Quantile regression-based seasonal adjustment

Coordinatore del Corso: Prof. Massimiliano Caporin

Supervisore: Prof. Massimiliano Caporin

Dottorando/a: Mohammed Elseidi

30/09/2019

Abstract

Time series of different nature might be characterised by the presence of deterministic and/or stochastic seasonal patterns. By seasonality, we refer to periodic fluctuations affecting not only the mean but also the shape, the dispersion and in general the density of the variable of interest over time. Using traditional approaches for seasonal adjustment might not be efficient because they do not ensure, for instance, that the adjusted data are free from periodic behaviours in, say, higher-order moments. We introduce a seasonal adjustment method based on quantile regression that is capable of capturing different forms of deterministic and/or stochastic seasonal patterns. Given a variable of interest, by describing its seasonal behaviour over an approximation of the entire conditional distribution, we are capable of removing seasonal patterns affecting the mean and/or the variance, or seasonal patterns varying over quantiles of the conditional distribution. In the first part of this work, we provide a proposed approach to deal with the deterministic seasonal pattern cases. We provide empirical examples based on simulated and real data where we compare our proposal to least-squares approaches. The results are in favour of the proposed approach in case if the seasonal patterns change across quantiles. In the second part of this work, we improve the proposed approach flexibly to account for the essential effect of the structural breaks in the time series. Again, we compare the proposed methods to segmented-least squares and provide several empirical examples based on simulated and real data that are affected by both the structural breaks and seasonal patterns. The results, in case of stochastic periodic behaviour, are in favour of the proposed approaches especially when the seasonal patterns change across quantiles.

Dedicated to my family and my supervisor

Acknowledgements

First of all, I would like to thank my great God the Merciful for giving me the strength to complete my PhD.

I would like to express my sincere thanks and gratitude to my supervisor and the course coordinator professor Massimiliano Caporin from the core of my heart for his support, wisdom, encouragement and precious advice and guidance throughout this journey of research.

I would like to extend my thanks to the previous course coordinators professor Monica Chiogna and professor Nicola Sartori for their help in all administrative aspects of the doctoral study.

I also extend my sincere thanks to all the distinguished professors who taught me during the first year of the PhD. Special thanks also go to Patrizia Piacentini for her assistance to me and all my colleagues before and after our arrival in Padova.

My sincere thanks and gratitude to my PhD colleagues for the wonderful time we spent together and for the active participation as a team to successfully complete this stage.

My sincere thanks and appreciation to the wonderful and great person don Roberto Ravazzolo, director of the Centro Universitario Padovano, for his permanent assistance, and for the outstanding accommodation that he provided me in the residence of the doctoral students.

Contents

List of Figures

List of Tables

Introduction

Overview

By seasonality we refer to periodic fluctuations affecting not only the mean but also the shape, the dispersion and in general the density of the variable of interest over time. Notable examples are weather-related times series (temperature, solar radiation, wind speed, etc.), economic time series (industrial production, retail sales, new orders, etc.), financial time series (intra-daily traded volume, intra-daily financial asset returns, etc.) and energy data (energy prices, electricity consumption, gas consumption, etc.). Seasonal (or periodic – we use the two words as synonyms) variation is thus a widespread phenomenon common to many research fields. Moving to specific examples, retail sales tend to peak during the Christmas season and then decline after the holidays. Consequently, retail sales time series will typically show increasing sales from September through December and declining sales in January and February. The most widely cited feature of temperature is its seasonal variation. In the northern hemisphere, it reaches, in general, its minimum during the months of January and February, gradually increases from the month of March and usually peaks in July and August. This pattern is repeated annually and is superimposed to an intra-daily pattern due to the day/night alternation.

In general, the higher the frequency of the seasonal data, the more complex the seasonality we might observe. In fact, the periodic behaviour of the data might be a consequence of several seasonal factors that are superimposed one to the other. A simple example is given by energy or electricity time series, whose periodic oscillation at the monthly frequency might depend on the alternation of seasons, which leads to changes in the consumption of electricity in the summer compared to the consumption during the winter. If we move to a daily frequency, a further periodic component emerges, as the time series fluctuations also depend on the day of the week, with the week-end being characterised by a behaviour different from that of the working days. Furthermore, if we

move at an intra-daily level, a third element appears due to the differences associated with the day and night alternation.

In many cases, the periodic behaviour, due to its relevance or even just due to its presence, hides the underlying role and relevance of stochastic elements characterising the evolution of the time series of interest. Therefore, we need criteria and methods for filtering out the seasonal pattern in order to highlight the underlying stochastic behaviour. Such approaches are referred to as seasonal adjustment methods, whose main goal is to determine the presence, estimate and then filter out the seasonal effect of a time series. The statistics literature includes a variety of methods that we might use to estimate and remove a periodic (or seasonal) component. These include simple approaches based on time series regression over dummy variables or periodic waves, the use of moving average filters (additive or multiplicative) or more complex smoothing approaches. They might also require the estimation of a model, such as autoregressive integrated moving average (ARIMA) models with seasonal behaviour, or structural time series methods. These approaches can handle purely deterministic seasonal patterns, seasonal variations characterised by stochastic behaviour and combinations of the two. Of the available methods, we focus in the first chapter on those pointing at removing purely deterministic periodic fluctuations. While in the second chapter we deal with the stochastic seasonal behavior. Regarding the models of the deterministic seasonal patterns, the most common examples are those based on linear regressions over time and seasonal dummies or over periodic functions (i.e. harmonic functions).

These approaches make several assumptions, starting from the fact the deterministic periodic function is invariant over time (otherwise, a more flexible model with a structural break should be taken into account) and that the periodic component impacts, usually, the observations' location and/or the dispersion of the variable of interest. The presence of periodic patterns in the mean and variance characterises, for instance, the time series of temperature, where we observe seasonality in both the mean and in the variance, with the latter differing from the mean because winter temperature vary more than summer temperature; see Caporin and Preś (2013). While the violation of the first assumption, time invariance of the periodic components, can be easily handled, the violation of the second assumption, impact on location and/or shape, is much more challenging, as it opens the door for potential model misspecification. In fact, using traditional approaches for seasonal adjustment might not be efficient because they do not ensure, for instance, that the adjusted data are free from periodic behaviours in, say, higher order moments. Even in the simple case of temperature data, where two periodic patterns can be easily identified, the first in the mean and the second in the variance,

several seasonal adjustment methods point at adjusting the mean of the time series, but the seasonal pattern might remain very clear in the higher order moments, starting from the second-order moment of the seasonally (mean-)adjusted time series. Depending on the purpose of the analysis, the seasonal behaviour on higher order moments should also be removed. This calls for multi-step seasonal adjustment approaches, which might still leave some unexplained periodic behaviour in moments above the second-order one and might also be exposed to a loss of efficiency in the estimation approaches due to the use of multi-step methods.

The literature includes examples of periodic patterns in, at least, the second-order moment. For instance, Alexandridis and Zapranis (2014) present the autocorrelation function of the squared residuals of modelling detrended and deseasonalised daily average wind speed data in Figure 9.6. Their Figure shows a clear seasonal pattern in the variance of the seasonally adjusted time series. A similar problem, leading to heteroskedasticity in the variance of a wind time series, is presented in Figure 5 of Caporin and Preś (2012).

We provide here a further motivating example building on the daily average temperature time series of Ada (see Section 4.1 for details on the data). We applied a harmonic regression approach to perform the seasonal adjustment. In Figure 1, we observe that the daily average temperature (Figure1(a)) and its auto-correlation function (Figure 1(b)) show a clear seasonal pattern. Figure 1(c) shows the seasonally adjusted series, which is free from the periodic behaviour in the mean. On the contrary, in Figure 1(d), the auto-correlation function of the de-meaned and squared seasonally adjusted series (a proxy of a time-varying second-order moment) shows evidence of a clear deterministic seasonal pattern. Therefore, the use of a simple adjustment tool might leave deterministic seasonal components in the series. Standard practice calls for additional regressions/adjustments based on transformed series. However, the multi-step approach is exposed to possible estimation errors, loss of efficiency and does not ensure that the residual series is properly seasonally adjusted. For instance, if we take the de-meaned squared seasonally adjusted data and, again by means of a harmonic regression, filter out the periodic behaviour,[1] we might still find mild evidence of seasonal variation in higher order moments.

---

[1]In this case, we apply a log transformation on the squared seasonally adjusted time series. This leads to a multiplicative seasonal component acting on the shape of the distribution. The adjustment in the mean is thus additive while for the variance we use a ration. On the resulting series the first and second order moments are no longer characterised by periodic behaviour, but seasonal patterns still appear in proxies of third- and fourth-order conditional moments.

These empirical evidences motivate the introduction of a more flexible seasonal adjustment approach. Accordingly, we propose a quantile regression-based seasonal adjustment. The advantage of our method lies in the ability of quantile regression to describe the entire conditional distribution of the dependent variable, which makes it possible to determine a specific (deterministic) periodic component for each quantile of the conditional distribution. In turn, for each observation of a time series, we are able to identify its location under the conditional distribution and then recover the best periodic behaviour, which can then be easily removed.
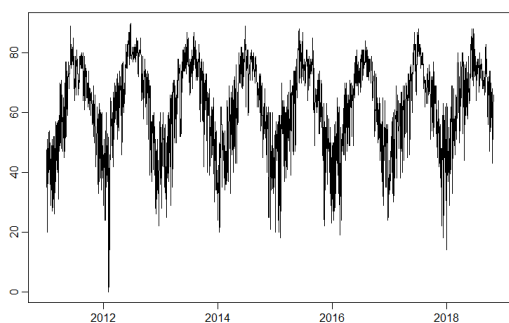
(a) Daily temperature time series.
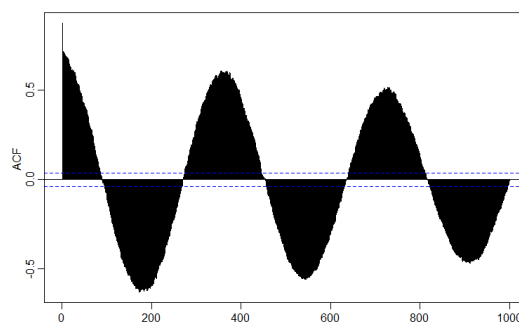
(b) ACF of the time series.
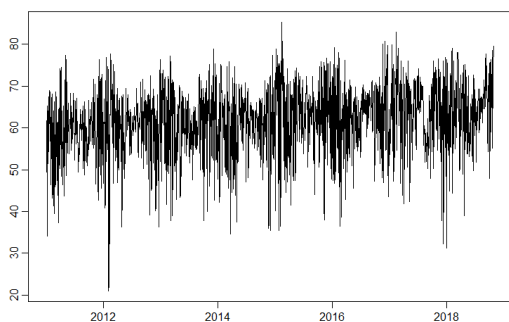
(c) Seasonally adjusted time series using least squares.
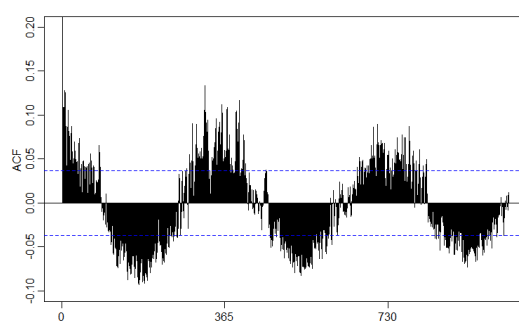
(d) ACF of the squared demeaned seasonally adjusted time series using least squares.
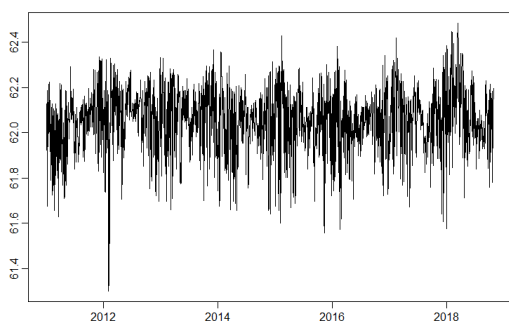
(e) The seasonally adjusted series with variance adjustment made by least squares.
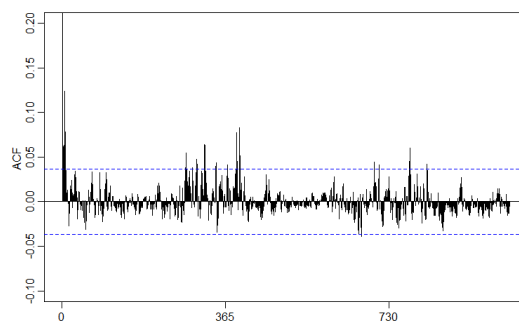
(f) The ACF of the third-order power of the seires included in figure (e).

FIGURE 1: Daily temperature of Ada station in USA and its seasonal adjustment using least squares.

Main contributions of the thesis

In the first chapter, we introduce a more flexible seasonal adjustment approach namely the quantile regression-based seasonal adjustment. The advantage of our method lies in the ability of quantile regression to describe the entire conditional distribution of the dependent variable, which makes possible to determine a specific (deterministic) periodic

component for each quantile of the conditional distribution. In turn, for each observation of a time series, we can identify its location under the conditional distribution and then recover the best periodic behavior, which can then be easily removed. In this thesis, we evaluate the proposed approach using simulations of different data-generating process, considering both a case in which the periodic fluctuations affect only the location and/or the dispersion as well as a case with quantile-based seasonal patterns. We also evaluate the performances of our proposal if the data-generating processes allow for serial dependence in the error term. Further, we show how to make use of quantile stability tests as a tool for deciding whether the seasonal adjustment should be based on linear regression or quantile regression approaches. The introduction of serial dependence in the error term does not alter the previous findings. We also provide examples on real data, focusing on daily temperature time series and five-minute returns of a financial asset. The simulations show that the quantile-based seasonal adjustment is comparable to simpler approaches when the periodic components are not changing over quantiles. Conversely, when the data-generating process includes quantile-based periodic components, we observe that seasonal adjustment based on standard regression-based techniques leaves periodic components in the adjusted series while quantile regression seasonal adjustment correctly removes the deterministic periodic pattern.

In the second chapter, we introduce the seasonal adjustment based quantile regression approach in the presence of the structural breaks in the time series. The significant effect caused by the structural breaks motivates us to improve the quantile regression-based seasonal adjustment approach flexibly to account for structural breaks. We thus propose two approaches, the first is based on the rolling analysis using a rolling window of fixed size. The second is based on implementing a structural break test to determine the locations of the change-points in the time series, then we can perform what we call segmented quantile regression-based seasonal adjustment. Actually, the concept of the piecewise quantile regression approach has been implemented in different studies (Aue *et al.* (2017); Lahiani (2019); Chen *et al.* (2017)). We evaluate the proposed approaches using simulations of different data generating processes, considering the presence of both the seasonal patterns and the structural breaks. For the seasonal patterns, we provide the two cases where the seasonal patterns affect only the location and/or the dispersion as well as a case with quantile-based seasonal patterns. We also presented the structural breaks in different ways which include introducing the break in the seasonal pattern of the data generating process, in the variance of the innovations of the data generating process and through the intercept of the data generating process. We also provide examples on real data focusing on the industrial production index of the USA.

The simulations in case of structural breaks show that both the rolling and segmented quantile-based seasonal adjustment are comparable to simpler approaches that account for the structural breaks when the periodic components are not changing over quantiles. While in the case when both the seasonal pattern changes across quantiles and the time series is affected by the structural break, the suitable approaches are the rolling and segmented quantile-based seasonal adjustment.

Chapter 1

Quantile regression-based seasonal adjustment

1.1 Quantile regression-based seasonal adjustment

Quantile regression, as introduced by Koenker and Bassett (1978), makes it possible to perform estimations and make inferences about conditional quantile functions. By evaluating conditional quantile functions over a wide range of quantiles, we are capable of providing a more complete statistical analysis of the stochastic relationships among random variables (Koenker and Xiao (2002)). In our setting, we specifically focus on the relation variation across conditional quantiles of the role and relevance of deterministic and periodic functions.

A further feature of quantile regression, compared to conditional mean estimation via regression, is that is allows the estimation of the entire conditional distribution of the dependent variable without requiring a prior analysis of the distribution of errors. In fact, an estimate of the conditional distribution for a variable of interest might be obtained by proper smoothing of a collection of estimated conditional quantile functions. We will build on this aspect to derive our proposal for the seasonal adjustment.

The flexibility of quantile regression, and in particular the latter feature, inspired the idea of using this approach for seasonal adjustment as an alternative to the existing seasonal adjustment methods based on the modelling of specific (conditional) moments of a random variable. We introduce a seasonal adjustment method that modifies the conditional distribution form, without focusing on a single moment (even conditional) of a random variable. This turn out to be relevant in cases where the traditional methods might face the risk of leaving some seasonal variation in the higher order moments, as shown in the introduction. The peculiar form of seasonality in higher order moments,

after the seasonal adjustment of the mean and/or the variance, might signal the inappropriateness of traditional seasonal adjustment tools, and this could be associated with specific features of the variable of interest. For instance, it is possible that the seasonal behaviour affects the conditional distribution of the target random variable but with different intensities over quantiles. Therefore, our approach builds on a quantile-specific seasonal adjustment in which we choose for each observation of the time series an optimal estimated seasonal component recovered from a (large) family of estimated conditional quantiles. The choice of the optimal quantile will depend on the minimisation of an intuitive criterion function, the absolute value of the difference between the observed target variable and the estimated conditional quantile. Consequently, our method focuses on adjusting each observation in the time series according to its location under the estimated conditional density, the latter being recovered from a collection of conditional quantile functions.

To perform the seasonal adjustment based on quantile regression, we suggest the following steps. Our proposal starts from the implementation of a collection of quantile regressions to model the observed data. We make a simplifying assumption that the periodic behavior is purely deterministic and might thus be represented by resorting to functions of time. Among the various possible functions we might consider, we opt for the use of time trends and of a collection of harmonics. This is coherent with the choices made in several areas, including remote sensing (Neuenschwander and Crews (2008)), eye movements (Lindsey *et al.* (1978)), accidental deaths series (Brockwell *et al.* (2002)) and airline passenger series (Young *et al.* (1999)). Consequently, the conditional quantile of order τ might be estimated by quantile regression on the following linear model:

$$Q_\tau(y_t) \;=\; \alpha_{0,\tau} + \sum_{i=1}^{p} \alpha_{i,\tau} t + \sum_{j=1}^{m} [\delta_{j,\tau} \cos(2\pi f_j t) + \phi_{j,\tau} \sin(2\pi f_j t)] \qquad (1.1)$$

where $\alpha_0, \alpha_i,\ i = 1, 2, \ldots, p,\ \delta_j,\ \gamma_j,\ j = 1, \ldots, m$ are the parameters to be estimated, $f_j,\ j = 1, \ldots, m$ are the known frequencies of the harmonics . Clearly, to estimate the conditional quantiles, we first need to recover the frequencies. In turn, these might be easily identified by means of spectral analysis or by resorting to the features characterising the reference data. For instance, if we focus on weather data, the amplitudes of possible oscillations affecting the time series evolution is known and might point to the season alternation, the day and night alternation or the moon cycle. Differently, for economic data, we might have quarterly, monthly or weekly amplitudes. These hypothesised amplitudes and the associated frequencies might be validated through the analysis

of the periodogram (i.e. the sample counterpart of the power spectrum) of the target variable. Within the periodogram, the frequencies to be used for the specification of the harmonics corresponds to the highest values, the spikes of the periodogram, as the power spectrum is infinite at the frequencies associated with a deterministic seasonal component. Given that the periodogram is an estimate of the power spectrum and that the precision in frequencies and amplitude estimation depends on the sample size, there might be the need to perform some rounding, as we will show later in the empirical examples. In addition, the identification of the optimal number of harmonics and the inclusion of a time trend should be carefully taken into account in the analysis of the series features as well as of the periodogram of the seasonally adjusted time series. The literature has also condidered the quantilogram Linton and Whang (2007), Dette *et al.* (2015) as a possible choice for the selection of the harmonics.

Given a choice for the time trend order and the number of harmonics, together with their frequencies, we need the estimated conditional quantiles for a collection of τ values. We suggest compromising between the sample size, which has a relevant impact on the evaluation of extreme quantiles, and the precision required in the seasonal adjustment procedure.

