.4). It should also be remarked, that in both formulations (4.3) and (4.4), the first season seasonal effect, θ_1, is the effect of the season the first observation falls into. In some cases we may want to change the ordering of the seasons. This can be easily achieved by replacing the indicator function in (4.4) by $I((t + s_1 - 1) \bmod \theta = k)$ with s_1 the season of the first observation. We will ignore this reordering until we turn to the application in section 4.4.

Using the compact formulation involving the modulo, the time varying mean in the additive decomposition model of (4.2) is given by

$$\begin{aligned} \mu_{t,T} &= T_t + S_t \\ &= m_0(\frac{t}{T}) + \sum_{k=0}^{\theta-1} \theta_k I(t \bmod \theta = k) \quad \forall t \in 1, \ldots, T. \end{aligned}$$

The model for $\{Y_{t,T}\}$ is then given by

$$Y_{t,T} = m_0(\frac{t}{T}) + \sum_{k=0}^{\theta-1} \theta_k I(t \bmod \theta = k) + Z_t \quad \forall t \in 1, \ldots, T. \tag{4.5}$$

for $\{Z_t\}$ a zero mean stationary process.

We see from (4.5), that the observatoin time point is essentially used twice as a regressor, once as the argument for the trend and once as the argument for the seasonal component. It seems quite natural to do this as we are using two distinct pieces of information about each time point. Firstly, we are using the fact that time progresses linearly to justify its use in the trend function. And secondly we are using the seasonally recurring informational content in the seasonal component. These two separate pieces of information contained in the observation time point are illustrated in Figure 4.1. Each small tick denotes an observation time point. As is customary in nonparametric trend estimation, the observed time points have been rescaled to the unit interval. Thus the first observation occurs at $\frac{1}{T}$, the second at $\frac{2}{T}$ and so on until the last observation which occurs at $\frac{T}{T} = 1$. Hence, all the observations are $\frac{1}{T}$ apart from each other.

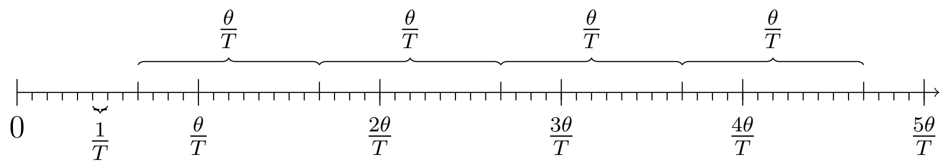


Figure 4.1: Illustration of observations in rescaled time and their seasonal links.

The first piece of information contained in the observation time point, namely the order within the sample, is illstrated by the arrangement of the observations within the unit interval. The braces above the time line indicate the second piece of information contained in the observation time point, namely its link to other time points in the same season.

The modelling framework in (4.5) uses these two pieces of information on the observation time point to disentangle the seasonal and trend components by requiring them to be additive, i.e. overlaying them. The seasonal component is assumed to be a periodic function, linking observations in the same season. The smooth trend component uses the first piece of information by linking each observation with observations close to it in time. The resulting overall time varying mean in (4.5) is the sum of the two. In the next section, we will introduce a modelling framework that also utilizes the two distinct pieces of information contained in the observation time point. However, we will refrain from imposing an additive structure as in (4.5).

4.2 Model

In this section we will use the two distinct pieces of information contained in the observation time point, namely its order over time and the season it is in, to construct a model that does not rely on an additive decomposition of seaonal and trend effects. This will result in a model with a season-trend function. In order to interpret this function we will rearrage the data, so that the season-trend function can be interpreted as a regression function. The idea is to consider the time varying mean of the process as a function that combines both the seasonal and the trend component in a function $m(\cdot,\cdot)$ defined on $[0,1] \times \{0,\dots,\theta-1\}$. This leads to the model for the real-valued process $\{Y_{t,T}\}$ given by[2]

$$Y_{t,T} = m(\frac{t}{T}, t \bmod \theta) + Z_t \quad \forall t = 1,\dots,T. \tag{4.6}$$

with $\{Z_t\}$ a zero mean stationary error process.

In order to interpret the season-trend function $m(\cdot,\cdot)$ we will rearrange the data as illustrated in Figure 4.2 for $\theta = 12$. The observation time points are denoted by points in the figure. The vertical axis keeps track of the season the observation was made in, whereas the horizontal axis gives the observed time point in rescaled time. All observation time points are still $\frac{1}{T}$ apart in the rescaled time direction. The points in a given season are also still $\frac{\theta}{T}$ apart in rescaled time direction, but one season apart in the season direction.[3]

Using a data arrangement as in Figure 4.2 allows us to view $m(\cdot,\cdot)$ in (4.6) as a regression function. Given the zero mean stationary error $\{Z_t\}$ we see that

$$\mathbb{E}[Y_{t,T}|\frac{t}{T} = u, t \bmod \theta = k] = m(u,k) \tag{4.7}$$

for $u \in [0,1]$ and $k \in \{0,1,\dots,\theta-1\}$. The regression function $m(u,k)$ can thus be interpreted as the determinstic trend at rescaled time point u in season k. This includes combinations of (u,k) that are not observed. With this interpretation of $m(\cdot,\cdot)$, the estimation of $m(\cdot,\cdot)$ looks like a mean regression problem with a continuous covariate supported on $[0,1]$ and a categrorical covariate taking values in $\{0,\dots,\theta-1\}$. This is the approach we will follow, when constructing an estimator for the season-trend function $m(\cdot,\cdot)$ in the next section.

[2]Similarly to the remark after (4.4), one can change the second argument of m to $(t + s_1 - 1) \bmod \theta$ with s_1 the season the first observation was made in so as to rearrange the seasonal effects..

[3]Although not visible from the graph this will also be true for observations in season 11 and 0 once an appropriate distance measure in the seasonal direction is used.

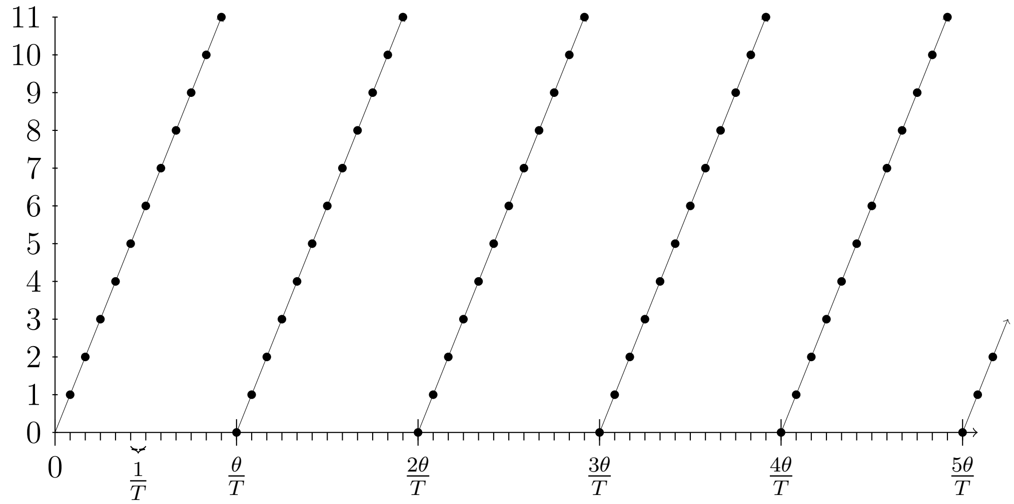


Figure 4.2: Data Arrangement in season-time space with $\theta = 12$.

4.3 Estimation

We will construct our estimator for $m(\cdot,\cdot)$ using the interpretation of $m(\cdot,\cdot)$ as a mean regression function in a setting with a continuous covariate supported on $[0,1]$ and a categrorical covariate taking values in $\{0,\dots,\theta-1\}$. We will base our estimator on the one considered in Hall et al. (2007) for the regression with mixed data in an i.i.d. setting. Following their suggestion we estimate the regression mean $m(\cdot,\cdot)$ using kernel methods by smoothing in the rescaled time direction, i.e in the direction of the continuous regressor, using the kernel $K_h(u,\frac{t}{T}) = \frac{1}{h}K(\frac{u-\frac{t}{T}}{h})$. Smoothing in the seasonal direction, i.e. in the direction of the discrete regressor, is done using the kernel $L_\lambda(x, t \bmod \theta) = \lambda^{d(x,t \bmod \theta)}$ with the exponent $d(x, t \bmod \theta)$ measuring the distance between x and $t \bmod \theta$. The bandwidths are given by h and λ. Combining these kernels into a product kernel we can define a local constant estimator for $m(\cdot,\cdot)$ by

$$\widehat{m}(u,x) = \underset{m(u,x)}{\arg\min} \sum_{t=1}^{T} (Y_{t,T} - m(u,x))^2 K_h(u,\frac{t}{T}) L_\lambda(x, t \bmod \theta). \qquad (4.8)$$

for $u \in [0,1]$ and $x \in \{0,\dots,\theta-1\}$.

The use of the kernel $L_\lambda(x, t \bmod \theta)$ including the distance measure in the exponant is recommended in Hall et al. (2007) for the case of ordered categorical covariates, which is clearly the case here. In fact, we have even more structure in our categorical covariate. Not only do we know that they are ordered, but we also know that they are circular, with the "lowest" season following the "highest". For example, take $\theta = 12$ and denote the seasons by months. The ordering of the months is clear and it is also obvious that December, the "highest" month, is followed by January, the "lowest" month. This additional structure in the covariate is incorporated by defining the distance measure $d(x, t \bmod \theta)$ by

$$d(x, t \bmod \theta) = \min\{|x - t \bmod \theta|, |x + \theta - t \bmod \theta|\}.$$

Note that due to the circularity, $d(x, t \bmod \theta) \in \{0, \dots \lfloor\frac{\theta}{2}\rfloor\}$, with $\lfloor x \rfloor$ the largest integer, smaller than or equal to x. The closed form solution to (4.8) is given by

$$\widehat{m}(u, x) = \frac{\frac{1}{T}\sum_{t=1}^{T} Y_{t,T} K_h(u, \frac{t}{T}) L_\lambda(x, t \bmod \theta)}{\frac{1}{T}\sum_{t=1}^{T} K_h(u, \frac{t}{T}) L_\lambda(x, t \bmod \theta)}$$

Denoting the product kernel by $W_{(h,\lambda)}(u, x, t) = K_h(u, \frac{t}{T}) L_\lambda(x, t \bmod \theta)$ we can rewrite the estimator as

$$\widehat{m}(u, x) = \frac{\sum_{t=1}^{T} W_{(h,\lambda)}(u, x, t) Y_{t,T}}{\sum_{t=1}^{T} W_{(h,\lambda)}(u, x, t)}.$$

Given the estimated season-trend function $\widehat{m}(u, x)$ we can retrieve an estimator for the error process. This is given by

$$\widehat{Z}_t = Y_{t,T} - \widehat{m}(\frac{t}{T}, s_t) \quad \forall t = 1, \dots, T, \tag{4.9}$$

with s_t denoting the season that observation t was made in.

4.4 Data Illustration using Weather Data

In this section we will illustrate our estimation procedure using weather data obtained from the German weather service, the DWD[4]. We will show how to interpret the estimate and briefly discuss some issues in implementing the estimator. The focus will be on using our estimator as an initial data analytic step. In doing so we hope to be able to see possible violations of any additionally imposed structure on the season and trend specification. Specifically, we will use our estimate to investigate possible deviations from imposing additivity as in (4.2).

4.4.1 The Data

The data used to illustrate our estimation procedure are monthly air temperature measurements from the weather station on the Zugspitze, Germany's highest mountain. The data run from August 1900 until December 2013 and consist of the monthly average of the average daily temperature measurements[5]. Thus, our total sample consists of $T = 1349$ monthly observations[6]. Denote the monthly average temperature at time t by $Temp_t$. The data are plotted in Figure 4.3.

By purely looking at the data plot it is difficult to judge, whether there is a seasonal pattern or not. Furthermore, it is virtually impossible in this case to discern whether it is plausible to assume an additive structure for the season and trend function as in (4.2). In the following, we will see that estimating our season-trend function may help in judging whether such a specification seems at all plausible.

[4]The DWD provides a vast amount of historical weather and climate data at http://www.dwd.de/datenservice.

[5]On each day, three measurements were taken. One in the morning, one at midday and one in the evening. The daily measurement times were changed in January 1987 and again in April 2004. For more information see the relevant pages at http://www.dwd.de

[6]The values for May, June, July and August 1945 seem to have been imputed by the DWD as the station was under US control in this period and no daily values are available.

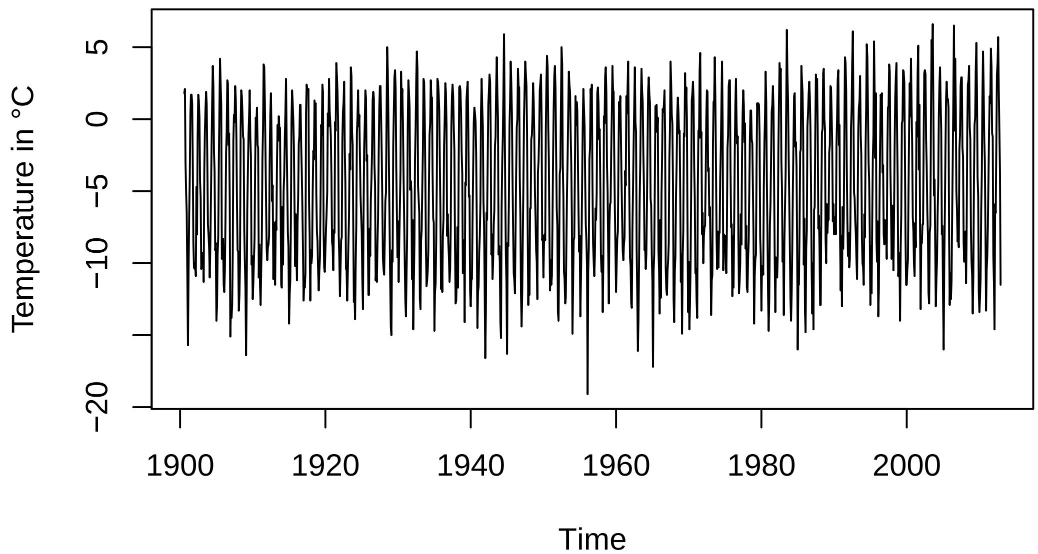


Figure 4.3: Plot of $Temp_t$, the monthly average of the average daily measured temperature on the Zugspitze.

4.4.2 The Model

We will set the number of seasons to $\theta = 12$, i.e. equal to the number of months in a year. As the first observation is in August, i.e. $s_1 = 8$, we will use the reordering of the seasons so that the first seasonal effect θ_1 in the additive model (4.2) corresponds to January. Thus, the model we will use for our monthly mean temperature is

$$Temp_{t,T} = m(\frac{t}{T}, (t+7) \bmod \theta) + Z_t \quad \forall t = 1, \ldots, T = 1349 \tag{4.10}$$

with $\{Z_t\}$ a zero mean stationary process and $m(\cdot,\cdot)$ defined on $[0,1]\times\{0,\ldots,11\}$. The corresponding additive model as in (4.2) is obtained by setting

$$m(\frac{t}{T}, t \bmod \theta) = m_0(\frac{t}{T}) + m_\theta(t)$$

with m_0 a smooth deterministic trend component and m_θ the periodic component, which using the dummy variable approach as in (4.4) is given by

$$m_\theta(t) = \sum_{k=0}^{\theta-1} \theta_k I((t+7) \bmod \theta = k).$$

Hence, the additive model we will use for the monthly mean temperature is given by

$$Temp_{t,T} = m_0(\frac{t}{T}) + \sum_{k=0}^{\theta-1} \theta_k I((t+7) \bmod \theta = k) + Z_t \quad \forall t = 1,\ldots,T \qquad (4.11)$$

for $\{Z_t\}$ a zero mean stationary process and m_0 a smooth deterministic trend function.

4.4.3 The Estimate

In this subsection, we will present the estimate of the season-trend function $m(u,x)$ for the temperature model in (4.10). The estimate $\widehat{m}(u,x)$ was obtained as described in section 4.3 using the bandwidths $(h,\lambda) = (0.24, 0.06)$. We will comment on how these bandwidths were chosen in section 4.4.4. The estimate is calculated over the grid $\{(u,x) : u \in \{\frac{1}{T}, \frac{2}{T}, \ldots, \frac{t-1}{T}, 1\}$ and $x \in \{0,1,\ldots,11\}\}$. At first we will present two ways to illustrate the estimated seasonal-trend function $\widehat{m}(u,x)$. In comparing the estimate of our model with the estimate of the additive model in (4.11) we will also make use of the interpretation of the estimated seasaon-trend function $\widehat{m}(u,x)$ as an estimated seasonal curve for every rescaled time point u or as the estimated time trend for every season x.

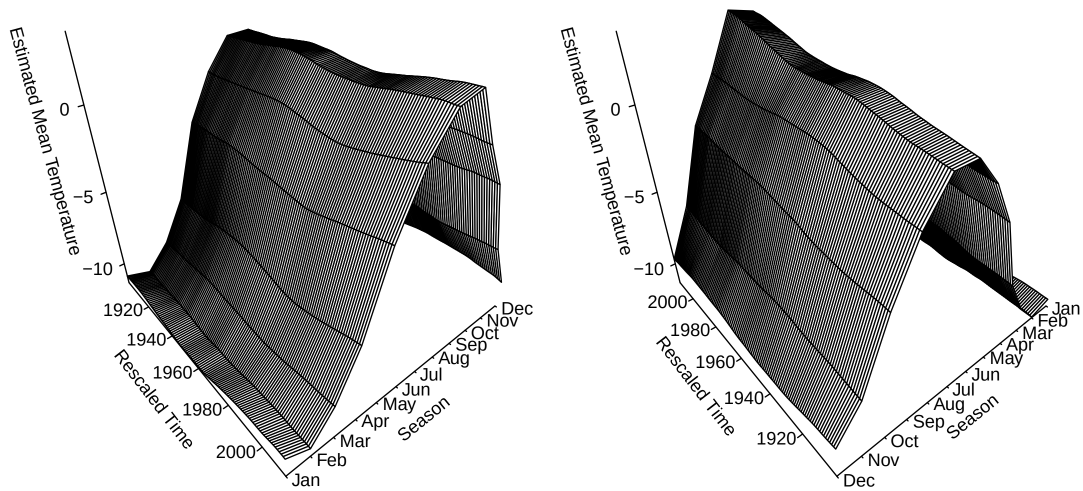


Figure 4.4: Perspective plots of estimated season-trend function $\widehat{m}(u,x)$ with bandwidth choice $(h,\lambda) = (0.24, 0.06)$. Left Panel: View from January 2013. Right Panel: View from December 1900.

The first way to illustrate the estimate $\widehat{m}(u,x)$ is given in Figure 4.4. Note, that although $m(\cdot,\cdot)$ is only defined at the seasons $x \in \{0,1,\dots,11\}$, for ease of interpretation the estimate is depicted using perspective plots[7]. The labelling on the rescaled time axis has been done using the time points prior to rescaling and the season indicators have been labelled using the corresponding month abbreviations. The perspective plot in the right panel is obtained from the one in the left panel by rotating it through 180°. This enables us to see the whole season-trend function estimate. Returning to the two interpretations of the estimate, the seasonal curve

[7]The plot only uses the estimates at yearly intervals from June 1901, corresponding to $u = 11/T$ up until June 2012, which corresponds to $u = 1331/T$. Reducing the amount of points in the perspective plots was needed as using all time points would have made the perspective plot so dense that no shape would be visible.

for every rescaled time point is obtained by taking slices parallel to the season axis. The other interpretation of the estimate as an estimated time trend for every season can be seen by taking slices parallel to the rescaled time axis. These estimated trends for each season are depicted by the 'horizontal' lines on the estimated surface. In general, the mean monthy temperature seems to have increased slightly, peaking in about 1950 before dropping down to a trough in about 1970 and then increasing again until today. This rough pattern seems to be common across all seasons, although the increases and drops vary in magnitude.

An alternative way to present the estimate of the season-trend function is to look at the contour plot corresponding to the above perspective plots, which is given in Figure 4.5.

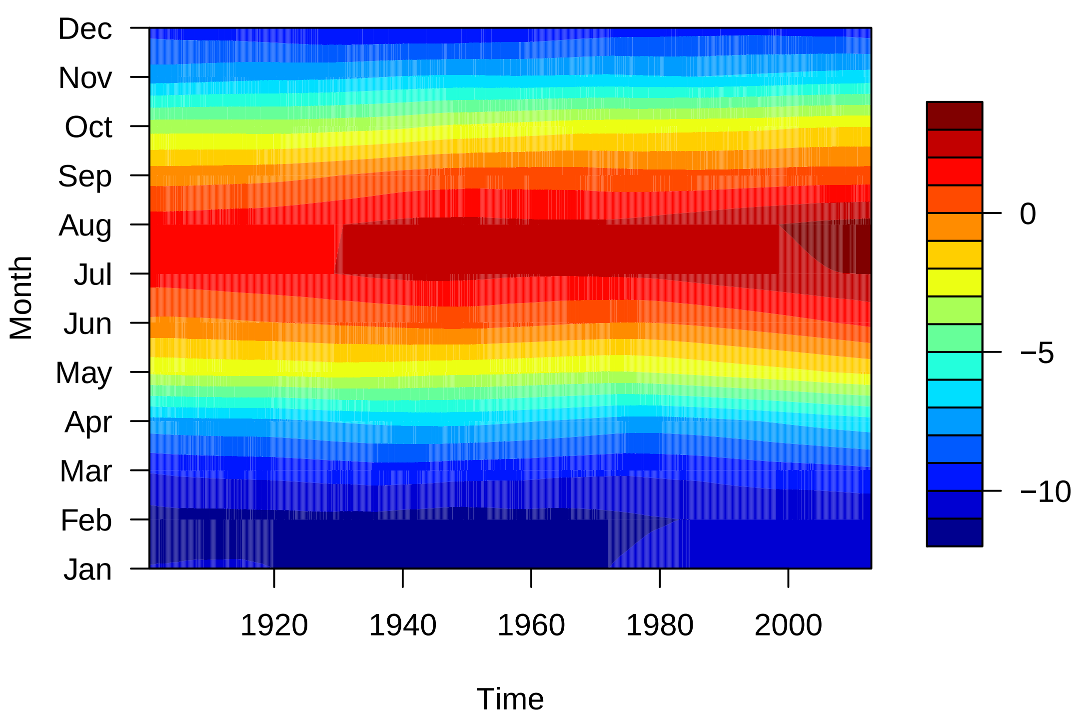


Figure 4.5: Contour plot of estimated seasonal trend function with bandwidth choice $(h, \lambda) = (0.24, 0.06)$.

Taking vertical slices one obtains the estimated seasonal curve at the respective rescaled time point, whereas taking horizontal slices yields the estimated time trends for the respective season. The general pattern of a mean temperature rise until about 1950, a subsequent fall until about 1970 and the continuing rise until the present day is again visible. However, one can now we begin to see that this pattern is not the same over all seasons. One difference is with respect to the maximal estimated temperature difference within a season. Comparing the horizontal slices of the estimated contour for March and the summer months of July and August, one can see that the change in March over the entire period is less than $1°C$, whereas it is approaching $2°C$ for Juli and August. This observed difference in the range of the estimated mean monthly temperature by month is made more explicit in table 4.1. The first row (*Max.*) provides the largest estimated mean monthly temperature by month over the observation period. The second row (*Min.*) gives the corresponding smallest estimated mean monthly temperature. Finally, their differenc is in the third row (*Range*) from which we see that the maximal difference of the estimated mean monthly temperature is $0.72°C$ in March and $1.85°C$ in August.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Max.	-10.16	-10.88	-9.19	-6.35	-1.80	1.27	3.03	3.34	0.54	-2.04	-6.54	-9.38
Min.	-11.43	-11.49	-9.91	-7.43	-3.05	-0.22	1.45	1.49	-0.43	-3.51	-7.54	-10.00
Range	1.27	0.61	0.72	1.08	1.25	1.49	1.58	1.85	0.97	1.47	1.00	0.62

Table 4.1: Maximum (*Max.*), Minimum (*Min.*) and Range (*Range*) of estimated mean monthly temperature by month.

Not only does the maximal difference of the estimated mean monthly temperature vary by month, but the shape of the estimated trend itself als varies by month. This can be seen more clearly by plotting the estimated seasonal trends by month as in Figure 4.6. The shape of the trends in March, April, May and June are very similar. The trends in July and August are also very similar and differ from the four preceding months by the absence of a drop in the middle of the observation period. The remaining six trends cannot be grouped so easily and are quite different to the other trends in terms of shape.

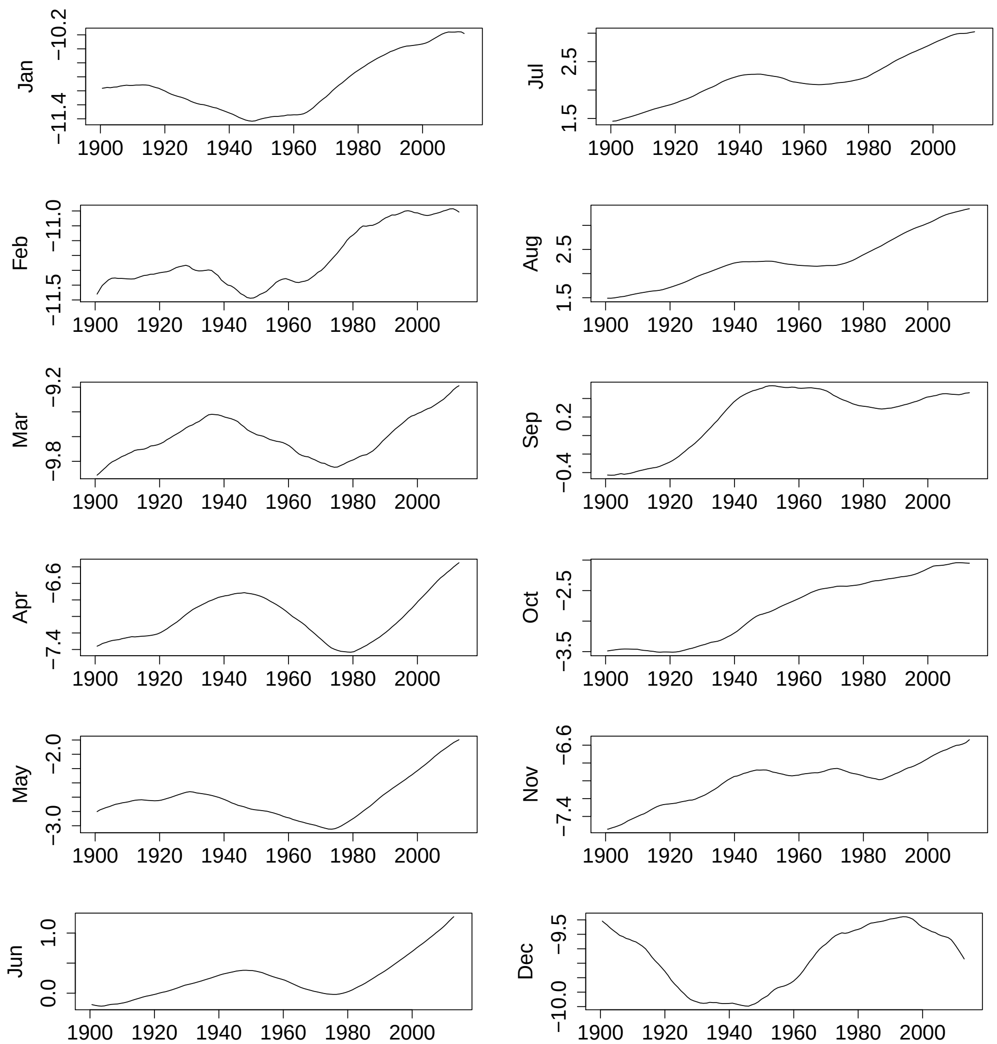


Figure 4.6: Plot of estimated seasonal trends by month with bandwidth choice $(h, \lambda) = (0.24, 0.06)$ for January until June (left hand panel) as well as July until December (right hand panel).