However, given that we need to estimate many conditional quantiles, a further issue emerges. In fact, we need to estimate the model in 1.1 by quantile regression over different choices of τ while simultaneously imposing the absence of quantile crossing, see (Bondell *et al.* (2010)). Therefore, we must recover a collection of conditional quantiles satisfying

$$\widehat{Q_\tau(y_t)} \;=\; \hat{\alpha}_{0,\tau} + \sum_{i=1}^{p} \hat{\alpha}_{i,\tau} t + \sum_{j=1}^{m} [\hat{\delta}_{j,\tau} \cos(2\pi f_j t) + \hat{\phi}_{j,\tau} \sin(2\pi f_j t)], \qquad (1.2)$$

subject to $\widehat{Q_{\tau_l}(y_t)} \geq \widehat{Q_{\tau_{l-1}}(y_t)}$ for $l = 2, ..., q$, where q is the number of estimated conditional quantiles and $0 < \tau_1 < \tau_2 < \ldots < \tau_q < 1$. We achieve that objective by following the estimation approach of Bondell *et al.* (2010).

Given the collection of estimated conditional quantiles, we proceed by identifying the optimal quantile from among the τ_l quantiles available for each observation. The vine of estimated non-crossing quantiles makes it possible to recover a semi-parametric estimate of the density for each observation of y_t. Thus, we identify the location of y_t under the estimated density, identify the optimal quantile and then recover the proper seasonal pattern.

For every point in time there are q estimated quantiles, from which we chose one

specific quantile as the best estimate of the location for the observation at time t. The criteria we follow for choosing the best quantile is the minimum absolute value of deviations, that is

$$x_t \;=\; argmin_\tau |y_t - \hat{Q}_\tau(y_t)|. \tag{1.3}$$

We stress that we evaluate this minimisation for all points in time, and we thus recover a time series of optimal quantiles. Using the optimal quantiles for each point in time, as obtained in 1.3, we recover the optimal periodic pattern for each y_t. We evaluate such a component by extracting the pure seasonal deterministic part from the estimated conditional quantiles; that is, we compute

$$p_t \;=\; \sum_{j=1}^{m} [\hat{\delta}_{j,x_t} \cos(2\pi f_j t) + \hat{\phi}_{j,x_t} \sin(2\pi f_j t)]. \tag{1.4}$$

Note that the periodic component p_t is a function of the optimal quantile at time t, and the estimates of both the trend and the constant are not included, as we are interested in removing the seasonal component only and in preserving the trend, if present, and the series level. Further, as the seasonal coefficients change according to the change of the optimal quantile, the periodic pattern takes into account both the seasonal evolution and the location of the target variable over its conditional density.

Finally, the seasonal adjustment could be performed by subtracting the estimated periodic pattern in 1.4 from the original time series y_t. The seasonally adjusted series z_t (1.5) thus equals

$$z_t \;=\; y_t - p_t. \tag{1.5}$$

We highlight that the criteria we follow identifies an optimal quantile from the grid of quantiles we adopt for the estimation of non-crossing quantile curves. However, this is a simplifying choice, the precision of which depends on the total number of estimated quantiles. Clearly, other approaches might be used. For instance, for each y_t, we could estimate a conditional density by smoothing the estimated quantiles and then estimate with greater precision the location of y_t under the density. However, this has a further consequence on the derivation of the seasonal component. In fact, the latter comes from a collection of quantiles, for which we do have coefficients. If we need to recover

the seasonal component for a quantile that has not been estimated, we need a further smoothing of the estimated harmonic coefficients. Such an approach is clearly feasible, but we left it to further works, and we concentrate here on a simplified procedure that provides satisfactory results, as we will show later.

Through the introduction of the seasonal adjustment method based on the quantile regression approach, as described in the previous steps, it is possible to capture different forms of seasonal patterns, including those affecting only, for example, the mean of the variable of interest and including, differently from traditional methods, more general forms of seasonal variation that impact the entire conditional density of the variable of interest. We verify the performance of the approach we introduce through simulations and with real data. Figure 1.1 represents a summary diagram of the proposed seasonal adjustment method.
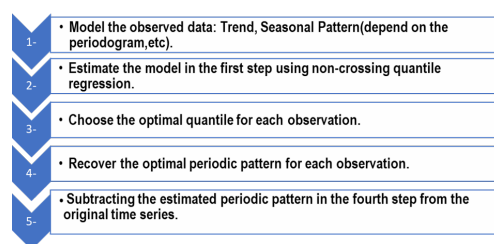
FIGURE 1.1: Summary diagram of the seasonal adjustment method

1.2 Simulations

1.2.1 Data-generating processes

We consider eight data-generating processes differing in the form of seasonality and allowing, in some cases, for serial dependence in the innovations. In all data-generating processes, except the fourth and the eighth processes, the seasonal adjustment can be properly performed by means of least square methods. Conversely, in the fourth and the eighth data-generating processes, the most appropriate approach for seasonal adjustment builds on quantile regression. In all cases, for simplicity, we limit the periodic patterns to be equal to a single harmonic. The last four data-generating processes differ from the previous ones, as we assume a serial dependence structure in the innovations. In particular, we introduce an autoregressive term of order one.

In the first data-generating process, the seasonal effect S_t only impacts the location of a random variable. Therefore, the seasonal pattern is part of the conditional mean of a time series:

$$
\begin{aligned}
y_t &= \mu + S_t + \epsilon_t, &\text{(1.6)}\\
S_t &= \delta \cos(2\pi f t) + \phi \sin(2\pi f t),
\end{aligned}
$$

where μ is an intercept, f is a known frequency and $\epsilon_t \sim N(0,1)$ (the same density is used in the following). In this case, we can estimate, and then remove, the seasonal pattern by focusing on the linear regression of the variable of interest over an intercept and two sinusoidal functions. This case corresponds to the presence of an additive seasonal pattern.

The second data-generating process, the seasonal effect, now defined as κ_t, only impacts the scale of the distribution

$$
\begin{aligned}
y_t &= \mu + \kappa_t \epsilon_t, &\text{(1.7)}\\
\ln(\kappa_t)^2 &= \alpha \cos(2\pi f t) + \beta \sin(2\pi f t).
\end{aligned}
$$

The simulated series thus includes a multiplicative seasonal component that might be captured and removed by focusing again on a standard linear regression, where the dependent variable becomes $\ln(y_t - \mu)^2$ and we use two sinusoidal functions as explanatory variables.

The third data generating process included two seasonal patterns, S_t and κ_t, affecting the location and the scale of y_t, respectively

$$
\begin{aligned}
y_t &= \mu + S_t + \kappa_t \epsilon_t, &\text{(1.8)}\\
S_t &= \delta \cos(2\pi f_1 t) + \phi \sin(2\pi f_2 t),\\
\ln(\kappa_t)^2 &= \alpha \cos(2\pi f_2 t) + \beta \sin(2\pi f_2 t),
\end{aligned}
$$

where f_1 and f_2 are two known frequencies, μ is an intercept and $\epsilon_t \sim N(0,1)$. This process corresponds to a case where two seasonal patterns, with possibly different amplitudes, impact the time evolution of a variable of interest. In this case, the seasonal adjustment process commonly adopted requires a two-step procedure in which we first remove the periodic component from the mean with a first linear regression over two sinusoidal functions. Then, we run a second regression on the first step residuals to identify the multiplicative seasonal pattern.

In the fourth data-generating process, we build a structure where the seasonal adjustment requires the use of quantile regression. We use the following model to simulate time series with a seasonal component that changes across quantiles:

$$Q_\tau (y_t) = \Phi_\tau c_t + Q_\tau (\epsilon_t), \qquad (1.9)$$
$$\Phi_\tau = \Delta_0 + \Delta_1 \tau,$$
$$c_t = \cos(2\pi f t),$$

where $\Phi_\tau c_t$ is a zero-mean periodic function (across all possible values of τ) with an associated f frequency, and $\epsilon_t \sim N(0,1)$. Consequently, the conditional quantile intercept corresponds to the unconditional quantile of a standardised normal density. Finally (Δ_0, Δ_1) are chosen in such a way that the conditional quantile curves do not cross. We simulate y_t by generating from a Uniform (between 0 and 1) the quantiles τ and then building the corresponding y_t quantile.

In the fifth, sixth and seventh data-generating processes, the innovation ϵ_t follows an autoregressive process of order one

$$\epsilon_t = \varphi \epsilon_{t-1} + \eta_t, \qquad (1.10)$$

where η_t is i.i.d. according to a standardised normal. We then use these serially correlated innovations to modify the first, second and third data-generating processes of equations (1.6), (1.7) and (1.8).

Finally, the last data-generating process introduced serial correlation in the innovations of the fourth process, and, as the latter, its design requires the use of the quantile regression approach for performing the seasonal adjustment. In this case, the data-generating process becomes

$$Q_\tau (y_t) = \Phi_\tau c_t + Q_\tau (\epsilon_t), \qquad (1.11)$$

with innovations following equation (1.10). Consequently, to generate the values of y_t, we first simulate the innovations and then evaluate the τ sequence over the conditional density of the innovations.

For each data-generating process, we perform 1000 simulations, and each simulation includes 1000 observations. Furthermore, we use the settings in Table 1.1 for the

simulations.[1]

DGP	Seasonality	Coefficients of each DGP	Period
1	In mean	$\delta = 0.5, \phi = 0.6$	100
2	In variance	$\alpha = 0.5, \beta = 0.6$	50
3	In mean and variance	$\delta = 0.5, \phi = 0.6, \alpha = 0.7, \beta = 0.4$	100 and 50
4	Changing across quantiles	$\Delta_0 = 1, \Delta_1 = 3$	100
5–8	As in 1–4	As in 1–4 and $\varphi = 0.7$	As in 1–4

TABLE 1.1: Settings of each data-generating process.

For each simulated series, we perform the seasonal adjustment by resorting to both the least square methods and quantile regression. Under the first and the fifth data-generating processes, we run a single linear regression on the simulated series. For the second and the sixth data-generating processes, we consider a linear regression on the log-squared simulated series. In the third, fourth, seventh and the eighth data-generating processes we apply two linear regressions, first on the simulated series and then on the log-squared mean seasonally adjusted data. Under all data-generating processes we estimate a collection of quantile regressions, assuming the perfect knowledge of the frequency of the periodic components.

1.2.2   Validation criteria

To determine the appropriateness of the seasonal adjustment methods we apply, we consider three validation criteria.

First, we search for the presence of seasonal patterns by focusing on the autocorrelations of the adjusted series $\hat{\epsilon}_t$. Therefore, we employ the Ljung-box test on the series $\hat{\epsilon}_t$ for a number of lags equal to twice the length of the amplitude of the seasonal oscillation. We test the null of the absence of serial correlation for all lags, and we graphically report the frequency of rejection of the null. Large frequencies will signal an inefficiency in the seasonal adjustment approach. Note that we also report this graphical evidence for the squared values of $\hat{\epsilon}_t$ when the data-generating process includes a seasonal pattern in the variances or to highlight the inefficiency of least squares methods when the appropriate approach requires quantile regression.

The second validation criterion involves the use of loss functions. We use a simple loss function, namely the squared difference between the true innovations ϵ_t and the

---

[1]In addition to what is reported in Table 1.1, in section 1.2.4 we also consider a case where both periods are set equal to 100 observations.

estimated one, i.e. $l_t = (\epsilon_t - \hat{\epsilon}_t)^2$. If the time series is seasonally adjusted in a proper way, the values of the losses are expected to be approximately zero. On the contrary, larger values of the loss functions indicate inefficiencies in the seasonal adjustment process.

In the third validation criteria, we perform a two-sample Kolmogorov–Smirnov goodness-of-fit test to determine if the true innovations and the estimated one appear to come from the same distribution. The test is based on the maximum vertical distance between the two cumulative distribution functions, with a null hypothesis that the cumulative distribution function of both the true innovations and the estimated one are equal. We implement the test on both the seasonally adjusted series in the mean and the seasonally adjusted series in the mean and variance. All validation criteria are used to evaluate the first four data-generating processes. Meanwhile, we only use the first validation criteria for the last four data-generating processes.

We expect that in the all data-generating processes, except the fourth and the eighth, least squares-based seasonal adjustment turns out to be the most efficient method. In fact, the seasonal pattern in these models is nothing more than a change in the location or the scale (or in both of them). Furthermore, we also expect that quantile regression-based seasonal adjustment performs reasonably well, as the existence of a unique seasonal pattern across all quantiles is a special case of a more general situation where the seasonal behavior varies across quantiles. Consequently, for the first three data-generating processes, we expect that the Ljung-Box test and the Kolmogorov–Smirnov test will provide similar performances for both least squares and quantile regression. Finally, for the fourth and eighth data-generating processes, we expect a preference for quantile regression-based seasonal adjustment for all validation criteria.

1.2.3   Results

Figure 1.2 presents a summary of the validation criteria for the first data-generating process. The scatter plot of the average losses obtained by least squares and quantile regression seasonal adjustments (upper left panel) provides evidence in favor of least squares, as this method leads to smaller losses. However, we also note that the two approaches do not result in huge differences in terms of average loss. For the Kolmogorov–Smirnov test (upper right panel) we observe broadly similar patterns for the p-values associated with the two approaches, with the average p-values equal to 0.99 for both quantile regression and least squares. This is an indication that both least squares- and quantile regression-based seasonal adjustment have an appropriate fit. The plot of the Ljung-Box test (lower panel) confirms that the two approaches for removing the seasonal pattern in the mean are very similar to each other. Therefore, under the presence of an

additive seasonal pattern, both least squares and quantile regression provide a proper adjustment, although the former method leads to smaller losses, as expected.

(a) Losses

(b) K-S test on the seasonally adjusted series in the mean.

(c) LB test
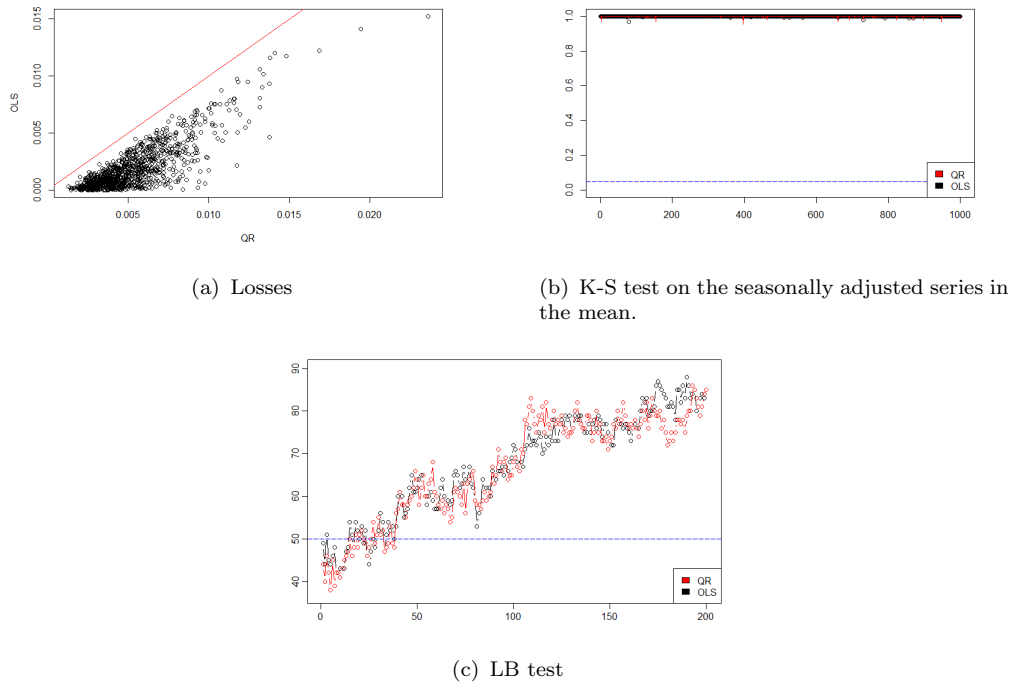
FIGURE 1.2: Data-generating process: seasonal component in mean; see Table 1.1. The upper left panel reports the average losses associated with the two seasonal adjustment methods (i.e. least squares and quantile regression). The upper right panel refers to the p-values of the Kolmogorov–Smirnov test of each simulation and for the two seasonal adjustment methods (i.e. least squares and quantile regression). The lower panel includes the frequency of rejection of the null for the Ljung-Box test over lags from 1 to 200 on the seasonally adjusted series obtained from least squares adjustment or from quantile regression adjustment. Number of simulations: 1000. Series length: 1000 observations.

Figure 1.3 includes the summary results for the second data-generating process, where we introduce a multiplicative seasonal pattern. We report the scatter plot for the average losses for the two approaches as well as the Ljung-Box test and the Kolmogorov–Smirnov test evaluated on the innovations (i.e. the seasonally adjusted series). The empirical evidence we recover from the losses and the Ljung-Box test is similar to that from the first data-generating process: when pointing at the serial correlation, both approaches lead to very similar results without any evidence of seasonal behaviors; the losses are small in both cases but slightly lower for the least square method. For the Kolmogorov–Smirnov test, although the test results indicate that both methods perform well, the acceptance probability of the goodness-of-fit hypothesis using the quantile regression approach is higher; the average p-value is 0.96 for quantile regression and 0.80 for least squares. The results of the third data-generating process (Figure 1.4), where we have both an additive and a multiplicative seasonal component, are in line with the two previous

cases. Again the Ljung-Box results are similar for least squares seasonal adjustment and quantile regression-based seasonal adjustment; the losses are close but smaller for least squares. The p-values of the Kolmogorov–Smirnov test show the appropriateness of both methods with, again, some preference for the quantile regression approach. In this case, the average p-value is 0.93 for quantile regression and 0.88 for the mean and variance seasonally adjusted series with least squares.
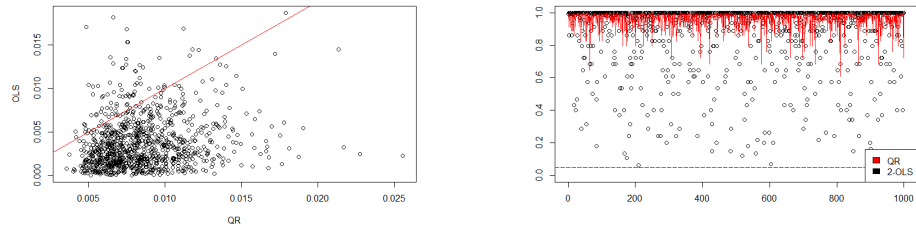
(a) losses

(b) K-S test on the seasonally adjusted series.
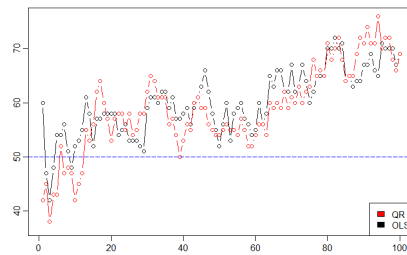
(c) LB test (squared SA series))

FIGURE 1.3: data-generating process: seasonal component in variance; see Table 1.1. The panel (a) reports the average losses associated with the two seasonal adjustment methods (i.e. least squares and quantile regression). Panel (b) refers to the p-values of the Kolmogorov–Smirnov test of each simulation and for the two seasonal adjustment methods (i.e. least squares and quantile regression) evaluated on seasonally adjusted data. Panel (c) reports the frequency of rejection of the null for the Ljung-Box test over lags from 1 to 200 on the squared seasonally adjusted series (to detect periodic components in the variances). Again the adjusted series might be obtained from least squares variance adjustment or from quantile regression adjustment. Number of simulations: 1000. Series length: 1000 observations.
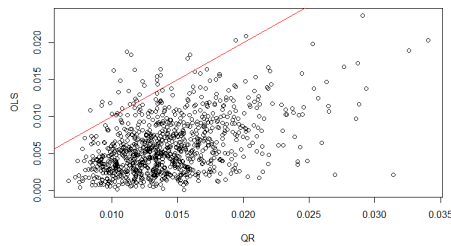
(a) losses
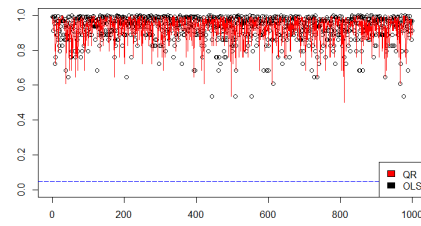
(b) K-S test on the seasonally adjusted series in the mean.
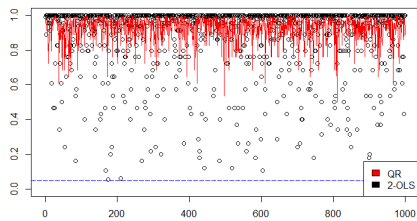
(c) K-S test on the seasonally adjusted series in the mean and variance.
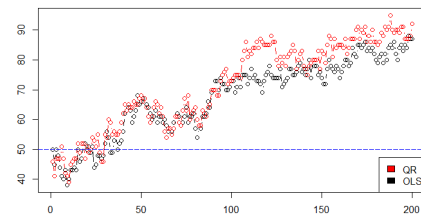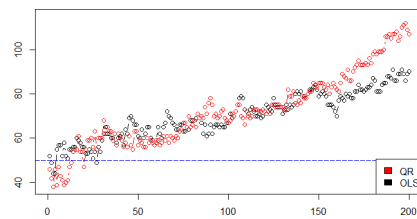
(d) LB test

(e) LB test (squared SA series)

FIGURE 1.4: Data-generating process: seasonal component in mean and variance; see Table 1.1. Panel (a) reports the average losses associated with the two seasonal adjustment methods (i.e. least squares and quantile regression). Panel (b) refers to the p-values of the Kolmogorov–Smirnov test of each simulation and for the two seasonal adjustment methods (i.e. least squares and quantile regression) where the least squares accounts only for additive seasonal components. Panel (c) refers to the p-values of the Kolmogorov–Smirnov test evaluated on seasonally adjusted series, where for least squares we account for mean and variance periodic components. Panel (d) includes the frequency of rejection of the null for the Ljung-Box test over lags from 1 to 200 on the seasonally adjusted series obtained from the least squares mean and variance adjustment or from quantile regression adjustment. Panel (e) reports the frequency of rejection of the null for the Ljung-Box test over lags from 1 to 200 on the squared seasonally adjusted series (to detect periodic components in the variances). Again, the adjusted series might be obtained from the least squares mean and variance adjustment or from quantile regression adjustment. Number of simulations: 1000. Series length: 1000 observations.

In the fourth data-generating process, the seasonal adjustment requires the use of quantile regression because the seasonal patterns have different intensities over quantiles.

Figure 1.5(a) reports the scatter plot of the average losses. We note that the losses are smaller when the adjustment is made with the aid of quantile regression. Figures 1.5(b) and 1.5(c) report the p-values of the Kolmogorov–Smirnov tests. It is clear that the pattern of the p-values using quantile regression is comparable with the previous cases, as the average p-value is 0.9. On the contrary, using least squares for the seasonal adjustment in the mean and two-step least squares for seasonal adjustment in the mean and variance lead to a highly significant rejection of the goodness-of-fit hypothesis, as the average p-value is zero in both cases. Figure 1.5(d) shows the Ljung-Box rejection rates for the seasonally adjusted time series, contrasting the least square residuals of a mean regression on a sinusoidal function with the quantile regression adjustment, while Figure 1.5(e) focuses on the Ljung-Box for the squared innovations of the seasonally adjusted series using quantile regression, least squares and two-least squares regressions. Notably, while the pattern with quantile regression is similar to those of the previous cases, with least squares we see different behaviour, suggesting that a periodic pattern is left in the second order moment and there is a need to perform a second least squares to remove the seasonal pattern from the variance. Figure 1.5(f) shows the Ljung-Box for the third-order power of the innovations of the seasonally adjusted series using quantile regression and two-least squares, and the results are very similar to those in Figure 1.5(d).
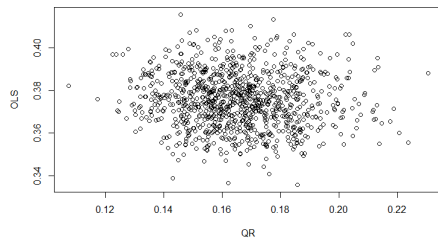
(a) losses
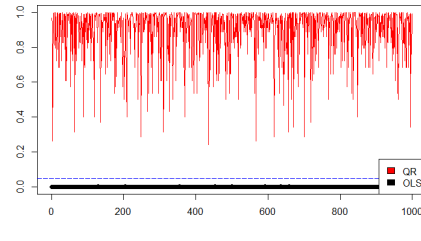
(b) K-S test on the seasonally adjusted series in the mean.
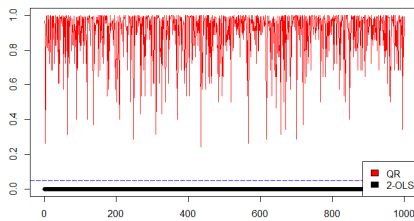
(c) K-S test on the seasonally adjusted series in the mean and variance.
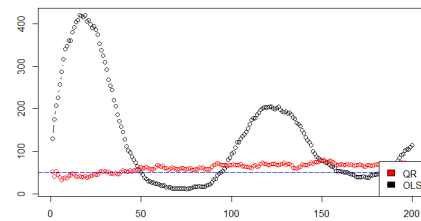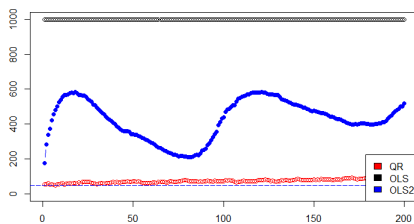
(d) LB test

(e) LB test (squared SA series)
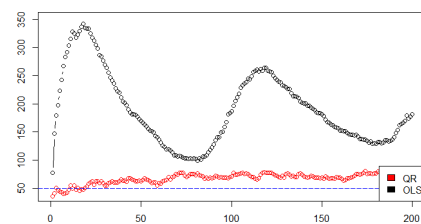
(f) LB test (cubed SA series)

FIGURE 1.5: Data-generating process: seasonal component changing across quantiles; see Table 1.1. Panel (a) reports the average losses associated with the two seasonal adjustment methods (i.e. least squares and quantile regression). Panel (b) refers to the p-values of the Kolmogorov–Smirnov test of each simulation and for the two seasonal adjustment methods (i.e. least squares and quantile regression) evaluated on seasonally adjusted series in the mean. Panel (c) refers to the p-values of the Kolmogorov–Smirnov test evaluated on seasonally adjusted series in the mean and variance. Panel (d) includes the frequency of rejection of the null for the Ljung-Box test over lags from 1 to 200 on the seasonally adjusted series obtained from least squares mean adjustment or from quantile regression adjustment. Panel (e) reports the frequency of rejection of the null for the Ljung-Box test over lags from 1 to 200 on the squared seasonally adjusted series (to detect periodic components in the variances). The plot includes three cases, the quantile seasonal adjustment, the least squares mean seasonal adjustment and the least squares mean variance seasonal adjustment. Panel (f) includes the frequency of rejection of the null for the Ljung-Box test over lags from 1 to 200 on the cubed seasonally adjusted series (to detect periodic components in the variances) for both the mean and variance least squares adjustment and the quantile regression adjustment. Number of simulations: 1000. Series length: 1000 observations.

Figure 1.6 shows the average losses for the last four data-generating processes, where innovations include serial dependence. The plots are comparable to those in Figures 1.2, 1.3, 1.4 and 1.5. We do find a confirmation that when the data-generating process includes an additive and/or multiplicative seasonal component which is invariant across quantiles, the least squares seasonal adjustment provides smaller losses, but quantile regression is close. On the contrary, the quantile regression performs much better than least squares if the periodic component has a different impact across quantiles.
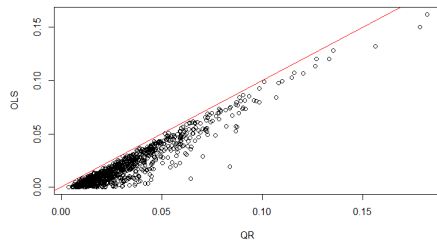
(a) DGP 5.
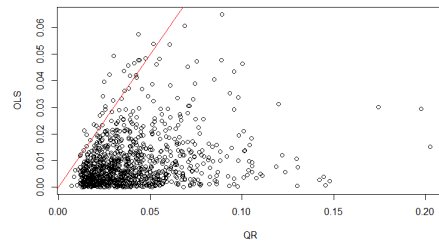
(b) DGP 6.
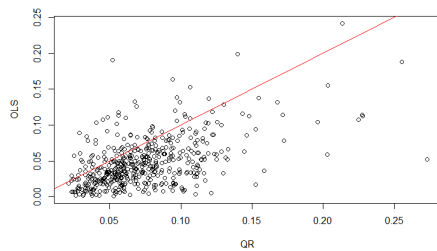
(c) DGP 7.
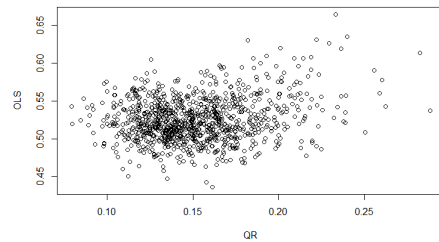
(d) DGP 8.

FIGURE 1.6: The figure reports the average losses of the last four data-generating processes where innovations include serial dependence. Panel (a) reports the average losses of the fifth data-generating process with an additive seasonal component. Panel (b) reports the average losses of the sixth data-generating process with a multiplicative seasonal component. Panel (c) reports the average losses of the seventh data-generating process with additive and multiplicative seasonal components. Panel (d) reports the average losses of the eighth data-generating process with quantile-varying seasonal behavior. Number of simulations: 1000. Series length: 1000 observations.

1.2.4   Least squares or quantile regression?
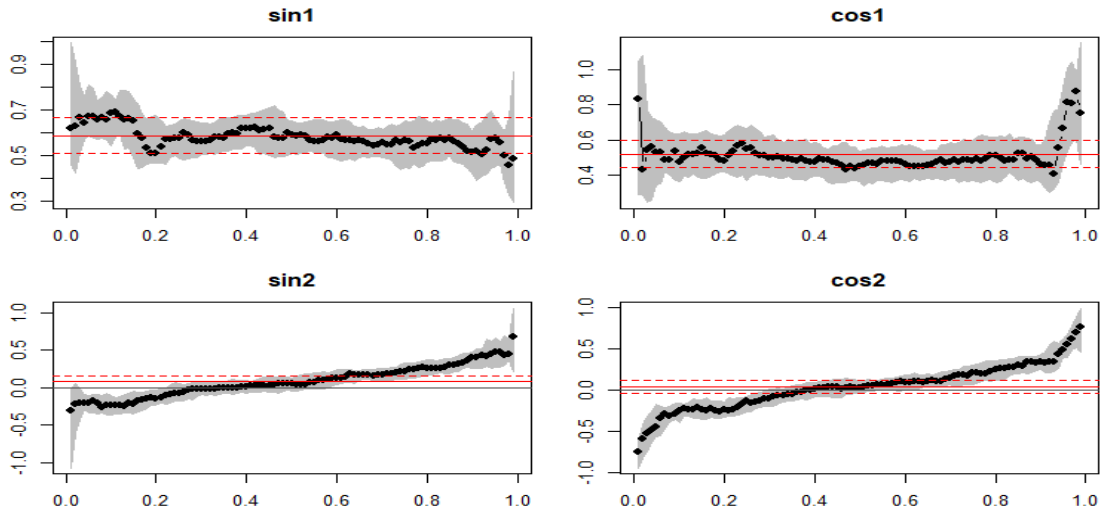
FIGURE 1.7: Quantile regression coefficients for a simulated series under the third data-generating process. The four panels report the estimated coefficients associated with the two harmonics. The upper coefficients refer to the frequency of the additive seasonal component, while the lower coefficients refer to the frequency of the multiplicative seasonal component.

An important issue that we shall take into account is how to distinguish between the cases where there is the need for quantile regression-based seasonal adjustment or where a least-squares based seasonal adjustment could be used. The distinction between the two cases is not immediate, as can be seen in Figure 1.7. In this graph, we report the quantile coefficient process for a single simulated series under the third data-generating process. The coefficients associated with the seasonal pattern affecting the mean are, roughly, stable across quantiles and not significantly different from the least squares estimates. On the other hand, the quantile regression coefficients associated with the second frequency, the one impacting the variances, vary over quantiles and are significantly different from the least squares estimates. The graphical evidence might thus suggest the need of a quantile-based seasonal adjustment. However, the estimated graphical analysis assumes perfect knowledge of the two frequencies associated with the mean and variance periodic components. This might be reasonable in some fields, where the amplitudes of possibly superimposed seasonal components are known (such as for weather variables), but in other cases the frequencies must be identified on the basis of the periodogram. If this is the case, the choice between least squares and quantile regression might not be immediate. In order to choose the appropriate method, we suggest a multi-step procedure.

First, we recover the reference frequencies of the seasonal patterns from the periodogram of the series of interest. With these frequencies, we proceed to the estimation

of a set of non-crossing quantile regressions. Based on the estimates, we verify the null hypothesis of coefficient stability across quantiles. Note that the size and power of the test depend on the number of quantiles and on the sample size, as we will show later. If we reject the null, the appropriate method is the one based on quantile regression. On the contrary, if we accept the null, we must validate the choice of the least squares approach by checking that the second-order conditional moment of the mean seasonally adjusted series is devoid of seasonal patterns. In fact, there might be cases where additional periodic components affect only the variances of a time series, as in our third data-generating process. In addition, if we do not observe periodic components on the series level, we should consider the possible presence of periodic components in the variances, as in our second data-generating process (the following section also includes an example with real data). To validate the seasonal adjustment choice made on the series level, or to verify the existence of periodic components in the variances, we suggest evaluating the periodogram of a transformation of the mean seasonally adjusted series or performing a transformation of the series of interest if no periodic components are identified on the series level. Possible choices for the transformation are the absolute value, the squares or the log-squares. If the periodogram of the transformed series show evidence of the presence of a periodic component, a second set of quantile regressions must be considered, and again the null hypothesis of coefficient stability has to be verified. If the null is rejected, then quantile regression must be used, while if we accept the null, we should decide if we should stop here and focus on least squares approaches for seasonal adjustment or proceed to a further check on, for example, the third-order power of the mean and variance adjusted series. Ideally, the procedure could go farther, continuing beyond the third-order power, but the choice depends on the data features (and source) and on the interpretation one could recover from seasonal patterns appearing only on higher order moments.

We verified such a procedure for the data-generating processes in the previous subsection. However, the number of quantiles and the sample length could impact the size and power of the coefficient stability test across quantiles. To control for that, we consider different sample sizes (500, 1000 and 2000 observations) and two different sets of quantiles (9 quantiles, i.e. from 10% to 90% with a 10% step, and 19 quantiles, from 5% to 95% with a 5% step). Table 1.2 reports summary measures of our analyses of the first four data-generating processes, while Table 1.3 reports the same summary of the last four data generating processes.

For the first data-generating process, which is characterised by an additive periodic

Ind.	NQ	T	Model 1	Model 2	Model 3.a	Model 3.b	Model 4
				Original series y_t			
P1		500	0.010 (0.000)	0.246 (0.145)	0.010 (0.000)	0.010 (0.000)	0.010 (0.000)
P1		1000	0.010 (0.000)	0.247 (0.146)	0.010 (0.000)	0.010 (0.000)	0.010 (0.000)
P1		2000	0.010 (0.000)	0.257 (0.147)	0.010 (0.000)	0.010 (0.000)	0.010 (0.000)
P2		500	—	—	0.248 (0.145)	—	—
P2		1000	—	—	0.248 (0.146)	—	—
P2		2000	—	—	0.257 (0.147)	—	—
R	9	500	0.060	—	0.064	0.999	1.000
R	9	1000	0.053	—	0.059	1.000	1.000
R	9	2000	0.042	—	0.035	1.000	1.000
R	19	500	0.136	—	0.148	1.000	1.000
R	19	1000	0.108	—	0.102	1.000	1.000
R	19	2000	0.060	—	0.068	1.000	1.000
				Log-squared series $log(y_t^2)$			
P1		500	—	0.025 (0.042)	0.023 (0.033)	0.108 (0.151)	—
P1		1000	—	0.020 (0.000)	0.020 (0.000)	0.034 (0.081)	—
P1		2000	—	0.020 (0.000)	0.020 (0.000)	0.011 (0.014)	—
R	9	500	—	0.078	0.212	0.083	—
R	9	1000	—	0.056	0.139	0.070	—
R	9	2000	—	0.059	0.112	0.117	—
R	19	500	—	0.139	0.677	0.167	—
R	19	1000	—	0.103	0.440	0.136	—
R	19	2000	—	0.074	0.296	0.172	—
				Log-squared mean-adjusted series $log\left(\left(y_t - \hat{S}_t\right)^2\right)$			
P1		500	—	—	0.025 (0.038)	0.017 (0.048)	0.010 (0.000)
P1		1000	—	—	0.020 (0.000)	0.010 (0.000)	0.010 (0.000)
P1		2000	—	—	0.020 (0.000)	0.010 (0.000)	0.010 (0.000)
R	9	500	—	—	0.069	0.065	0.994
R	9	1000	—	—	0.068	0.061	1.000
R	9	2000	—	—	0.054	0.065	1.000
R	19	500	—	—	0.142	0.157	0.999
R	19	1000	—	—	0.112	0.106	1.000
R	19	2000	—	—	0.089	0.075	1.000

TABLE 1.2: Summary results for the identification of the most appropriate seasonal adjustment approach (quantile regression versus least squares). All results are based on 1000 simulations. Column 1 specifies the indicator reported in the rows; P1 and P2 stands for the frequencies associated with the highest and second highest values, respectively, of the periodogram; R stands for rejection frequency for the null hypothesis of coefficient stability across quantiles at the 5% confidence level. The second column reports, if relevant, the number of quantiles used to evaluate the null hypothesis of stability of estimated coefficients across quantiles. The third column reports the sample size of the simulated time series. Columns four to eight include results for the four data-generating processes we consider. For the estimated frequencies, we report the average and the standard deviation (in parenthesis) across simulations. The table includes three panels referring to the series levels, the series log-squared levels and the log-squared values of the mean seasonally adjusted series.

component, and thus invariant across quantiles, the periodogram (evaluated on the simulated series levels) allows perfect detection of the seasonal pattern frequency. The subsequent non-crossing quantile regressions show evidence of rejection frequencies in line with our expectations: with increasing sample sizes, the rejection frequency converges to the nominal level, with results that worsen (in particular with smaller samples) as the number of quantiles we consider increases. We thus correctly identify least squares as the appropriate approach.

The second data-generating process includes a multiplicative periodic component. As we already noted, this component only impacts the variances and should not be treated with quantile regression but with least squares on a transformation of the original series. The periodogram evaluated on the simulated series levels does not show any evidence of the presence of a seasonal pattern. In fact, the average frequency across all simulations converges to 0.25, the expected level, as, in the absence of deterministic and stochastic components, the maximum of the periodogram has a uniform distribution in $[0, 0.5]$. In this case, to verify the presence of a periodic component in the variances, we evaluate the periodogram on the log-squared series. We now have a perfect identification of the seasonal frequency. When estimating the non-crossing quantile regressions, the rejection frequency converges to the nominal level as the sample size increases. Our procedure thus correctly suggests that the least squares, adapted on a transformed time series, should be used to remove the periodic component.