Of course this difference in the estimated trend over each season cannot be seen when one estimates the additive model (4.11). Instead, $\widehat{m}_0$, the estimated trend in the additive model, is a weighted average of the estmiated seasonal trends in Figure 4.6. In Figure 4.7 one can see that the estimated trend in the additive model has a shape similar to the one seen for the July and August trends in Figure 4.6. This is not so surprising given that the shape of the trends for March to June are fairly similar and the range of the estimated trends is largest in July and August, as seen in table 4.1.

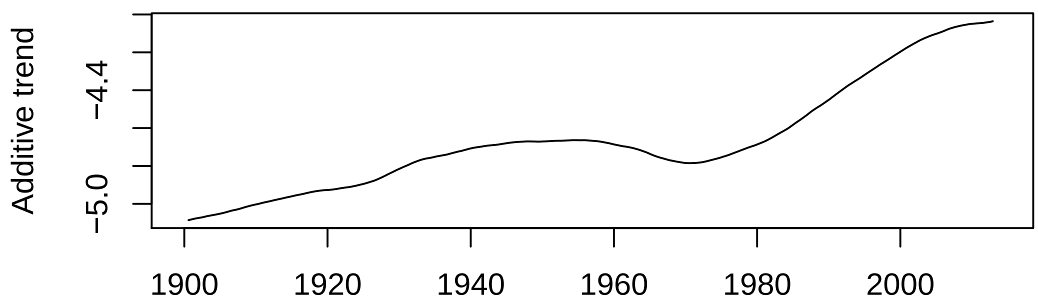


Figure 4.7: Plot of estimated trend $\widehat{m}_0$ from additive model (4.11) with bandwidth $h = 0.2$.

Up until now, we have focused mainly on the interpretation of the season-trend function as giving the estimated time trend for every season. Let us now turn to the other interpretation of it providing a seasonal curve for every rescaled time point. From the estimated season-trend function $\widehat{m}(u, x)$ the estimated season curve at rescaled time point u_0 is given by

$$\{\widehat{m}(u_0, x) : x \in \{0, \ldots, \theta - 1\}\}.$$

Figure 4.8 provides a plot of these estimated seasonal curves for seven different time points. If the season-trend function $m(u,x)$ were additive as in (4.11), then we would expect the estimated seasonal curves $\{\widehat{m}(u,x) : x \in \{0,\ldots,\theta-1\}\}$ to have a similar shape and roughly be parallel.

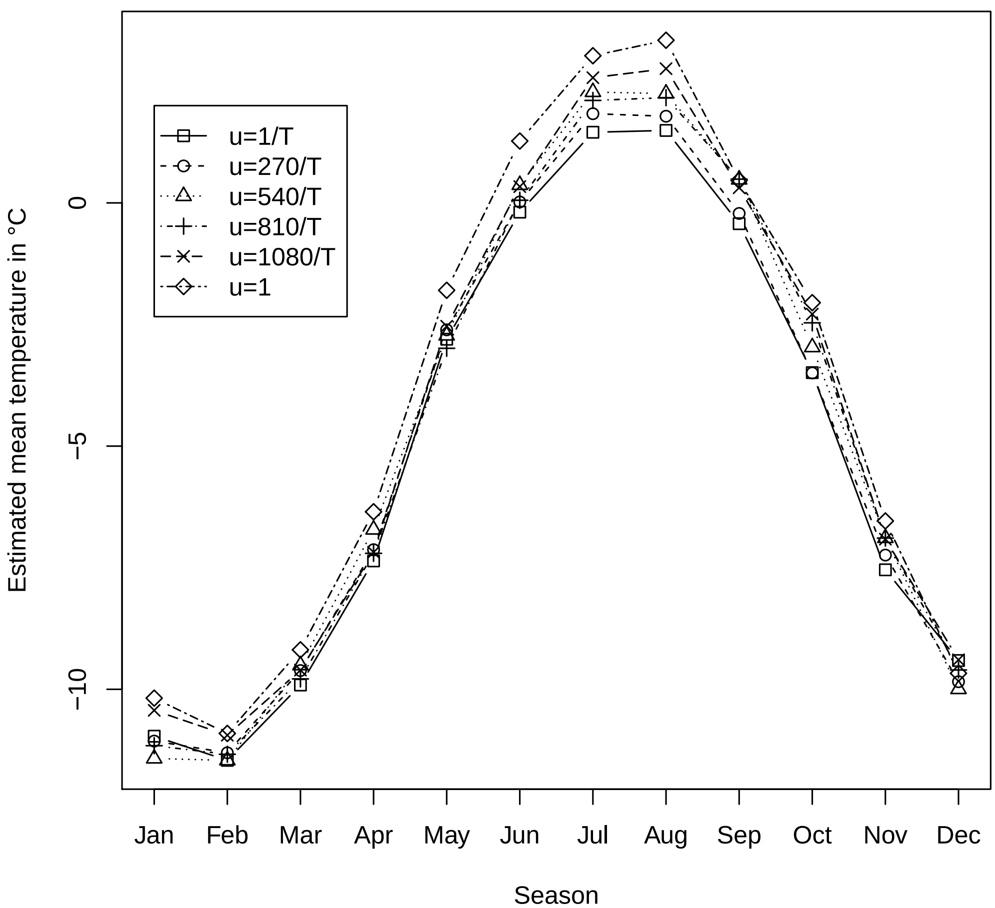


Figure 4.8: Plot of estimated seasonal trends with bandwidth choice $(h,\lambda) = (0.24, 0.06)$ for $u \in \{1/T, 270/T, 540/T, 810/T, 1080/T, 1\}$.

As we can see for the seven chosen time points the seasonal curves do indeed have the same shape. However, we again observe that the seasonal curves have changed over time, with the estimated effect of the summer months having increased, whereas the estimated seasonal effect in December having hardly changed at all.

Comparison of estimate to actual data

In this subsection, we see how our estimate compares to the actual data series. To do so we compare the fit of our model at the observed combinations of time and season with the actual series. The fit of our model is given by $\widehat{m}(\frac{t}{T}, (t+7) \bmod 12)$. In Figure (4.9) we see a plot of the fit along with the actually observed monthly mean Temperature ($Temp_t,$).

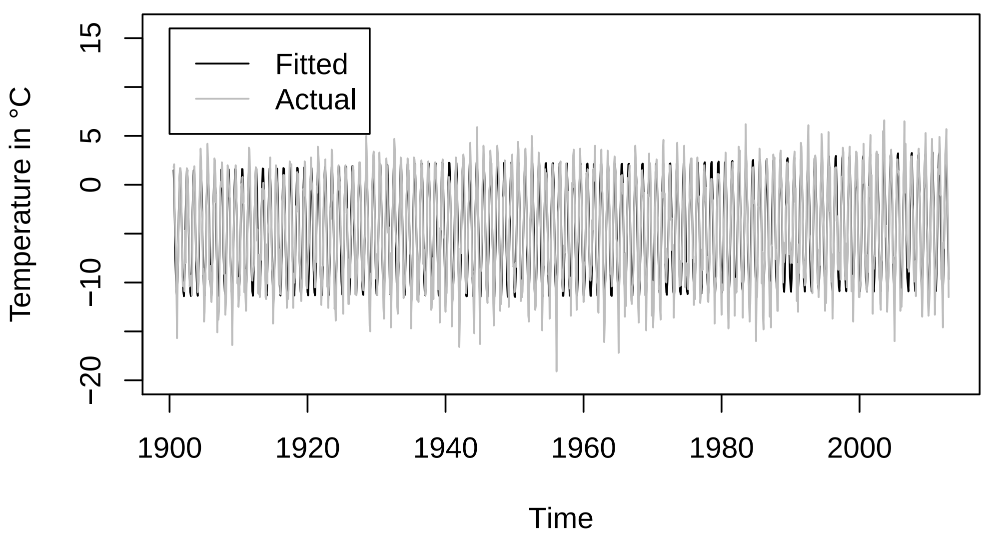


Figure 4.9: Plot of fitted versus actual mean monthly temperature with bandwidth choice $(h, \lambda) = (0.24, 0.06)$.

The fit is given by the black line plotted in the background. The actual monthly temperature is plotted using the grey line in the foreground. The fit seems to capture the seasonal fluctation quite well. The model fit also displays a slight upward movement particularly in the summer months. However, the peaks are not fitted so well in particular those in the negative direction.

We will take a closer look at the discrepancy between the estimated and the actual mean temperature by analysing the residual process, i.e. the estimated error process, given by

$$\widehat{Z}_t = Temp_t - \widehat{m}(\frac{t}{T}, (t+7) \bmod 12) \quad \forall t = 1, \ldots, T.$$

The interest will be in determining what additional structure remains in the residual process. The plot of the residual process over time is given in Figure 4.10.

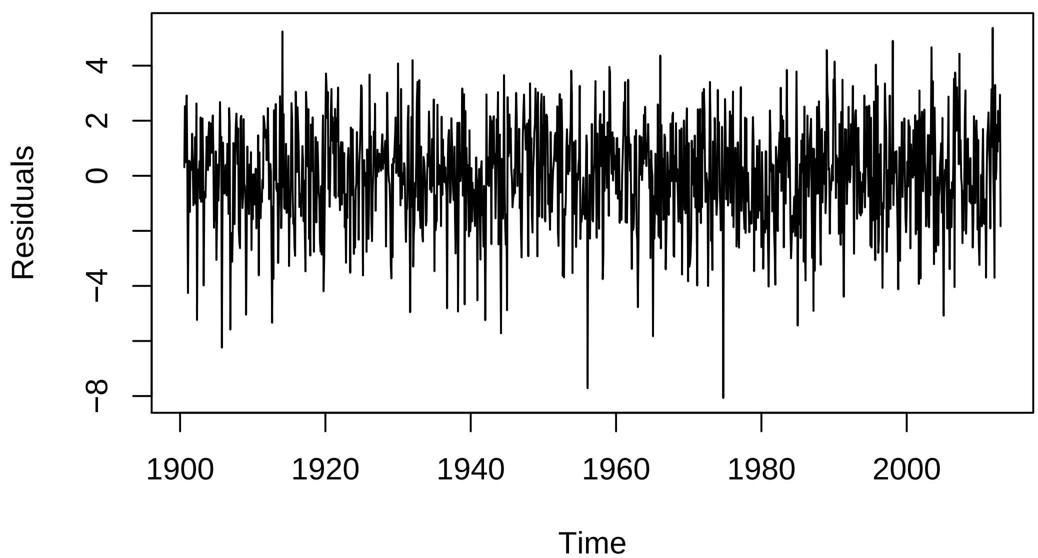


Figure 4.10: Plot of residual process over time.

We can see that the residual process displays occasional large outliers in particular in the negative direction, which given the fitted versus actual plot in Figure 4.9 most likely result from overestimating the monthly mean temperature for months with large negative values of $Temp_t$. From the residual plot we can also see that there seems to be some slight positive persistence and possibly an increase in variation over time.

Figure 4.11 provides the estimated autocorrelations up to lag 72 of the residual process. The dashed lines in the graph correspond to $1.96/\sqrt{T}$ and $-1.96/\sqrt{T}$, which would be the asymptotic 95% pointwise confidence bands if the residual process were independent white noise (see Brockwell and Davis (1991)).

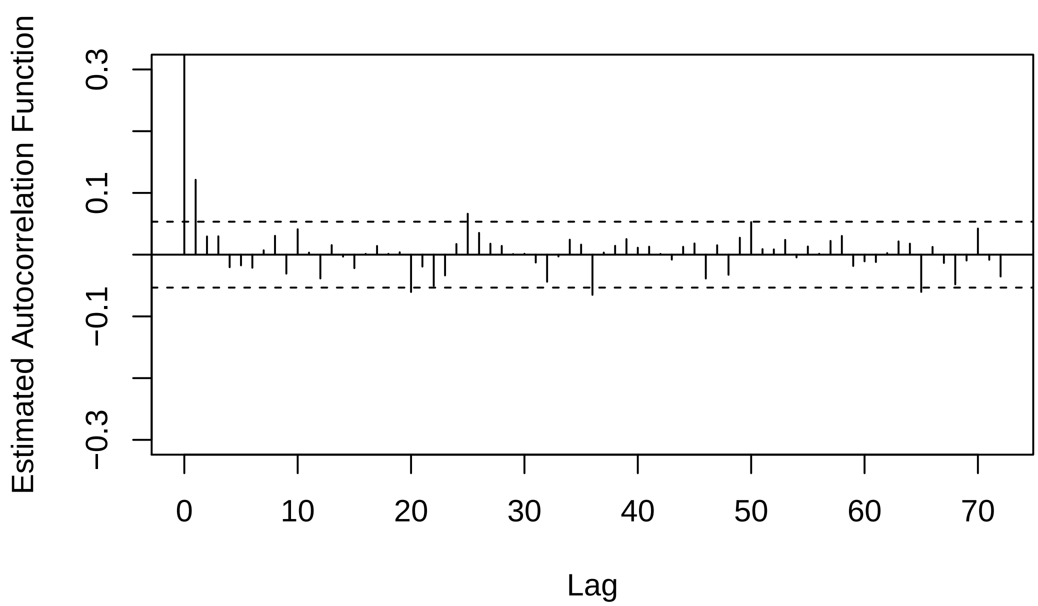


Figure 4.11: Estimated autocorrelation function of residual process.

Looking at the estimated autocorrelation function in Figure 4.11 two things become apparent. Firstly, the seemingly positive persistence of the residual process is confirmed by the estimated autocorrelation of the first lag being somewhat larger than the remaining lags. Secondly, the estimated autocorrelation function does not seem to indicate any additional periodicity, which would result in a sinusoidal shape. As we will see in the next section, these two findings have a lot to do with the choice of the smoothing bandwidths.

4.4.4 Comment on bandwidth selection

When implementing the estimate in Subsection 4.4.3 we used the bandwidth choice $(h, \lambda) = (0.24, 0.06)$. In this subsection, we will explain how we arrived at this particular choice. As in all nonparametric estimation settings, the choice of the smoothing parameters is crucial. However, without distributional results for our estimator, we cannot use the popular method of plug-in bandwidths, which rely on asymptotic expansions of the respective estimators. Furthermore, it is known that in trend estimation, standard crossvalidation procedures do not work when the error term is autocorrelated, in particular when the error process is positively correlated (see for example Altman (1990),Altman (1993), Hart (1991), Hart (1994), Hall and Keilegom (2003) or Herrmann et al. (1992)). All the aforementioned studies look at bandwidth selection in nonparametric trend estimation without any seasonality. In our setting, crossvalidation also appears to result in undersmoothing in the time direction when the error is positively autocorrelated.

Without guidance from statistical theory to help choose our bandwidths, we followed a modelling principal mentioned in Mudelsee (2010) for climate data models, namely that the residual process should contain little structure. Thus we estimated our model for a large number of bandwidths and then chose the bandwidths that in some way minimized the structure in the residual process. The bandwidth in the rescaled time direction (h) was taken from $\{0.05, 0.055, 0.06, \dots, 0.29, 0.295, 0.3\}$ and the one in the seasonal direction (λ) was taken from $\{0, 0.02, \dots, 0.28, 0.3\}$. In total, this resulted in the need to estimate 816 models, one for each pair of bandwidths (h, λ). For each choice of bandwidths we then computed the estimated autocorrelation function of the residual process, denoted by $\{\widehat{\rho}_k(h, \lambda) : k = 0, 1, 2, \dots\}$.

To capture the idea of minimizing the structure in the residual process, we chose to minimize the sum of squared estimated autocorrelations up to some order p, i.e. set

$$(h_p, \lambda_p) = \arg\min_{h,\lambda} \sum_{k=1}^{p} \widehat{\rho}_k(h, \lambda)^2,$$

with $\widehat{\rho}_k$ denoting the estimated lag k autocorrelation of the residual process and p the number of lags included in the sum. Using this criterion, we computed (h_p, λ_p) for different values of p. The results are given in Table 4.2. The choice of $(h, \lambda) = (0.24, 0.06)$ was made due to the closeness of the values in the last three columns of Table 4.2 to this pair of values.

p	12	24	36	48	60	72	84	96
h_p	0.18	0.24	0.24	0.24	0.215	0.24	0.245	0.25
λ_p	0.00	0.00	0.04	0.04	0.08	0.06	0.06	0.06

Table 4.2: Bandwidth choice that maximizes sum of estimated squared autorcorrelations of error up to order of lag in first column.

To finish this comment on bandwidth choice, we want to highlight the effect of the bandwidths on the estimate and the residual process. When comparing the estimate over all 816 model one can see that the overall shape of the estimates are similar. Increasing λ, the smoothing parameter in the seasonal direction leads to a dampening of the estimated season-trend function. Increasing h, the bandwidth in the rescaled time direction leads the seasonal trends shown in Figure 4.6 to become smoother. In contrast ‘wigglier’ seasonal trends are obtained for smaller bandwidths h.

The effect of the bandwidth choice on the residual process is even more pronounced. Firstly, by eye there is hardly any recognizable difference when varying h in [0.14,0.3] and λ in [0.00,0.14]. However, for smaller values of h, the estimated autocorrelations at multiples of lag 12 start to decrease markedly. This is seen in the upper panel of Figure 4.12, which gives the estimated autocorrelation function for the residual process with the extreme bandwidth choice $(h, \lambda) = (0.05, 0.06)$. The larger negative autocorrelations at lags 12, 24, 36 and 48 are clearly visible.

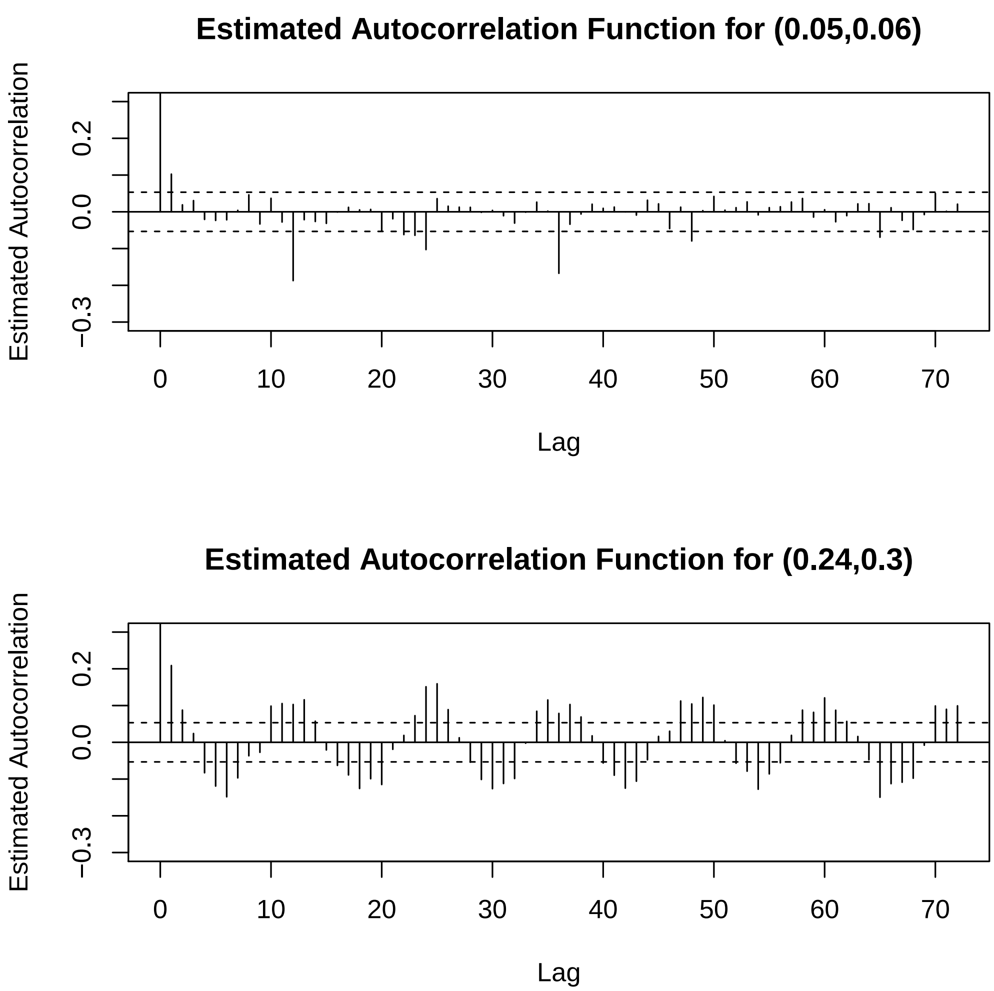


Figure 4.12: Upper panel: Estimated autocorrelation function of residual process for $(h, \lambda) = (0.0.5, 0.06)$. Lower panel: Estimated autocorrelation function of residual process for $(h, \lambda) = (0.24, 0.30)$.

The other clearly visible impact of the bandwidth choice on the residual process pertains to large values of λ, which lead to a sinusoidal pattern in the estimated autocorrelation function as illustrated in the lower panel of Figure 4.12 for the bandwidth choice $(h, \lambda) = (0.24, 0.30)$.

4.4.5 Comparison to other models

In this final subsection, we will compare the estimate from our model with estimates obtained from three competing models. We will consider the nonparametric additive model in (4.11) and two parametric additive models: One using a linear time trend, and the other a more flexible cubic time trend. The seasonal component will be modelled using the dummy variable approach in all of the models. Using $\theta = 12$ and the fact that the first observation is in August the competing models are:

$$\text{(I)} \quad Temp_{t,T} = m_0(\frac{t}{T}) + \sum_{k=0}^{11} \theta_k I((t+7) \bmod 12 = k) + Z_t$$

$$\text{(II)} \quad Temp_{t,T} = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \sum_{k=0}^{11} \theta_k I((t+7) \bmod 12 = k) + Z_t$$

$$\text{(II)} \quad Temp_{t,T} = \beta_0 + \beta_1 t + \sum_{k=0}^{11} \theta_k I((t+7) \bmod 12 = k) + Z_t$$

In all three models we follow the estimation procedure in Chapter 2 to first estimate the parameters of the seasonal component. Subtracting the estimated seasonal component from the monthly mean temperature we then estimated the respective trend components. The parametric trends are estimated by least squares. The trend in the nonparametric additive component model in (I) is estimated using a local constant estimator.

We will focus solely on comparing the fits and the residual processes of the competing models with ours. As it is difficult to discern much by overlaying the fits we have plotted the difference of the fit from our model to the ones for the three competing models in Figure (4.13).

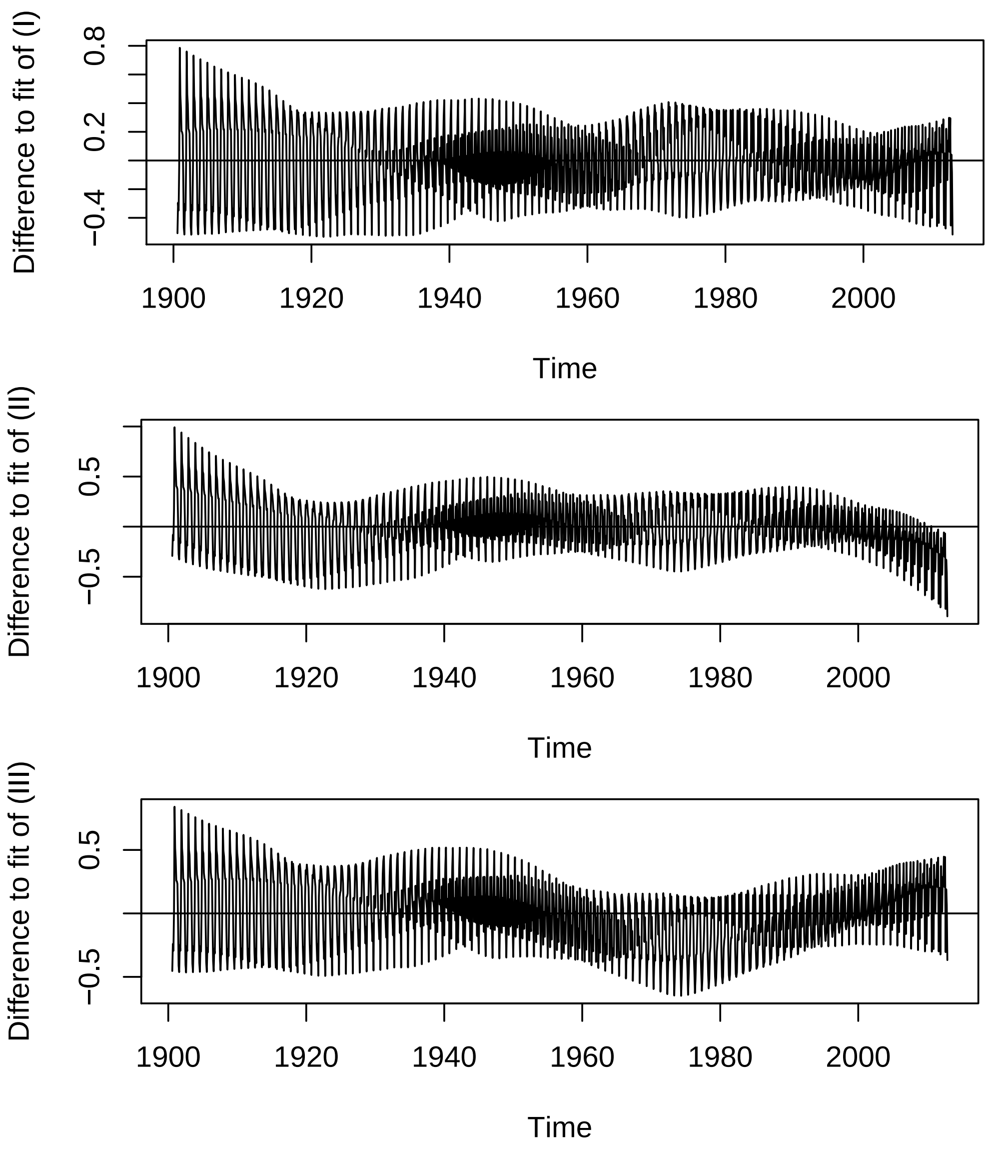


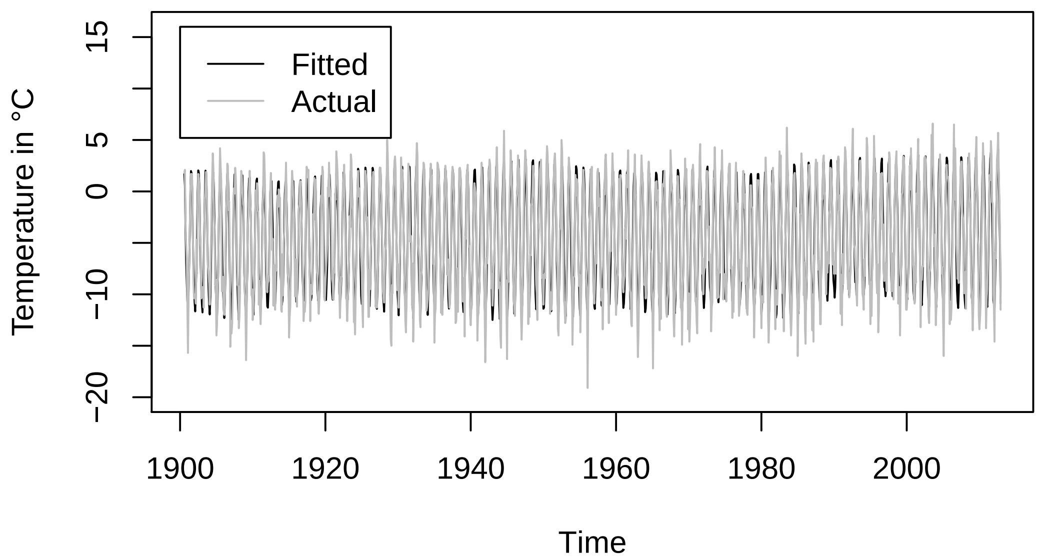
Figure 4.13: Upper panel: Difference of fits between our model and model (I). Middle panel: Difference of fits between our model and model (II). Lower panel: Difference of fits between our model and model (III).

As we have already seen the additive model fit is in close accordance with our model fit. This visible again in the top panel of Figure (4.13). The largest difference between the two fits is at the beginning of the sample with our model fitting an at most $0.8°C$ higher temperature than the additive model. The difference in the fits then reduces to below $0.5°C$. It is also seen to be quite symmetric without one of the models systematically etimating a higher temperature than the other. In contrast modelling the trend using a cubic time trend results in a much less symmetric difference of fits. Especially at the end of the observation period the fit using the model with the cubic trend is systematically higher than the one from our model. The behaviour of the fit of the linear trend model in the bottom panel seems to be intermediate to the other two fits.

In total, all the fits are fairly close to one another with none deviating from our fit by more than $1°C$ at any point in time. Furthermore, there seems to be little difference in the quality of the fits for all the models as the residual processes and their respective estimated autocorellation functions are virtually identical.

4.5 Concluding Remarks

We have introduced a model that avoids decomposing the time series under study into a periodic component, a trend component and a noise component. Instead, the model is characterised by a season-trend function that can be interpreted as a regression function when rearranging the data. We have illustrated how to interpret the estimate obtained from the model by appplying it to a German temperature series. There it was seen that the behaviour of our estimate is fairly similar to the one obtained from the additive specifications. This in turn is strongly dependent on the chosen bandwidth as we will show next. To do so, we have plotted the fitted temperature versus the actual mean monthly temperature for our model (4.10) and the additive model (4.11) in Figure 4.14 for a different choice of bandwidths than in Section 4.4. For the additive model we have set $h = 0.05$ and for our model we have taken $(h, \lambda) = (0.05, 0.0.6)$.



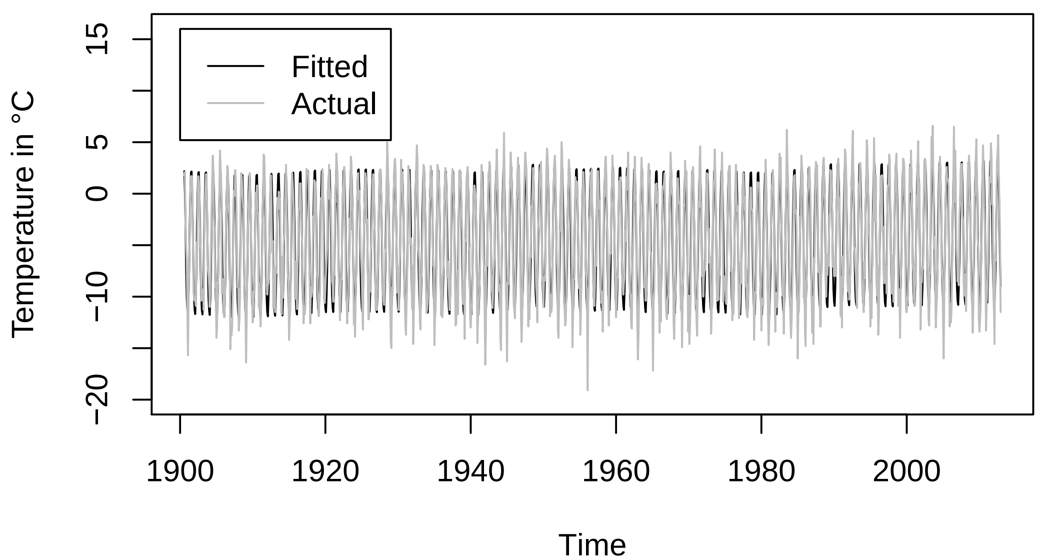


Figure 4.14: Upper panel: Fit versus actual mean monthly temperature for our model with bandwidth choice $(h, \lambda) = (0.05, 0.06)$. Lower panel: Fit versus actual mean monthly temperature for addtive model with bandwidth choice $h = 0.05$.

The fit in the upper panel seems to capture the "medium" swings in the monthly temperature much better than the fit for the additive model in the lower panel. Hence, although this bandwidth choice leaves additional structure in the residual process as is evident from Figure 4.10, it seems to produce a model that fits the mean monthly temperature much better than the one in Section 4.4. This highlights how important the bandwidth choice is in our model. Consequently, it would be desirable to have a data dependent bandwidth selection method. One possibility may be to use a local linear estimator in the rescaled time direction and select the bandwidths by adapting the one sided cross validation approach introduced by Hart and Yi (1998), which is well behaved for autocorrelated errors as shown in Hart and Lee (2005).

Given a reliable bandwidth selection method one could apply our model to seasonally unadjusted economic time series. This would be especially interesting given the extensive use of additive decomposition models in data agencies. In this context one would also want to be able to test for additivity.

Lastly, one could look at the choice of θ, in particular with respect to the robustness of the estimate and the possible estimation of θ.

Appendices

Appendix A

Proofs for Chapter 2

In this appendix we have collected all the proofs needed to establish the results in Chapter 2 on the additive model.

A.1 Proof of Theorem 2.4.2

In this appendix, we prove Theorem 2.4.2, which describes the asymptotic behaviour of our smooth backfitting estimates. For the proof, we split up the estimates into a "stochastic" part and a "bias" part. In Theorem A.1.1, we provide a uniform expansion of the stochastic part. This result is an extension of a related expansion given in Mammen and Park (2005) in the context of bandwidth selection in additive models. The bias part is treated in Theorem A.1.2. The proof of both theorems requires uniform convergence results for the kernel smoothers that enter the backfitting procedure as pilot estimates. These results are summarized in Appendix A.3. Note that the two theorems A.1.1 and A.1.2 are not only needed for the second estimation step but also for the derivation of the asymptotics of the AR estimates in the third step. Throughout this appendix, we use the symbol C to denote a finite real constant which may take a different value on each occurrence.

Proof of Theorem 2.4.2

We decompose the backfitting estimates $\widetilde{m}_j$ into a stochastic part $\widetilde{m}_j^A$ and a bias part $\widetilde{m}_j^B$ according to

$$\widetilde{m}_j(x_j) = \widetilde{m}_j^A(x_j) + \widetilde{m}_j^B(x_j).$$

The two components are defined by

$$\widetilde{m}_j^S(x_j) = \widehat{m}_j^S(x_j) - \sum_{k \neq j} \int_0^1 \widetilde{m}_k^S(x_k) \frac{\widehat{p}_{k,j}(x_k, x_j)}{\widehat{p}_j(x_j)} \, dx_k - \widetilde{m}_c^S \tag{A.1}$$

for $S = A$, B. Here, $\widehat{m}_k^A$ and $\widehat{m}_k^B$ denote the stochastic part and the bias part of the Nadaraya-Watson pilote estimates defined as

$$\widehat{m}_j^A(x_j) = \frac{1}{T} \sum_{t=1}^T K_h(x_j, X_t^j) \varepsilon_t / \widehat{p}_j(x_j) \tag{A.2}$$

$$\begin{aligned} \widehat{m}_j^B(x_j) = \frac{1}{T} \sum_{t=1}^T K_h(x_j, X_t^j) \big[(m_\theta(t) - \widetilde{m}_\theta(t)) \\ + m_0\Big(\frac{t}{T}\Big) + \sum_{k=1}^d m_k(X_t^k) \big] / \widehat{p}_j(x_j) \end{aligned} \tag{A.3}$$

for $j = 0, \ldots, d$, where we set $X_t^0 = \frac{t}{T}$ to shorten the notation. Furthermore, $\widetilde{m}_c^A = \frac{1}{T} \sum_{t=1}^T \varepsilon_t$ and $\widetilde{m}_c^B = \frac{1}{T} \sum_{t=1}^T \{ (m_\theta(t) - \widetilde{m}_\theta(t)) + m_0(\frac{t}{T}) + \sum_{k=1}^d m_k(X_t^k) \}$. We now analyse the convergence behaviour of $\widetilde{m}_j^A$ and $\widetilde{m}_j^B$ separately.

We first provide a higher-order expansion of the stochastic part $\widetilde{m}_j^A$. The following result extends Theorem 6.1 in Mammen and Park (2005) (in particular their equation (6.3)) to our setting.

Theorem A.1.1. *Suppose that assumptions* (A1) – (A5) *apply and that the bandwidth h satisfies* (A6)(a) *or* (A6)(b). *Then*

$$\sup_{x_j \in [0,1]} \Big| \widetilde{m}_j^A(x_j) - \widehat{m}_j^A(x_j) - \frac{1}{T} \sum_{t=1}^T r_{j,t}(x_j) \varepsilon_t \Big| = o_p\Big(\frac{1}{\sqrt{T}} \Big),$$

where $r_{j,t}(\cdot) := r_j(\frac{t}{T}, X_t, \cdot)$ are absolutely uniformly bounded functions with

$$|r_{j,t}(x_j') - r_{j,t}(x_j)| \leq C |x_j' - x_j|$$

for a constant $C > 0$.

Proof. As Mammen and Park (2005) work in an i.i.d. setting, we cannot apply their Theorem 6.1 directly. In what follows, we outline the arguments needed to extend their proof to our framework. For an additive function $g(x) = g_0(x_0) + \ldots + g_d(x_d)$, let

$$\widehat{\psi}_j g(x) = g_0(x_0) + \ldots + g_{j-1}(x_{j-1}) + g_j^*(x_j) + g_{j+1}(x_{j+1}) + \ldots + g_d(x_d)$$

with

$$g_j^*(x_j) = -\sum_{k \neq j} \int_0^1 g_k(x_k) \frac{\widehat{p}_{j,k}(x_j, x_k)}{\widehat{p}_j(x_j)} dx_k + \sum_{k=0}^{d} \int_0^1 g_k(x_k) \widehat{p}_k(x_k) dx_k.$$

Using the uniform convergence results from Appendix A.3 and exploiting our model assumptions, we can show that Lemma 3 in Mammen et al. (1999) applies in our case. For $\widetilde{m}^A(x) = \widetilde{m}_0^A(x_0) + \ldots + \widetilde{m}_d^A(x_d)$, we therefore have the expansion

$$\widetilde{m}^A(x) = \sum_{r=0}^{\infty} \widehat{S}^r \widehat{\tau}(x),$$

where $\widehat{S} = \widehat{\psi}_d \cdots \widehat{\psi}_0$ and $\widehat{\tau}(x) = \widehat{\psi}_d \cdots \widehat{\psi}_1 [\widehat{m}_0^A(x_0) - \widehat{m}_{c,0}^A] + \ldots + \widehat{\psi}_d [\widehat{m}_{d-1}^A(x_{d-1}) - \widehat{m}_{c,d-1}^A] + [\widehat{m}_d^A(x_d) - \widehat{m}_{c,d}^A]$ with $\widehat{m}_{c,j}^A = \int_0^1 \widehat{m}_j^A(x_j)\widehat{p}_j(x_j)dx_j$. Now decompose $\widetilde{m}^A(x)$ according to

$$\widetilde{m}^A(x) = \widehat{m}^A(x) - \widehat{m}_c^A + \sum_{r=0}^{\infty} \widehat{S}^r (\widehat{\tau}(x) - (\widehat{m}^A(x) - \widehat{m}_c^A)) + \sum_{r=1}^{\infty} \widehat{S}^r (\widehat{m}^A(x) - \widehat{m}_c^A)$$

with $\widehat{m}^A(x) = \widehat{m}_0^A(x_0) + ... + \widehat{m}_d^A(x_d)$ and $\widehat{m}_c^A = \widehat{m}_{c,0}^A + \ldots + \widehat{m}_{c,d}^A$. We show that there exist absolutely bounded functions $a_t(x)$ with $|a_t(x) - a_t(y)| \leq C\|x - y\|$ for a constant C s.t.

$$\sum_{r=1}^{\infty} \widehat{S}^r (\widehat{m}^A(x) - \widehat{m}_c^A) = \frac{1}{T} \sum_{t=1}^{T} a_t(x)\varepsilon_t + o_p\Big(\frac{1}{\sqrt{T}}\Big) \tag{A.4}$$

uniformly in x. A similar claim holds for the term $\sum_{r=0}^{\infty} \widehat{S}^r (\widehat{\tau}(x) - (\widehat{m}^A(x) - \widehat{m}_c^A))$. As $\widehat{m}_c^A = (d+1)\frac{1}{T}\sum_{t=1}^{T} \varepsilon_t$, this implies the result.

The idea behind the proof of (A.4) is as follows: From the definition of the operators $\widehat{\psi}_j$, it can be seen that

$$\widehat{S}(\widehat{m}^A(x) - \widehat{m}_c^A) = \sum_{j=0}^{d-1} \widehat{\psi}_d \cdots \widehat{\psi}_{j+1} \Big(\sum_{k=j+1}^{d} S_{j,k}(x_j) \Big) \tag{A.5}$$

with

$$S_{j,k}(x_j) = -\int_0^1 \frac{\widehat{p}_{j,k}(x_j, x_k)}{\widehat{p}_j(x_j)} (\widehat{m}_k^A(x_k) - \widehat{m}_{c,k}^A) dx_k.$$

In what follows, we show that the terms $S_{j,k}(x_j)$ have the representation

$$S_{j,k}(x_j) = -\frac{1}{T} \sum_{t=1}^{T} \Big(\frac{p_{j,k}(x_j, X_t^k)}{p_j(x_j) p_k(X_t^k)} - 1 \Big) \varepsilon_t + o_p\Big(\frac{1}{\sqrt{T}} \Big) \tag{A.6}$$

uniformly in x_j. Thus, they essentially have the desired form $\frac{1}{T} \sum_t w_{t,k}(x_j) \varepsilon_t$ with some weights $w_{t,k}$. This allows us to infer that

$$\widehat{S}(\widehat{m}^A(x) - \widehat{m}_c^A) = \frac{1}{T} \sum_{t=1}^{T} b_t(x) \varepsilon_t + o_p\Big(\frac{1}{\sqrt{T}} \Big) \tag{A.7}$$

uniformly in x with some absolutely bounded functions b_t satisfying $|b_t(x) - b_t(y)| \leq C\|x - y\|$ for some $C > 0$. Moreover, using the uniform convergence results from Appendix A.3, it can be shown that

$$\sum_{r=0}^{\infty} \widehat{S}^r (\widehat{m}^A(x) - \widehat{m}_c^A) = \sum_{r=0}^{\infty} S^{r-1} \widehat{S}(\widehat{m}^A(x) - \widehat{m}_c^A) + o_p\Big(\frac{1}{\sqrt{T}} \Big) \tag{A.8}$$

uniformly in x, where S is defined analogously to $\widehat{S}$ with the density estimates replaced by the true densities. Combining (A.7) and (A.8) completes the proof.

To show (A.6), we exploit the mixing behaviour of the variables X_t. Plugging the definition of $\widehat{m}_k^A$ into the term $S_{j,k}$, we can write

$$S_{j,k}(x_j) = -\frac{1}{T} \sum_{t=1}^{T} \Big(\int_0^1 \frac{\widehat{p}_{j,k}(x_j, x_k)}{\widehat{p}_j(x_j) \widehat{p}_k(x_k)} K_h(x_k, X_t^k) dx_k - 1 \Big) \varepsilon_t.$$

Then applying the uniform convergence results from Appendix A.3, we can replace the density estimates in the above expression by the true densities. This yields

$$S_{j,k}(x_j) = -\frac{1}{T}\sum_{t=1}^{T}\Big(\int_0^1 \frac{p_{j,k}(x_j,x_k)}{p_j(x_j)p_k(x_k)}K_h(x_k,X_t^k)dx_k - 1\Big)\varepsilon_t + o_p\Big(\frac{1}{\sqrt{T}}\Big)$$
$$=: S^*_{j,k}(x_j) + o_p\Big(\frac{1}{\sqrt{T}}\Big)$$

uniformly for $x_j \in [0,1]$. In the final step, we show that

$$S^*_{j,k}(x_j) = -\frac{1}{T}\sum_{t=1}^{T}\Big(\frac{p_{j,k}(x_j,X_t^k)}{p_j(x_j)p_k(X_t^k)} - 1\Big)\varepsilon_t + o_p\Big(\frac{1}{\sqrt{T}}\Big)$$

again uniformly in x_j. This is done by applying a covering argument together with an exponential inequality for mixing variables. The employed techniques are similar to those used to establish the results of Appendix A.3. □

We now turn to the bias part $\widetilde{m}_j^B$.

Theorem A.1.2. *Suppose that* (A1) – (A5) *hold. If the bandwidth h satisfies* (A6)(a), *then*

$$\sup_{x_j \in I_h} |\widetilde{m}_j^B(x_j) - m_j(x_j)| = O_p(h^2) \tag{A.9}$$
$$\sup_{x_j \in I_h^c} |\widetilde{m}_j^B(x_j) - m_j(x_j)| = O_p(h) \tag{A.10}$$

for $j = 0, \dots, d$. If the bandwidth satisfies (A6)(b), *we have*

$$\sup_{x_j \in I_h} \Big|\widetilde{m}_j^B(x_j) + \frac{1}{T}\sum_{t=1}^{T} m_j(X_t^j) - m_j(x_j)\Big| = O_p(h^2) \tag{A.11}$$
$$\sup_{x_j \in I_h^c} \Big|\widetilde{m}_j^B(x_j) + \frac{1}{T}\sum_{t=1}^{T} m_j(X_t^j) - m_j(x_j)\Big| = O_p(h) \tag{A.12}$$

for $j = 0, \dots, d$.

Proof. The result follows from Theorem 3 in Mammen et al. Mammen et al. 1999. To make sure that the latter theorem applies in our case, we have to show that the high-order conditions (A1) – (A5), (A8), and (A9) from Mammen et al. 1999 are fulfilled in our setting.[1] This can be achieved by using the results from Appendix A.3, in particular the expansion of $\widehat{m}_j^B$ given in Lemma A.3.3, and by following the arguments for the proof of Theorem 4 in Mammen et al. 1999. To see that (A.9) – (A.10) have to be replaced by (A.11) – (A.12) in the undersmoothing case with $h = O(T^{-(\frac{1}{4}+\delta)})$, note that

$$\int_0^1 \alpha_{T,j}(x_j)\widehat{p}_j(x_j)dx_j = \frac{1}{T}\sum_{t=1}^{T} m_j(X_t^j) + O_p(h^2)$$

with $\frac{1}{T}\sum_{t=1}^{T} m_j(X_t^j) = O_p(\frac{1}{\sqrt{T}})$, where $\alpha_{T,j}(x_j)$ is defined in Lemma A.3.3. Using this in the proof of Theorem 3 of Mammen et al. 1999 instead of $\int_0^1 \alpha_{T,j}(x_j)\widehat{p}_j(x_j)dx_j = \gamma_{T,j} + o_p(h^2)$ with $\gamma_{T,j} = O(h^2)$ gives (A.11) – (A.12). □

By combining Theorems A.1.1 and A.1.2, it is now straightforward to complete the proof of Theorem 2.4.2. □

A.2 Proofs of Theorems 2.4.3 and 2.4.4

This appendix contains the proofs of Theorems 2.4.3 and 2.4.4, which show consistency and asymptotic normality of the AR estimates. By far the most difficult part is the proof of asymptotic normality. After giving some auxiliary results and proving consistency, we run through the main steps of the normality proof postponing the major technical difficulties to a series of lemmas. The main challenge of the proof is to derive a stochastic expansion of $\frac{1}{\sqrt{T}}\frac{\partial \widetilde{l}_T(\phi^*)}{\partial \phi}$. This expansion is given in Lemmas A.2.1 – A.2.4. Note that as in Appendix A.1, C denotes a finite real constant which may take a different value on each occurrence.

[1] Note that (A6) is not needed for the proof of Theorem 3 as opposed to the statement in Mammen et al. 1999.

Auxiliary Results

Before we come to the proofs, we list some simple facts that are frequently used throughout this section. For ease of notation, we work with the likelihood functions

$$l_T(\phi) = -\sum_{t=1}^{T} \big(\varepsilon_t - \varepsilon_t(\phi)\big)^2$$
$$\widetilde{l}_T(\phi) = -\sum_{t=1}^{T} \big(\widetilde{\varepsilon}_t - \widetilde{\varepsilon}_t(\phi)\big)^2,$$

where $\varepsilon_t(\phi) = \sum_{i=1}^{p} \phi_i \varepsilon_{t-i}$ and $\widetilde{\varepsilon}_t(\phi) = \sum_{i=1}^{p} \phi_i \widetilde{\varepsilon}_{t-i}$. These differ from the functions defined in (2.17) and (2.19) only in that the sum over t starts at the time point $t = 1$ rather than at $t = p+1$. Trivially, the error resulting from this modification can be neglected in the proofs.

To bound the distance between l_T and $\widetilde{l}_T$, the following facts are useful: From the convergence results on the estimates $\widetilde{m}_\theta, \widetilde{m}_0, \ldots, \widetilde{m}_d$, it is easily seen that

$$\max_{t=1,\ldots,T} |\varepsilon_t - \widetilde{\varepsilon}_t| = O_p(h). \tag{R1}$$

Using (R1), we can immediately infer that

$$\max_{t=1,\ldots,T} \sup_{\phi\in\Phi} |\varepsilon_t(\phi) - \widetilde{\varepsilon}_t(\phi)| = O_p(h). \tag{R2}$$

Moreover, noting that $\frac{\partial \varepsilon_t(\phi)}{\partial \phi_i} = \varepsilon_{t-i}$ and analogously $\frac{\partial \widetilde{\varepsilon}_t(\phi)}{\partial \phi_i} = \widetilde{\varepsilon}_{t-i}$, we get

$$\max_{t=1,\ldots,T} \sup_{\phi\in\Phi} \Big| \frac{\partial \varepsilon_t(\phi)}{\partial \phi_i} - \frac{\partial \widetilde{\varepsilon}_t(\phi)}{\partial \phi_i} \Big| = O_p(h). \tag{R3}$$

Proof of Theorem 2.4.3

Let $l_T(\phi)$ and $\widetilde{l}_T(\phi)$ be the likelihood functions introduced in the previous subsection. We show that

$$\sup_{\phi\in\Phi} \Big| \frac{1}{T}\widetilde{l}_T(\phi) - \frac{1}{T} l_T(\phi) \Big| = o_p(1). \tag{A.13}$$

This together with standard arguments yields consistency of $\widetilde{\phi}$. In order to prove (A.13), we decompose $\frac{1}{T}\widetilde{l}_T(\phi) - \frac{1}{T}l_T(\phi)$ into

$$\frac{1}{T}\widetilde{l}_T(\phi) - \frac{1}{T}l_T(\phi) = \frac{1}{T}\sum_{t=1}^{T}\big(\varepsilon_t^2 - \widetilde{\varepsilon}_t^2\big) + \frac{2}{T}\sum_{t=1}^{T}\big(\widetilde{\varepsilon}_t - \varepsilon_t\big)\widetilde{\varepsilon}_t(\phi) + \frac{2}{T}\sum_{t=1}^{T}\varepsilon_t\big(\widetilde{\varepsilon}_t(\phi) - \varepsilon_t(\phi)\big) + \frac{1}{T}\sum_{t=1}^{T}\big(\varepsilon_t^2(\phi) - \widetilde{\varepsilon}_t^2(\phi)\big).$$

Using (R1) – (R3), it is straightforward to show that the four terms on the right-hand side of the above equation are all $o_p(1)$ uniformly in ϕ. This shows (A.13). □

Proof of Theorem 2.4.4

By the usual Taylor expansion argument, we obtain

$$0 = \frac{1}{T}\frac{\partial\widetilde{l}_T(\widetilde{\phi})}{\partial\phi} = \frac{1}{T}\frac{\partial\widetilde{l}_T(\phi^*)}{\partial\phi} + \frac{1}{T}\frac{\partial^2\widetilde{l}_T(\bar{\phi})}{\partial\phi\partial\phi^T}(\widetilde{\phi} - \phi^*)$$

with some intermediate point $\bar{\phi}$ between ϕ^* and $\widetilde{\phi}$. Rearranging and premultiplying by $\sqrt{T}$ yields

$$\sqrt{T}(\widetilde{\phi} - \phi^*) = -\Big(\frac{1}{T}\frac{\partial^2\widetilde{l}_T(\bar{\phi})}{\partial\phi\partial\phi^T}\Big)^{-1}\frac{1}{\sqrt{T}}\frac{\partial\widetilde{l}_T(\phi^*)}{\partial\phi}.$$

In what follows, we show that

$$\frac{1}{T}\frac{\partial^2\widetilde{l}_T(\bar{\phi})}{\partial\phi\partial\phi^T} \xrightarrow{P} H \tag{A.14}$$

$$\frac{1}{\sqrt{T}}\frac{\partial\widetilde{l}_T(\phi^*)}{\partial\phi} \xrightarrow{d} N(0, \Psi) \tag{A.15}$$

with $\Psi = 4W + 4\Omega$ and $H = -2\Gamma_p$, where Γ_p is the autocovariance matrix of the AR process $\{\varepsilon_t\}$, $W = (\mathbb{E}[\eta_0^2\varepsilon_{-i}\varepsilon_{-j}])_{i,j=1,\ldots,p}$ and Ω is given in (A.24). This completes the proof.