We now move to the third model, which is more challenging because it includes both additive and multiplicative periodic components. We consider two cases for this data-generating process. The first corresponds to the same model adopted in the previous subsection, with different amplitudes for the mean and variance periodic components, while the second has mean and variance seasonal components with the same amplitude (with a length of 100 observations). We start from the first case. The periodogram on the series level shows evidence of the presence of a single seasonal component. In fact, the average frequency of the second highest value of the periodogram again converges to 0.25. In this case, we evaluate a set of non-crossing quantile regressions characterised by the presence of a single periodic component. The associated stability test suggests that least squares should be used to remove the additive periodic component (the results are comparable to those of the first data-generating process). One might argue that additional periodic components impacting, for example, the variance, could be identified by means of the periodogram evaluated on transformed data. This is exactly the case shown in Table 2 in the second panel: we correctly identify the frequency of the periodic component impacting the variances. However, the performances of the stability test

worsen, particularly when increasing the number of quantiles we consider. This could
be a byproduct of the presence of an additive periodic component that is not taken into
account. Furthermore, one might test the original series to determine the effect of re-
moving both periodic components with quantile regression. However, this would lead to
an improper selection of the model for seasonal adjustment. We also consider this case,
the results of which are not reported in Table 1.2, and when we evaluate non-crossing
quantile regressions with two periodic components, the rejection frequency of the sta-
bility test converges to 100%. The choice is clearly incorrect, and it is driven by the
joint presence of additive and multiplicative seasonal components. Consequently, when
running non-crossing quantile regressions on the series level, we suggest only including
the periodic components identified on the series level. Following such a procedure on
the seasonally adjusted series, we correctly identify the frequency of the second periodic
component, and the rejection frequencies of the stability test converge to the nominal
level, although with a somewhat lower convergence compared to previous cases (see the
third panel of Table 1.2). Our procedure could thus lead to the proper selection of the
seasonal adjustment approach. When deterministic additive and multiplicative seasonal
components are both present, and when they are also characterised by the same period,
in order to appropriately select least squares, we must consider the analysis on the orig-
inal series as well as on the mean-adjusted data. In fact, limiting the analysis to the
series level will always cause the quantile stability test to reject the null.

In the fourth data-generating process, the periodogram allows correct identification
of the seasonal component frequency, and non-crossing quantile regressions clearly lead
to a rejection of the null. However, if we disregard this test and proceed to a least
squares estimation, the search for additional periodic components on the adjusted series
show evidence of seasonal variation. Further, rejection frequencies still lead to a clear
preference for quantile regression compared to least squares. These elements should
be viewed as evidence in favour of quantile regression-based seasonal adjustment on
the original series. Further, we note that contrasting the third model (case b) and the
fourth one, the quantile stability test indicates different behaviour on the mean-adjusted
series. This different behavior is crucial for the distinction between quantile regression
and least squares seasonal adjustment.

Finally, the results of the last four data-generating processes in Table 1.3 (i.e. when
innovations include serial dependence) only partially differ from those in Table 1.2.
First, we observe a worsening of the rejection frequencies of the quantile stability test,
particularly when the periodic component is additive. We link this to the presence of the
serial correlation on the innovations, but we also observe that the increase in the sample

size alleviates this problem. Second, in models 6 and 7.a, we observe that the frequencies with the maximal periodogram value oscillate around 0.02. This does not indicate that the periodic behavior for the variances appears in the original series; rather, it is a byproduct of the auto-regressive dynamic. In fact, the evaluation of the periodogram for a single series does not show evidence of periodic components but rather of serial dependence.

Ind.	NQ	T	Model 5	Model 6	Model 7.a	Model 7.b	Model 8
				Original series y_t			
P1		500	0.010 (0.004)	0.022 (0.017)	0.011 (0.005)	0.011 (0.006)	0.010 (0.000)
P1		1000	0.010 (0.000)	0.020 (0.015)	0.010 (0.001)	0.010 (0.001)	0.010 (0.000)
P1		2000	0.010 (0.000)	0.018 (0.013)	0.010 (0.000)	0.010 (0.000)	0.010 (0.000)
P2		500	—	—	0.023 (0.016)	—	—
P2		1000	—	—	0.020 (0.014)	—	—
P2		2000	—	—	0.018 (0.013)	—	—
R	9	500	0.225	—	0.227	0.996	1.000
R	9	1000	0.212	—	0.220	1.000	1.000
R	9	2000	0.182	—	0.206	1.000	1.000
R	19	500	0.371	—	0.389	1.000	1.000
R	19	1000	0.283	—	0.327	1.000	1.000
R	19	2000	0.205	—	0.271	1.000	1.000
				Log-squared series $log(y_t^2)$			
P1		500	—	0.023 (0.017)	0.023 (0.014)	0.020 (0.029)	—
P1		1000	—	0.020 (0.000)	0.020 (0.005)	0.012 (0.012)	—
P1		2000	—	0.020 (0.000)	0.020 (0.000)	0.010 (0.001)	—
R	9	500	—	0.081	0.256	0.104	—
R	9	1000	—	0.084	0.182	0.092	—
R	9	2000	—	0.072	0.140	0.096	—
R	19	500	—	0.199	0.725	0.204	—
R	19	1000	—	0.128	0.519	0.146	—
R	19	2000	—	0.097	0.304	0.100	—
				Log-squared mean-adjusted series $log\left(\left(y_t - \hat{S}_t\right)^2\right)$			
P1		500	—	—	0.024 (0.019)	0.015 (0.024)	0.010 (0.000)
P1		1000	—	—	0.020 (0.004)	0.010 (0.008)	0.010 (0.000)
P1		2000	—	—	0.020 (0.000)	0.010 (0.000)	0.010 (0.000)
R	9	500	—	—	0.070	0.063	0.998
R	9	1000	—	—	0.072	0.075	1.000
R	9	2000	—	—	0.067	0.070	1.000
R	19	500	—	—	0.169	0.170	0.997
R	19	1000	—	—	0.120	0.128	1.000
R	19	2000	—	—	0.102	0.096	1.000

TABLE 1.3: Summary results for the identification of the most appropriate seasonal adjustment approach (quantile regression versus least squares). All results are based on 1000 simulations. Column 1 specifies the indicator reported in the rows; P1 and P2 stand for the frequencies associated with the highest and second highest values, respectively, of the periodogram; R stands for rejection frequency for the null hypothesis of coefficient stability across quantiles at the 5% confidence level. The second column reports, if relevant, the number of quantiles used to evaluate the null hypothesis of stability of estimated coefficients across quantiles. The third column reports the sample size of the simulated time series. Columns four to eight include results for the last four data-generating processes we consider. For the estimated frequencies, we report the average and the standard deviation (in parenthesis) across simulations. The table includes three panels referring to the series levels, the series log-squared levels and the log-squared values of the mean seasonally adjusted series.

1.3 Application to real data

1.3.1 Daily average temperature data

In this section, we apply our quantile regression seasonal adjustment approach and compare it to a least squares adjustment method. We work here on a temperature time series including 2854 daily observations from January 2010 to November 2017. These observations refer to the average temperature for Ada, in the USA. Data have been recovered from the North Dakota Agricultural Weather Network. Seasonal adjustment of temperature time series is important for better recovering time trends as well as extreme events.

We present the time series plot in Figure 1 (a) and its autocorrelation function for 1000 lags in Figure 1 (b). We report the partial autocorrelation function for 100 lags in Figure 1.8 (a) and the periodogram in Figure 1.8 (b). Visual inspection of these graphs shows clear evidence of the existence of a seasonal pattern in the time series. Given that the seasonal variation might be due to a number of seasonal factors, we sort out the essential frequencies from the periodogram, where we chose the three frequencies that, to our understanding, are the most relevant for the seasonality in the temperature time series; namely, yearly, half-yearly and monthly. In Table 1.4, the first two columns represent the chosen empirical periods and their corresponding frequencies, while the third and fourth columns include the theoretical periods and the corresponding frequencies, respectively. The values of the periodogram differ from the theoretical frequencies due to the discretisation introduced in the evaluation of the sample power spectrum, where the latter would highlight the existence of periodic components associated with the frequencies that match natural phenomena, namely, the period of revolution of the earth. The table reports an approximation of the length of the period of revolution and adopts fractions of this length (one half for the half-year and one twelfth for the month). In our empirical analyses, we also verified that the inclusion of additional frequencies does not lead to improvement in the model fit (both for least squares and quantile regression approaches).

Period (days)	Frequency	Adopted period	Adopted frequency
356.625	0.0028	365.250	0.00273
178.310	0.0056	182.625	0.0054
30.670	0.0326	30.4375	0.0328

TABLE 1.4: Essential frequencies of the daily average temperature data.
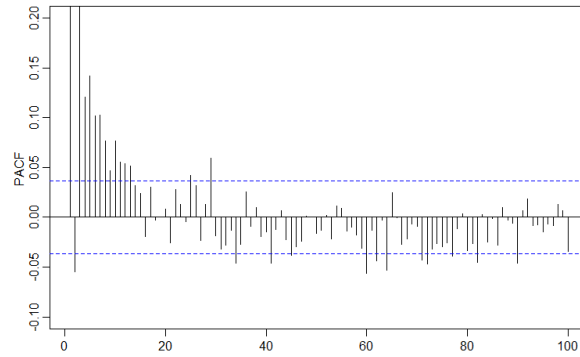
(a) Pacf of the daily average temperature.
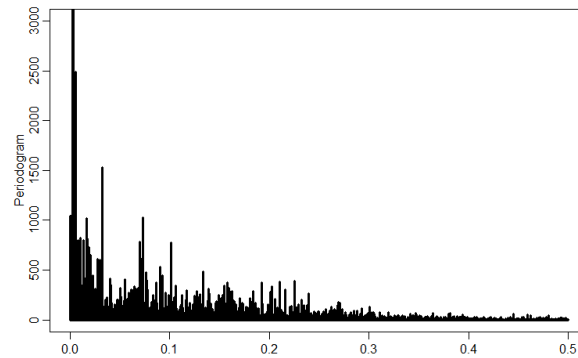
(b) Periodogram of daily average temperature.

FIGURE 1.8: Daily average temperature time series.

The deterministic function we adopt for the quantile regression and linear regression specifications also includes a time trend that we might link to a global warming effect. The overall model we consider for the temperature time series x_t is thus

$$x_t \;=\; \alpha_0 + \alpha_1 t + \sum_{j=1}^{3} \left[\delta_j cos\left(2\pi f_j\right) + \gamma_j sin\left(2\pi f_j\right)\right] + \varepsilon_t, \tag{1.12}$$

with the frequencies f_j reported in Table 1.4. In Figure 1.9, we graph the estimated parameters for the non-crossing quantile regression curves over the 99 quantiles we consider. The figure contains the patterns across quantiles for all coefficients, intercept and

trend included. Notably, all coefficients show relevant variation over quantiles, thus supporting the existence of seasonal patterns that change over quantiles of the temperature data. Along the conditional distribution of temperature data, the seasonal oscillations have varying impacts, thus supporting the introduction of quantile regression-based seasonal adjustment. As a further confirmation of our choice, we perform the coefficient stability test on the original series as well as on the least squares mean-adjusted series using 9 and 19 quantiles, obtaining a p-value of zero in the four cases. These results verify the use of the quantile regression approach, particularly since we reject the stability test on the seasonally adjusted series using the least squares approach. Finally, for comparison purposes, we also report least squares estimation of the model in (1.12) and the associated confidence intervals. It is clear that the quantile regression coefficients differ in a significant way from the least squares ones.
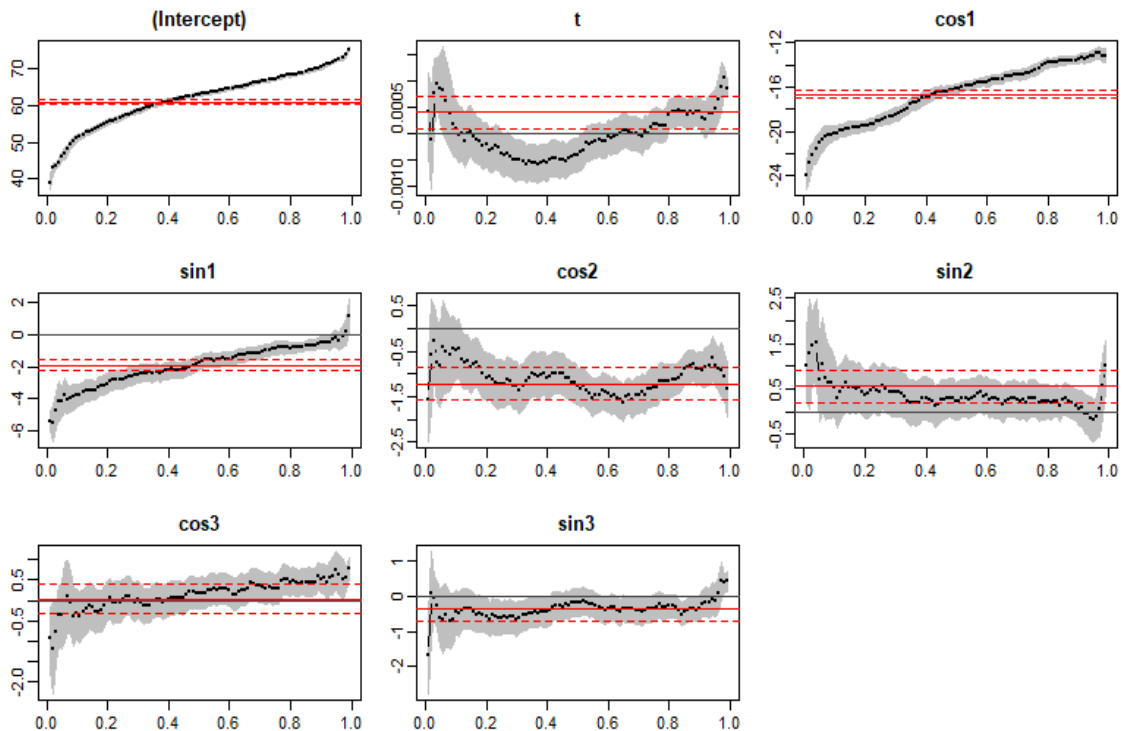
FIGURE 1.9: Quantile regression estimation.

To contrast the quantile regression and least squares seasonal adjustment approaches, we compare the seasonally adjusted temperature time series by focusing on autocorrelation functions. For the quantile regression case, we report (Figure 1.10) the autocorrelation function of the seasonally adjusted time series level, squared values and third-order power. For the least squares case, we focus on the mean and variance adjusted series and consider the autocorrelation function of the third-order power, which is included

in Figure 1. Notably, in the quantile regression case, the adjusted time series does not show any evidence of seasonal variation.
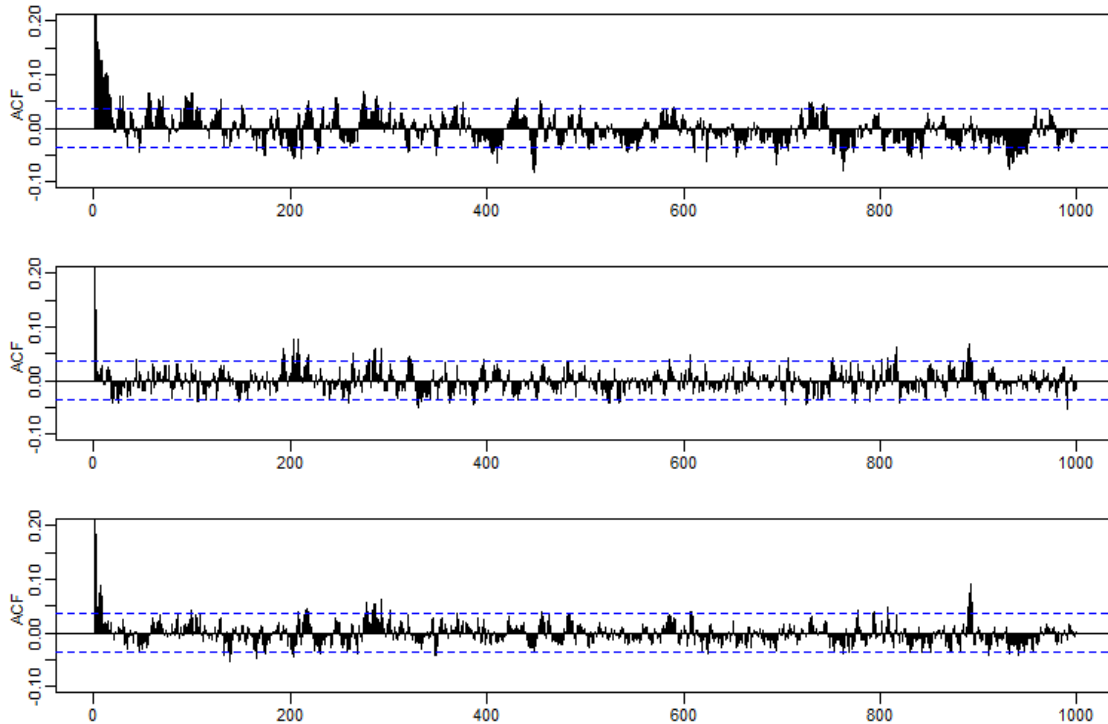
FIGURE 1.10: The autocorrelation function of the seasonally adjusted time series level, squared values and third-order power respectively.

1.3.2   Apple five-minute returns data

We now apply quantile-based seasonal adjustment and the least squares seasonal adjustment on a financial time series, namely, the five-minute returns data of Apple equity recorded from 03/01/2011 to 04/06/2015. We focus on the regular market trading hours, lasting from 9:30 in the morning until 4 in the afternoon. Consequently, we have 78 observations per day and a time series with a total length of 86113 observations.

Figure 1.11 reports the time series plot while Figure 1.12 includes, in the upper panel, the autocorrelation function of the returns. As expected, the two graphs do not show evidence of seasonal variation. On the contrary, when taking the squares, the autocorrelation function clearly indicates the existence of a periodic component; see the lower panel of Figure 1.12. This behaviour is well known in the financial econometrics literature and relates to the increase in volatility that is observed just after the market opening and before market closing. The periodogram (Figure 1.13) confirms the existence of deterministic seasonal fluctuations, which we associate to the frequencies reported in Table 1.5. The table focuses on the first three frequencies, the most relevant

ones in the periodogram, which are associated with periods of one day, half a day and one-third of a day. Note that the discretisation effect on the frequencies is no longer present given the huge sample size.

Period	Frequency	Adopted period	Adopted frequency
78	0.0128	78	0.0128
39	0.0256	39	0.0256
26	0.0385	26	0.0385

TABLE 1.5: Essential frequencies of Apple data.
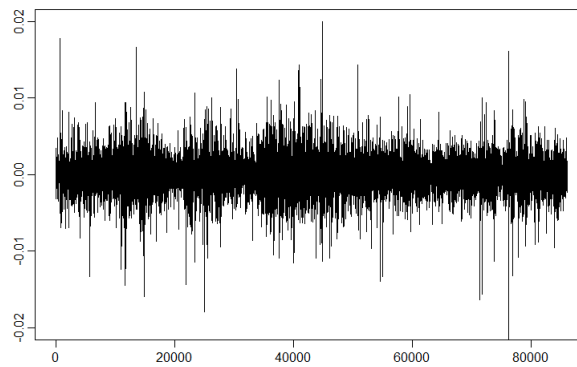
FIGURE 1.11: Apple data.
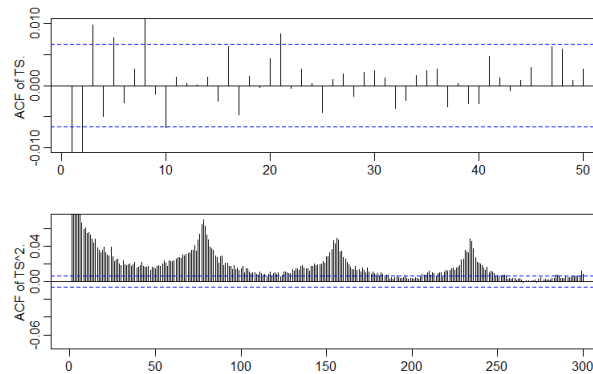
FIGURE 1.12: ACF of apple time series and its squares.
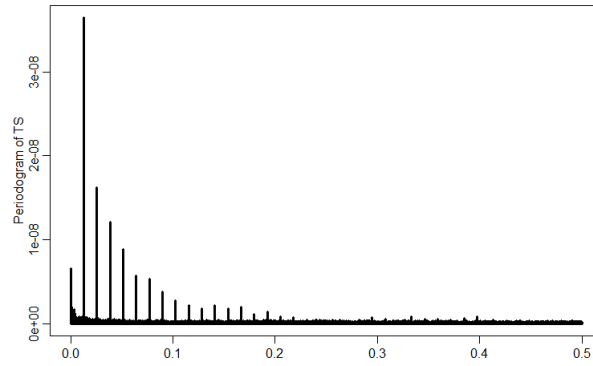
FIGURE 1.13: Periodogram of apple time series squares.