Proof of (A.14). By straightforward calculations it can be seen that

$$\sup_{\phi\in\Phi}\Big|\frac{1}{T}\frac{\partial^2\widetilde{l}_T(\phi)}{\partial\phi\partial\phi^T} - \frac{1}{T}\frac{\partial^2 l_T(\phi)}{\partial\phi\partial\phi^T}\Big| = o_p(1)$$

and $\frac{1}{T}\frac{\partial^2 l_T(\bar{\phi})}{\partial\phi\partial\phi^T} \xrightarrow{P} H$. This yields $(A.14)$. □

Proof of (A.15). We write

$$\frac{1}{\sqrt{T}}\frac{\partial \widetilde{l}_T(\phi^*)}{\partial \phi_i} = \frac{1}{\sqrt{T}}\frac{\partial l_T(\phi^*)}{\partial \phi_i} + \Big(\frac{1}{\sqrt{T}}\frac{\partial \widetilde{l}_T(\phi^*)}{\partial \phi_i} - \frac{1}{\sqrt{T}}\frac{\partial l_T(\phi^*)}{\partial \phi_i}\Big).$$

Introducing the notation $\phi_0^* = -1$, we obtain that

$$\begin{aligned}\frac{1}{\sqrt{T}}\frac{\partial \widetilde{l}_T(\phi^*)}{\partial \phi_i} - \frac{1}{\sqrt{T}}\frac{\partial l_T(\phi^*)}{\partial \phi_i} &= \sum_{k=0}^{p} 2\phi_k^* \Big(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\varepsilon_{t-k} - \widetilde{\varepsilon}_{t-k})\varepsilon_{t-i}\Big) \\ &\quad + \sum_{k=0}^{p} 2\phi_k^* \Big(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\varepsilon_{t-i} - \widetilde{\varepsilon}_{t-i})\widetilde{\varepsilon}_{t-k}\Big) \\ &= \sum_{k=0}^{p} 2\phi_k^* \Big(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\varepsilon_{t-k} - \widetilde{\varepsilon}_{t-k})\varepsilon_{t-i}\Big) \\ &\quad + \sum_{k=0}^{p} 2\phi_k^* \Big(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\varepsilon_{t-i} - \widetilde{\varepsilon}_{t-i})\varepsilon_{t-k}\Big) + o_p(1),\end{aligned} \tag{A.16}$$

where the last equality follows from the fact that $(\varepsilon_{t-i} - \widetilde{\varepsilon}_{t-i})(\widetilde{\varepsilon}_{t-k} - \varepsilon_{t-k}) = O_p(h^2) = o_p(\sqrt{T})$ uniformly in t, k, and i by (R1). In what follows, we derive a stochastic expansion of the terms

$$Q_T = Q_T^{[k,i]} := \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\varepsilon_{t-k} - \widetilde{\varepsilon}_{t-k})\varepsilon_{t-i}.$$

By symmetry this also gives us an expansion for $Q_T^{[i,k]}$ and thus by (A.16) also for the difference $\frac{1}{\sqrt{T}}\frac{\partial \widetilde{l}_T(\phi^*)}{\partial \phi_i} - \frac{1}{\sqrt{T}}\frac{\partial l_T(\phi^*)}{\partial \phi_i}$.

Introducing the shorthand $X_t^0 = \frac{t}{T}$, we have

$$\varepsilon_t - \widetilde{\varepsilon}_t = \big(\widetilde{m}_\theta(t) - m_\theta(t)\big) + \sum_{j=0}^{d}\big(\widetilde{m}_j(X_t^j) - m_j(X_t^j)\big).$$

From Appendix A.1, we know that the backfitting estimates $\widetilde{m}_j(x_j)$ can be decomposed into a stochastic part $\widetilde{m}_j^A(x_j)$ and a bias part $\widetilde{m}_j^B(x_j)$. This allows us to rewrite the term Q_T as

$$Q_T = Q_{T,\theta} + \sum_{j=0}^{d} Q_{T,V,j} + \sum_{j=0}^{d} Q_{T,B,j} \tag{A.17}$$

with

$$\begin{aligned}
Q_{T,\theta} &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_{t-i} \Big[\widetilde{m}_\theta(t-k) - m_\theta(t-k) - \sum_{j=0}^{d} \frac{1}{T} \sum_{s=1}^{T} m_j(X_s^j) \Big] \\
Q_{T,V,j} &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_{t-i} \widetilde{m}_j^A(X_{t-k}^j) \\
Q_{T,B,j} &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_{t-i} \Big[\widetilde{m}_j^B(X_{t-k}^j) + \frac{1}{T} \sum_{s=1}^{T} m_j(X_s^j) - m_j(X_{t-k}^j) \Big]
\end{aligned}$$

for $j = 0, \ldots, d$. In Lemmas A.2.3 and A.2.4, we will show that

$$Q_{T,\theta} = o_p(1) \tag{A.18}$$

$$Q_{T,B,j} = o_p(1) \quad \text{for } j = 0, \ldots, d. \tag{A.19}$$

Moreover, Lemmas A.2.1 and A.2.2 establish that

$$Q_{T,V,0} = o_p(1) \tag{A.20}$$

$$Q_{T,V,j} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} g_j\Big(\frac{t}{T}, X_t\Big) \varepsilon_t + o_p(1) \quad \text{for } j = 1, \ldots, d, \tag{A.21}$$

where $g_j = g_j^{[k,i]}$ are deterministic functions whose exact forms are given in the statement of Lemma A.2.1. These functions are easily seen to be absolutely bounded by a constant independent of T. Inserting the above results in (A.17), we obtain

$$Q_T = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Big[\sum_{j=1}^{d} g_j\Big(\frac{t}{T}, X_t\Big) \Big] \varepsilon_t + o_p(1).$$

Using this together with (A.16) now yields

$$\frac{1}{\sqrt{T}} \frac{\partial \widetilde{l}_T(\phi^*)}{\partial \phi_i} - \frac{1}{\sqrt{T}} \frac{\partial l_T(\phi^*)}{\partial \phi_i} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} h_i\Big(\frac{t}{T}, X_t\Big) \varepsilon_t + o_p(1) \tag{A.22}$$

with the absolutely bounded function

$$h_i\Big(\frac{t}{T}, X_t\Big) = \sum_{j=1}^{d} \sum_{k=0}^{p} 2\phi_k^* \Big[g_j^{[k,i]}\Big(\frac{t}{T}, X_t\Big) + g_j^{[i,k]}\Big(\frac{t}{T}, X_t\Big) \Big], \tag{A.23}$$

where we suppress the dependence of h_i on the parameter vector ϕ^* in the notation. As a result,

$$\begin{aligned}
\frac{1}{\sqrt{T}} \frac{\partial \widetilde{l}_T(\phi^*)}{\partial \phi_i} &= \frac{1}{\sqrt{T}} \frac{\partial l_T(\phi^*)}{\partial \phi_i} + \frac{1}{\sqrt{T}} \sum_{t=1}^{T} h_i\Big(\frac{t}{T}, X_t\Big) \varepsilon_t + o_p(1) \\
&= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Big[2\eta_t \varepsilon_{t-i} + h_i\Big(\frac{t}{T}, X_t\Big) \varepsilon_t \Big] + o_p(1) \\
&=: \frac{1}{\sqrt{T}} \sum_{t=1}^{T} U_{t,T} + o_p(1),
\end{aligned}$$

i.e. the term of interest can be written as a normalized sum of random variables $U_{t,T}$ plus a term which is asymptotically negligible. Using the mixing assumptions in (A1), it is straightforward to see that the variables $\{U_{t,T}, t = 1, \ldots, T\}$ form an α-mixing array with mixing coefficients that decay exponentially fast to zero. We can thus apply a central limit theorem for mixing arrays to obtain that

$$\frac{1}{\sqrt{T}} \frac{\partial \widetilde{l}_T(\phi^*)}{\partial \phi_i} \xrightarrow{d} N(0, \psi_{ii})$$

with $\psi_{ii} = \lim_{T\to\infty} \mathbb{E}(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} U_{t,T})^2$. Using the Cramer-Wold device, it is now easy to show that

$$\frac{1}{\sqrt{T}} \frac{\partial \widetilde{l}_T(\phi^*)}{\partial \phi} \xrightarrow{d} N(0, \Psi)$$

with $\Psi = (\psi_{ij})_{i,j=1,\ldots,p}$, where $\Psi = 4W + 4\Omega$ and $\Omega = (\omega_{ij})_{i,j=1,\ldots,p}$ with

$$\begin{aligned}
\omega_{ij} = \frac{1}{2} \sum_{l=-\infty}^{\infty} \mathbb{E}\Big[\eta_0 \varepsilon_{-i} \varepsilon_l \int_0^1 h_j(u, X_l) du \Big] + \frac{1}{2} \sum_{l=-\infty}^{\infty} \mathbb{E}\Big[\eta_0 \varepsilon_{-j} \varepsilon_l \int_0^1 h_i(u, X_l) du \Big] \\
+ \frac{1}{4} \sum_{l=-\infty}^{\infty} \mathbb{E}\Big[\varepsilon_0 \varepsilon_l \int_0^1 h_i(u, X_0) h_j(u, X_l) du \Big].
\end{aligned} \tag{A.24}$$

□

In order to complete the proof of asymptotic normality, we still need to show that equations (A.18) – (A.21) are fulfilled for the terms $Q_{T,\theta}$, $Q_{T,V,j}$, and $Q_{T,B,j}$. We begin with the expansion of the variance components $Q_{T,V,j}$ for $j = 1, \dots, d$, as this is the technically most interesting part.

Lemma A.2.1. *It holds that*

$$Q_{T,V,j} = \frac{1}{\sqrt{T}} \sum_{s=1}^{T} g_j\Big(\frac{s}{T}, X_s\Big)\varepsilon_s + o_p(1)$$

for $j = 1, \dots, d$. The functions g_j are given by

$$g_j\Big(\frac{s}{T}, X_s\Big) = g_j^{NW}(X_s^j) + g_j^{SBF}\Big(\frac{s}{T}, X_s\Big)$$

with

$$g_j^{NW}(X_s^j) = \mathbb{E}_{-s}\Big[\frac{K_h(X_{-k}^j, X_s^j)\varepsilon_{-i}}{\int_0^1 K_h(X_{-k}^j, w)dw\ p_j(X_{-k}^j)}\Big]$$
$$g_j^{SBF}\Big(\frac{s}{T}, X_s\Big) = \mathbb{E}_{-s}[r_{j,s}(X_{-k}^j)\varepsilon_{-i}],$$

where $\mathbb{E}_{-s}[\,\cdot\,]$ is the expectation with respect to all variables except for those depending on the index s and the functions $r_{j,s}(\cdot) = r_j(\frac{s}{T}, X_s, \cdot)$ are defined in Theorem A.1.1 of Appendix A.1.

Proof. By Theorem A.1.1, the stochastic part $\widetilde{m}_j^A$ of the smooth backfitting estimate $\widetilde{m}_j$ has the expansion

$$\widetilde{m}_j^A(x_j) = \widehat{m}_j^A(x_j) + \frac{1}{T}\sum_{s=1}^{T} r_{j,s}(x_j)\varepsilon_s + o_p\Big(\frac{1}{\sqrt{T}}\Big)$$

uniformly in x_j, where $\widehat{m}_j^A$ is the stochastic part of the Nadaraya-Watson pilot estimate and $r_{j,s}(\cdot) = r_j(\frac{s}{T}, X_s, \cdot)$ is Lipschitz continuous and absolutely bounded. With this result, we can decompose $Q_{T,V,j}$ as follows:

$$\begin{aligned} Q_{T,V,j} &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{t-i}\widehat{m}_j^A(X_{t-k}^j) + \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{t-i}\Big[\frac{1}{T}\sum_{s=1}^{T} r_{j,s}(X_{t-k}^j)\varepsilon_s\Big] + o_p(1) \\ &=: Q_{T,V,j}^{NW} + Q_{T,V,j}^{SBF} + o_p(1). \end{aligned}$$

In the following, we will give the arguments needed to treat $Q^{NW}_{T,V,j}$. The line of argument for $Q^{SBF}_{T,V,j}$ is essentially identical although some of the steps are easier due to the properties of the $r_{j,s}$ functions.

Plugging the definition (A.2) of the estimate $\widehat{m}^A_j(x_j)$ into the term $Q^{NW}_{T,V,j}$, we get

$$Q^{NW}_{T,V,j} = \frac{1}{\sqrt{T}} \sum_{s=1}^{T} \Big(\frac{1}{T} \sum_{t=1}^{T} \frac{K_h(X^j_{t-k}, X^j_s)}{\frac{1}{T}\sum_{v=1}^{T} K_h(X^j_{t-k}, X^j_v)} \varepsilon_{t-i} \Big) \varepsilon_s. \tag{A.25}$$

In a first step, we show that

$$Q^{NW}_{T,V,j} = \frac{1}{\sqrt{T}} \sum_{s=1}^{T} \Big(\frac{1}{T} \sum_{t=1}^{T} K_h(X^j_{t-k}, X^j_s) \mu_t \Big) \varepsilon_s + o_p(1), \tag{A.26}$$

where $\mu_t := q_j^{-1}(X^j_{t-k})\varepsilon_{t-i}$ with $q_j(x_j) = \int_0^1 K_h(x_j, w)dw\, p_j(x_j)$. To do so, decompose $\frac{1}{T}\sum_{v=1}^{T} K_h(x_j, X^j_v)$ as $\frac{1}{T}\sum_{v=1}^{T} K_h(x_j, X^j_v) = q_j(x_j) + B_j(x_j) + V_j(x_j)$ with

$$B_j(x_j) = \frac{1}{T} \sum_{v=1}^{T} \mathbb{E}[K_h(x_j, X^j_v)] - q_j(x_j)$$
$$V_j(x_j) = \frac{1}{T} \sum_{v=1}^{T} \big(K_h(x_j, X^j_v) - \mathbb{E}[K_h(x_j, X^j_v)] \big).$$

Notice that $\sup_{x_j \in [0,1]} |B_j(x_j)| = O_p(h)$ and $\sup_{x_j \in [0,1]} |V_j(x_j)| = O_p(\sqrt{\log T / Th})$. Using a second order Taylor expansion of $f(z) = (1+z)^{-1}$ we arrive at

$$\frac{1}{\frac{1}{T}\sum_{v=1}^{T} K_h(x_j, X^j_v)} = \frac{1}{q_j(x_j)} \Big(1 + \frac{B_j(x_j) + V_j(x_j)}{q_j(x_j)} \Big)^{-1}$$
$$= \frac{1}{q_j(x_j)} \Big(1 - \frac{B_j(x_j) + V_j(x_j)}{q_j(x_j)} + O_p(h^2) \Big)$$

uniformly in x_j. Plugging this decomposition into $(A.25)$, we obtain

$$Q^{NW}_{T,V,j} = \frac{1}{\sqrt{T}} \sum_{s=1}^{T} \frac{1}{T} \sum_{t=1}^{T} \frac{K_h(X^j_{t-k}, X^j_s)}{q_j(X^j_{t-k})} \varepsilon_{t-i} \varepsilon_s - Q^{NW,B}_{T,V,j} - Q^{NW,V}_{T,V,j} + o_p(1)$$

with

$$Q^{NW,B}_{T,V,j} = \frac{1}{\sqrt{T}} \sum_{s=1}^{T} \frac{1}{T} \sum_{t=1}^{T} K_h(X^j_{t-k}, X^j_s) \frac{B_j(X^j_{t-k})}{q^2_j(X^j_{t-k})} \varepsilon_{t-i}\varepsilon_s$$
$$Q^{NW,V}_{T,V,j} = \frac{1}{\sqrt{T}} \sum_{s=1}^{T} \frac{1}{T} \sum_{t=1}^{T} K_h(X^j_{t-k}, X^j_s) \frac{V_j(X^j_{t-k})}{q^2_j(X^j_{t-k})} \varepsilon_{t-i}\varepsilon_s.$$

All that is required to establish (A.26) is to show that both $Q^{NW,B}_{T,V,j}$ and $Q^{NW,V}_{T,V,j}$ are $o_p(1)$. As $\sup_{x_j \in I_h} |B_j(x_j)| = O_p(h^2)$ and $\sup_{x_j \in I_h^c} |B_j(x_j)| = O_p(h)$, we can use Markov's inequality together with (A9) to get that $Q^{NW,B}_{T,V,j} = o_p(1)$. In order to show that $Q^{NW,V}_{T,V,j} = o_p(1)$, let $\mathbb{E}_v[\cdot]$ denote the expectation with respect to the variables indexed by v. Then

$$\begin{aligned}
\big|Q^{NW,V}_{T,V,j}\big| &= \Big| \frac{1}{\sqrt{T}} \sum_{s=1}^{T} \frac{1}{T} \sum_{t=1}^{T} \frac{K_h(X^j_{t-k}, X^j_s)}{q^2_j(X^j_{t-k})} \varepsilon_{t-i} \\
&\qquad \times \Big(\frac{1}{T} \sum_{v=1}^{T} (K_h(X^j_{t-k}, X^j_v) - \mathbb{E}_v[K_h(X^j_{t-k}, X^j_v)]) \Big) \varepsilon_s \Big| \\
&\le \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{|\varepsilon_{t-i}|}{q^2_j(X^j_{t-k})} \sup_{x_j \in [0,1]} \Big| \frac{1}{T} \sum_{s=1}^{T} K_h(x_j, X^j_s)\varepsilon_s \Big| \\
&\qquad \times \sup_{x_j \in [0,1]} \Big| \frac{1}{T} \sum_{v=1}^{T} (K_h(x_j, X^j_v) - \mathbb{E}_v[K_h(x_j, X^j_v)]) \Big| \\
&= O_p\Big(\frac{\log T}{Th}\Big) \Big(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{|\varepsilon_{t-i}|}{q^2_j(X^j_{t-k})} \Big) = O_p\Big(\frac{\log T}{Th}\sqrt{T}\Big) = o_p(1),
\end{aligned}$$

as $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} |\varepsilon_{t-i}| \; q_j^{-2}(X^j_{t-k}) = O_p(\sqrt{T})$ by Markov's inequality.

In the next step, we replace the inner sum over t in (A.26) by a deterministic function that only depends on X^j_s and show that the resulting error can be asymptotically neglected. Define

$$\psi_{t,s} = K_h(X^j_{t-k}, X^j_s)\mu_t - \mathbb{E}_{-s}[K_h(X^j_{t-k}, X^j_s)\mu_t],$$

where $\mathbb{E}_{-s}[\cdot]$ is the expectation with respect to all variables except for those depending on the index s. With the above notation at hand, we can rewrite (A.26) as

$$Q^{NW}_{T,V,j} = \frac{1}{\sqrt{T}} \sum_{s=1}^{T} \Big(\frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{-s}[K_h(X^j_{t-k}, X^j_s)\mu_t] \Big) \varepsilon_s + R^{NW}_{T,V,j} + o_p(1),$$

where

$$R^{NW}_{T,V,j} = \frac{1}{\sqrt{T}} \sum_{s=1}^{T} \frac{1}{T} \sum_{t=1}^{T} \psi_{t,s}\varepsilon_s. \tag{A.27}$$

Once we show that $R^{NW}_{T,V,j} = o_p(1)$, we are left with

$$\begin{aligned} Q^{NW}_{T,V,j} &= \frac{1}{\sqrt{T}} \sum_{s=1}^{T} \Big(\frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{-s}[K_h(X^j_{t-k}, X^j_s)\mu_t] \Big) \varepsilon_s + o_p(1) \\ &= \frac{1}{\sqrt{T}} \sum_{s=1}^{T} \mathbb{E}_{-s}[K_h(X^j_{-k}, X^j_s)\mu_0]\varepsilon_s + o_p(1) \\ &=: \frac{1}{\sqrt{T}} \sum_{s=1}^{T} g^{NW}_j(X^j_s)\varepsilon_s + o_p(1) \end{aligned}$$

with $\mu_0 = q_j^{-1}(X^j_{-k})\varepsilon_{-i}$ and $q_j(X^j_{-k}) = \int_0^1 K_h(X^j_{-k}, w)dw\ p_j(X^j_{-k})$.

Thus it remains to prove that $R^{NW}_{T,V,j} = o_p(1)$. To do so, define

$$P := \mathbb{P}\Big(\Big| \frac{1}{\sqrt{T}} \sum_{s=1}^{T} \frac{1}{T} \sum_{t=1}^{T} \psi_{t,s}\varepsilon_s \Big| > \delta \Big)$$

for a fixed $\delta > 0$. Then by Chebychev's inequality

$$\begin{aligned} P &\le \frac{1}{T^3\delta^2} \sum_{s,s'=1}^{T} \sum_{t,t'=1}^{T} \mathbb{E}\Big[\psi_{t,s}\varepsilon_s\psi_{t',s'}\varepsilon_{s'} \Big] \\ &= \frac{1}{T^3\delta^2} \sum_{(s,s',t,t')\in S} \mathbb{E}\Big[\psi_{t,s}\varepsilon_s\psi_{t',s'}\varepsilon_{s'} \Big] + \frac{1}{T^3\delta^2} \sum_{(s,s',t,t')\in S^c} \mathbb{E}\Big[\psi_{t,s}\varepsilon_s\psi_{t',s'}\varepsilon_{s'} \Big] \\ &=: P_S + P_{S^c}, \end{aligned}$$

where S is the set of tuples (s, s', t, t') with $1 \le s, s', t, t' \le T$ such that (at least) one index is separated from the others and S^c is its complement. We say that an index, for instance t, is separated from the others if $\min\{|t - t'|, |t - s|, |t - s'|\} > C_2 \log T$, i.e. if it is further away from the other indices than $C_2 \log T$ for a constant C_2 to be chosen later on. We now analyse P_S and P_{S^c} separately.

(a) First consider P_{S^c}. If a tuple (s,s',t,t') is an element of S^c, then no index is separated from the others. Since the index t is not separated, there exists an index, say t', such that $|t-t'| \leq C_2 \log T$. Now take an index different from t and t', for instance s. Then by the same argument, there exists an index, say s', such that $|s-s'| \leq C_2 \log T$. As a consequence, the number of tuples $(s,s',t,t') \in S^c$ is smaller than $CT^2(\log T)^2$ for some constant C. Using (A8), this suffices to infer that

$$\big|P_{S^c}\big| \leq \frac{1}{T^3\delta^2} \sum_{(s,s',t,t')\in S^c} \frac{C}{h^2} \leq \frac{C}{\delta^2}\frac{(\log T)^2}{Th^2} \to 0.$$

(b) The term P_S is more difficult to handle. First note that S can be written as the union of the disjoint sets

$$S_1 = \{(s,s',t,t') \in S \mid \text{the index } t \text{ is separated}\}$$
$$S_2 = \{(s,s',t,t') \in S \mid (s,s',t,t') \notin S_1 \text{ and the index } s \text{ is separated}\}$$
$$S_3 = \{(s,s',t,t') \in S \mid (s,s',t,t') \notin S_1 \cup S_2 \text{ and the index } t' \text{ is separated}\}$$
$$S_4 = \{(s,s',t,t') \in S \mid (s,s',t,t') \notin S_1 \cup S_2 \cup S_3 \text{ and the index } s' \text{ is separated}\}.$$

Thus, $P_S = P_{S_1} + P_{S_2} + P_{S_3} + P_{S_4}$ with

$$P_{S_r} = \frac{1}{T^3\delta^2} \sum_{(s,s',t,t')\in S_r} \mathbb{E}\Big[\psi_{t,s}\varepsilon_s \psi_{t',s'}\varepsilon_{s'}\Big].$$

for $r=1,\dots,4$. In what follows, we show that $P_{S_r} \to 0$ for $r=1,\dots,4$. As the four terms can be treated in exactly the same way, we restrict attention to the analysis of P_{S_1}.