Figure 1.14 includes the estimated coefficients from the non-crossing quantile regression on the Apple data returns times series, while Figure 1.15 focuses on the same regression but on the log-squared returns. We run both estimates because, according to our procedure for deciding between least squares and quantile regression, we should focus on the log-squared case. This is because, with the Apple data, the returns periodogram does not show evidence of a seasonal component, which emerges only on the log-squared returns. Consequently, the quantile stability test run on the log-squared case shows evidence of instability, thus suggesting the need for a quantile regression-based seasonal adjustment. We perform the latter by focusing on the estimates shown in Figure 1.14.
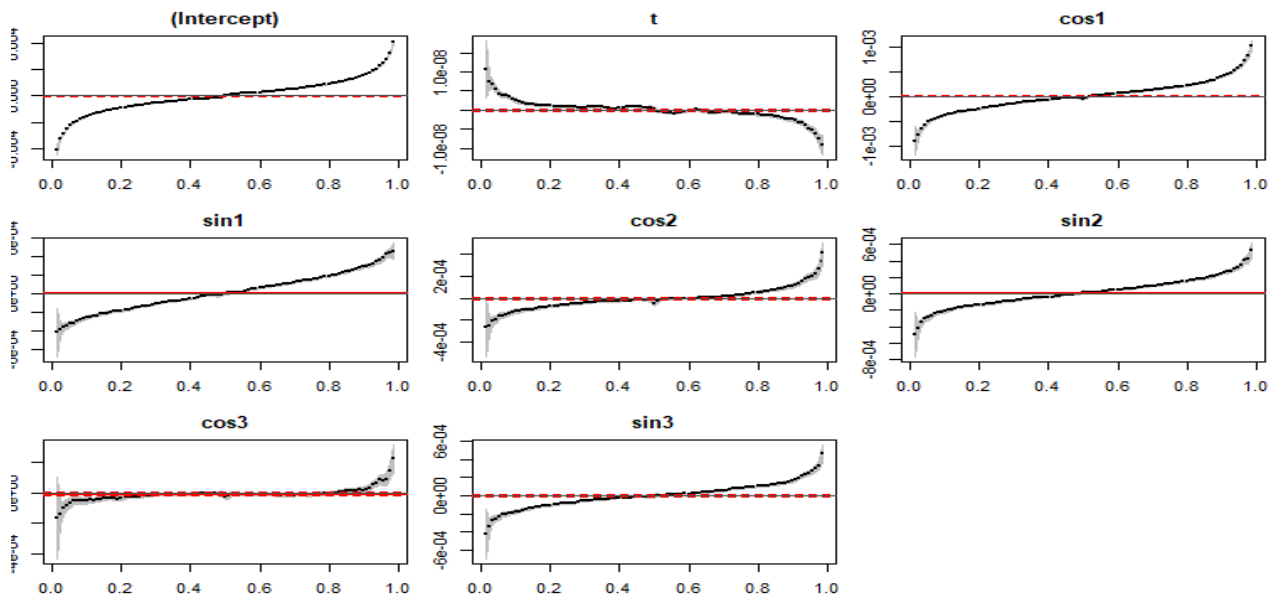
FIGURE 1.14: Process of coefficients across 99 quantiles for the Apple returns.
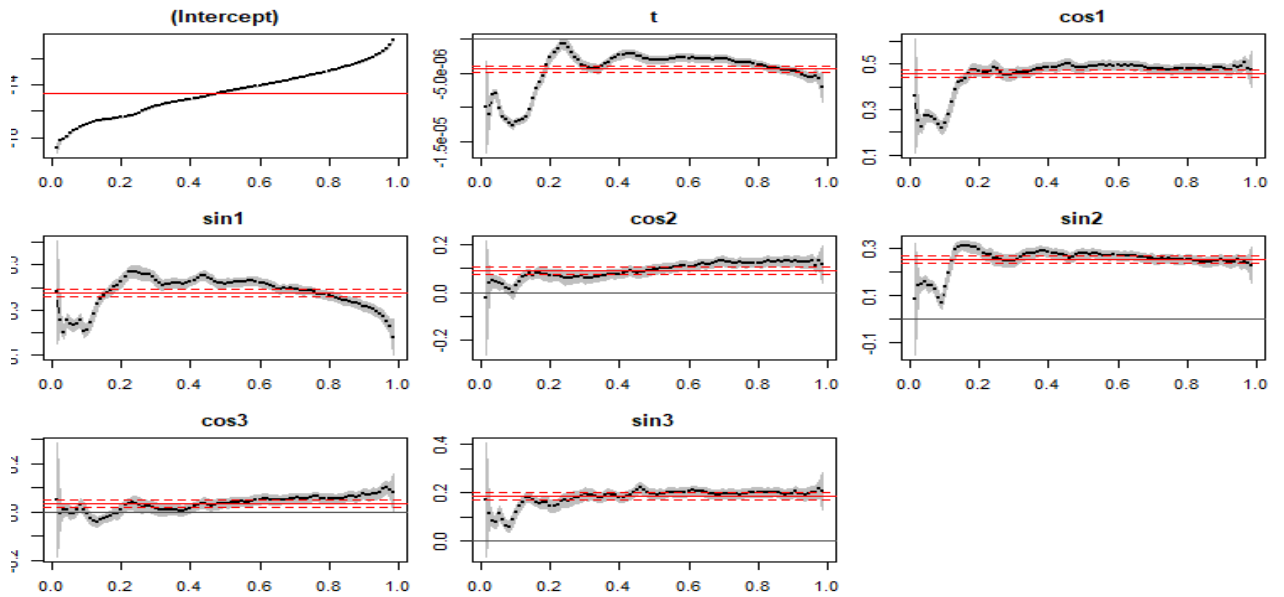
FIGURE 1.15: Process of coefficients across quantiles for the log squared Apple returns.

Similarly to the temperature data, we also contrast the quantile regression adjustment to the least squares adjustment. Figure 1.16 reports the autocorrelation functions of the squared values for of the quantile-based seasonally adjusted returns and of the variance-adjusted (using least squares) returns (left and right panels, respectively). The correlogram for the quantile regression-based seasonal adjustment shows a very stable pattern, similar to those of the long-memory process. Figure 1.17 includes the autocorrelation functions and the periodogram of the third-order powers of the same series. In this case, we note that, by removing the periodic component with quantile regression, the third-order power of the adjusted data seems to be closer to a random walk sequence compared to the least squares adjustment case.

(a) ACF of the squared seasonally adjusted time series using quantile regression.

(b) ACF of the squared seasonally adjusted series with variance adjustment made by least squares.
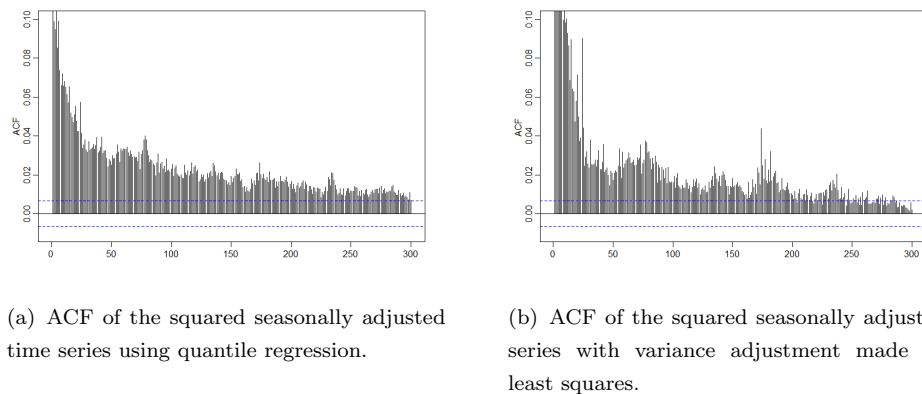
FIGURE 1.16: Analysis of the squared seasonally adjusted Apple returns data.
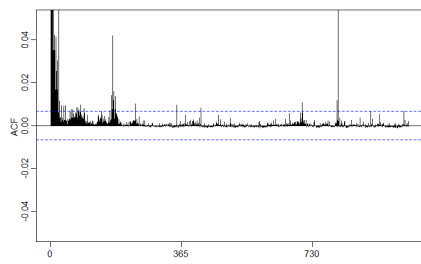
(a) ACF of the third-order power of the least squares seasonally adjusted series.
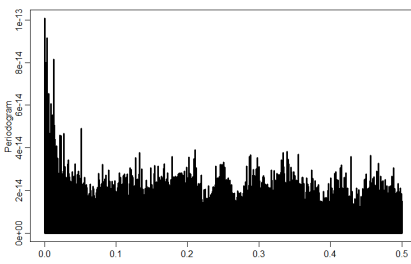
(b) Periodogram of the third-order power of the least squares seasonally adjusted series.
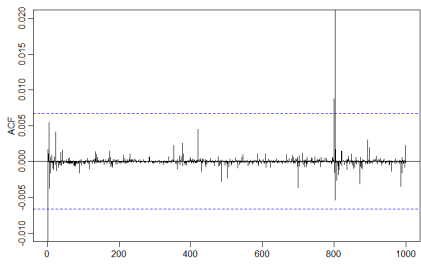
(c) ACF of the third-order power of the quantile regression seasonally adjusted series.
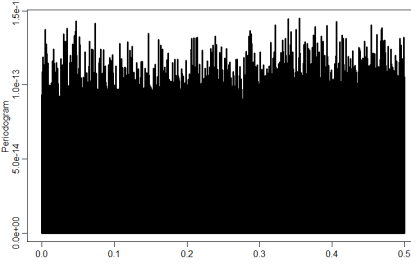
(d) Periodogram of the third-order power of the quantile regression seasonally adjusted series.

FIGURE 1.17: Analysis of the third-order power of seasonally adjusted Apple returns data.

1.4   Conclusion

We propose a seasonal adjustment method, based on quantile regression, that ensures proper adjustment of the seasonal pattern when periodic deterministic components might have varying impacts on the conditional density of a variable of interest. We compare the proposed approach with a more conventional method for performing a seasonal adjustment, namely, the use of least squares based on harmonics. We apply both models to environmental and economic time series as well as to simulated data generated from eight different models. The findings of the first, second, third, fifth, sixth and seventh data generating processes ,when the data-generating process includes additive and/or multiplicative seasonal components (invariant across quantiles), show that our approach is only marginally inferior to a least squares-based seasonal adjustment. The results of all validation criteria of these processes are very close to each other using the two approaches. This is clear from the results of the loss function which has a small value and proves the appropriateness of both methods for performing the seasonal adjustment. Also, the goodness of fit test shows a well fit using both approaches. The rejection frequencies of the Ljung-box test are also near the nominal level and at

the same time very close using the two approaches. However the findings of the fourth and the eighth data generating processes, when the periodic component changes across quantiles, show that our proposal outperforms the least squares-based approach. The results of all validation criteria of these processes are in favor of the proposed approach. This is clear from the results of the loss function which have very small values using the proposed approach in comparison with least squares. Also, the goodness of fit test shows the well fit only using the proposed approach. The rejection frequencies of the Ljung-box test are near the nominal level using only the proposed approach. We also introduce a procedure for choosing the most appropriate seasonal adjustment method between quantile regression and least squares. The empirical examples confirm the flexibility of our approach for both environmental and economics data. This work aims to develop a seasonal adjustment procedure. We didn't consider the issue of the prediction, however, the seasonally adjusted time series properties will allow for the use of a variety of time series methods, including ARMA, quantile regression and score-driven approaches. A further possible extension of this part can be achieved by performing the comparison of forecasts made on the seasonally adjusted series from both linear and quantile regression approaches.

Chapter 2

Seasonal adjustment by quantile regression in the presence of structural break

2.1   Introduction

In the first chapter, it was assumed that the periodic function is invariant across the entire period. However, many time series, especially in economics and finance, do not hold to this assumption as there are cases where an unexpected change over time in the parameters of regression models could happen. This is called a structural break in the time series in which the change in the data-generating process could be due to many reasons, such as major changes in the business or social environment, or changes in the legal and administrative systems that in such cases produce the data. In this case, a structural break model would be more flexible in incorporating the change points in the parameters of the models. Large practical problems that can be found in different fields include the stock market analysis Hsu (1982). For example, the daily oscillation of any stock price changes normally as explained by the economic theory, however, there are some unusual changes which require the awareness of the investors. Another application is in the quality control Hawkins *et al.* (2003) as with statistical process control, the goal is to determine and interpret the cases in which the process has gone out of the statistical control, this important to monitor the stability of the quality of the products. Time-series models are implemented for different purposes like predicting future events, understanding the pattern of the time series in the past and making the economic policy proposals. The violation of the assumption of the parameter's invariance over time reduces the capability of a model to deal with any of

these purposes. Literature shows that essential and commonly used economic indicators contain structural breaks. Therefore, failure to determine and model structural breaks may cause imprecise predictions and illogical conclusions. Being able to detect when the structure of the time series changes, can provide insights into the problem we are studying. Structural break tests help us to determine whether and when there is a significant change in our data.

In this chapter, we are interested in performing the seasonal adjustment in the presence of the structural breaks in the time series using the quantile regression approach. It was evidenced in the first chapter the need for the quantile regression-based seasonal adjustment approach in the cases when the seasonal pattern affects not only observations' location and/or the dispersion of the variable of interest but also the higher-order moments. The significant effect that caused by the structural breaks motivates us to improve the quantile regression-based seasonal adjustment approach flexibly to account for the effect of the structural breaks. In this part we make an assumption about how changes in model parameters occur which is parameters shift immediately at a specific breakpoint. We focused on this assumption because the sudden changes in the time series are commonly observed in many applications. We thus propose two approaches: the first is based on the rolling analysis using a rolling window of fixed size. The second is based on implementing a structural break test to determine the locations of the change-points in the time series, then we can perform what we call segmented quantile regression-based seasonal adjustment. Actually, the concept of the piecewise quantile regression approach has been implemented in different studies (Aue *et al.* (2017); Lahiani (2019); Chen *et al.* (2017)). There are further possibilities for modeling the dynamic change in the time series, for instance, we can allow for the dynamic which depending on the conditional autoregressive value at risk (CAViaR) model Engle and Manganelli (2004). We evaluate the proposed approaches using simulations of different data generating processes, considering the presence of both the seasonal patterns and the structural breaks. For the seasonal patterns, we provide the two cases where the seasonal patterns affect only the location and/or the dispersion of the time series as well as a case with quantile-based seasonal patterns. We also introduce the structural breaks in different ways which include introducing the break in the seasonal pattern of the data generating process, in the variance of the innovations of the data generating process and through the intercept of the data generating process. We end this part with an analysis of real data on the industrial production index.

The chapter proceeds as follows. Section 2 describes our proposals for quantile regression-based seasonal adjustment in the presence of structural breaks. Section 3

is devoted to simulations. Section 4 focuses on one case study namely the industrial production index. Section 5 provides a conclusion to the study.

2.2   Models

2.2.1   Rolling quantile regression based seasonal adjustment.

One of the important issues in the time series framework is to ascertain the stability of the model over time. The main hypothesis of most research on economic time series is that the coefficients are constant in terms of time. The significant changes in economic environments usually show that this key assumption isn't prudent. To overcome this problem, the so-called rolling analysis can be implemented. The idea behind this methodology is to compute the estimates of the model coefficients across a rolling window of fixed size. If these coefficients are stable over time, then the estimates shouldn't be significantly different from each other. Otherwise, if they are too different due to some reasons like structural breaks, then we can consider the coefficients of the interest as time-varying parameters in which the rolling estimates should capture this instability Zivot and Wang (2007). It should be mentioned that the choice of the window size is one of the drawbacks of using this method because of its essential effect on the behavior of the estimates over time. Also, the obtained estimates are not for all the entire sample as it just for the fixed window size Zanin and Marra (2012). (Pascual, 2003) recommended the use of rolling regression with a fixed window instead of using recursive regression as this maintains the test power fixed. The implementation of the rolling quantile regression approach has been applied in previous studies which mainly focuses on finance (Guler *et al.* (2017); (Aretz and Arisoy, 2016); (Cenesizoglu and Timmermann, 2007); (Bonaccolto, 2016); (Xu and Childs, 2013); (Cho *et al.*, 2015); (Chen *et al.*, 2016)). Here we implement the quantile regression-based seasonal adjustment approach in the framework of the rolling analysis. The estimation window is rolled over almost the data-set period, keeping the window period constant, starting at the beginning and moving the window forward one observation at a time. Another approach is to divide the entire sample into non-overlapping subsets, but in this case, we have to pay attention that each sub-sample is large enough to get accurate estimates of the statistics Hällman (2017). Following the rolling analysis with fixed window size, we can write the rolling quantile regression-based seasonal adjustment procedures as following:

In the same manner, as the first chapter, we first model the seasonality of the data using a collection of harmonics obtained from the spectral analysis.

$$Q_\tau(y_t) \;=\; \alpha_{0,\tau} + \sum_{i=1}^{p} \alpha_{i,\tau} t + \sum_{j=1}^{m} [\delta_{j,\tau} \cos(2\pi f_j t) + \phi_{j,\tau} \sin(2\pi f_j t)], \qquad (2.1)$$

where α_0, α_i, $i = 1, 2, \ldots, p$, δ_j, γ_j, $j = 1, \ldots, m$ are the parameters to be estimated, f_j, $j = 1, \ldots, m$ are the known frequencies of the harmonics. Secondly, we need to estimate the model in 2.1 by rolling quantile regression over different choices for τ.

$$\widehat{Q_\tau(y_t)} \;=\; \hat\alpha_{0,\tau,t} + \sum_{i=1}^{p} \hat\alpha_{i,\tau,t} t + \sum_{j=1}^{m} [\hat\delta_{j,\tau,t} \cos(2\pi f_j t) + \hat\phi_{j,\tau,t} \sin(2\pi f_j t)]. \qquad (2.2)$$

We notice in model 2.2 that the parameters α_0, α_i, $i = 1, 2, \ldots, p$, δ_j, ϕ_j, $j = 1, \ldots, m$ are time varying parameters due to the use of the rolling analysis. The third step in the proposed seasonal adjustment approach, as explained in chapter 1.1, is identifying the best quantile for each observation using the criteria of the minimum absolute value of deviations, that is

$$x_t \;=\; argmin_\tau |y_t - \hat{Q}_\tau(y_t)|. \qquad (2.3)$$

Using the optimal quantiles for each point in time, as obtained in 2.3, we move to the fourth step by recovering the optimal periodic pattern for each y_t. We evaluate such a component by extracting the pure seasonal deterministic part from the estimated time varying conditional quantiles, that is, we compute

$$p_t \;=\; \sum_{j=1}^{m} [\hat\delta_{j,x_t,t} \cos(2\pi f_j t) + \hat\phi_{j,x_t,t} \sin(2\pi f_j t)]. \qquad (2.4)$$

Finally we compute the seasonally adjusted time series by subtracting the estimated periodic pattern in 2.4 from the original time series y_t. The seasonally adjusted series z_t (2.5) thus equals

$$z_t \;=\; y_t - p_t. \qquad (2.5)$$

To overcome the aforementioned challenges of using the rolling analysis framework, we propose a method for determining the best window size. This method is built on the Fisher g-test (Fisher, 1929) and associate the choice of the window size with the purpose

of the analysis which is doing the seasonal adjustment. Firstly, we determine an interval of the window sizes to choose the best through them. This could be by choosing an interval starting from the main seasonal amplitude that affects the time series and ending with one of its multipliers. Secondly, we implement the rolling quantile regression-based seasonal adjustment using each window in the interval and compute the seasonally adjusted time series. Thirdly, on the seasonally adjusted time series, we test the null hypothesis of Gaussian white noise against the alternative of an added deterministic periodic component of unspecified frequency using Fisher g-test. The basic idea behind the test is to reject the null hypothesis if the periodogram contains a value significantly larger than the average value. Finally, we choose the window size that leads to the largest non-significant p-value of the test.

2.2.2 Segmented quantile regression.

If we assumed that there are structural breaks in the time series, then, in this case, there is a need for testing the existence and detecting the locations of the change points. In applications, we don't know the exact number of change-points. The detection of these points is essential in methodological and practical aspects such as the validation of untested scientific assumptions, Controlling and evaluation of safety-critical process and the validation of the modeling assumptions Eckley *et al.* (2011). In the literature, there are many changepoint search algorithms have been suggested to deal with this problem. we focus here on the offline change point detection in which we have the full-time series is available for the analysis. The most frequently used algorithms include the binary segmentation method (Scott and Knott (1974); Sen and Srivastava (1975)); the segment neighborhood method (Auger and Lawrence (1989);Bai and Perron (1998)); the PELT algorithm Killick *et al.* (2012). In this paper, we applied the binary segmentation approach, as this method has advantages of determining the number of change-points and its locations at the same time with saving time in the computational process. We use the detection method here in both the mean and the variance and with assuming the normal distribution for the data. Assuming that $x_1, x_2, ..., x_n$ is a sequence of independent normal random variables with parameters $(\mu_1, \sigma_1^2), (\mu_2, \sigma_2^2), ...(\mu_n, \sigma_n^2)$. We are concerned here with testing the hypothesis

$$H_0 : \mu_1 = ... = \mu_k = \mu_{k+1} = ...\mu_n$$

and

$$\sigma_1^2 = ... = \sigma_k^2 = \sigma_{k+1}^2 = ...\sigma_n^2.$$

against

$$H_1 : \mu_1 = ... = \mu_k \neq \mu_{k+1} = ...\mu_n$$

and

$$\sigma_1^2 = ... = \sigma_k^2 \neq \sigma_{k+1}^2 = ...\sigma_n^2.$$

We use here the likelihood ratio test statistics Lehmann and Romano (2006)

$$\lambda_n = \max_{2 \leq k \leq n-2} \frac{\hat{\sigma^n}}{\hat{\sigma_1^k}\hat{\sigma_n^{n-k}}}$$

A general description of the binary segmentation technique in the detection of the changes followed by implementing the seasonal adjustment can be summarized in the following steps.

Step 1: Test the null hypothesis that there is no change-point against the alternative that there is a single change-point. If the null is not rejected, then there is not any change point and the procedure is stopped here. On the other hand, if the null hypothesis is rejected we move to the second step.

Step 2: The change-point divides the sequence into two sub-sequences, so we test for the existence of the change point in each sub-sequence.

Step 3: Repeat step two until there are no further change-points.

Step 4: Denote the detected change-points as $\hat{k}_1, \hat{k}_2, ..., \hat{k}_q$ to represent a q change points, then split the entire data-set into $q + 1$ sub-periods.

Step 5: for each sub-period, apply the quantile regression-based seasonal adjustment approach explained in section 1.1:

$$Q_\tau(y_t) = \alpha_{0,\tau} + \sum_{i=1}^{p} \alpha_{i,\tau} t + \sum_{j=1}^{m} [\delta_{j,\tau} \cos(2\pi f_j t) + \phi_{j,\tau} \sin(2\pi f_j t)], \quad (2.6)$$

$$\widehat{Q_\tau(y_t)} = \hat{\alpha}_{0,\tau} + \sum_{i=1}^{p} \hat{\alpha}_{i,\tau} t + \sum_{j=1}^{m} [\hat{\delta}_{j,\tau} \cos(2\pi f_j t) + \hat{\phi}_{j,\tau} \sin(2\pi f_j t)]. \quad (2.7)$$

$$x_t = argmin_\tau |y_t - \hat{Q}_\tau(y_t)|. \quad (2.8)$$

$$p_t \;\; = \;\; \sum_{j=1}^{m} [\hat{\delta}_{j,x_t} \cos(2\pi f_j t) + \hat{\phi}_{j,x_t} \sin(2\pi f_j t)]. \tag{2.9}$$

$$z_t \;\; = \;\; y_t - p_t. \tag{2.10}$$

Figure 2.1 represents a summary diagram of the proposed seasonal adjustment method.
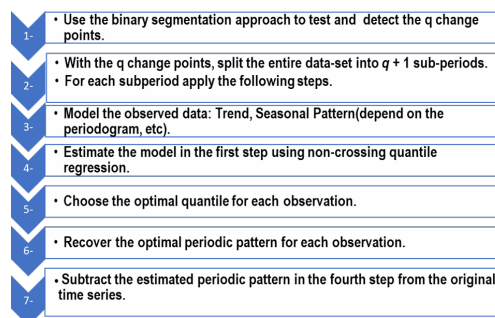
FIGURE 2.1: Summary diagram of the seasonal adjustment method

For more information about the used test statistics and the binary segmentation approach see Chen and Gupta (2011).

2.3   Simulations

We consider nine data-generating processes, differing in the form of the seasonality and the structural breaks that affect the time series. In all data-generating processes except the fourth, the eighth and the ninth processes, the seasonal adjustment can be properly performed using segmented least squares method. In contrast, in the fourth, the eighth and the ninth data-generating processes, the most appropriate approach for seasonal adjustment builds on the proposed quantile regression approaches in the second chapter. We assume that the structural changes in the time series shift immediately at a specific breakpoint. The simulations have been performed through three different classes of structural breaks. The first class includes the first four data-generating processes in which we introduce the structural break as a change in the coefficients of the seasonal covariates after a specific observation. The introduced heteroskedasticity, in this case, is due to the change in the variance of the seasonal pattern. While the structural change in the following four data-generating processes, representing the second class, is introduced through the variance of the innovations after a specific observation which causes a heteroskedasticity due to the change in the variance of the innovations. In the

last data-generating process, the third class, the structural change is introduced through the intercept of the time series after a specific observation.

In the first data-generating process, the seasonal effect S_t impacts only on the location of a random variable. Therefore, the seasonal pattern is part of the conditional mean of a time series, at the same time S_t allows for a single structural break in the coefficients of the seasonal harmonics as follows :

$$
\begin{aligned}
y_t &= \mu + S_t + \epsilon_t, & (2.11) \\
S_t &= \delta_1 d_1 \cos(2\pi f t) + \phi_1 d_1 \sin(2\pi f t) + \delta_2 d_2 \cos(2\pi f t) + \phi_2 d_2 \sin(2\pi f t),
\end{aligned}
$$

where μ is an intercept, f is a known frequency, and $\epsilon_t \sim N(0,1)$ (the same density is used in the following). d_i is a dummy variable that takes the value 1 if t $\in$ sub-period i as determined by the break date and 0 otherwise(the same dummy variables are used in the following). In this case, we can estimate, and then remove, the seasonal pattern by detecting the structural break location and implementing a segmented linear regression of the variable of interest over an intercept and two sinusoidal functions. This case corresponds to the presence of an additive seasonal pattern.

In the second data-generating process, the seasonal effect, now defined as κ_t which allows for a structural change in the coefficients of the seasonal harmonics, the seasonal pattern impacts only on the scale of the distribution

$$
\begin{aligned}
y_t &= \mu + \kappa_t \epsilon_t, & (2.12) \\
\ln(\kappa_t)^2 &= \alpha_1 d_1 \cos(2\pi f t) + \beta_1 d_1 \sin(2\pi f t) + \alpha_2 d_2 \cos(2\pi f t) + \beta_2 d_2 \sin(2\pi f t).
\end{aligned}
$$

The simulated series thus includes a multiplicative seasonal component that might be captured and removed by focusing again on a segmented standard linear regression where the dependent variable becomes $\ln(y_t - \mu)^2$ and we use two sinusoidal functions as explanatory variables.

The third data-generating process included two seasonal patterns S_t and κ_t, affecting the location and the scale of y_t, respectively, and allow for structural changes in both seasonal patterns.

$$
\begin{aligned}
y_t &= \mu + S_t + \kappa_t \epsilon_t, && (2.13) \\
S_t &= \delta_1 d_1 \cos(2\pi f t) + \phi_1 d_1 \sin(2\pi f t) + \delta_2 d_2 \cos(2\pi f t) + \phi_2 d_2 \sin(2\pi f t), \\
\ln(\kappa_t)^2 &= \alpha_1 d_1 \cos(2\pi f t) + \beta_1 d_1 \sin(2\pi f t) + \alpha_2 d_2 \cos(2\pi f t) + \beta_2 d_2 \sin(2\pi f t),
\end{aligned}
$$

where f_1 and f_2 are two known frequencies, μ is an intercept and $\epsilon_t \sim N(0,1)$. This process corresponds to a case where two seasonal patterns, with possibly different amplitudes, impact on the time evolution of a variable of interest. In this case, the seasonal adjustment process commonly adopted requires a two-step procedure where we first remove the periodic component from the mean with a first segmented linear regression over two sinusoidal functions. Then, we run a second segmented regression on the first step residuals to identify the multiplicative seasonal pattern.

In the fourth data-generating process, we build a structure where the seasonal adjustment requires the use of the proposed quantile regression approaches in the second chapter. We use the following model to simulate time series with a seasonal component that changes across quantiles and allows for a structural break in the coefficients of the seasonal harmonics:

$$
\begin{aligned}
Q_\tau(y_t) &= \Phi_\tau c_t + Q_\tau(\epsilon_t), && (2.14) \\
\Phi_\tau &= \Delta_0 d_1 + \Delta_1 d_1 \tau + \Delta_2 d_2 + \Delta_3 d_2 \tau, \\
c_t &= \cos(2\pi f t).
\end{aligned}
$$

where $\Phi_\tau c_t$ is a zero-mean periodic function (across all possible values of τ) with an associated f frequency, and $\epsilon_t \sim N(0,1)$. Consequently, the conditional quantile intercept corresponds to the unconditional quantile of a standardized Normal density. Finally, $(\Delta_0, \Delta_1, \Delta_2, \Delta_3)$, are chosen in such a way that the conditional quantile curves do not cross. We simulate y_t by generating from a uniform (between 0 and 1) the quantiles τ and then building the corresponding y_t quantile.

In the fifth, sixth, seventh and eighth data-generating processes, the innovation $\epsilon_t \sim N(0, \sigma_i^2)$, σ_i^2 indicate the variance of the innovations in the sub-period i which is defined by the change-point.

In the fifth data-generating process, the seasonal effect S_t impacts only on the location of a random variable. with the structural break in the variance of the innovations σ_i^2.

$$
\begin{aligned}
y_t &= \mu + S_t + \epsilon_t, & \text{(2.15)} \\
S_t &= \delta \cos(2\pi f t) + \phi \sin(2\pi f t).
\end{aligned}
$$

The sixth data-generating process introduces a model in which the seasonal pattern κ_t impacts only on the scale of the distribution and the structural break is introduced in the variance of the innovations σ_i^2.

$$
\begin{aligned}
y_t &= \mu + \kappa_t \epsilon_t, & \text{(2.16)} \\
\ln(\kappa_t)^2 &= \alpha \cos(2\pi f t) + \beta \sin(2\pi f t).
\end{aligned}
$$

In the seventh data-generating process, two seasonal patterns S_t and κ_t, affecting the location and the scale of y_t, respectively. The structural break again is introduced through the variance of the innovations σ_i^2.

$$
\begin{aligned}
y_t &= \mu + S_t + \kappa_t \epsilon_t, & \text{(2.17)} \\
S_t &= \delta \cos(2\pi f_1 t) + \phi \sin(2\pi f_2 t), \\
\ln(\kappa_t)^2 &= \alpha \cos(2\pi f_2 t) + \beta \sin(2\pi f_2 t),
\end{aligned}
$$

The eighth data-generating process requires the use of the proposed quantile regression approaches in the second chapter. Again the structural break introduced here in the variance of the innovations σ_i^2.

$$
\begin{aligned}
Q_\tau(y_t) &= \Phi_\tau c_t + Q_\tau(\epsilon_t), & \text{(2.18)} \\
\Phi_\tau &= \Delta_0 + \Delta_1 \tau, \\
c_t &= \cos(2\pi f t).
\end{aligned}
$$

In the last data-generating process, we introduce the structural change through the intercept of the time series, and we focus here only on the case in which the seasonal

pattern changes across quantiles.

$$Q_\tau(y_t) = d_1\mu_1 + d_2\mu_2 + \Phi_\tau c_t + Q_\tau(\epsilon_t), \qquad (2.19)$$
$$\Phi_\tau = \Delta_0 + \Delta_1\tau,$$
$$c_t = \cos(2\pi f t).$$

Finally, we implement the simulations using a single structural break at observation 701 which divides the data-set into two sub-periods. For all data-generating processes, we implement three intensities for the introduced structural breaks, to evaluate its impact on the seasonal adjustment process.

2.3.1   Evaluation

To determine the appropriateness of the seasonal adjustment methods in the presence of the structural breaks, we consider two validation criteria. Also, we consider another validation criterion for evaluating the used structural break test.

Firstly, we search for the presence of seasonal patterns by focusing on the autocorrelations of the seasonally adjusted series $\hat{\epsilon}_t$. Therefore, we employ the Ljung-box test on the series $\hat{\epsilon}_t$ for many lags equal to twice the length of the amplitude of the seasonal oscillation. We test the null of the absence of serial correlation for all lags and we graphically report the frequency of rejection of the null. Large frequencies will signal an inefficiency in the seasonal adjustment approach. Note that we report this graphical evidence also for the squared values of $\hat{\epsilon}_t$ when the data-generating process includes a seasonal pattern in the variances or to highlight the inefficiency of segmented least squares method when the appropriate approach requires quantile regression.

The second validation criteria involve the use of loss functions. We use a simple loss function, namely the squared difference between the true innovations ϵ_t and the estimated one, i.e. $l_t = (\epsilon_t - \hat{\epsilon}_t)^2$. If the time series is seasonally adjusted properly, the values of the losses are expected to be approximately zero. On the contrary, larger values of the loss functions indicate inefficiencies in the seasonal adjustment process.

In the third validation criteria, we evaluate the binary segmentation method, the used approach in change-point detection, in determining the correct structural break location.

We implement the simulations using a single structural break at observation 701 which divide the data-set into two sub-periods. For all data-generating processes, we implement three intensities for the introduced structural breaks, to evaluate the impact of this increase on the seasonal adjustment process. We compare the proposed

approaches, the rolling analysis and the segmented quantile regression, with the segmented least squares and the static quantile regression which was explained in the first chapter.

We expect that in all data-generating processes, except the fourth, the eighth and the ninth processes, segmented least-squares based seasonal adjustment turns out to be the most efficient method. The seasonal pattern in these models is nothing more than a change in the location or on the scale (or in both of them). Furthermore, we also expect that both rolling and segmented quantile regression-based seasonal adjustment performs reasonably well, as the existence of a unique seasonal pattern across all quantiles is a special case of a more general situation where the seasonal behavior varies across quantiles. Finally, for the fourth, the eighth and the ninth data-generating processes, we expect a preference for rolling and segmented quantile regression-based seasonal adjustment, and this for the first two validation criteria. For the last criterion, we expect that the greater the intensity of the structural break, the more precise the detection of the correct change-point location, and this for all data generating processes.

2.3.2   Results

Figure 2.2 presents a summary of the validation criteria for the first data-generating process. The histograms of the structural break test in Panels c, f and i show a right detection of the known structural break location. The frequency of the correct detection is positively correlated to the intensity of the introduced structural break. Panels a, d and g represent the box plots of the average of losses obtained by rolling quantile regression, segmented least squares, segmented quantile regression and static quantile regression seasonal adjustments for the three levels of the structural break. The results show evidence in favor of segmented least squares, as this method leads to smaller losses. However, we also note that both segmented and rolling quantile regressions are comparable to the segmented least squares.

The static quantile regression shows a large difference from the other methods due to the neglect of the structural break effect in the modeling step. Increasing the intensity of the structural break leads to the same pattern of the average of losses between the approaches with an increase in the scale of the average of losses. The results in Panels b, e, and h of the Ljung-Box test match the results obtained from the losses' criterion. Both segmented approaches and the rolling analysis for removing the seasonal pattern in the mean are very close one to another and to the nominal level of the rejection frequency (five percentage). The static quantile regression approach shows the inefficiency of this approach in the presence of the structural break as the rejection frequencies reach ninety

percent. Again, by increasing the intensity of the structural break, we obtain the same pattern between the approaches with an increase in the level of the rejection frequencies. Therefore, under the presence of an additive seasonal pattern, segmented least squares, segmented quantile regression and rolling quantile regression-based seasonal adjustments provide a proper adjustment even if the segmented least squares method leads to smaller losses, as expected.

Figure 2.3 includes the summary results for the second data-generating process, where we introduce a structural break in the multiplicative seasonal pattern. We report again, for the three levels of the structural break, the histogram of the structural break test, the box plots for the average losses and the Ljung-Box test. The empirical evidence we recover from the validation criteria is similar to those of the first data-generating process. The structural break test' results are more concentrated around the correct value by increasing the intensity of the structural break. On the other hand, in the case of losses, the segmented least squares show small noticeable differences than the segmented quantile regression and the rolling analysis. Again, static quantile regression approach results in the highest average of losses. Increasing the level of the structural break results in the same pattern of the average of losses between the approaches with an increase in the scale of the average of losses. When considering the serial correlation, both segmented approaches and rolling analysis lead to very similar results without any evidence of seasonal behaviors. This contrasts with the static quantile regression which is evidently affected by the serial correlation across all lags due to the remain seasonal behavior in the seasonally adjusted series. Again, by increasing the intensity of the structural break, we obtain the same pattern between the approaches with an increase in the level of the rejection frequencies.

The results of the third data-generating process (Figure 2.4), where we have structural breaks in both additive and multiplicative seasonal components, are in line with the two previous cases: again the structural break test results are concentrated around the correct breakpoint, and are much better with increasing the intensity of the structural break. The Ljung-Box test' results in both the mean and the variance are similar for the segmented approaches and the rolling analysis. The static approach is completely inefficient due to the neglect of the effect of the structural break. On the other hand, the segmented approaches and the rolling analysis are very close in terms of losses but it still smaller for segmented least squares. The static quantile regression is very high in terms of the average of losses compared to the other approaches. Again, by increasing the intensity of the structural break, we obtain the same pattern between the approaches with an increase in the level of the rejection frequencies and the average of

losses.

In the fourth data-generating process (Figure 2.5) the seasonal adjustment requires the use of the segmented or rolling quantile regressions as the seasonal patterns have different intensities over quantiles, alongside the effect of the structural break. The results of the structural break test in Figures 2.5 (d),(h) and (l) are in line with the previous cases and are much better with increasing the intensity of the structural break. Figures 2.5 (a),(e) and (i) report the box plots of the average of losses. We note that the losses are smaller when the adjustment is made with the aid of both segmented and rolling quantile regressions. This is because these approaches account simultaneously for the seasonal pattern that changes across quantiles and the structural break effect. Figure 2.5 (b),(f) and (j) show the Ljung-Box rejection rates for the seasonally adjusted time series, contrasting the segmented approaches with the static and rolling quantile regression adjustments, while Figures 2.5 (c),(g) and (k) focus on the Ljung-Box for the squared innovations of the seasonally adjusted series. Notably, while with segmented, static and rolling quantile regressions the patterns are comparable to those of the previous cases, for the segmented least squares a different behaviour is observed, suggesting that a periodic pattern is left in the higher-order moment. Again, by increasing the intensity of the structural break, we obtain the same pattern between the approaches with an increase in the level of the rejection frequencies and the average of losses.

In the fifth data-generating process (Figure 2.6) the structural break is introduced in the variance of the innovations of the time series. Regarding the structural break test in Figures 2.6 (c),(f) and (i), it is clear that the test detects the correct breakpoint at observation 701, the more we increase the intensity of the heteroskedasticity the more concentration of the histogram of the test around the correct breakpoint. The results of the losses in Figures 2.6 (a),(d) and (g) are better in case of using segmented least squares, segmented quantile regression and rolling analysis respectively. While the static quantile regression shows a small difference than the other approaches. By increasing the intensity of the structural break, we get the same pattern of the losses but with an increase in the value of the losses. The results of the Ljung-Box test are in Figures 2.6 (b),(e) and (h). The patterns for both the segmented approaches, rolling analysis and the static one are almost similar. The results of the static quantile regression are logical here because the structural break is introduced in the innovations and not in the seasonal pattern. In this case, the results show that the static approach can adjust correctly the seasonal behavior despite the effect of the structural break in this data-generating process.