We start by taking a cover $\{I_m\}_{m=1}^{M_T}$ of the compact support $[0,1]$ of X^j_{t-k}. The elements I_m are intervals of length $1/M_T$ given by $I_m = [\frac{m-1}{M_T}, \frac{m}{M_T})$ for $m=1,\dots,M_T-1$ and $I_{M_T} = [1-\frac{1}{M_T}, 1]$. The midpoint of the interval I_m is denoted by x_m. With this, we can write

$$K_h(X^j_{t-k}, X^j_s) = \sum_{m=1}^{M_T} I(X^j_{t-k} \in I_m) \times \big[K_h(x_m, X^j_s) + (K_h(X^j_{t-k}, X^j_s) - K_h(x_m, X^j_s))\big]. \tag{A.28}$$

Using (A.28), we can further write

$$\begin{aligned}\psi_{t,s} &= \sum_{m=1}^{M_T} \Big\{ I(X^j_{t-k} \in I_m) K_h(x_m, X^j_s)\mu_t \\ &\qquad - \mathbb{E}_{-s}[I(X^j_{t-k} \in I_m) K_h(x_m, X^j_s)\mu_t] \Big\} \\ &\quad + \sum_{m=1}^{M_T} \Big\{ I(X^j_{t-k} \in I_m)(K_h(X^j_{t-k}, X^j_s) - K_h(x_m, X^j_s))\mu_t \\ &\qquad - \mathbb{E}_{-s}[I(X^j_{t-k} \in I_m)(K_h(X^j_{t-k}, X^j_s) - K_h(x_m, X^j_s))\mu_t] \Big\} \\ &=: \psi^A_{t,s} + \psi^B_{t,s}\end{aligned}$$

and

$$\begin{aligned}P_{S_1} &= \frac{1}{T^3\delta^2} \sum_{(s,s',t,t') \in S_1} \mathbb{E}\big[\psi^A_{t,s}\varepsilon_s \psi_{t',s'}\varepsilon_{s'}\big] + \frac{1}{T^3\delta^2} \sum_{(s,s',t,t') \in S_1} \mathbb{E}\big[\psi^B_{t,s}\varepsilon_s \psi_{t',s'}\varepsilon_{s'}\big] \\ &=: P^A_{S_1} + P^B_{S_1}.\end{aligned}$$

We first consider $P^B_{S_1}$. Set $M_T = CT(\log T)h^{-3}$ and exploit the Lipschitz continuity of the kernel K to get that $|K_h(X^j_{t-k}, X^j_s) - K_h(x_m, X^j_s)| \le \frac{C}{h^2}|X^j_{t-k} - x_m|$. This gives us

$$\begin{aligned}\big|\psi^B_{t,s}\big| &\le \frac{C}{h^2} \sum_{m=1}^{M_T} \Big(\underbrace{I(X^j_{t-k} \in I_m)|X^j_{t-k} - x_m|}_{\le I(X^j_{t-k} \in I_m) M_T^{-1}} |\mu_t| \\ &\qquad + \mathbb{E}\big[\underbrace{I(X^j_{t-k} \in I_m)|X^j_{t-k} - x_m|}_{\le I(X^j_{t-k} \in I_m) M_T^{-1}} |\mu_t| \big] \Big) \le \frac{C}{M_T h^2} \big(|\mu_t| + \mathbb{E}|\mu_t| \big).\end{aligned}$$

Plugging this into the expression for $P^B_{S_1}$, we arrive at

$$\big|P^B_{S_1}\big| \le \frac{1}{T^3\delta^2} \frac{C}{M_T h^2} \sum_{(s,s',t,t') \in S_1} \underbrace{\mathbb{E}\big[(|\mu_t| + \mathbb{E}|\mu_t|)|\varepsilon_s \psi_{t',s'}\varepsilon_{s'}|\big]}_{\le Ch^{-1}} \le \frac{C}{\delta^2 \log T} \to 0.$$

We next turn to $P^A_{S_1}$. Write

$$P^A_{S_1} = \frac{1}{T^3\delta^2} \sum_{(s,s',t,t') \in S_1} \Big(\sum_{m=1}^{M_T} \gamma_m \Big)$$

with

$$\gamma_m = \mathbb{E}\Big[\big\{I(X^j_{t-k} \in I_m)K_h(x_m, X^j_s)\mu_t - \mathbb{E}_{-s}[I(X^j_{t-k} \in I_m)K_h(x_m, X^j_s)\mu_t]\big\}\varepsilon_s\psi_{t',s'}\varepsilon_{s'}\Big].$$

By Davydov's inequality, it holds that

$$\begin{aligned}\gamma_m &= \mathrm{Cov}\Big(I(X^j_{t-k} \in I_m)\mu_t - \mathbb{E}[I(X^j_{t-k} \in I_m)\mu_t], K_h(x_m, X^j_s)\varepsilon_s\psi_{t',s'}\varepsilon_{s'}\Big) \\ &\leq \frac{C}{h^2}\big(\alpha(C_2 \log T)\big)^{1-\frac{1}{q}-\frac{1}{r}} \leq \frac{C}{h^2}\big(a^{C_2 \log T}\big)^{1-\frac{1}{q}-\frac{1}{r}} \leq \frac{C}{h^2}T^{-C_3}\end{aligned}$$

with some $C_3 > 0$, where q and r are chosen slightly larger than $\frac{4}{3}$ and 4, respectively. Note that we can make C_3 arbitrarily large by choosing C_2 large enough. From this, it is easily seen that $P^A_{S_1} \to 0$.

Combining (a) and (b) yields that $P \to 0$ for each fixed $\delta > 0$. As a result,

$$R^{NW,V}_{T,V,j} = o_p(1),$$

which completes the proof for the term $Q^{NW}_{T,V,j}$. As stated at the beginning of the proof, exactly the same arguments can be used to analyze the term $Q^{SBF}_{T,V,j}$. □

Lemma A.2.2. *It holds that*

$$Q_{T,V,0} = o_p(1).$$

Proof. As in Lemma A.2.1, we can write

$$\begin{aligned}Q_{T,V,0} &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{t-i}\widehat{m}^A_0\Big(\frac{t-k}{T}\Big) + \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{t-i}\Big[\frac{1}{T}\sum_{s=1}^{T}r_{0,s}\Big(\frac{t-k}{T}\Big)\varepsilon_s\Big] + o_p(1) \\ &=: Q^{NW}_{T,V,0} + Q^{SBF}_{T,V,0} + o_p(1).\end{aligned}$$

We again restrict attention to the arguments for $Q^{NW}_{T,V,0}$, those for $Q^{SBF}_{T,V,0}$ being essentially the same. Plugging the definition of $\widehat{m}^A_0(x_0)$ into the term $Q^{NW}_{T,V,0}$ yields

$$Q^{NW}_{T,V,0} = \frac{1}{\sqrt{T}}\sum_{s=1}^{T}\frac{1}{T}\sum_{t=1}^{T}w_{t,s}\varepsilon_{t-i}\varepsilon_s$$

with $w_{t,s} = K_h(\frac{t-k}{T}, \frac{s}{T})/\frac{1}{T}\sum_{v=1}^{T} K_h(\frac{t-k}{T}, \frac{v}{T})$. Now let $\{\rho_T\}$ be some sequence that slowly converges to zero, e.g. $\rho_T = (\log T)^{-1}$. By Chebychev's inequality,

$$\mathbb{P}\big(|Q_{T,V,0}^{NW}| > C\rho_T\big) \leq C\frac{\mathbb{E}(Q_{T,V,0}^{NW})^2}{\rho_T^2}$$

with

$$\mathbb{E}(Q_{T,V,j}^{NW})^2 = \frac{1}{T^3}\sum_{s,s',t,t'=1}^{T} w_{t,s}w_{t',s'}\mathbb{E}[\varepsilon_{t-i}\varepsilon_s\varepsilon_{t'-i}\varepsilon_{s'}].$$

The moments $\mathbb{E}[\varepsilon_{t-i}\varepsilon_s\varepsilon_{t'-i}\varepsilon_{s'}]$ can be written as covariances if one of the indices s, s', t, t' is different from the others. Exploiting our mixing assumptions, these covariances can be bounded by Davydov's inequality. With the help of the resulting bounds, it is straightforward to show that $\mathbb{E}(Q_{T,V,j}^{NW})^2/\rho_T^2$ goes to zero, which in turn yields that $Q_{T,V,j}^{NW} = o_p(1)$. □

Note that the above argument for $Q_{T,V,0}$ is much easier than that for $Q_{T,V,j}$ presented in Lemma A.2.1. The main reason is that the weights $w_{t,s}$ and $w_{t',s'}$ are deterministic allowing us to separate the expectations $\mathbb{E}[\varepsilon_{t-i}\varepsilon_s\varepsilon_{t'-i}\varepsilon_{s'}]$ from the weights. In contrast, in Lemma A.2.1 we have the situation that

$$Q_{T,V,j}^{NW} = \frac{1}{\sqrt{T}}\sum_{s=1}^{T}\frac{1}{T}\sum_{t=1}^{T} w_{t,s}\varepsilon_{t-i}\varepsilon_s$$

with $w_{t,s} = K_h(X_{t-k}^j, X_s^j)/\frac{1}{T}\sum_{v=1}^{T} K_h(X_{t-k}^j, X_v^j)$. In this case,

$$\mathbb{E}(Q_{T,V,j}^{NW})^2 = \frac{1}{T^3}\sum_{s,s',t,t'=1}^{T} \mathbb{E}[w_{t,s}w_{t',s'}\varepsilon_{t-i}\varepsilon_s\varepsilon_{t'-i}\varepsilon_{s'}]. \tag{A.29}$$

If the covariate process $\{X_t\}$ is independent of $\{\varepsilon_t\}$, then $\mathbb{E}[w_{t,s}w_{t',s'}\varepsilon_{t-i}\varepsilon_s\varepsilon_{t'-i}\varepsilon_{s'}] = \mathbb{E}[w_{t,s}w_{t',s'}]\mathbb{E}[\varepsilon_{t-i}\varepsilon_s\varepsilon_{t'-i}\varepsilon_{s'}]$ and similar arguments as those for the term $Q_{T,V,0}^{\mathrm{NW}}$ yield that $Q_{T,V,j}^{\mathrm{NW}} = o_p(1)$. However, if we allow X_t and ε_t to be dependent, then the expectations in (A.29) do not split up into two separate parts any more. Moreover, since the weights $w_{t,s}$ and $w_{t',s'}$ depend on all the X_t^j for $t = 1, \ldots, T$, applying covariance inequalities like Davydov's inequality to the expressions $\mathbb{E}[w_{t,s}w_{t',s'}\varepsilon_{t-i}\varepsilon_s\varepsilon_{t'-i}\varepsilon_{s'}]$ is of no use any more. This necessitates the much more subtle arguments of Lemma A.2.1 to exploit the covariance structure of the processes $\{X_t\}$ and $\{\varepsilon_t\}$.

We finally turn to the analysis of the terms $Q_{T,\theta}$ and $Q_{T,B,j}$.

Lemma A.2.3. *It holds that*

$$Q_{T,\theta} = o_p(1).$$

Proof. We write

$$\begin{aligned} Q_{T,\theta} &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_{t-i} \big[\widetilde{m}_\theta(t-k) - m_\theta(t-k) \big] \\ &\quad - \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_{t-i} \Big[\sum_{j=0}^{d} \frac{1}{T} \sum_{s=1}^{T} m_j(X_s^j) \Big] \\ &=: Q_{T,\theta,a} + Q_{T,\theta,b} \end{aligned}$$

and consider the two terms $Q_{T,\theta,a}$ and $Q_{T,\theta,b}$ separately. For $Q_{T,\theta,a}$, we have

$$\begin{aligned} Q_{T,\theta,a} &= \sum_{t_\theta=1}^{\theta} \frac{1}{\sqrt{T}} \sum_{r=1}^{K_{t_\theta,T}} \varepsilon_{t_\theta+(r-1)\theta-i} \big(\widetilde{m}_\theta(t_\theta-k) - m_\theta(t_\theta-k) \big) \\ &= \sum_{t_\theta=1}^{\theta} \underbrace{\big(\widetilde{m}_\theta(t_\theta-k) - m_\theta(t_\theta-k) \big)}_{=o_p(1)} \underbrace{\Big(\frac{1}{\sqrt{T}} \sum_{r=1}^{K_{t_\theta,T}} \varepsilon_{t_\theta+(r-1)\theta-i} \Big)}_{=O_p(1)} = o_p(1). \end{aligned}$$

Recalling the normalization of the functions m_j in (2.4), a similar argument yields that $Q_{T,\theta,b} = o_p(1)$ as well. □

Lemma A.2.4. *It holds that*

$$Q_{T,B,j} = o_p(1)$$

for $j = 0, \ldots, d$.

Proof. We start by considering the case $j \neq 0$: Let $I_h = [2C_1h, 1-2C_1h]$ and $I_h^c = [0, 2C_1h) \cup (1-2C_1h, 1]$ as defined in Theorem 2.4.2. Using the uniform convergence rates from Theorem A.1.2, we get

$$\begin{aligned}|Q_{T,B,j}| &= \Big|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{t-i}\Big[\widetilde{m}_j^B(X_{t-k}^j) + \frac{1}{T}\sum_{s=1}^{T}m_j(X_s^j) - m_j(X_{t-k}^j)\Big]\Big| \\ &\leq O_p(h^2)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}|\varepsilon_{t-i}|I(X_{t-k}^j \in I_h) + O_p(h)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}|\varepsilon_{t-i}|I(X_{t-k}^j \notin I_h).\end{aligned}$$

By Markov's inequality, the first term on the right-hand side is $O_p(h^2\sqrt{T}) = o_p(1)$. Recognizing that by (A9), $\mathbb{E}[|\varepsilon_{t-i}|I(X_{t-k}^j \notin I_h)] \leq Ch$ for a sufficiently large constant C, another appeal to Markov's inequality yields that the second term is $O_p(h^2\sqrt{T}) = o_p(1)$ as well. This completes the proof for $j \neq 0$.

The proof for $j = 0$ is essentially the same: We have

$$\begin{aligned}|Q_{T,B,0}| &= \Big|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{t-i}\Big[\widetilde{m}_0^B\Big(\frac{t-k}{T}\Big) + \frac{1}{T}\sum_{s=1}^{T}m_0\Big(\frac{s}{T}\Big) - m_0\Big(\frac{t-k}{T}\Big)\Big]\Big| \\ &\leq O_p(h^2)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}|\varepsilon_{t-i}|I\Big(\frac{t-k}{T} \in I_h\Big) + O_p(h)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}|\varepsilon_{t-i}|I\Big(\frac{t-k}{T} \in I_h^c\Big) \\ &= O_p(h^2\sqrt{T}) + O_p(h)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}|\varepsilon_{t-i}|I\Big(\frac{t-k}{T} \in I_h^c\Big).\end{aligned}$$

As $\sum_{t=1}^{T} I(\frac{t-k}{T} \in I_h^c) \leq CTh$ for a sufficiently large constant C, Markov's inequality yields that the second term on the right-hand side is $O_p(h^2\sqrt{T}) = o_p(1)$ as well. $\square$

A.3 Auxiliary Results

For completeness, we collect some standard type uniform convergence results in this appendix which were used to prove Theorem 2.4.2 in Appendix A.1. These can be shown by small modifications of standard arguments as given for example in Bosq (1998), Masry (1996) or Hansen (2008). We start with the kernel density estimates $\widehat{p}_j$ and $\widehat{p}_{j,k}$. Using the notation $p_0(x_0) = I(x_0 \in (0,1])$, we have the following result.

Lemma A.3.1. *Suppose that* (A1) – (A5) *hold and that the bandwidth h satisfies* (A6)(a) *or* (A6)(b)*. Then*

$$\sup_{x_j \in I_h} \left| \widehat{p}_j(x_j) - p_j(x_j) \right| = O_p\Big(\sqrt{\frac{\log T}{Th}}\Big) + o(h) \quad \text{(A.30)}$$

$$\sup_{0 \le x_j \le 1} \left| \widehat{p}_j(x_j) - \kappa_0(x_j) p_j(x_j) \right| = O_p\Big(\sqrt{\frac{\log T}{Th}}\Big) + O(h) \quad \text{(A.31)}$$

$$\sup_{x_j, x_k \in I_h} \left| \widehat{p}_{j,k}(x_j, x_k) - p_{j,k}(x_j, x_k) \right| = O_p\Big(\sqrt{\frac{\log T}{Th^2}}\Big) + o(h) \quad \text{(A.32)}$$

$$\sup_{0 \le x_j, x_k \le 1} \left| \widehat{p}_{j,k}(x_j, x_k) - \kappa_0(x_j)\kappa_0(x_k) p_{j,k}(x_j, x_k) \right| = O_p\Big(\sqrt{\frac{\log T}{Th^2}}\Big) + O(h) \quad \text{(A.33)}$$

for $j, k = 0, \dots, d$ with $j \neq k$, where $\kappa_0(v) = \int_0^1 K_h(v, w) dw$ and $I_h = [2C_1 h, 1 - 2C_1 h]$.

We next consider the convergence behaviour of the one-dimensional Nadaraya-Watson smoothers $\widehat{m}_j$ defined in (2.11) and (2.14). For the stochastic part $\widehat{m}_j^A$, we have

Lemma A.3.2. *Under* (A1) – (A5) *together with* (A6)(a) *or* (A6)(b)*,*

$$\sup_{x_j \in [0,1]} \left| \widehat{m}_j^A(x_j) \right| = O_p\Big(\sqrt{\frac{\log T}{Th}}\Big) \quad \text{(A.34)}$$

for all $j = 0, \dots, d$.

For the bias part $\widehat{m}_j^B$, we have the following expansion:

Lemma A.3.3. *Under* (A1) – (A5) *together with* (A6)(a) *or* (A6)(b)*,*

$$\sup_{x_j \in I_h} \left| \widehat{m}_j^B(x_j) - \widehat{\mu}_{T,0} - \widehat{\mu}_{T,j}(x_j) \right| = o_p(h^2) \quad \text{(A.35)}$$

$$\sup_{x_j \in I_h^c} \left| \widehat{m}_j^B(x_j) - \widehat{\mu}_{T,0} - \widehat{\mu}_{T,j}(x_j) \right| = O_p(h^2) \quad \text{(A.36)}$$

for all $j = 0, \ldots, d$, *where*

$$\widehat{\mu}_{T,0} = -\frac{1}{T}\sum_{t=1}^{T}\Big(\sum_{j=1}^{d} m_j(X_t^j) + \varepsilon_t\Big)$$

$$\widehat{\mu}_{T,j}(x_j) = \alpha_{T,0} + \alpha_{T,j}(x_j) + \sum_{k\neq j}\int_0^1 \alpha_{T,k}(x_k)\frac{\widehat{p}_{j,k}(x_j,x_k)}{\widehat{p}_j(x_j)}dx_k + h^2\int \beta(x)\frac{q(x)}{p_j(x_j)}dx_{-j}.$$

Here, $\alpha_{T,0} = 0$ *and*

$$\alpha_{T,k}(x_k) = m_k(x_k) + m_k'(x_k)\frac{h\kappa_1(x_k)}{\kappa_0(x_k)}$$

$$\beta(x) = \sum_{k=0}^{d}\int u^2 K(u)du\Big(\frac{\partial \log q(x)}{\partial x_k}m_k'(x_k) + \frac{1}{2}m_k''(x_k)\Big)$$

with $\kappa_0(x_k) = \int_0^1 K_h(x_k,w)dw$ *and* $\kappa_1(x_k) = \int_0^1 K_h(x_k,w)(\frac{w-x_k}{h})dw.$

Lemma A.3.3 can be proven by going along the lines of the arguments for Theorem 4 in Mammen et al. (1999). To see that

$$\widehat{\mu}_{T,0} = -\frac{1}{T}\sum_{t=1}^{T}\Big(\sum_{j=1}^{d} m_j(X_t^j) + \varepsilon_t\Big), \tag{A.37}$$

note that

$$\widehat{m}_j^B(x_j) = \frac{1}{T}\sum_{t=1}^{T} K_h(x_j, X_t^j)\big(m_\theta(t) - \widetilde{m}_\theta(t)\big)/\widehat{p}_j(x_j)$$
$$+ \frac{1}{T}\sum_{t=1}^{T} K_h(x_j, X_t^j)\Big[m_0\Big(\frac{t}{T}\Big) + \sum_{k=1}^{d} m_k(X_t^k)\Big]/\widehat{p}_j(x_j)$$

for $j = 0, \ldots, d$ with $X_t^0 = \frac{t}{T}$. Moreover,

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T} K_h(x_j, X_t^j)\big(m_\theta(t) - \widetilde{m}_\theta(t)\big)/\widehat{p}_j(x_j) \\
&= \sum_{t_\theta=1}^{\theta} \big(m_\theta(t_\theta) - \widetilde{m}_\theta(t_\theta)\big)\frac{1}{T}\sum_{k=1}^{K_{t_\theta,T}} K_h(x_j, X^j_{t_\theta+(k-1)\theta})/\widehat{p}_j(x_j) \\
&= \frac{1}{\theta}\sum_{t_\theta=1}^{\theta} \big(m_\theta(t_\theta) - \widetilde{m}_\theta(t_\theta)\big)\underbrace{\frac{1}{K_{t_\theta,T}}\sum_{k=1}^{K_{t_\theta,T}} K_h(x_j, X^j_{t_\theta+(k-1)\theta})}_{\xrightarrow{P}\kappa_0(x_j)p_j(x_j) \text{ uniformly in } x_j}/\widehat{p}_j(x_j) + o_p(h^2) \\
&= \frac{1}{\theta}\sum_{t_\theta=1}^{\theta} \big(m_\theta(t_\theta) - \widetilde{m}_\theta(t_\theta)\big) + o_p(h^2)
\end{aligned}$$

uniformly in x_j and

$$\begin{aligned}
&\frac{1}{\theta}\sum_{t_\theta=1}^{\theta} \big(m_\theta(t_\theta) - \widetilde{m}_\theta(t_\theta)\big) \\
&= -\frac{1}{\theta}\sum_{t_\theta=1}^{\theta}\frac{1}{K_{t_\theta,T}}\sum_{k=1}^{K_{t_\theta,T}}\Big(m_0\Big(\frac{t_\theta+(k-1)\theta}{T}\Big) + \sum_{j=1}^{d} m_j(X^j_{t_\theta+(k-1)\theta}) + \varepsilon_{t_\theta+(k-1)\theta}\Big) \\
&= -\frac{1}{\theta}\sum_{t_\theta=1}^{\theta}\frac{1}{K_{t_\theta,T}}\sum_{k=1}^{K_{t_\theta,T}}\Big(\sum_{j=1}^{d} m_j(X^j_{t_\theta+(k-1)\theta}) + \varepsilon_{t_\theta+(k-1)\theta}\Big) + o_p(h^2) \\
&= -\frac{1}{T}\sum_{t=1}^{T}\Big(\sum_{j=1}^{d} m_j(X_t^j) + \varepsilon_t\Big) + o_p(h^2).
\end{aligned}$$

Combining the above calculations with the arguments from the proof of Theorem 4 in Mammen et al. 1999 yields formula (A.37) for $\widehat{\mu}_{T,0}$.

Appendix B

Proofs for Chapter 3

In this appendix we have collected all the proofs needed to establish the results in Chapter 3 for the multiplicative volatility model.

B.1 Proof of Theorem 3.4.1

As mentioned before Theorem 3.4.1, the proof rests on the corresponding results for the additive model of Chapter 2. These were proven in Appendix A.1. The result in Theorem 3.4.1 is established by the smoothness of the reverse transform $\widetilde{\tau}_j^2 = \exp(\widetilde{m}_j)$ for $j = 0, \ldots, d$.

B.1.1 Restatement of results from Appendix A.1

Recall that the backfitting estimates $\widetilde{m}_j$ can be decomposed into a stochastic part $\widetilde{m}_j^A$ and a bias part $\widetilde{m}_j^B$ according to

$$\widetilde{m}_j(x_j) = \widetilde{m}_j^A(x_j) + \widetilde{m}_j^B(x_j).$$

The two components are defined by

$$\widetilde{m}_j^S(x_j) = \widehat{m}_j^S(x_j) - \sum_{k \neq j} \int \widetilde{m}_k^S(x_k) \frac{\widehat{p}_{k,j}(x_k, x_j)}{\widehat{p}_j(x_j)}\, dx_k - \widetilde{m}_c^S \tag{B.1}$$

for $S = A$, B. Here, $\widehat{m}_k^A$ and $\widehat{m}_k^B$ denote the stochastic part and the bias part of the Nadaraya-Watson pilote estimates defined as

$$\widehat{m}_j^A(x_j) = \frac{1}{T}\sum_{t=1}^{T} K_h(x_j, X_t^j)u_t/\widehat{p}_j(x_j) \tag{B.2}$$

$$\widehat{m}_j^B(x_j) = \frac{1}{T}\sum_{t=1}^{T} K_h(x_j, X_t^j)\Big[m_c + m_0\Big(\frac{t}{T}\Big) + \sum_{j=1}^{d} m_j(X_t^j)\Big]/\widehat{p}_j(x_j) \tag{B.3}$$

for $j = 0, \ldots, d$, where we set $X_t^0 = \frac{t}{T}$ to shorten the notation. Furthermore, $\widetilde{m}_c^A = \frac{1}{T}\sum_{t=1}^{T} u_t$ and $\widetilde{m}_c^B = \frac{1}{T}\sum_{t=1}^{T}\{m_c + m_0(\frac{t}{T}) + \sum_{j=1}^{d} m_j(X_t^j)\}$.

We next state the results of Appendix A.1 under the assumptions of our model in Chapter 3. As before, we first give the higher order expansion of the stochastic part $\widetilde{m}_j^A$. Then we state the corresponding expansion for the bias part $\widetilde{m}_j^B$.

Theorem B.1.1. *Suppose that assumptions* (V1) – (V9) *apply and that the bandwidth h satisfies* (V10)(a) *or* (V10)(b). *Then uniformly for* $0 \le x_j \le 1$,

$$\widetilde{m}_j^A(x_j) = \widehat{m}_j^A(x_j) + \frac{1}{T}\sum_{t=1}^{T} r_{j,t}(x_j)u_t + o_p\Big(\frac{1}{\sqrt{T}}\Big),$$

where $r_{j,t}(\cdot) := r_j(\frac{t}{T}, X_t, \cdot)$ *are absolutely uniformly bounded functions with*

$$|r_{j,t}(x_j') - r_{j,t}(x_j)| \le C|x_j' - x_j|$$

for a constant $C > 0$.

Theorem B.1.2. *Suppose that* (V1) – (V9) *hold. If the bandwidth h satisfies* (V10)(a), *then*

$$\sup_{x_j \in I_h} |\widetilde{m}_j^B(x_j) - m_j(x_j)| = O_p(h^2) \tag{B.4}$$

$$\sup_{x_j \in I_h^c} |\widetilde{m}_j^B(x_j) - m_j(x_j)| = O_p(h) \tag{B.5}$$

for $j = 0, \ldots, d$. *If the bandwidth satisfies* (b), *we have*

$$\sup_{x_j \in I_h} \left| \widetilde{m}_j^B(x_j) + \frac{1}{T}\sum_{t=1}^{T} m_j(X_t^j) - m_j(x_j) \right| = O_p(h^2) \tag{B.6}$$

$$\sup_{x_j \in I_h^c} \left| \widetilde{m}_j^B(x_j) + \frac{1}{T}\sum_{t=1}^{T} m_j(X_t^j) - m_j(x_j) \right| = O_p(h) \tag{B.7}$$

for $j = 0, \ldots, d$.