In the sixth data-generating process (Figure 2.7) the structural break is again causing

heteroskedasticity in the innovations of the time series. Regarding the structural break test in Figures 2.7 (c),(f) and (i), it is in line with the results of the fifth data-generating process and it is clear that the test detects the correct breakpoint at observation 701, the more we increase the intensity of the break the higher is the concentration of the histogram of the test around the correct breakpoint. The results of the losses in Figures 2.7 (a),(d) and (g) are better in case of using segmented least-squares approach. By increasing the intensity of the structural break, we get the same pattern of the losses but with an increase in the value of the losses. The Ljung-Box test in Figures 2.7 (b),(e) and (h) show that the pattern of the test is in line with the fifth data-generating process (Figure 2.6). There is no evidence that the introduced structural break here affects the behavior of the static approach.

In the seventh data-generating process (Figure 2.8),again, the structural break is introduced as a heteroskedasticity in the innovations of the time series. Regarding the structural break test in Figures 2.8 (d),(h) and (l), it is in line with the results of the previous data-generating process with correct detection of the breakpoint around observation 701. The results of the losses in Figures 2.8 (a),(e) and (i) are better in case of using segmented least squares approach than the other approaches. By increasing the intensity of the structural break, we get a similar pattern of the losses but with an increase in the value of the losses. Again we perform the Ljung-Box test on the seasonally adjusted time series, squared values of the seasonally adjusted time series. The results of all approaches in Figures 2.8 (b), (f), (j) (c), (g) and (k) are not different from each other and from the nominal level.

In the eighth data-generating process (Figure 2.9) the structural break is introduced, again, as a heteroskedasticity in the innovations of the time series. The seasonal adjustment requires the use of segmented or rolling quantile regressions as the seasonal patterns have different intensities over quantiles. Regarding the structural break test in Figures 2.9 (d),(h) and (l), it is in line with the results of the previous data-generating processes. It is clear that the test detects the correct breakpoint at observation 701, the more we increase the intensity of the heteroskedasticity the higher the concentration of the histogram of the test around the correct breakpoint. The results of the losses in Figures 2.9 (a), (e) and (i) are much better in case of using segmented and rolling quantile regressions approaches than the other approaches. By increasing the intensity of the structural break, the same pattern of the losses is obtained but with an increase in the value of the losses. Again we perform the Ljung-Box test on the seasonally adjusted time series and squared values of the seasonally adjusted time series. The best and the stable approaches that have the same pattern in the Figures 2.9 (b), (f), (j), (c), (g)

and (k) are the segmented and rolling quantile regressions. By using these approaches, we don't see any evidence of the seasonal pattern. The number of rejection frequencies in the static quantile regression is high because the structural break effect is neglected. While for the segmented least squares, it is high due to the seasonal pattern remained in the higher-order moment of the seasonally adjusted series.

In the ninth data-generating process, (Figure 2.10) in which the structural break is introduced through the intercept of the time series. The seasonal adjustment requires the use of segmented or rolling quantile regression as the seasonal patterns have different intensities over quantiles. The structural break test in Figures 2.10 (d),(h) and (l), is in line with the results of the previous data-generating processes. It is clear that the test detects the correct breakpoint at observation 701, the more we increase the intensity of the break the higher is the concentration of the histogram of the test around the correct breakpoint. The results of the losses in Figures 2.10 (a),(e) and (i) are better when using segmented and rolling quantile regression approaches. By increasing the intensity of the structural break, we get the same pattern of losses but the values are higher. We again perform the Ljung-Box test on the seasonally adjusted time series and squared values of the seasonally adjusted time series. Regarding the test on the seasonally adjusted time series, the best and the only stable approaches that have the same pattern in the Figures 2.10 (b), (f), (j), (c), (g) and (k) are the segmented and rolling quantile regressions. The number of rejections in the static quantile regression is high due to the neglect of the structural break effect. While for the segmented least squares it is high due to the seasonal pattern remained in the higher-order moment of the seasonally adjusted series.
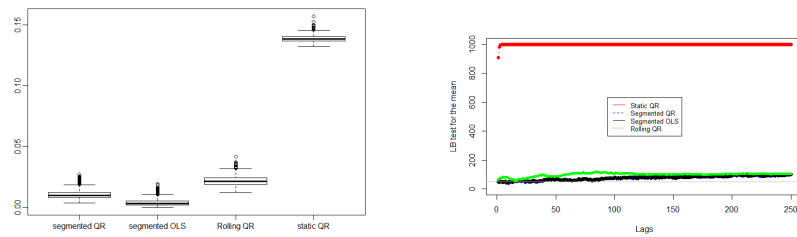
(a) Losses1

(b) lb mean1
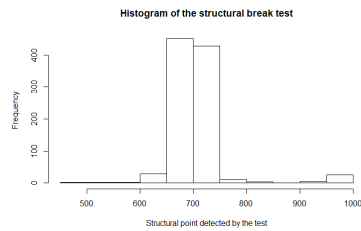
(c) test of structural break 1

(d) Losses 2

(e) lb mean 2

(f) test of structural break 2

(g) Losses 3

(h) lb mean 3

(i) test of structural break 3

FIGURE 2.2: First data-generating process: Structural break through the coefficients of the seasonal pattern in mean. Plots a, d and g refer to the losses with first, second and third specifications of the introduced structural break intensity respectively. Plots b,e and h report the frequency of the rejection of the null for the Ljung-box test results with the first,second and third specifications respectively and over lags from 1:250 on the seasonally adjusted series obtained from segmented least squares, segmented quantile regression, rolling quantile regression and static quantile regression adjustment. Plots c,f and i reports the histogram of the structural break test with first,second and third specifications respectively. Number of simulations: 1000. Series length: 1000 observations.
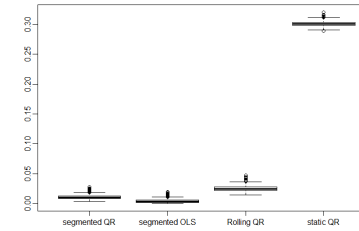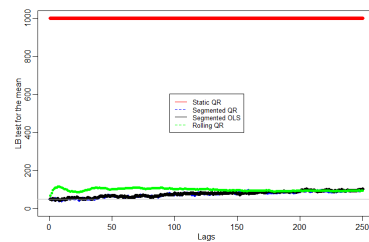
(a) Losses1

(b) lb variance1

(c) test of structural break 1

(d) Losses 2

(e) lb variance 2

(f) test of structural break 2

(g) Losses 3

(h) lb variance 3

(i) test of structural break 3

FIGURE 2.3: Second data-generating process: Structural break through the coefficients of the seasonal pattern in variance. Plots a, e and i refer to the losses with first, second and third specifications of the structural break intensity respectively. Plots b,e and h report the frequency of rejection of the null for the Ljung-box test results with first, second and third specifications respectively and over lags from 1:250 on the squared values of the seasonally adjusted series (to detect periodic components in the variances) obtained from segmented least squares, segmented quantile regression, rolling quantile regression and static quantile regression adjustment. Plots c,f and i report the histogram of the structural break test with first, second and third specifications respectively. Number of simulations: 1000. Series length: 1000 observations.

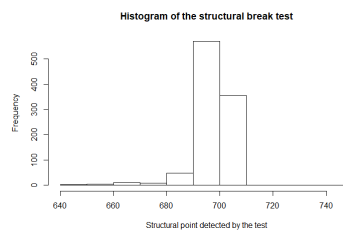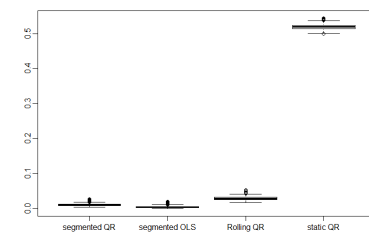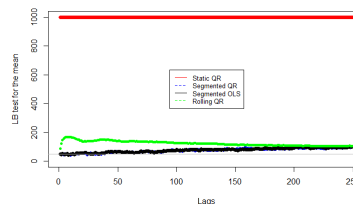(a) Losses1

(b) lb mean1

(c) lb variance1

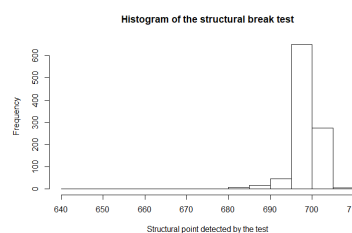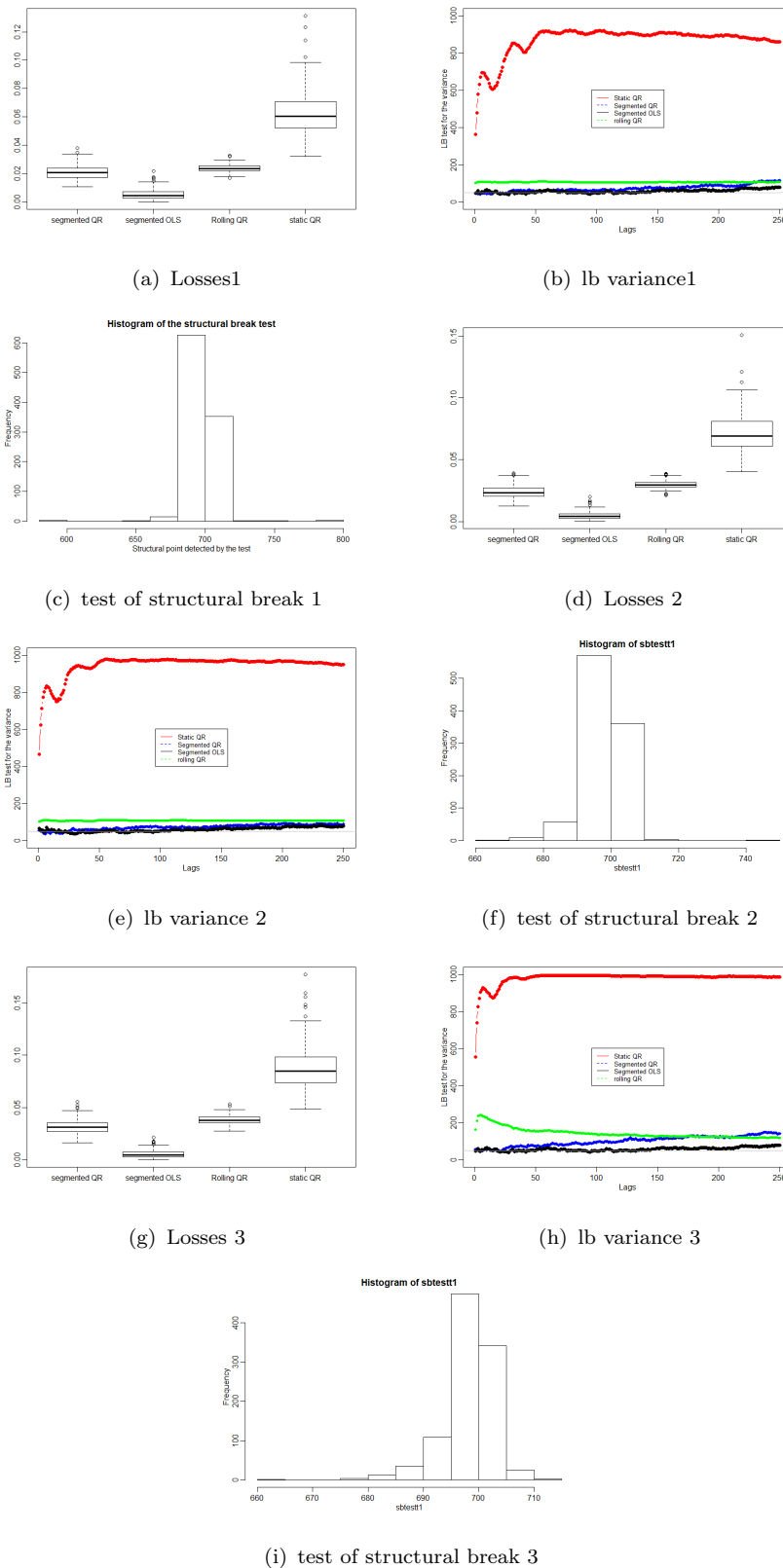
(d) test of structural break 1

(e) Losses 2

(f) lb mean 2

(g) lb variance2

(h) test of structural break 2

(i) Losses 3

(j) lb mean 3

(k) lb variance3

(l) test of structural break 3

FIGURE 2.4: Third data-generating process: Structural break through the coefficients of the seasonal pattern in the mean and variance. Panels a, e and i refer to the losses with first, second and third specifications respectively. Plots b,f and j reports the frequency of rejection of the null for the Ljung-box test results with first, second and third specifications respectively and over lags from 1:250 on the seasonally adjusted series obtained from segmented least squares, segmented quantile regression, rolling quantile regression and static quantile regression adjustments. Panels c,g and k report the frequency of rejection of the null for the Ljung-Box test over lags from 1 to 250 on the squared seasonally adjusted series (to detect periodic components in the variances). Plots d,h and l report the histogram of the structural break test with first, second and third specifications respectively. Number of simulations: 1000. Series length: 1000

(a) Losses1

(b) lb mean1

(c) lb variance1

(d) test of structural break 1

(e) Losses 2

(f) lb mean 2

(g) lb variance2

(h) test of structural break 2

(i) Losses 3

(j) lb mean 3

(k) lb variance3

(l) test of structural break 3

FIGURE 2.5: Fourth data-generating process: Structural break through the coefficients of the seasonal pattern that changes across quantiles. Panels a, e and i refer to the losses with first, second and third specifications respectively. Plots b,f and j report the frequency of rejection of the null for the Ljung-box test results with first, second and third specifications respectively and over lags from 1:250 on the seasonally adjusted series obtained from segmented least squares, segmented quantile regression, rolling quantile regression and static quantile regression adjustments. Panels c,g and k report the frequency of rejection of the null for the Ljung-Box test over lags from 1 to 250 on the squared seasonally adjusted series (to detect periodic components in the variances). Plots d,h and l report the histogram of the structural break test with first,second and third specifications respectively. Number of simulations: 1000. Series length: 1000
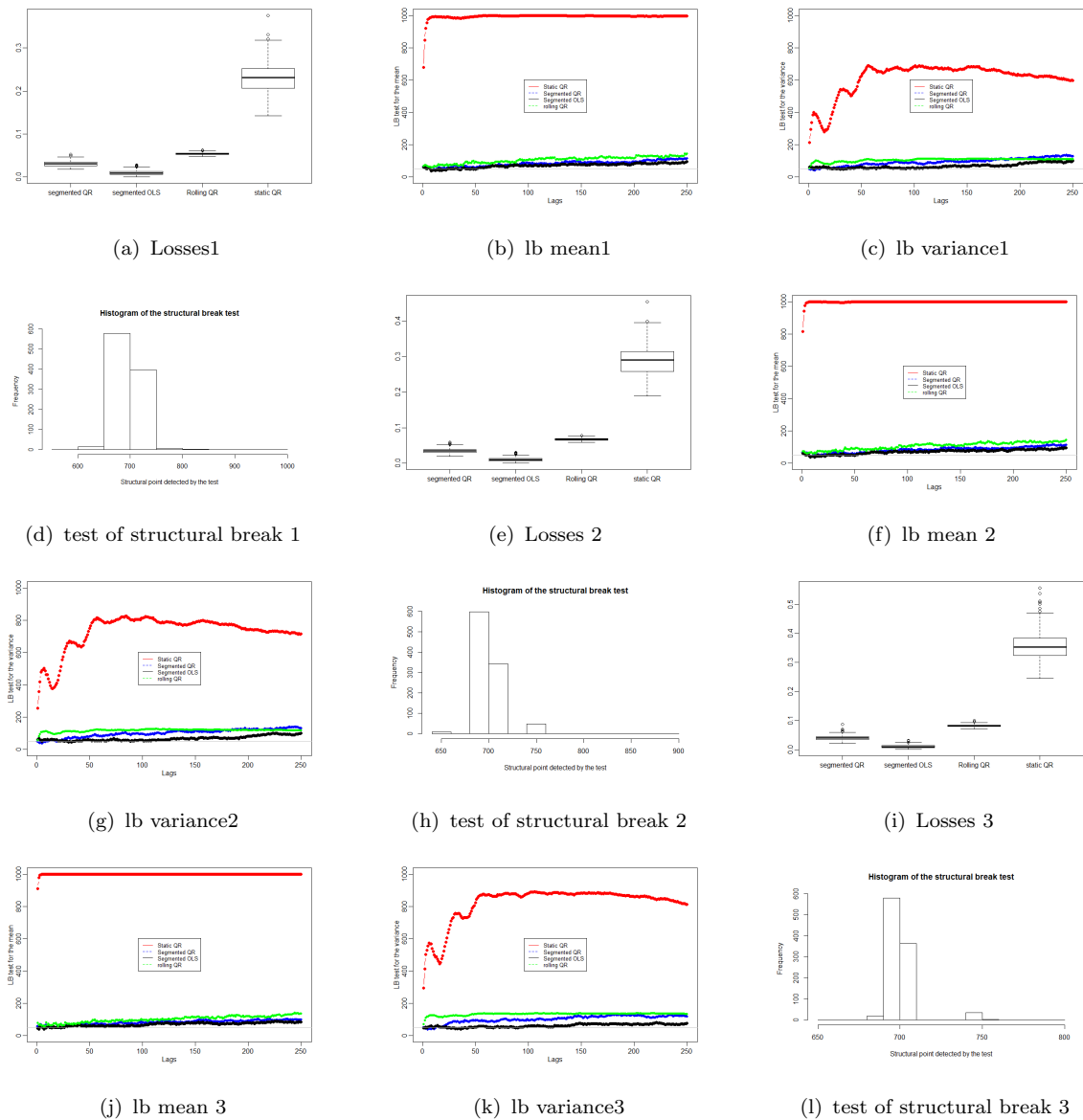
(a) Losses1