These expansions can be combined to show the equivalent of Theorem 2.4.2 for the transformed model (3.10). By a second order Taylor expansions of $\widetilde{\tau}_j^2 = \exp(\widetilde{m}_j)$ for $j = 0, \ldots, d$ the result of Theroem 3.4.1 follows due to the smoothness of the exponential function. □

B.2 Proofs of Theorems 3.4.2 and 3.4.3

This appendix contains the proofs of Theorems 3.4.2 and 3.4.3, which show consistency and asymptotic normality of the GARCH estimates. In particular the proof of asymptotic normality is rather involved. To establish the normality result we will thus start by providing the general idea of the proof which is based on an expansion of the likelihood. The subsequent steps needed to move from this expansion to establishing the asymptotic normality of our estimators contain the major challenges. These have been collected in a series of lemmas at the end of this appendix. As already pointed out in Section 3.4.2, the main difficulty is to derive a stochastic expansion of $\frac{1}{\sqrt{T}} \frac{\partial \widetilde{l}_T(\phi_0)}{\partial \phi}$. The expansion is given in Lemmas B.2.1 – B.2.3. Throughout this appendix C will again denote a finite real constant which may take a different value on each occurrence.

B.2.1 Auxiliary Results

To start with, we state some facts about the behaviour of the approximate GARCH variables $\widetilde{\varepsilon}_t$ and of the conditional volatilities $\widetilde{v}_t^2(\phi)$, which were defined in Subsection 3.3.2. As will become clear these will be used in the proofs Theorems 3.4.2

and 3.4.3. For ease of notation, we use the shorthand $\tau(x) = \prod_{j=0}^{d} \tau_j(x_j)$ in what follows.

(G1) We can express $\widetilde{\varepsilon}_t^2 - \varepsilon_t^2$ as

$$\widetilde{\varepsilon}_t^2 - \varepsilon_t^2 = \varepsilon_t^2 \Big[\frac{\tau^2(\frac{t}{T}, X_t) - \widetilde{\tau}^2(\frac{t}{T}, X_t)}{\tau^2(\frac{t}{T}, X_t)} + R_\varepsilon\Big(\frac{t}{T}, X_t\Big)\Big]$$

with $\sup_{x \in [0,1]^{d+1}} |R_\varepsilon(x)| = O_p(h^2)$.

(G2) The conditional volatility $v_t^2(\phi)$ has the expansion

$$v_t^2(\phi) = w \sum_{k=1}^{t-1} b^{k-1} + a \sum_{k=1}^{t-1} b^{k-1} \varepsilon_{t-k}^2 + b^{t-1} \frac{w}{1-b},$$

which yields that

$$\widetilde{v}_t^2(\phi) - v_t^2(\phi) = \sum_{k=1}^{t-1} a b^{k-1} (\widetilde{\varepsilon}_{t-k}^2 - \varepsilon_{t-k}^2).$$

(G3) It holds that

$$\max_{1 \le t \le T} \sup_{\phi \in \Phi} \left| \widetilde{v}_t^2(\phi) - v_t^2(\phi) \right| = O_p(h).$$

(G4) It holds that

$$\frac{1}{\widetilde{v}_t^2(\phi)} - \frac{1}{v_t^2(\phi)} = \frac{v_t^2(\phi) - \widetilde{v}_t^2(\phi)}{v_t^2(\phi) v_t^2(\phi)} + R_t(\phi)$$

with $\max_{1 \le t \le T} \sup_{\phi \in \Phi} |R_t(\phi)| = O_p(h^2)$.

(G5) The derivatives of $v_t^2(\phi)$ with respect to the parameters w, a, and b are given by

$$\frac{\partial v_t^2(\phi)}{\partial w} = \sum_{k=1}^{t-1} b^{k-1} + \frac{b^{t-1}}{1-b}$$

$$\frac{\partial v_t^2(\phi)}{\partial a} = \sum_{k=1}^{t-1} b^{k-1} \varepsilon_{t-k}^2$$

$$\frac{\partial v_t^2(\phi)}{\partial b} = w \Big(\sum_{k=1}^{t-1} (k-1) b^{k-2} + \frac{(t-1) b^{t-2}}{1-b} + \frac{b^{t-1}}{(1-b)^2} \Big) + a \sum_{k=1}^{t-1} (k-1) b^{k-2} \varepsilon_{t-k}^2.$$

The above facts are straightforward to verify. We thus omit the details.

B.2.2 Proof of Theorem 3.4.2

Let $l_T(\phi)$ and $\widetilde{l}_T(\phi)$ be the likelihood functions introduced in (2.17) and (2.19) and define

$$l(\phi) = \mathbb{E}\Big[\frac{1}{T} l_T(\phi)\Big].$$

By the triangle inequality,

$$\sup_{\phi\in\Phi} \big|\frac{1}{T}\widetilde{l}_T(\phi) - l(\phi)\big| \leq \sup_{\phi\in\Phi} \big|\frac{1}{T}\widetilde{l}_T(\phi) - \frac{1}{T} l_T(\phi)\big| + \sup_{\phi\in\Phi} \big|\frac{1}{T} l_T(\phi) - l(\phi)\big|.$$

From standard theory we know that

$$\sup_{\phi\in\Phi} \big|\frac{1}{T} l_T(\phi) - l(\phi)\big| = o_p(1)$$

and that $l(\phi)$ is a continuous function of ϕ with a unique maximum at ϕ_0. If we can further show that

$$\sup_{\phi\in\Phi} \big|\frac{1}{T}\widetilde{l}_T(\phi) - \frac{1}{T} l_T(\phi)\big| = o_p(1), \tag{B.8}$$

then standard theory on M-estimation implies $\widetilde{\phi} \xrightarrow{P} \phi_0$.

We will show (B.8) by decomposing $\frac{1}{T}\widetilde{l}_T(\phi) - \frac{1}{T} l_T(\phi)$ into the sum of three uniformly $o_p(1)$ terms.

$$\begin{aligned}
&\frac{1}{T}\widetilde{l}_T(\phi) - \frac{1}{T} l_T(\phi) \\
&= -\frac{1}{T}\sum_{t=1}^{T}\Big(\log \widetilde{v}_t^2(\phi) + \frac{\widetilde{\varepsilon}_t^2}{\widetilde{v}_t^2(\phi)}\Big) + \frac{1}{T}\sum_{t=1}^{T}\Big(\log v_t^2(\phi) + \frac{\varepsilon_t^2}{v_t^2(\phi)}\Big) \\
&= \frac{1}{T}\sum_{t=1}^{T}\big(\log v_t^2(\phi) - \log \widetilde{v}_t^2(\phi)\big) + \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2\Big(\frac{\widetilde{v}_t^2(\phi) - v_t^2(\phi)}{\widetilde{v}_t^2(\phi) v_t^2(\phi)}\Big) + \frac{1}{T}\sum_{t=1}^{T}\frac{1}{\widetilde{v}_t^2(\phi)}(\varepsilon_t^2 - \widetilde{\varepsilon}_t^2) \\
&=: (A) + (B) + (C).
\end{aligned}$$

In order to prove that the three terms (A), (B), and (C) are indeed uniformly $o_p(1)$, it suffices to show that

$$\max_{1\le t\le T}\sup_{\phi\in\Phi}\left|\widetilde{v}_t^2(\phi)-v_t^2(\phi)\right|=o_p(1) \tag{B.9}$$

$$\frac{1}{T}\sum_{t=1}^{T}\left|\widetilde{\varepsilon}_t^2-\varepsilon_t^2\right|=o_p(1) \tag{B.10}$$

$$v_t^2(\phi)\ge v_{\min}>0 \quad \text{and} \quad \widetilde{v}_t^2(\phi)\ge v_{\min}>0 \quad \text{for some constant } v_{\min}. \tag{B.11}$$

$(B.9)$ is implied by (G3). For the proof of $(B.10)$, we use (G1) together with Theorem 3.4.1 to obtain

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\left|\widetilde{\varepsilon}_t^2-\varepsilon_t^2\right| &\le \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2\left|\frac{\tau^2(\frac{t}{T},X_t)-\widetilde{\tau}^2(\frac{t}{T},X_t)}{\tau^2(\frac{t}{T},X_t)}+R_\varepsilon\Big(\frac{t}{T},X_t\Big)\right|\\ &=O_p(h)\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2=O_p(h).\end{aligned}$$

Finally, $(B.11)$ is automatically satisfied, as by (V11)

$$v_t^2(\phi)=w\sum_{k=1}^{t-1}b^{k-1}+a\sum_{k=1}^{t-1}b^{k-1}\varepsilon_{t-k}^2+b^{t-1}\frac{w}{1-b}\ge w\ge\underline{\kappa}>0.$$

The same holds true for $\widetilde{v}_t^2(\phi)$. □

B.2.3 Proof of Theorem 3.4.3

By the usual Taylor expansion argument, we obtain

$$0=\frac{1}{T}\frac{\partial\widetilde{l}_T(\widetilde{\phi})}{\partial\phi}=\frac{1}{T}\frac{\partial\widetilde{l}_T(\phi_0)}{\partial\phi}+\frac{1}{T}\frac{\partial^2\widetilde{l}_T(\bar{\phi})}{\partial\phi\partial\phi^T}(\widetilde{\phi}-\phi_0)$$

with some intermediate point $\bar{\phi}$ between ϕ_0 and $\widetilde{\phi}$. Rearranging and premultiplying by $\sqrt{T}$ yields

$$\sqrt{T}(\widetilde{\phi}-\phi_0)=-\Big(\frac{1}{T}\frac{\partial^2\widetilde{l}_T(\bar{\phi})}{\partial\phi\partial\phi^T}\Big)^{-1}\frac{1}{\sqrt{T}}\frac{\partial\widetilde{l}_T(\phi_0)}{\partial\phi}.$$

The proof will be completed upon showing that

$$\frac{1}{\sqrt{T}}\frac{\partial \widetilde{l}_T(\phi_0)}{\partial \phi} \xrightarrow{d} N(0,Q) \tag{B.12}$$

$$\frac{1}{T}\frac{\partial^2 \widetilde{l}_T(\bar{\phi})}{\partial \phi \partial \phi^T} \xrightarrow{P} J, \tag{B.13}$$

where Q is some covariance matrix to be specified later on and J is an invertible deterministic matrix. Thus we see that the asymptotic covariance matrix given in Theorem 3.4.3 is

$$\Sigma = J^{-1}QJ^{-1}. \tag{B.14}$$

Proof of (B.12). Let $v_t^2 = v_t^2(\phi_0)$ and $\widetilde{v}_t^2 = \widetilde{v}_t^2(\phi_0)$ in order to lighten notation. Writing out the i-th element of left hand side of (B.12) we get

$$\frac{1}{\sqrt{T}}\frac{\partial \widetilde{l}_T(\phi_0)}{\partial \phi_i} = -\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Big(1-\frac{\widetilde{\varepsilon}_t^2}{\widetilde{v}_t^2}\Big)\frac{\partial \widetilde{v}_t^2}{\partial \phi_i}\frac{1}{\widetilde{v}_t^2}$$

Succesively replacing the approximate expressions we can show that

$$\begin{aligned}
\frac{1}{\sqrt{T}}\frac{\partial \widetilde{l}_T(\phi_0)}{\partial \phi_i} =& -\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Big((1-\frac{\widetilde{\varepsilon}_t^2}{\widetilde{v}_t^2})-(1-\frac{\varepsilon_t^2}{v_t^2})\Big)\frac{\partial \widetilde{v}_t^2}{\partial \phi_i}\frac{1}{\widetilde{v}_t^2} && \text{(A)}\\
& -\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Big(1-\frac{\varepsilon_t^2}{v_t^2}\Big)\frac{\partial \widetilde{v}_t^2}{\partial \phi_i}\Big(\frac{1}{\widetilde{v}_t^2}-\frac{1}{v_t^2}\Big) && \text{(B)}\\
& -\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Big(1-\frac{\varepsilon_t^2}{v_t^2}\Big)\frac{1}{v_t^2}\Big(\frac{\partial \widetilde{v}_t^2}{\partial \phi_i}-\frac{\partial v_t^2}{\partial \phi_i}\Big) && \text{(C)}\\
& -\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Big(1-\frac{\varepsilon_t^2}{v_t^2}\Big)\frac{1}{v_t^2}\frac{\partial v_t^2}{\partial \phi_i} && \text{(D)}
\end{aligned}$$

Notice that the term in (D) is $\frac{1}{\sqrt{T}}\frac{\partial l_T(\phi_0)}{\partial \phi_i}$ and will thus by standard arguments contribute to the asymptotic distribution. In what follows, we will show that the term (A) will also contribute to the limiting distribution, whereas the terms (B) and (C) will be asymptotically negligible. We will deal with each term individually starting with (B) and (C). The results will be established by replacing the truncated conditional volatilities v_t^2 by σ_t^2 and then repeatedly appealing to the martingale difference structure of $\eta_t = \sigma_t^2 - \varepsilon_t^2$ and the results stated in Subsection B.2.1.

We will start with (C) as it is slightly more complicated than (B). Replacing v_t^2 by σ_t^2 we obtain

$$
\begin{aligned}
(C) = &- \frac{1}{\sqrt{T}} \sum_{t=1}^{T} [(v_t^2 - \varepsilon_t^2) - (\sigma_t^2 - \varepsilon_t^2)] \Big(\frac{\partial \widetilde{v}_t^2}{\partial \phi_i} - \frac{\partial v_t^2}{\partial \phi_i} \Big) \frac{1}{(v_t^2)^2} \\
&- \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\sigma_t^2 - \varepsilon_t^2) \Big(\frac{\partial \widetilde{v}_t^2}{\partial \phi_i} - \frac{\partial v_t^2}{\partial \phi_i} \Big) \Big(\frac{(\sigma_t^2)^2 - (v_t^2)^2}{(v_t^2 \sigma_t^2)^2} \Big) \\
&- \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\sigma_t^2 - \varepsilon_t^2) \Big(\frac{\partial \widetilde{v}_t^2}{\partial \phi_i} - \frac{\partial v_t^2}{\partial \phi_i} \Big) \frac{1}{(\sigma_t^2)^2}
\end{aligned}
$$

Using (G2), we can show that $|\sigma_t^2 - v_t^2| = b^{t-1} |\sigma_1^2 - \frac{w}{1-b}|$. This implies that the first two terms are negligible and we get

$$
(C) = -\frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\sigma_t^2 - \varepsilon_t^2) \Big(\frac{\partial \widetilde{v}_t^2}{\partial \phi_i} - \frac{\partial v_t^2}{\partial \phi_i} \Big) \frac{1}{(\sigma_t^2)^2} + o_p(1).
$$

As $\eta_t = \sigma_t^2 - \varepsilon_t^2$ is a martingale difference, we can use results from empirical process theory to show that $(C) = o_p(1)$. Analogously, we obtain that $(B) = o_p(1)$.

Next we will consider the term (A), which will be easier to analyse be splitting it in two:

$$
\begin{aligned}
(A) &= -\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Big((1 - \frac{\widetilde{\varepsilon}_t^2}{\widetilde{v}_t^2}) - (1 - \frac{\varepsilon_t^2}{v_t^2}) \Big) \frac{\partial \widetilde{v}_t^2}{\partial \phi_i} \frac{1}{\widetilde{v}_t^2} \\
&= -\frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\varepsilon_t^2 - \widetilde{\varepsilon}_t^2) \frac{1}{v_t^2} \frac{1}{\widetilde{v}_t^2} \frac{\partial \widetilde{v}_t^2}{\partial \phi_i} - \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \widetilde{\varepsilon}_t^2 \Big(\frac{1}{\widetilde{v}_t^2} - \frac{1}{v_t^2} \Big) \frac{1}{\widetilde{v}_t^2} \frac{\partial \widetilde{v}_t^2}{\partial \phi_i} \\
&=: (A_1) + (A_2).
\end{aligned}
$$

Next we will present the steps needed to deal with (A_2). The results for (A_1) are established in an analogous fashion. As before, replacing the approximate espressions by the exact ones and using the results from Subsection B.2.1 we obtain

$$(A_2) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_t^2 \Big(\frac{v_t^2 - \widetilde{v}_t^2}{v_t^2 v_t^2} \Big) \frac{\partial v_t^2}{\partial \phi_i} \frac{1}{v_t^2} + o_p(1)$$

Now replacing the occurences of the truncated conditional volatilities in the denominator by σ_t^2 results in

$$\begin{aligned}
(A_2) &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_t^2 \Big(\frac{v_t^2 - \widetilde{v}_t^2}{\sigma_t^2 \sigma_t^2} \Big) \frac{\partial v_t^2}{\partial \phi_i} \frac{1}{\sigma_t^2} + o_p(1) \\
&= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Big(\frac{v_t^2 - \widetilde{v}_t^2}{\sigma_t^2 \sigma_t^2} \Big) \frac{\partial v_t^2}{\partial \phi_i} + \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\varepsilon_t^2 - \sigma_t^2) \Big(\frac{v_t^2 - \widetilde{v}_t^2}{\sigma_t^2 \sigma_t^2} \Big) \frac{\partial v_t^2}{\partial \phi_i} \frac{1}{\sigma_t^2} + o_p(1) \\
&= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Big(\frac{v_t^2 - \widetilde{v}_t^2}{\sigma_t^2 \sigma_t^2} \Big) \frac{\partial v_t^2}{\partial \phi_i} + o_p(1)
\end{aligned}$$

with the last equality again due to the Martingale difference argument. Defining $G_t^{[i]} := \frac{\partial v_t^2}{\partial \phi_i} \frac{1}{\sigma_t^2 \sigma_t^2}$, using (G1) – (G3)) and writing $m(x) = m_c + m_0(x_0) + \ldots + m_d(x_d)$ for short, we can infer that

$$\begin{aligned}
(A_2) &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} G_t^{[i]} \sum_{k=1}^{t-1} ab^{k-1} (\varepsilon_{t-k}^2 - \widetilde{\varepsilon}_{t-k}^2) + o_p(1) \\
&= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} G_t^{[i]} \sum_{k=1}^{t-1} ab^{k-1} \varepsilon_{t-k}^2 \Big[\frac{\tau^2(\frac{t-k}{T}, X_{t-k}) - \widetilde{\tau}^2(\frac{t-k}{T}, X_{t-k})}{\tau^2(\frac{t-k}{T}, X_{t-k})} + O_p(h^2) \Big] + o_p(1) \\
&= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} G_t^{[i]} \sum_{k=1}^{t-1} ab^{k-1} \varepsilon_{t-k}^2 \Big[\frac{\exp(\xi_{t-k})[m(\frac{t-k}{T}, X_{t-k}) - \widetilde{m}(\frac{t-k}{T}, X_{t-k})]}{\exp(m(\frac{t-k}{T}, X_{t-k}))} \Big] + o_p(1) \\
&= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} G_t^{[i]} \sum_{k=1}^{t-1} ab^{k-1} \varepsilon_{t-k}^2 \Big[m\Big(\frac{t-k}{T}, X_{t-k} \Big) - \widetilde{m}\Big(\frac{t-k}{T}, X_{t-k} \Big) \Big] + o_p(1)
\end{aligned}$$

where the third equality is by a first order Taylor expansion with an intermediate point ξ_{t-k} between $m(\frac{t-k}{T}, X_{t-k})$ and $\widetilde{m}(\frac{t-k}{T}, X_{t-k})$. We are now in a position to use the stochastic expansion of our estimators in the additive model, which were given in Appendix B.1. To do so, split the regression function and the estimators into their additive components and use the expansion of the component estimators into their respective bias and variance parts, denoted by $(A^j_{2,B})$ and $(A^j_{2,V})$ to get

We finally split up the difference $m(\frac{t-k}{T}, X_{t-k}) - \widetilde{m}(\frac{t-k}{T}, X_{t-k})$ into its additive components and decompose the various components into their bias and stochastic parts. This yields

$$(D) = (D_c) - \sum_{j=0}^{d} (D_{V,j}) + \sum_{j=0}^{d} (D_{B,j}) + o_p(1)$$

with

$$(D_c) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} G_t \sum_{k=1}^{t-1} ab^{k-1} \varepsilon^2_{t-k} \Big[(m_c - \widetilde{m}_c) + \sum_{j=0}^{d} \frac{1}{T} \sum_{s=1}^{T} m_j(X^j_s) \Big]$$

$$(D_{V,j}) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} G_t \sum_{k=1}^{t-1} ab^{k-1} \varepsilon^2_{t-k} \widetilde{m}^A_j(X^j_{t-k})$$

$$(D_{B,j}) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} G_t \sum_{k=1}^{t-1} ab^{k-1} \varepsilon^2_{t-k} \Big[m_j(X^j_{t-k}) - \widetilde{m}^B_j(X^j_{t-k}) - \frac{1}{T} \sum_{s=1}^{T} m_j(X^j_s) \Big]$$

for $j = 0, \ldots, d$, where for ease of notation we have used the shorthand $X^0_{t-k} = \frac{t-k}{T}$. As in Appendix A, $\widetilde{m}^A_j$ denotes the stochastic part of the backfitting estimate $\widetilde{m}_j$ and $\widetilde{m}^B_j$ denotes the bias part.

In Lemmas B.2.1 – B.2.3, we will show that

$$(D_c) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} g_{c,D} u_t + o_p(1) \tag{B.15}$$

$$(D_{V,j}) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} g_{j,D}\Big(\frac{t}{T}, X_t\Big) u_t + o_p(1) \tag{B.16}$$

$$(D_{B,j}) = o_p(1) \tag{B.17}$$

for all $j = 0, \ldots, d$ with $u_t = \log(\varepsilon^2_t)$. Here, $g_{c,D}$ is a constant which is specified in Lemma B.2.2 and $g_{j,D}$ for $j = 0, \ldots, d$ are functions whose exact forms are given in Lemma B.2.1. Using (V12), these functions are easily seen to be absolutely bounded by a constant independent of T. To summarize, we obtain that

$$(D) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Big[g_{c,D} + \sum_{j=0}^{d} g_{j,D}\Big(\frac{t}{T}, X_t\Big)\Big] u_t + o_p(1).$$

Repeating the arguments from above, we can derive an analogous expression for (C). We thus get that

$$(C) + (D) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} g\Big(\frac{t}{T}, X_t\Big) u_t + o_p(1)$$

with a function $g(\frac{t}{T}, X_t) = g_c + \sum_{j=0}^{d} g_j(\frac{t}{T}, X_t)$ whose additive components are absolutely bounded. Recalling that $(A) = o_p(1)$ and $(B) = o_p(1)$, we finally obtain that

$$\frac{1}{\sqrt{T}} \frac{\partial \widetilde{l}_T(\phi_0)}{\partial \phi_i} - \frac{1}{\sqrt{T}} \frac{\partial l_T(\phi_0)}{\partial \phi_i} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} g\Big(\frac{t}{T}, X_t\Big) u_t + o_p(1) \tag{B.18}$$

with an absolutely bounded function g.

We next consider the term $\frac{1}{\sqrt{T}} \frac{\partial l_T(\phi_0)}{\partial \phi_i}$ more closely. W.l.o.g. we can take $\phi_i = a$. (The case $\phi_i = b$ runs analogously and the case $\phi_i = w$ is much easier to handle.) By similar arguments to before,

$$\begin{aligned}\frac{1}{\sqrt{T}} \frac{\partial l_T(\phi_0)}{\partial \phi_i} &= -\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Big(1 - \frac{\varepsilon_t^2}{v_t^2}\Big) \frac{\partial v_t^2}{\partial \phi_i} \frac{1}{v_t^2} \\ &= -\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Big(\frac{1 - \eta_t^2}{\sigma_t^2}\Big) \sum_{k=1}^{t-1} b^{k-1} \varepsilon_{t-k}^2 + o_p(1).\end{aligned}$$

Furthermore,

$$\begin{aligned}\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Big(\frac{1 - \eta_t^2}{\sigma_t^2}\Big) \sum_{k=1}^{t-1} b^{k-1} \varepsilon_{t-k}^2 &= \sum_{k=1}^{T-1} b^{k-1} \frac{1}{\sqrt{T}} \sum_{t=k+1}^{T} \Big(\frac{1 - \eta_t^2}{\sigma_t^2}\Big) \varepsilon_{t-k}^2 \\ &= \sum_{k=1}^{C_2 \log T} b^{k-1} \frac{1}{\sqrt{T}} \sum_{t=k+1}^{T} \Big(\frac{1 - \eta_t^2}{\sigma_t^2}\Big) \varepsilon_{t-k}^2 + o_p(1) \\ &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Big(\sum_{k=1}^{\min_{t,T}} b^{k-1} \varepsilon_{t-k}^2 \Big) \Big(\frac{1 - \eta_t^2}{\sigma_t^2}\Big) + o_p(1),\end{aligned}$$

where $C_2 > 0$ is a sufficiently large constant and $\min_{t,T} := \min\{t-1, C_2 \log T\}$. For the second equality, we have used the fact that the weights b^k and b^i converge exponentially fast to zero as $i, k \to \infty$. This implies that only the sums up to $C_2 \log T$ with some constant C_2 are asymptotically relevant. Summing up, we have that

$$\frac{1}{\sqrt{T}} \frac{\partial l_T(\phi_0)}{\partial \phi_i} = -\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Big(\sum_{k=1}^{\min_{t,T}} b^{k-1} \varepsilon_{t-k}^2 \Big) \Big(\frac{1-\eta_t^2}{\sigma_t^2} \Big) + o_p(1). \tag{B.19}$$

Combining (B.18) and (B.19) yields

$$\begin{aligned}
\frac{1}{\sqrt{T}} \frac{\partial \widetilde{l}_T(\phi_0)}{\partial \phi_i} &= \frac{1}{\sqrt{T}} \frac{\partial l_T(\phi_0)}{\partial \phi_i} + \frac{1}{\sqrt{T}} \sum_{t=1}^{T} g\Big(\frac{t}{T}, X_t\Big) u_t + o_p(1) \\
&= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Big\{ g\Big(\frac{t}{T}, X_t\Big) u_t - \Big(\sum_{k=1}^{\min_{t,T}} b^{k-1} \varepsilon_{t-k}^2 \Big) \Big(\frac{1-\eta_t^2}{\sigma_t^2} \Big) \Big\} + o_p(1) \\
&=: \frac{1}{\sqrt{T}} \sum_{t=1}^{T} Z_{t,T} + o_p(1),
\end{aligned}$$

i.e. the term of interest can be written as a normalized sum of random variables $Z_{t,T}$ plus a term which is asymptotically negligible.

We now apply a central limit theorem for mixing arrays to the term $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} Z_{t,T}$. In particular, we employ the theorem of Francq & Zakoïan (2005), which allows the mixing coefficients of the array $\{Z_{t,T}\}$ to depend on the sample size T. Verifying the conditions of this theorem, we can conclude that

$$\frac{1}{\sqrt{T}} \frac{\partial \widetilde{l}_T(\phi_0)}{\partial \phi_i} \to N(0, \sigma^2)$$

with

$$\begin{aligned}
\sigma^2 = \mathbb{E}\Big[\lambda_2(X_0) u_0\Big] &- 2\mathbb{E}\Big[\lambda_1(X_0) u_0 \Big(\sum_{k=1}^{\infty} b^{k-1} \varepsilon_{-k}^2 \Big) \Big(\frac{1-\eta_0^2}{\sigma_0^2} \Big) \Big] \\
&+ \mathbb{E}\Big[\Big(\sum_{k=1}^{\infty} b^{k-1} \varepsilon_{-k}^2 \Big)^2 \Big(\frac{1-\eta_0^2}{\sigma_0^2} \Big)^2 \Big] + 2\mathbb{E}\big[\lambda_{1,1}(X_0, X_l) u_0 u_l\big]
\end{aligned}$$

$$- 2\mathbb{E}\Big[\lambda_1(X_0)u_0\Big(\sum_{k=1}^{\infty} b^{k-1}\varepsilon_{l-k}^2\Big)\Big(\frac{1-\eta_l^2}{\sigma_l^2}\Big)\Big]$$
$$- 2\mathbb{E}\Big[\lambda_1(X_l)u_l\Big(\sum_{k=1}^{\infty} b^{k-1}\varepsilon_{-k}^2\Big)\Big(\frac{1-\eta_0^2}{\sigma_0^2}\Big)\Big],$$

where we use the shorthand $\lambda_1(x) = \int_0^1 g(w,x)dw$, $\lambda_2(x) = \int_0^1 g^2(w,x)dw$, and $\lambda_{1,1}(x,x') = \int_0^1 g(w,x)g(w,x')dw$. Using the Cramer-Wold device, it is now easy to show that

$$\frac{1}{\sqrt{T}}\frac{\partial \widetilde{l}_T(\phi_0)}{\partial \phi} \to N(0,Q).$$

The entries of the matrix Q can be calculated similarly to the expression σ^2. We omit the details as the formulas are rather lengthy and complicated. □

Proof of (A.14). By straightforward but tedious calculations it can be seen that

$$\sup_{\phi\in\Phi}\Big|\frac{1}{T}\frac{\partial^2 \widetilde{l}_T(\phi)}{\partial\phi\partial\phi^T} - \frac{1}{T}\frac{\partial^2 l_T(\phi)}{\partial\phi\partial\phi^T}\Big| = o_p(1).$$

From standard theory for GARCH models, we further know that

$$\frac{1}{T}\frac{\partial^2 l_T(\bar{\phi})}{\partial\phi\partial\phi^T} \xrightarrow{P} J$$

with some invertible deterministic matrix J. This yields $(A.14)$. □

In order to complete the proof of asymptotic normality of the GARCH estimates we still need to show that equations (B.15) – (B.17) are fulfilled for the terms (D_c), $(D_{V,j})$, and $(D_{B,j})$. We begin with the expansion of the variance components $(D_{V,j})$, as this is the technically most interesting part.

Lemma B.2.1. *It holds that*

$$(D_{V,j}) = \frac{1}{\sqrt{T}}\sum_{s=1}^{T} g_{j,D}\Big(\frac{s}{T}, X_s\Big)u_s + o_p(1)$$

with

$$g_{j,D}\Big(\frac{s}{T}, X_s\Big) = g_{j,D}^{NW}(X_s^j) + g_{j,D}^{SBF}\Big(\frac{s}{T}, X_s\Big)$$

for $j = 0,\dots,d$. The functions $g_{j,D}^{NW}$ and $g_{j,D}^{SBF}$ are absolutely bounded. Their exact form is given in the proof (see (B.24) and (B.27) – (B.29)).

Proof. We start by giving a detailed exposition of the proof for $j \neq 0$. By Theorem B.1.1, the stochastic part $\widetilde{m}_j^A$ of the smooth backfitting estimate $\widetilde{m}_j$ has the expansion

$$\widetilde{m}_j^A(x_j) = \widehat{m}_j^A(x_j) + \frac{1}{T}\sum_{s=1}^T r_{j,s}(x_j)u_s + o_p\Big(\frac{1}{\sqrt{T}}\Big)$$

uniformly in x_j, where $\widehat{m}_j^A$ is the stochastic part of the Nadaraya-Watson pilot estimate and the function $r_{j,s}(\cdot) = r_j(\frac{s}{T}, X_s, \cdot)$ is Lipschitz continuous and absolutely bounded.

With this result, we can decompose $(D_{V,j})$ as follows:

$$\begin{aligned}
(D_{V,j}) &= \frac{1}{\sqrt{T}}\sum_{t=1}^T \frac{\partial v_t^2}{\partial \phi_i}\frac{1}{\sigma_t^2\sigma_t^2}\sum_{k=1}^{t-1} ab^{k-1}\varepsilon_{t-k}^2\widetilde{m}_j^A(X_{t-k}^j) \\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^T\sum_{k=1}^{t-1} ab^{k-1}\varepsilon_{t-k}^2\frac{\partial v_t^2}{\partial \phi_i}\frac{1}{\sigma_t^2\sigma_t^2}\widehat{m}_j^A(X_{t-k}^j) \\
&\quad + \frac{1}{\sqrt{T}}\sum_{t=1}^T\sum_{k=1}^{t-1} ab^{k-1}\varepsilon_{t-k}^2\frac{\partial v_t^2}{\partial \phi_i}\frac{1}{\sigma_t^2\sigma_t^2}\Big[\frac{1}{T}\sum_{s=1}^T r_{j,s}(X_{t-k}^j)u_s\Big] + o_p(1) \\
&=: (D_{V,j}^{NW}) + (D_{V,j}^{SBF}) + o_p(1).
\end{aligned}$$

In the following, we will give the exact arguments needed to treat $(D_{V,j}^{NW})$. The line of argument for $(D_{V,j}^{SBF})$ is essentially identical although some of the steps are easier due to the properties of the $r_{j,s}$ functions.

W.l.o.g set $\phi_i = a$ and let $m_{i,k} = \max\{k+1, i+1\}$. Using $\partial v_t^2/\partial a = \sum_{i=1}^{t-1} b^{i-1}\varepsilon_{t-i}^2$ and $\widehat{m}_j^A(x_j) = \frac{1}{T}\sum_{s=1}^T K_h(x_j, X_s^j)u_s / \frac{1}{T}\sum_{v=1}^T K_h(x_j, X_v^j)$, we get

$$\begin{aligned}
(D_{V,j}^{NW}) &= \sum_{k=1}^{T-1} ab^{k-1}\sum_{i=1}^{T-1} b^{i-1} \qquad\qquad \text{(B.20)} \\
&\quad \times \Big[\frac{1}{\sqrt{T}}\sum_{s=1}^T \frac{1}{T}\sum_{t=m_{i,k}}^T \frac{K_h(X_{t-k}^j, X_s^j)}{\frac{1}{T}\sum_{v=1}^T K_h(X_{t-k}^j, X_v^j)}\frac{1}{\sigma_t^2\sigma_t^2}\varepsilon_{t-k}^2\varepsilon_{t-i}^2 u_s\Big].
\end{aligned}$$

In a first step, we replace the sum $\frac{1}{T}\sum_{v=1}^T K_h(X_{t-k}^j, X_v^j)$ in (B.20) by a term which only depends on X_{t-k}^j and show that the resulting error is asymptotically negligible. Let $q_j(x_j) = \int_0^1 K_h(x_j, w)dw\ p_j(x_j)$. Furthermore define

$$
\begin{aligned}
B_j(x_j) &= \frac{1}{T}\sum_{v=1}^{T} \mathbb{E}[K_h(x_j, X_v^j)] - q_j(x_j) \\
V_j(x_j) &= \frac{1}{T}\sum_{v=1}^{T} \big(K_h(x_j, X_v^j) - \mathbb{E}[K_h(x_j, X_v^j)]\big).
\end{aligned}
$$

Notice that $\sup_{x_j\in[0,1]} |B_j(x_j)| = O_p(h)$ and $\sup_{x_j\in[0,1]} |V_j(x_j)| = O_p(\sqrt{\log T/Th})$. From the identity $\frac{1}{T}\sum_{v=1}^{T} K_h(x_j, X_v^j) = q_j(x_j) + B_j(x_j) + V_j(x_j)$ and a second order Taylor expansion of $(1+x)^{-1}$ we arrive at

$$
\begin{aligned}
\frac{1}{\frac{1}{T}\sum_{v=1}^{T} K_h(x_j, X_v^j)} &= \frac{1}{q_j(x_j)}\Big(1 + \frac{B_j(x_j) + V_j(x_j)}{q_j(x_j)}\Big)^{-1} \\
&= \frac{1}{q_j(x_j)}\Big(1 - \frac{B_j(x_j) + V_j(x_j)}{q_j(x_j)} + O_p(h^2)\Big)
\end{aligned}
\tag{B.21}
$$

uniformly in x_j. Plugging this decomposition into $(B.20)$, we obtain

$$
\begin{aligned}
(D_{V,j}^{NW}) = &\sum_{k=1}^{T-1} ab^{k-1} \sum_{i=1}^{T-1} b^{i-1}\Big[\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\frac{1}{T}\sum_{t=m_{i,k}}^{T} \frac{K_h(X_{t-k}^j, X_s^j)}{q_j(X_{t-k}^j)}\frac{1}{\sigma_t^2\sigma_t^2}\varepsilon_{t-k}^2\varepsilon_{t-i}^2 u_s\Big] \\
&- (D_{V,j}^{NW,B}) - (D_{V,j}^{NW,V}) + o_p(1)
\end{aligned}
$$

with

$$
\begin{aligned}
(D_{V,j}^{NW,B}) &= \sum_{k=1}^{T-1} ab^{k-1} \sum_{i=1}^{T-1} b^{i-1}\Big[\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\frac{1}{T}\sum_{t=m_{i,k}}^{T} K_h(X_{t-k}^j, X_s^j)\frac{B_j(X_{t-k}^j)}{q_j^2(X_{t-k}^j)}\frac{1}{\sigma_t^2\sigma_t^2}\varepsilon_{t-k}^2\varepsilon_{t-i}^2 u_s\Big] \\
(D_{V,j}^{NW,V}) &= \sum_{k=1}^{T-1} ab^{k-1} \sum_{i=1}^{T-1} b^{i-1}\Big[\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\frac{1}{T}\sum_{t=m_{i,k}}^{T} K_h(X_{t-k}^j, X_s^j)\frac{V_j(X_{t-k}^j)}{q_j^2(X_{t-k}^j)}\frac{1}{\sigma_t^2\sigma_t^2}\varepsilon_{t-k}^2\varepsilon_{t-i}^2 u_s\Big].
\end{aligned}
$$

As $\sup_{x_j\in I_h} |B_j(x_j)| = O_p(h^2)$ and $\sup_{x_j\in I_h^c} |B_j(x_j)| = O_p(h)$, we can proceed similarly to the proof of Lemma $B.2.3$ later on to show that $(D_{V,j}^{NW,B}) = o_p(1)$. Next we will show that $(D_{V,j}^{NW,V}) = o_p(1)$. Let $\mathbb{E}_v[\cdot]$ denote the expectation with respect to the variables indexed by v, then

$$\begin{aligned}
\big|(D_{V,j}^{NW,V})\big| &= \Big| \sum_{k=1}^{T-1} ab^{k-1} \sum_{i=1}^{T-1} b^{i-1} \Big[\frac{1}{\sqrt{T}} \sum_{s=1}^{T} \frac{1}{T} \sum_{t=m_{i,k}}^{T} \frac{K_h(X_{t-k}^j, X_s^j)}{q_j^2(X_{t-k}^j)} \frac{1}{\sigma_t^2 \sigma_t^2} \varepsilon_{t-k}^2 \varepsilon_{t-i}^2 \\
&\qquad \times \Big(\frac{1}{T} \sum_{v=1}^{T} (K_h(X_{t-k}^j, X_v^j) - \mathbb{E}_v[K_h(X_{t-k}^j, X_v^j)]) \Big) u_s \Big] \Big| \\
&\leq \sum_{k=1}^{T-1} ab^{k-1} \sum_{i=1}^{T-1} b^{i-1} \Big(\frac{1}{\sqrt{T}} \sum_{t=m_{i,k}}^{T} \Big| \frac{1}{q_j^2(X_{t-k}^j)} \frac{1}{\sigma_t^2 \sigma_t^2} \varepsilon_{t-k}^2 \varepsilon_{t-i}^2 \Big| \\
&\qquad \times \sup_{x_j \in [0,1]} \Big| \frac{1}{T} \sum_{v=1}^{T} (K_h(x_j, X_v^j) - \mathbb{E}_v[K_h(x_j, X_v^j)]) \Big| \\
&\qquad \times \sup_{x_j \in [0,1]} \Big| \frac{1}{T} \sum_{s=1}^{T} K_h(x_j, X_s^j) u_s \Big| \Big) \\
&= O_p\Big(\frac{\log T}{Th}\Big) \sum_{k=1}^{T-1} ab^{k-1} \sum_{i=1}^{T-1} b^{i-1} \underbrace{\Big(\frac{1}{\sqrt{T}} \sum_{t=m_{i,k}}^{T} \Big| \frac{1}{q_j^2(X_{t-k}^j)} \frac{1}{\sigma_t^2 \sigma_t^2} \varepsilon_{t-k}^2 \varepsilon_{t-i}^2 \Big| \Big)}_{=O_p(\sqrt{T}) \text{ by Markov's inequality}} \\
&= O_p\Big(\frac{\log T}{Th} \sqrt{T}\Big) = o_p(1).
\end{aligned}$$

Together with the fact that $(D_{V,j}^{NW,B}) = o_p(1)$, this yields

$$(D_{V,j}^{NW}) = \sum_{k=1}^{T-1} ab^{k-1} \sum_{i=1}^{T-1} b^{i-1} \Big[\frac{1}{\sqrt{T}} \sum_{s=1}^{T} \frac{1}{T} \sum_{t=m_{i,k}}^{T} K_h(X_{t-k}^j, X_s^j) \mu_t^{i,k} u_s \Big] + o_p(1), \quad \text{(B.22)}$$

where we use the shorthand $\mu_t^{i,k} = (q_j(X_{t-k}^j)\sigma_t^2\sigma_t^2)^{-1}\varepsilon_{t-k}^2\varepsilon_{t-i}^2$.

In the next step, we replace the inner sum over t in (B.22) by a term that only depends on X_s^j and show that the resulting error can be asymptotically neglected. Define

$$\xi(X_{t-k}^j, X_s^j) := \xi_t^{i,k}(X_{t-k}^j, X_s^j) := K_h(X_{t-k}^j, X_s^j)\mu_t^{i,k} - \mathbb{E}_{-s}[K_h(X_{t-k}^j, X_s^j)\mu_t^{i,k}],$$

where $\mathbb{E}_{-s}[\cdot]$ is the expectation with respect to all variables except for those depending on the index s. With the above notation at hand, we can write

$$(D_{V,j}^{NW}) = \sum_{k=1}^{T-1} ab^{k-1} \sum_{i=1}^{T-1} b^{i-1} \Big[\frac{1}{\sqrt{T}} \sum_{s=1}^{T} \frac{1}{T} \sum_{t=m_{i,k}}^{T} \mathbb{E}_{-s}[K_h(X_{t-k}^j, X_s^j)\mu_t^{i,k}] u_s \Big] + (R_{V,j}^{NW}) + o_p(1),$$

where

$$(R_{V,j}^{NW}) = \sum_{k=1}^{T-1} ab^{k-1} \sum_{i=1}^{T-1} b^{i-1} \Big[\frac{1}{\sqrt{T}} \sum_{s=1}^{T} \frac{1}{T} \sum_{t=m_{i,k}}^{T} \xi(X_{t-k}^j, X_s^j) u_s \Big] \tag{B.23}$$

$$= \sum_{k=1}^{C_2 \log T} ab^{k-1} \sum_{i=1}^{C_2 \log T} b^{i-1} \Big[\frac{1}{\sqrt{T}} \sum_{s=1}^{T} \frac{1}{T} \sum_{t=m_{i,k}}^{T} \xi(X_{t-k}^j, X_s^j) u_s \Big] + o_p(1)$$

for some sufficiently large constant $C_2 > 0$. Once we show that $(R_{V,j}^{NW}) = o_p(1)$, we are left with

$$(D_{V,j}^{NW}) = \sum_{k=1}^{T-1} ab^{k-1} \sum_{i=1}^{T-1} b^{i-1} \Big[\frac{1}{\sqrt{T}} \sum_{s=1}^{T} \frac{1}{T} \sum_{t=m_{i,k}}^{T} \mathbb{E}_{-s}[K_h(X_{t-k}^j, X_s^j)\mu_t^{i,k}] u_s \Big] + o_p(1)$$

$$= \frac{1}{\sqrt{T}} \sum_{s=1}^{T} \Big(\sum_{k=1}^{T-1} ab^{k-1} \sum_{i=1}^{T-1} b^{i-1} \frac{T - m_{i,k}}{T} \mathbb{E}_{-s}[K_h(X_{-k}^j, X_s^j)\mu_0^{i,k}] \Big) u_s + o_p(1).$$

As the terms with $i, k \geq C_2 \log T$ are asymptotically negligible, we can expand the i and k sums to infinity, which yields

$$(D_{V,j}^{NW}) = \frac{1}{\sqrt{T}} \sum_{s=1}^{T} \Big(\sum_{k=1}^{\infty} ab^{k-1} \sum_{i=1}^{\infty} b^{i-1} \mathbb{E}_{-s}[K_h(X_{-k}^j, X_s^j)\mu_0^{i,k}] \Big) u_s + o_p(1) \tag{B.24}$$

$$=: \frac{1}{\sqrt{T}} \sum_{s=1}^{T} g_{j,D}^{NW}(X_s^j) u_s + o_p(1)$$

with

$$\mu_0^{i,k} = \frac{1}{q_j(X_{-k}^j)} \frac{1}{\sigma_0^2 \sigma_0^2} \varepsilon_{-k}^2 \varepsilon_{-i}^2$$

$$q_j(X_{-k}^j) = \int_0^1 K_h(X_{-k}^j, w) dw \; p_j(X_{-k}^j).$$

Thus it remains to show that $(R^{NW}_{V,j}) = o_p(1)$, which requires a lot of care. We will prove that the term in square brackets in (B.23) is $o_p(1)$ uniformly over $i, k \le C_2 \log T$, which yields the desired result. It is easily seen that

$$\begin{aligned} P &:= \mathbb{P}\Big(\max_{i,k \le C_2 \log T} \Big| \frac{1}{\sqrt{T}} \sum_{s=1}^{T} \frac{1}{T} \sum_{t=m_{i,k}}^{T} \xi(X^j_{t-k}, X^j_s) u_s \Big| > \delta \Big) \\ &\le \sum_{k=1}^{C_2 \log T} \sum_{i=1}^{C_2 \log T} \underbrace{\mathbb{P}\Big(\Big| \frac{1}{\sqrt{T}} \sum_{s=1}^{T} \frac{1}{T} \sum_{t=m_{i,k}}^{T} \xi(X^j_{t-k}, X^j_s) u_s \Big| > \delta \Big)}_{=:P_{i,k}} \end{aligned}$$

for a fixed $\delta > 0$. Then by Chebychev's inequality

$$\begin{aligned} P_{i,k} &\le \frac{1}{T^3 \delta^2} \sum_{s,s'=1}^{T} \sum_{t,t'=m_{i,k}}^{T} \mathbb{E}\Big[\xi(X^j_{t-k}, X^j_s) u_s \xi(X^j_{t'-k}, X^j_{s'}) u_{s'} \Big] \\ &= \frac{1}{T^3 \delta^2} \sum_{(s,s',t,t') \notin \Gamma_{i,k}} \mathbb{E}\Big[\xi(X^j_{t-k}, X^j_s) u_s \xi(X^j_{t'-k}, X^j_{s'}) u_{s'} \Big] \\ &\quad + \frac{1}{T^3 \delta^2} \sum_{(s,s',t,t') \in \Gamma_{i,k}} \mathbb{E}\Big[\xi(X^j_{t-k}, X^j_s) u_s \xi(X^j_{t'-k}, X^j_{s'}) u_{s'} \Big] \\ &=: P^1_{i,k} + P^2_{i,k}, \end{aligned}$$

where $\Gamma_{i,k}$ is the set of tuples (s, s', t, t') with $1 \le s, s' \le T$ and $m_{i,k} \le t, t' \le T$ such that one index is separated from the others. We say that an index, for instance t, is separated from the others if $\min\{|t - t'|, |t - s|, |t - s'|\} > C_3 \log T$, i.e. if it is further away from the other indices than $C_3 \log T$ for a constant C_3 to be chosen later on. We now analyse $P^1_{i,k}$ and $P^2_{i,k}$ separately.

(a) First consider $P^1_{i,k}$. If a tuple (s, s', t, t') is not an element of $\Gamma_{i,k}$, then no index can be separated from the others. Since the index t cannot be separated, there exists an index, say t', such that $|t - t'| \le C_3 \log T$. Now take an index different from t and t', for instance s. Then by the same argument, there exists an index, say s', such that $|s - s'| \le C_3 \log T$. As a consequence, the number of tuples $(s, s', t, t') \notin \Gamma_{i,k}$ is smaller than $CT^2 (\log T)^2$ for some constant C. Using (C(V12)), this suffices to infer that

$$\left|P_{i,k}^1\right| \le \frac{1}{T^3\delta^2} \sum_{(s,s',t,t')\notin\Gamma_{i,k}} \frac{C}{h^2} \le \frac{C}{\delta^2}\frac{(\log T)^2}{Th^2}.$$

Hence, $|P_{i,k}^1| \le C\delta^{-2}(\log T)^{-3}$ uniformly in i and k.

(b) The term $P_{i,k}^2$ is more difficult to handle. We start by taking a cover $\{I_m\}_{m=1}^{M_T}$ of the compact support $[0,1]$ of X_{t-k}^j. The elements I_m are intervals of length $1/M_T$ given by $I_m = [\frac{m-1}{M_T}, \frac{m}{M_T})$ for $m = 1, \dots, M_T - 1$ and $I_{M_T} = [1 - \frac{1}{M_T}, 1]$. The midpoint of the interval I_m is denoted by x_m. With this, we can write

$$K_h(X_{t-k}^j, X_s^j) = \sum_{m=1}^{M_T} I(X_{t-k}^j \in I_m) \times \big[K_h(x_m, X_s^j) + (K_h(X_{t-k}^j, X_s^j) - K_h(x_m, X_s^j))\big]. \tag{B.25}$$

Using (B.25), we can further write

$$\begin{aligned}\xi(X_{t-k}^j, X_s^j) &= \sum_{m=1}^{M_T} \Big\{ I(X_{t-k}^j \in I_m) K_h(x_m, X_s^j)\mu_t^{i,k} \\ &\qquad - \mathbb{E}_{-s}[I(X_{t-k}^j \in I_m) K_h(x_m, X_s^j)\mu_t^{i,k}]\Big\} \\ &\quad + \sum_{m=1}^{M_T} \Big\{ I(X_{t-k}^j \in I_m)(K_h(X_{t-k}^j, X_s^j) - K_h(x_m, X_s^j))\mu_t^{i,k} \\ &\qquad - \mathbb{E}_{-s}[I(X_{t-k}^j \in I_m)(K_h(X_{t-k}^j, X_s^j) - K_h(x_m, X_s^j))\mu_t^{i,k}]\Big\} \\ &=: \xi_1(X_{t-k}^j, X_s^j) + \xi_2(X_{t-k}^j, X_s^j)\end{aligned}$$

and

$$\begin{aligned}P_{i,k}^2 &= \frac{1}{T^3\delta^2} \sum_{(s,s',t,t')\in\Gamma_{i,k}} \mathbb{E}\big[\xi_1(X_{t-k}^j, X_s^j) u_s \xi(X_{t'-k}^j, X_{s'}^j) u_{s'}\big] \\ &\quad + \frac{1}{T^3\delta^2} \sum_{(s,s',t,t')\in\Gamma_{i,k}} \mathbb{E}\big[\xi_2(X_{t-k}^j, X_s^j) u_s \xi(X_{t'-k}^j, X_{s'}^j) u_{s'}\big] \\ &=: P_{i,k}^{2,1} + P_{i,k}^{2,2}.\end{aligned}$$

We first consider $P^{2,2}_{i,k}$. Set $M_T = CT(\log T)^3 h^{-3}$ and exploit the Lipschitz continuity of the kernel K to get that $|K_h(X^j_{t-k}, X^j_s) - K_h(x_m, X^j_s)| \leq \frac{C}{h^2}|X^j_{t-k} - x_m|$. This gives us

$$\begin{aligned}\big|\xi_2(X^j_{t-k}, X^j_s)\big| &\leq \frac{C}{h^2}\sum_{m=1}^{M_T}\Big(\underbrace{I(X^j_{t-k} \in I_m)|X^j_{t-k} - x_m|}_{\leq I(X^j_{t-k}\in I_m)M_T^{-1}}\mu^{i,k}_t \\ &\qquad + \mathbb{E}\big[\underbrace{I(X^j_{t-k} \in I_m)|X^j_{t-k} - x_m|}_{\leq I(X^j_{t-k}\in I_m)M_T^{-1}}\mu^{i,k}_t\big]\Big) \\ &\leq \frac{C}{M_T h^2}\big(\mu^{i,k}_t + \mathbb{E}[\mu^{i,k}_t]\big).\end{aligned} \tag{B.26}$$

Plugging (B.26) into the expression for $P^{2,2}_{i,k}$, we arrive at

$$\begin{aligned}\big|P^{2,2}_{i,k}\big| &\leq \frac{1}{T^3\delta^2}\sum_{(s,s',t,t')\in\Gamma_{i,k}}\mathbb{E}\Big[\big|\xi_2(X^j_{t-k}, X^j_s)\big|\big|u_s\xi(X^j_{t'-k}, X^j_{s'})u_{s'}\big|\Big] \\ &\leq \frac{1}{T^3\delta^2}\frac{C}{M_T h^2}\sum_{(s,s',t,t')\in\Gamma_{i,k}}\underbrace{\mathbb{E}\big[(\mu^{i,k}_t + \mathbb{E}[\mu^{i,k}_t])|u_s\xi(X^j_{t'-k}, X^j_{s'})u_{s'}|}_{\leq Ch^{-1}}\big] \\ &\leq \frac{C}{\delta^2}\frac{1}{(\log T)^3}.\end{aligned}$$

We next turn to $P^{2,1}_{i,k}$. Write

$$P^{2,1}_{i,k} = \frac{1}{T^3\delta^2}\sum_{(s,s',t,t')\in\Gamma_{i,k}}\Big(\sum_{m=1}^{M_T} S_m\Big)$$

with

$$\begin{aligned}S_m = \mathbb{E}\Big[&\big\{I(X^j_{t-k} \in I_m)K_h(x_m, X^j_s)\mu^{i,k}_t - \mathbb{E}_{-s}[I(X^j_{t-k} \in I_m)K_h(x_m, X^j_s)\mu^{i,k}_t]\big\} \\ &\times u_s\xi(X^j_{t'-k}, X^j_{s'})u_{s'}\Big]\end{aligned}$$

and assume that an index, w.l.o.g. t, can be separated from the others.

Choosing $C_3 \gg C_2$, we get

$$\begin{aligned} S_m &= \mathrm{Cov}\Big(I(X^j_{t-k} \in I_m)\mu^{i,k}_t - \mathbb{E}[I(X^j_{t-k} \in I_m)\mu^{i,k}_t], \\ &\qquad K_h(x_m, X^j_s)u_s\xi(X^j_{t'-k}, X^j_{s'})u_{s'}\Big) \\ &\leq \frac{C}{h^2}(\alpha([C_3 - C_2]\log T))^{1-\frac{2}{p}} \leq \frac{C}{h^2}(a^{(C_3-C_2)\log T})^{1-\frac{2}{p}} \\ &\leq \frac{C}{h^2}T^{-C_4} \end{aligned}$$

with some $C_4 > 0$ by Davydov's inequality, where p is chosen slightly larger than 2. Note that the above bound is independent of i and k and that we can make C_4 arbitrarily large by choosing C_3 large enough. This shows that $|P^{2,1}_{i,k}| \leq C\delta^{-2}(\log T)^{-3}$ uniformly in i and k with some constant C.

Combining (a) and (b) yields that $P \to 0$ for each fixed $\delta > 0$. This implies that

$$(R^{NW,V}_{V,j}) = o_p(1),$$

which completes the proof for the term $(D^{NW}_{V,j})$.

As stated at the beginning of the proof, the term $(D^{SBF}_{V,j})$ can be treated in exactly the same way. Following analogous arguments as above, one obtains

$$(D^{SBF}_{V,j}) = \sum_{k=1}^{T-1} ab^{k-1} \sum_{i=1}^{T-1} b^{i-1}\Big[\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\frac{1}{T}\sum_{t=m_{i,k}}^{T} \mathbb{E}_{-s}[r_{j,s}(X^j_{t-k})\zeta^{i,k}_t]\ u_s\Big] + o_p(1) \tag{B.27}$$

$$\begin{aligned} &= \frac{1}{\sqrt{T}}\sum_{s=1}^{T}\Big(\sum_{k=1}^{\infty} ab^{k-1}\sum_{i=1}^{\infty} b^{i-1}\mathbb{E}_{-s}[r_{j,s}(X^j_{-k})\zeta^{i,k}_0]\Big)u_s + o_p(1) \\ &=: \frac{1}{\sqrt{T}}\sum_{s=1}^{T} g^{SBF}_{j,D}\Big(\frac{s}{T}, X_s\Big)u_s + o_p(1) \end{aligned}$$

with $\zeta^{i,k}_t = (\sigma^2_t\sigma^2_t)^{-1}\varepsilon^2_{t-k}\varepsilon^2_{t-i}$.

Finally, the proofs for $j = 0$ are very similar but somewhat simpler and are thus omitted here. For completeness we provide the functions $g_{0,D}^{NW}$ and $g_{0,D}^{SBF}$:

$$g_{0,D}^{NW}\Big(\frac{s}{T}\Big) = \Big(\sum_{k=1}^{\infty} ab^{k-1} \sum_{i=1}^{\infty} b^{i-1}\mathbb{E}\big[\frac{1}{\sigma_0^2\sigma_0^2}\varepsilon_{-k}^2\varepsilon_{-i}^2\big]\Big)\int_0^1 \frac{K_h(\frac{s}{T},v)}{\int_0^1 K_h(v,w)dw}dv \tag{B.28}$$

$$g_{0,D}^{SBF}\Big(\frac{s}{T},X_s\Big) = \Big(\sum_{k=1}^{\infty} ab^{k-1} \sum_{i=1}^{\infty} b^{i-1}\mathbb{E}\big[\frac{1}{\sigma_0^2\sigma_0^2}\varepsilon_{-k}^2\varepsilon_{-i}^2\big]\Big)\int_0^1 r_{0,s}(w)dw. \tag{B.29}$$

□

Lemma B.2.2. *It holds that*

$$(D_c) = \frac{1}{\sqrt{T}}\sum_{s=1}^{T} g_{c,D}u_s$$

with

$$g_{c,D} = \sum_{k=1}^{\infty} ab^{k-1} \sum_{i=1}^{\infty} b^{i-1}\mathbb{E}\Big[\frac{1}{\sigma_0^2\sigma_0^2}\varepsilon_{-i}^2\varepsilon_{-k}^2\Big].$$

Proof. Using the fact that

$$\widetilde{m}_c = \frac{1}{T}\sum_{s=1}^{T} Z_{s,T} = m_c + \frac{1}{T}\sum_{s=1}^{T} m_0\Big(\frac{s}{T}\Big) + \sum_{j=1}^{d}\frac{1}{T}\sum_{s=1}^{T} m_j(X_s^j) + \frac{1}{T}\sum_{s=1}^{T} u_s,$$

we arrive at

$$(D_c) = -\Big(\frac{1}{T}\sum_{t=1}^{T} G_t \sum_{k=1}^{t-1} ab^{k-1}\varepsilon_{t-k}^2\Big)\Big(\frac{1}{\sqrt{T}}\sum_{s=1}^{T} u_s\Big)$$

with $G_t = \frac{\partial v_t^2}{\partial \phi_i}(\sigma_t^2\sigma_t^2)^{-1}$. Now let $m_{i,k} = \max\{k+1, i+1\}$ and assume w.l.o.g. that $\phi_i = a$. Then

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T} G_t \sum_{k=1}^{t-1} ab^{k-1}\varepsilon_{t-k}^2 &= \frac{1}{T}\sum_{t=1}^{T}\Big(\sum_{i=1}^{t-1} b^{i-1}\varepsilon_{t-i}^2\Big)\frac{1}{\sigma_t^2\sigma_t^2}\sum_{k=1}^{t-1} ab^{k-1}\varepsilon_{t-k}^2 \\ &= \sum_{k=1}^{C_2\log T} ab^{k-1} \sum_{i=1}^{C_2\log T} b^{i-1}\frac{1}{T}\sum_{t=m_{i,k}}^{T}\frac{1}{\sigma_t^2\sigma_t^2}\varepsilon_{t-i}^2\varepsilon_{t-k}^2 + o_p(1)\end{aligned}$$

with some sufficiently large constant C_2. Using Chebychev's inequality and exploiting the mixing properties of the variables involved, one can show that

$$\max_{i,k\le C_2\log T}\frac{1}{T}\sum_{t=m_{i,k}}^{T}\Big(\frac{1}{\sigma_t^2\sigma_t^2}\varepsilon_{t-i}^2\varepsilon_{t-k}^2-\mathbb{E}\Big[\frac{1}{\sigma_t^2\sigma_t^2}\varepsilon_{t-i}^2\varepsilon_{t-k}^2\Big]\Big)=o_p(1).$$

This allows us to infer that

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}G_t\sum_{k=1}^{t-1}ab^{k-1}\varepsilon_{t-k}^2 &= \sum_{k=1}^{C_2\log T}ab^{k-1}\sum_{i=1}^{C_2\log T}b^{i-1}\frac{1}{T}\sum_{t=m_{i,k}}^{T}\mathbb{E}\Big[\frac{1}{\sigma_t^2\sigma_t^2}\varepsilon_{t-i}^2\varepsilon_{t-k}^2\Big]+o_p(1)\\ &= \sum_{k=1}^{\infty}ab^{k-1}\sum_{i=1}^{\infty}b^{i-1}\mathbb{E}\Big[\frac{1}{\sigma_0^2\sigma_0^2}\varepsilon_{-i}^2\varepsilon_{-k}^2\Big]+o_p(1),\end{aligned}$$

which completes the proof. □

Lemma B.2.3. *It holds that*

$$(D_{B,j})=o_p(1)$$

for $j=0,\dots,d$.

Proof. We start by considering the case $j=0$: Define

$$\begin{aligned}J_h&=\{t\in\{1,\dots,T\}:C_1h\le\frac{t}{T}\le 1-C_1h\}\\ J_{h,c}^u&=\{t\in\{1,\dots,T\}:1-C_1h<\frac{t}{T}\}\\ J_{h,c}^l&=\{t\in\{1,\dots,T\}:\frac{t}{T}<C_1h\},\end{aligned}$$

where $[-C_1,C_1]$ is the support of K. Using the uniform convergence rates from Theorem B.1.2 and assuming w.l.o.g. that $\phi_i=a$, we get

$$\begin{aligned}|(D_{B,0})|=\Big|&\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial v_t^2}{\partial a}\frac{1}{\sigma_t^2\sigma_t^2}\sum_{k=1}^{t-1}ab^{k-1}\varepsilon_{t-k}^2\\ &\times\Big[m_0\Big(\frac{t-k}{T}\Big)-\widetilde{m}_0^B\Big(\frac{t-k}{T}\Big)-\frac{1}{T}\sum_{s=1}^{T}m_0\Big(\frac{s}{T}\Big)\Big]\Big|\end{aligned}$$

$$\begin{aligned}
&\leq O_p(h)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{i=1}^{t-1}b^{i-1}\sum_{k=1}^{t-1}ab^{k-1}\varepsilon_{t-i}^2\varepsilon_{t-k}^2 I(t-k\in J_{h,c}^l)\\
&\quad+O_p(h)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{i=1}^{t-1}b^{i-1}\sum_{k=1}^{t-1}ab^{k-1}\varepsilon_{t-i}^2\varepsilon_{t-k}^2 I(t-k\in J_{h,c}^u)\\
&\quad+O_p(h^2)\frac{C}{\sqrt{T}}\sum_{t=1}^{T}\sum_{i=1}^{t-1}b^{i-1}\sum_{k=1}^{t-1}ab^{k-1}\varepsilon_{t-i}^2\varepsilon_{t-k}^2 I(t-k\in J_h)\\
&=: (D_{B,0}^{J_{h,c}^l})+(D_{B,0}^{J_{h,c}^u})+(D_{B,0}^{J_h}).
\end{aligned}$$

By Markov's inequality, $(D_{B,0}^{J_h}) = O_p(h^2\sqrt{T}) = o_p(1)$. Recognizing that

(i) $I(t-k\in J_{h,c}^u)\leq I(t\in J_{h,c}^u)$ for all $k\in\{0,\dots,t-1\}$

(ii) $\sum_{t=1}^{T} I(t\in J_{h,c}^u)\leq C_1Th$,

we get $(D_{B,0}^{J_{h,c}^u}) = O_p(h^2\sqrt{T}) = o_p(1)$ by another appeal to Markov's inequality. This just leaves $(D_{B,0}^{J_{h,c}^l})$, which is a bit more tedious. By a change of variable $j=t-k$,

$$\begin{aligned}
(D_{B,0}^{J_{h,c}^l}) &\leq O_p(h)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{i=1}^{t-1}b^{i-1}\varepsilon_{t-i}^2\sum_{j=1}^{t-1}ab^{t-j-1}\varepsilon_j^2 I(j\in J_{h,c}^l)\\
&= O_p(h)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{i=1}^{t-1}b^{i-1}\varepsilon_{t-i}^2 I\Big(\Big[\frac{t}{2}\Big]\in J_{h,c}^l\Big)\sum_{j=1}^{t-1}ab^{t-j-1}\varepsilon_j^2 I(j\in J_{h,c}^l)\\
&\quad+O_p(h)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{i=1}^{t-1}b^{i-1}\varepsilon_{t-i}^2 I\Big(\Big[\frac{t}{2}\Big]\notin J_{h,c}^l\Big)\sum_{j=1}^{t-1}ab^{t-j-1}\varepsilon_j^2 I(j\in J_{h,c}^l)\\
&=: (A)+(B),
\end{aligned}$$

where $[x]$ denotes the smallest integer larger than x. Realizing that $[t/2]\in J_{h,c}^l$ only if $t<2C_1hT$, we get $(A) = O_p(h^2\sqrt{T}) = o_p(1)$ once again by Markov's inequality. In (B) we can truncate the summation over j at $[t/2]-1$, as $I(j\in J_{h,c}^l)=0$ for

$j \geq [t/2]$ if $[t/2] \notin J_{h,c}^l$. We thus obtain

$$(B) \leq O_p(h) \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \sum_{i=1}^{t-1} b^{i-1} \varepsilon_{t-i}^2 \sum_{j=1}^{[t/2]-1} a b^{t-j-1} \varepsilon_j^2$$

$$= O_p(h) \frac{1}{\sqrt{T}} \sum_{t=1}^{T} b^{[t/2]} \sum_{i=1}^{t-1} b^{i-1} \sum_{j=1}^{[t/2]-1} a b^{t-j-1-[t/2]} \varepsilon_{t-i}^2 \varepsilon_j^2.$$

By a final appeal to Markov's inequality we arrive at

$$(B) = O_p(h) O_p\Big(\frac{1}{\sqrt{T}}\Big) = o_p(1),$$

thus completing the proof for $j = 0$.

Next consider the case $j \neq 0$. Similarly to before, we have

$$\begin{aligned} |(D_{B,j})| &\leq O_p(h^2) \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \sum_{i=1}^{t-1} b^{i-1} \sum_{k=1}^{t-1} a b^{k-1} \varepsilon_{t-i}^2 \varepsilon_{t-k}^2 I(X_{t-k}^j \in I_h) \\ &\quad + O_p(h) \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \sum_{i=1}^{t-1} b^{i-1} \sum_{k=1}^{t-1} a b^{k-1} \varepsilon_{t-i}^2 \varepsilon_{t-k}^2 I(X_{t-k}^j \notin I_h) \\ &= O_p(h^2 \sqrt{T}) + O_p\Big(\frac{h}{\sqrt{T}}\Big) \underbrace{\sum_{t=1}^{T} \sum_{i=1}^{t-1} b^{i-1} \sum_{k=1}^{t-1} a b^{k-1} \varepsilon_{t-i}^2 \varepsilon_{t-k}^2 I(X_{t-k}^j \notin I_h)}_{=: R_T} \end{aligned}$$

with $I_h = [2C_1 h, 1 - 2C_1 h]$ as defined in Theorem 2.4.2. Using (V12), it is easy to see that $R_T = O_p(h)$, which yields the result for $j \neq 0$. $\square$

Bibliography

Altman, Naomi S. (1990). "Kernel smoothing of data with correlated errors". In: *Journal of the American Statistical Association* 85.411, pp. 749–759.

— (1993). "Estimating error correlation in nonparametric regression". In: *Statistics & Probability Letters* 18.3, pp. 213–218.

Bollerslev, Tim (1986). "Generalized Autoregressive Conditional Heteroskedasticity". In: *Journal of Econometrics* 31.3, pp. 307–327.

Bosq, Denis (1998). *Nonparametric statistics for stochastic processes : estimation and prediction*. 2nd. Lecture notes in statistics 110. New York: Springer.

Box, George and Gwilym Jenkins (1970). *Time series analysis: Forecasting and control*. San Fransisco: Holden-Day.

Brockwell, Peter J. and Richard A. Davis (1991). *Time Series: Theory and Methods*. 2nd. Springer Series in Statistics. New York: Springer.

— (2002). *Introduction to Time Series and Forecasting*. 2nd. Springer Texts in Statistics. New York: Springer.

Canova, Fabio (1998). "Detrending and business cycle facts". In: *Journal of Monetary Economics* 41.3, pp. 475–512.

Dahlhaus, Rainer (1996a). "Asymptotic statistical inference for nonstationary processes with evolutionary spectra". In: *Athens conference on applied probability and time series analysis*. Ed. by P. Robinson and M. Rosenblatt. Vol. 2. New York: Springer.

— (1996b). "On the Kullback-Leibler information divergence of locally stationary processes". In: *Stochastic Processes and their Applications* 62.1, pp. 139–168.

Dahlhaus, Rainer (1997). "Fitting time series models to nonstationary processes". In: *The Annals of Statistics* 25.1, pp. 1–37.

— (2012). "Locally Stationary Processes". In: *Handbook of Statistics, Vol 30: Time Series-Methods and Applications*. Ed. by Tata Subba Rao, Suhasini Subba Rao, and Calyampudi Radhakrishna Rao. Oxford: Elsevier-North Holland, pp. 351–414.

Dahlhaus, Rainer and Suhasini Subba Rao (2006). "Statistical inference for time-varying ARCH processes". In: *The Annals of Statistics* 34.3, pp. 1075–1114.

Destatis (2004). "Komponentenzerlegung und Saisonbereinigung ökonomischer Zeitreihen mit dem Verfahren BV4.1". In: *Methodenberichte* 3.

Engle, Robert F. and Jose G. Rangel (2008). "The Spline-GARCH model for low-frequency volatility and its global macroeconomic causes". In: *The Review of Financial Studies* 21.3, pp. 1187–1222.

Engle, Robert F., Eric Ghysels, and Bumjean Sohn (2008). "On the Economic Sources of Stock Market Volatility (August 31, 2008)". In: *AFA 2008 New Orleans Meetings Paper*. URL: http://ssrn.com/abstract=9713104.

Eurostat (2009). "ESS Guidelines on Seasonal Adjustment". In: *Methodologies & Working Papers*. DOI: 10.2785/20152.

Feng, Yuanhua (2004). "Simultaneously Modelling Conditional Heteroskedasticity and Scale Change". In: *Econometric Theory* 20.3, pp. 563–596.

Fuller, Wayne A. (1996). *Introduction to Statistical Time Series*. 2nd. Wiley Series in Probability and Statistics. New York: Wiley.

Gijbels, Irène, Alun Pope, and Matt P. Wand (1999). "Understanding exponential smoothing via kernel regression". In: *Journal of the Statistical Royal Society. Series B* 61.1, pp. 39–50.

Gonçalves, Sílvia and Lutz Kilian (2004). "Bootstrapping autoregressions with conditional heteroskedasticity of unknown form". In: *Journal of Econometrics* 123.1, pp. 89–120.

Hafner, Christian M. and Oliver Linton (2010). "Efficient estimation of a multivariate multiplicative volatility model". In: *Journal of Econometrics* 159.1, pp. 55–73.

Hall, Peter and Ingrid Van Keilegom (2003). "Using Difference-Based Methods for Inference in Nonparametric Regression with Time Series Errors". In: *Journal of the Statistical Royal Society. Series B* 65.2, pp. 443–456.

Hall, Peter, Qi Li, and Jeffrey S. Racine (2007). "Nonparametric estimation of regression functions in the presence of irrelevant regressors". In: *The Review of Economics and Statistics* 89.4, pp. 784–789.

Hamilton, James D. (1994). *Time Series Analysis*. New Jersey: Princeton University Press.

Hansen, Bruce E. (2008). "Uniform Convergence Rates for Kernel Estimation with Dependent Data". In: *Econometric Theory* 24.3, pp. 726–748.

Härdle, Wolfgang, Helmut Lütkepohl, and Rong Chen (1997). "A Review of Nonparametric Time Series Analysis". In: *International Statistical Review* 65.1, pp. 49–72.

Hart, Jeffrey D. (1991). "Kernel regression estimation with time series errors". In: *Journal of the Statistical Royal Society. Series B (Methodological)* 53.1, pp. 173–187.

— (1994). "Automated Kernel Smoothing of Dependent Data by Using Time Series Cross-Validation". In: *Journal of the Statistical Royal Society. Series B (Methodological)* 56.3, pp. 529–542.

Hart, Jeffrey D. and Cherng-Luen Lee (2005). "Robustness of one-sided cross-validation to autocorrelation". In: *Journal of Multivariate Analysis* 92.1, pp. 77–96.

Hart, Jeffrey D. and Seongbaek Yi (1998). "One-Sided Cross-Validation". In: *Journal of the American Statistical Association* 93.442, pp. 620–631.

Harvey, Andrew C. (1985). "Trends and Cycles in Macroeconomics Time Series". In: *Journal of Business & Economic Statistics* 3.3, pp. 216–227.

— (1989). *Forecasting, structural time series models and the Kalman Filter*. Cambridge: Cambridge University Press.

Harvey, Andrew C. and Albert Jaeger (1993). "Detrending, stylized facts and the business cycle". In: *Journal of Applied Econometrics* 8.3, pp. 231–247.

Heiler, Siegfried (2001). "Nonparametric Time Series Analysis: Nonparametric Regression, Locally Weighted Regression,Autoregression, and Quantile Regression". In: *A Course in Time Series Analysis*. Ed. by Daniel Peña, George C. Tiao, and Ruey S. Tsay. Wiley Series in Probability and Statistics. New York: John Wiley Sons, Inc.

Hendry, David F. (1995). *Dynamic Econometrics*. Advanced texts in econometrics. New York: Oxford University Press.

Herrmann, Eva, Theo Gasse, and Alois Kneip (1992). "Choice of bandwidth for kernel regression when residuals are correlated". In: *Biometrika* 79.4, pp. 783–795.

Hughes, Gillian L., Suhasini Subba Rao, and Tata Subba Rao (2007). "Statistical analysis and time-series models for minimum/maximum temperatures in the Antarctic Peninsula". In: *Proceedings of the Royal Society A* 463.2077, pp. 241–259.

Lamoureux, Christopher G. and William D. Lastrapes (1990). "Persistence in Variance, Structural Change, and the GARCH Model". In: *Journal of Business & Economic Statistics* 8.2, pp. 225–234.

Lin, Tsung-Chih, Mohsen Pourahmadi, and Anton Schick (1999). "Regression Models with Time Series Errors". In: *Journal of Time Series Analysis* 20.4, pp. 425–433.

Machiwal, D., M.K. Jha, and Z. Zhang (2012). *Hydrologic Time Series Analysis: Theory and Practice*. New York: Springer.

Mammen, Enno and Byeong U. Park (2005). "Bandwidth selection for smooth backfitting in additive models". In: *The Annals of Statistics* 33.3, pp. 1260–1294.

— (2006). "A simple smooth backfitting method for additive models". In: *The Annals of Statistics* 34.5, pp. 2252–2271.

Mammen, Enno, Oliver Linton, and Jens-Perch Nielsen (1999). "The existence and asymptotic properties of a backfitting projection algorithm under weak conditions". In: *The Annals of Statistics* 27.5, pp. 1443–1490.

Masry, Elias (1996). "Multivariate local polynomial regression for time series: Uniform strong consistency and rates". In: *Journal of Time Series Analysis* 17.6, pp. 571–599.

Mikosch, Thomas and Catalin Stărică (2000). "Is it really long memory we see in financial returns?" In: *Extremes and Integrated Risk Management*. Ed. by Paul Embrechts. London: Risk Books, pp. 149–168.

— (2003). "Long-Range Dependence Effects and ARCH Modelling". In: *Theory and Applications of Long Range Dependence*. Ed. by Paul Doukhan, George Oppenheim, and Murrad S. Taqqu. Boston: Birkhäuser, pp. 439–459.

— (2004). "Non-stationarities in financial time series, the long-range dependence, and IGARCH effects". In: *The Review of Economics and Statistics* 86.1, pp. 378–390.

Mills, Terence C. and Raphael N. Markellos (2008). *The Econometric Modelling of Financial Time Series*. 3rd. Cambridge: Cambridge University Press.

Mudelsee, Manfred (2010). *Climate Time Series Analysis: Classical Statistical and Bootstrap Methods*. Atmospheric and oceanographic sciences library. New York: Springer.

Nielsen, Jens-Perch and Stefan Sperlich (2003). "Prediction of stock returns: A new way to look at it". In: *ASTIN Bulletin* 33.2, pp. 399–417.

Quinn, Barry G. and Edward J. Hannan (2001). *The Estimation and Tracking of Frequency*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge: Cambridge University Press.

Schick, Anton (1994). "Estimation of the autocorrelation coefficient in the presence of a regression trend". In: *Statistics & Probability Letters* 21.5, pp. 371–380.

Shao, Qin and L.J. Yang (2011). "Autoregressive coefficient estimation in nonparametric analysis". In: *Journal of Time Series Analysis* 32.6, pp. 587–597.

Sun, Ying, Jeffrey D. Hart, and Marc G. Genton (2012). "Nonparametric inference for periodic sequences". In: *Technometrics* 54, pp. 83–96.

Tjøstheim, Dag (2012). "Modelling Nonlinear and Nonstationary Time Series". In: *Handbook of Statistics, Vol 30: Time Series-Methods and Applications*. Ed. by

Tata Subba Rao, Suhasini Subba Rao, and Calyampudi Radhakrishna Rao. Oxford: Elsevier-North Holland, pp. 64–97.

Truong, Young K. (1991). "Nonparametric curve estimation with time series errors". In: *Journal of Statistical Planning and Inference* 28.2, pp. 167–183.

Truong, Young K. and Charles J. Stone (1994). "Semiparametric time series regression". In: *Journal of Time Series Analysis* 15.4, pp. 405–428.

Turner, John et al. (2002). "Recent temperature trends in the Antarctic". In: *Nature* 15, pp. 291–292.

Turner, John et al. (2005). "Antarctic climate change during the past 50 years". In: *International Journal of Climatology* 25, pp. 279–294.

Vogt, Michael and Oliver Linton (2014). "Nonparametric estimation of a periodic sequence in the presence of a smooth trend". In: *Biometrika* 101.1, pp. 121–140.

Wildi, Marc (2006). *Signal Extraction: Efficient Estimation, 'Unit Root'-Tests and Early Detection of Turning Points.* Lecture Notes in Economics and Mathematical Systems. Berlin: Springer.

Yu, Kyusang, Enno Mammen, and Byeong U. Park (2011). "Semi-parametric regression: Efficiency gains from modeling the nonparametric part". In: *Bernoulli* 17.2, pp. 736–748.

Lebenslauf

Ausbildung

10/2007-05/2014	Promotion in Volkswirtschaftslehre an der Universität Mannheim,
	Mitglied des Graduiertenkollegs VWL und CDSE
09/2007	Abschluß als M.A. Economics
09/2006-09/2007	Studium der Volkswirtschaftslehre
	an der Ludwig-Maximilians-Universität München
07/2005	Abschluß als B.Sc Econometrics and Mathematical Economics
09/2002–07/2005	Studium an der London School of Economics
07/2002	Abitur
08/1993–06/2002	Martin-Schongauer Gymnasium in Breisach am Rhein