(b) lb mean1

(c) test of structural break 1

(d) Losses 2

(e) lb mean 2

(f) test of structural break 2

(g) Losses 3

(h) lb mean 3

(i) test of structural break 3

FIGURE 2.6: Fifth data-generating process: Structural break through the variance of the innovations of the time series with seasonal pattern in the mean. Plots a, d and g refer to the losses with first, second and third specifications of the structural break intensity respectively. Plots b,e and h report the frequency of rejection of the null for the Ljung-box test results with first, second and third specifications respectively and over lags from 1:250 on the seasonally adjusted series obtained from segmented least squares, segmented quantile regression, rolling quantile regression and static quantile regression adjustments. Panels c,f and i report the histogram of the structural break test with first, second and third specifications respectively. Number of simulations: 1000. Series length: 1000 observations.
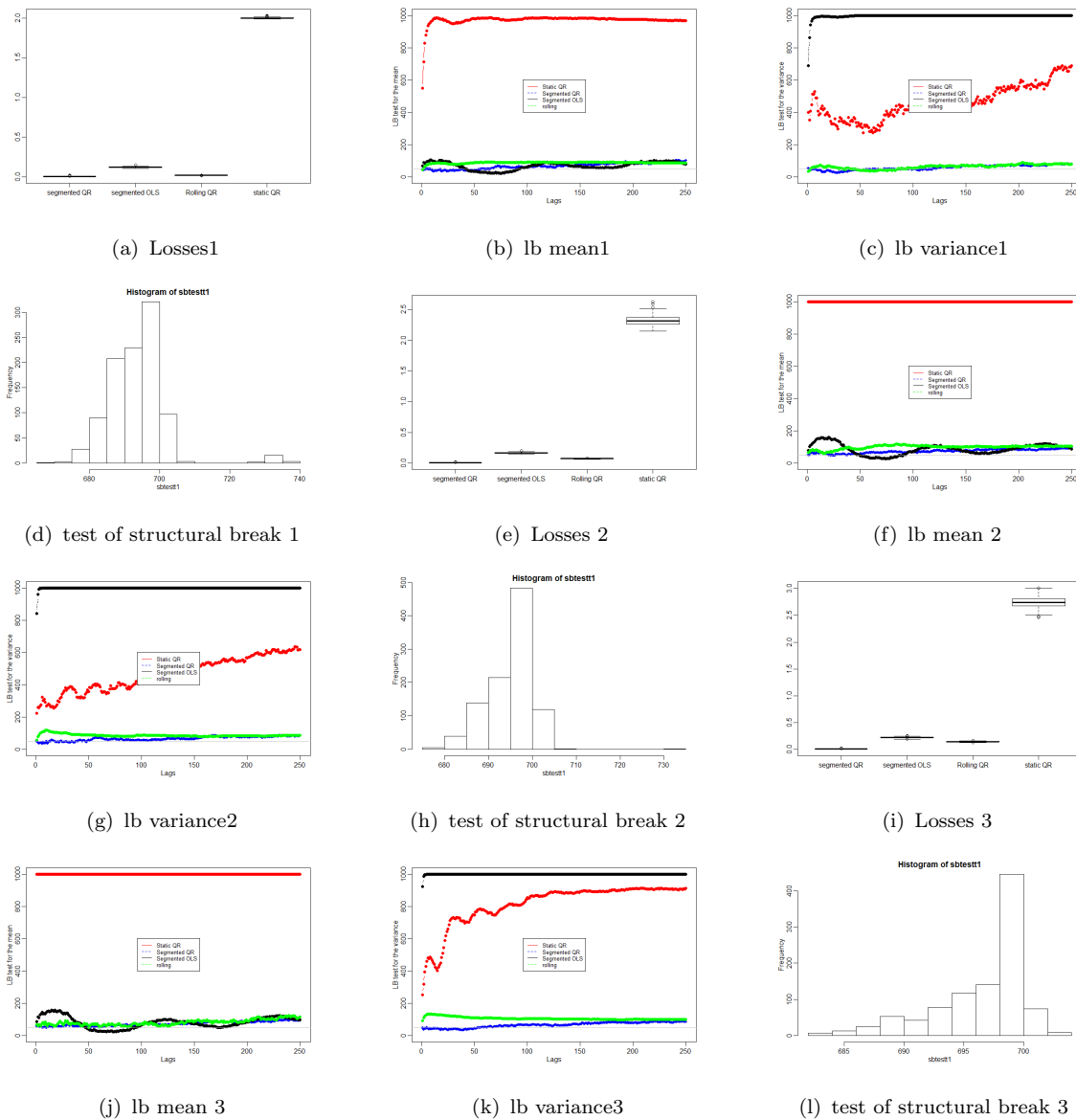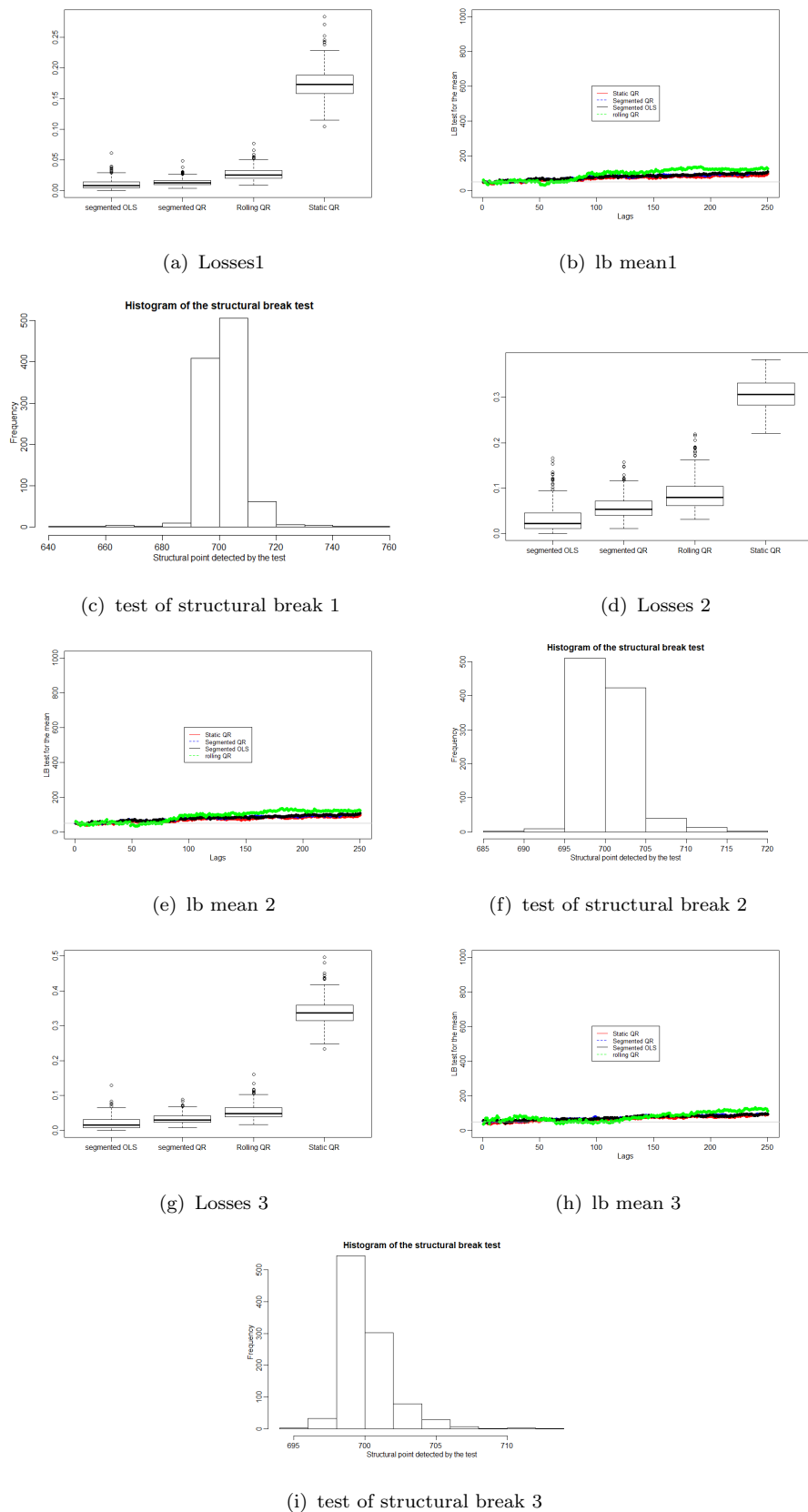
(a) Losses1

(b) lb variance1

(c) test of structural break 1

(d) Losses 2

(e) lb variance 2

(f) test of structural break 2

(g) Losses 3

(h) lb variance 3
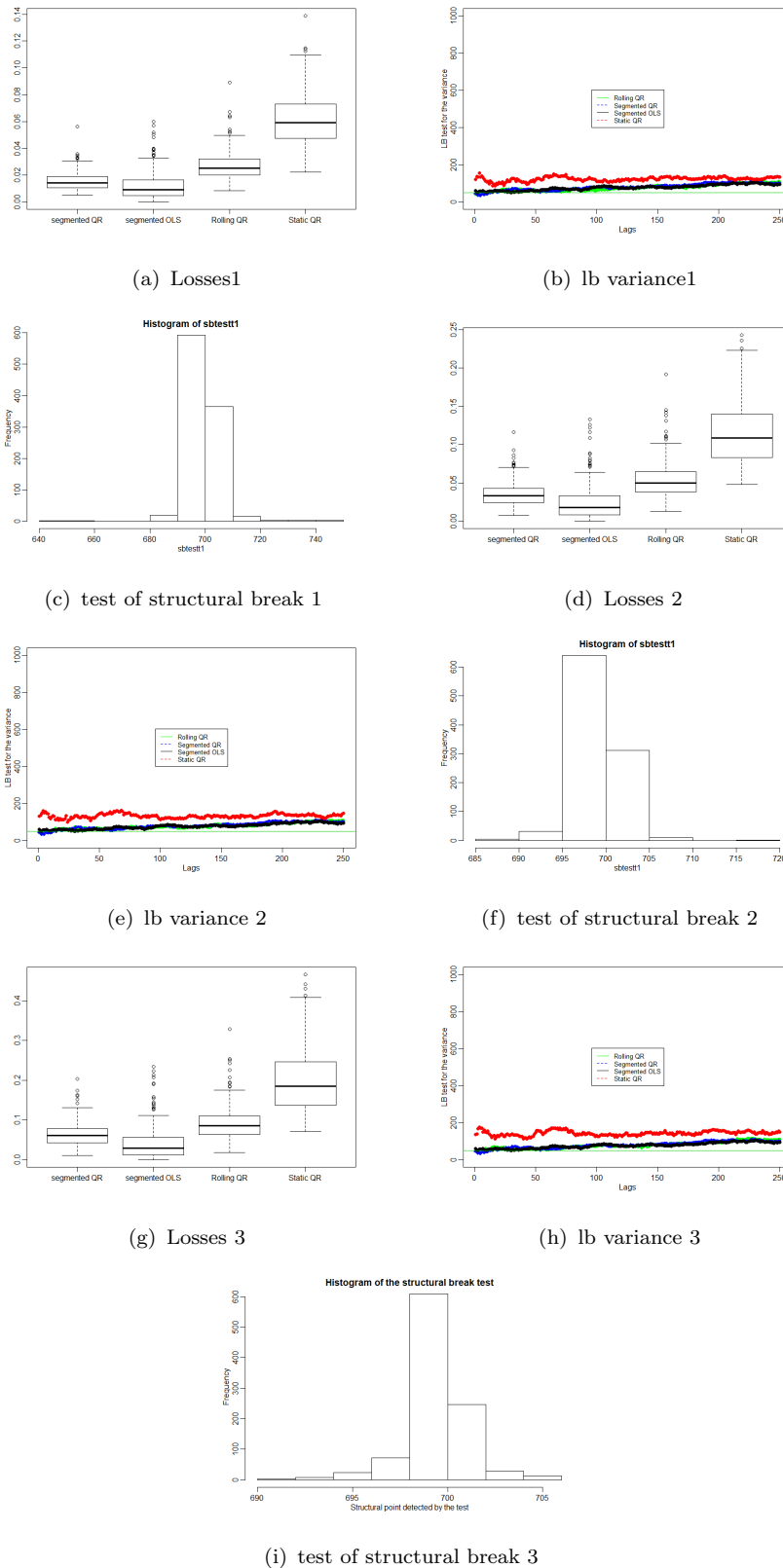

(i) test of structural break 3

FIGURE 2.7: Sixth data-generating process: Structural break through the variance of the innovations of the data-generating process with seasonal pattern in the variance. Plots a, d and g refer to the losses with first, second and third specifications of the structural break intensity respectively. Plots b, e and h report the frequency of rejection of the null for the Ljung-box test results with first, second and third specifications respectively and over lags from 1:250 on the squared seasonally adjusted series (to detect periodic components in the variances) obtained from segmented least squares, segmented quantile regression, rolling quantile regression and static quantile regression adjustments. Panels c,f and i report the histogram of the structural break test with first, second and third specifications respectively. Number of simulations: 1000. Series length: 1000 observations.

(a) Losses1

(b) lb mean1

(c) lb variance1

(d) test of structural break 1

(e) Losses 2

(f) lb mean 2

(g) lb variance2

(h) test of structural break 2

(i) Losses 3

(j) lb mean 3

(k) lb variance3
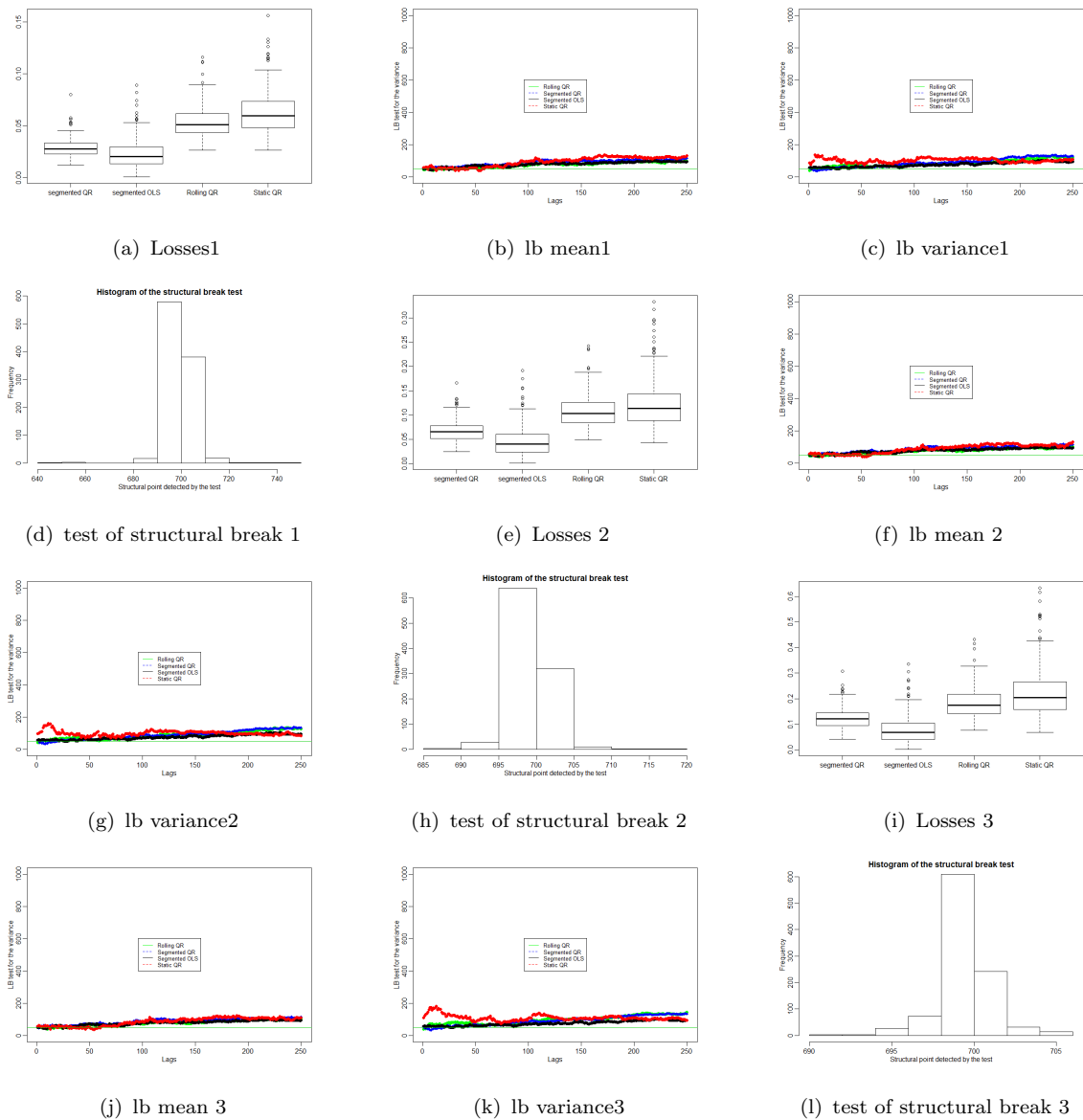
(l) test of structural break 3

FIGURE 2.8: seventh data-generating process: Structural break in the variance of the innovations of a data-generating process with a seasonal pattern in the mean and variance. Panels a, e and i refer to the losses with first, second and third specifications of the structural break intensity respectively. Plots b,f and j reports the frequency of rejection of the null for the Ljung-box test results with first, second and third specifications respectively and over lags from 1:250 on the seasonally adjusted series obtained from segmented least squares, segmented quantile regression, rolling quantile regression and static quantile regression adjustments. Panels c,g and k report the frequency of rejection of the null for the Ljung-Box test over lags from 1 to 250 on the squared seasonally adjusted series (to detect periodic components in the variances). Plots d,h and l report the histogram of the structural break test with first,second and third specifications respectively. Number of simulations: 1000. Series length: 1000 observations.

(a) Losses1

(b) lb mean1

(c) lb variance1

(d) test of structural break 1

(e) Losses 2

(f) lb mean 2

(g) lb variance2

(h) test of structural break 2

(i) Losses 3

(j) lb mean 3

(k) lb variance3
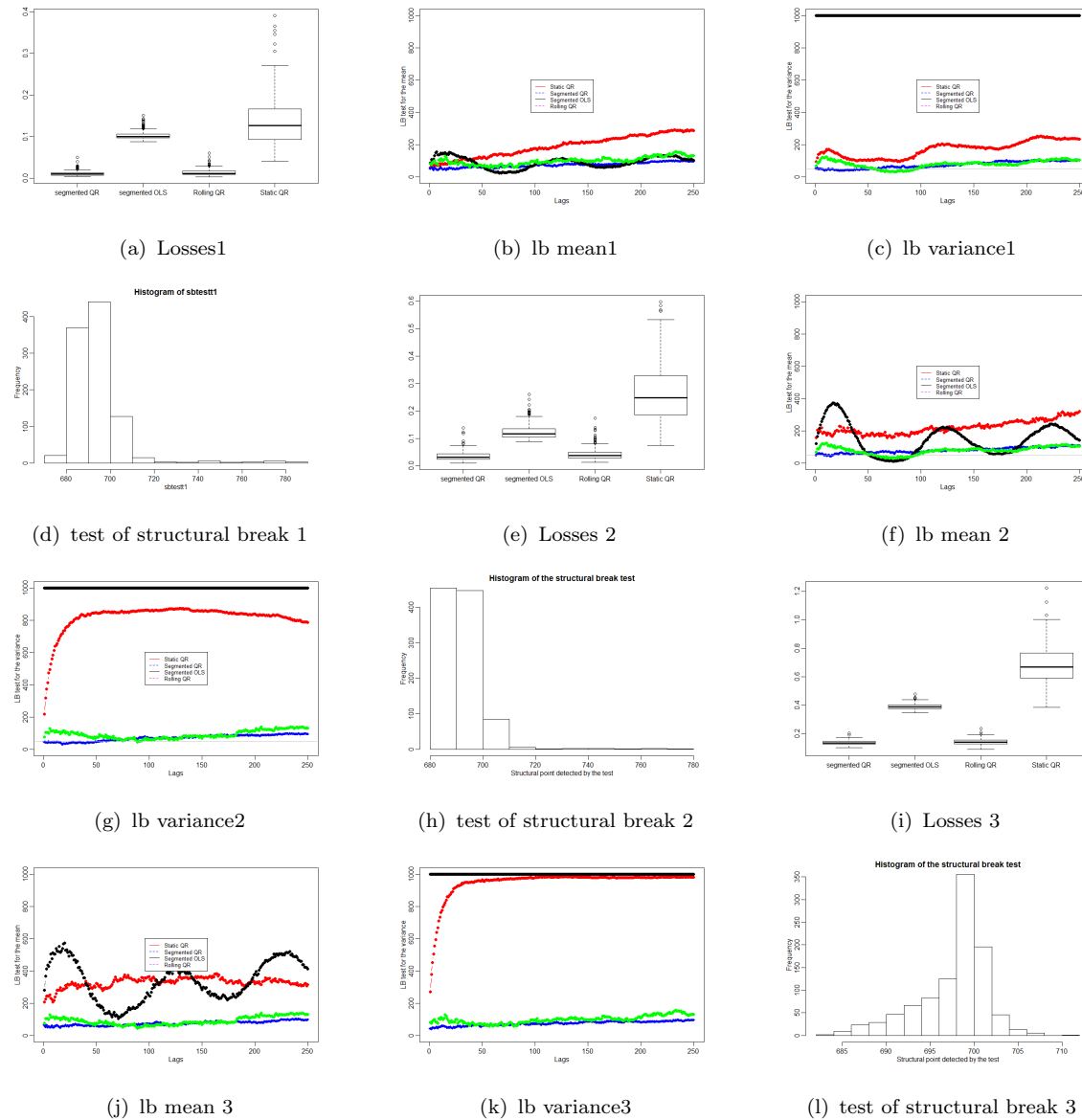
(l) test of structural break 3

FIGURE 2.9: Eighth data-generating process: Structural break in the variance of the innovations of the data-generating process with a seasonal pattern changes across quantiles. Panels a, e and i refer to the losses with first, second and third specifications respectively. Plots b,f and j reports the frequency of rejection of the null for the Ljung-box test results with first, second and third specifications respectively and over lags from 1:250 on the seasonally adjusted series obtained from segmented least squares, segmented quantile regression, rolling quantile regression and static quantile regression adjustments. Panels c,g and k report the frequency of rejection of the null for the Ljung-Box test over lags from 1 to 250 on the squared seasonally adjusted series (to detect periodic components in the variances). Plots d,h and l reports the histogram of the structural break test with first, second and third specifications respectively. Number of simulations: 1000. Series length: 1000

(a) Losses1

(b) lb mean1

(c) lb variance1

(d) test of structural break 1

(e) Losses 2

(f) lb mean 2

(g) lb variance2

(h) test of structural break 2

(i) Losses 3

(j) lb mean 3
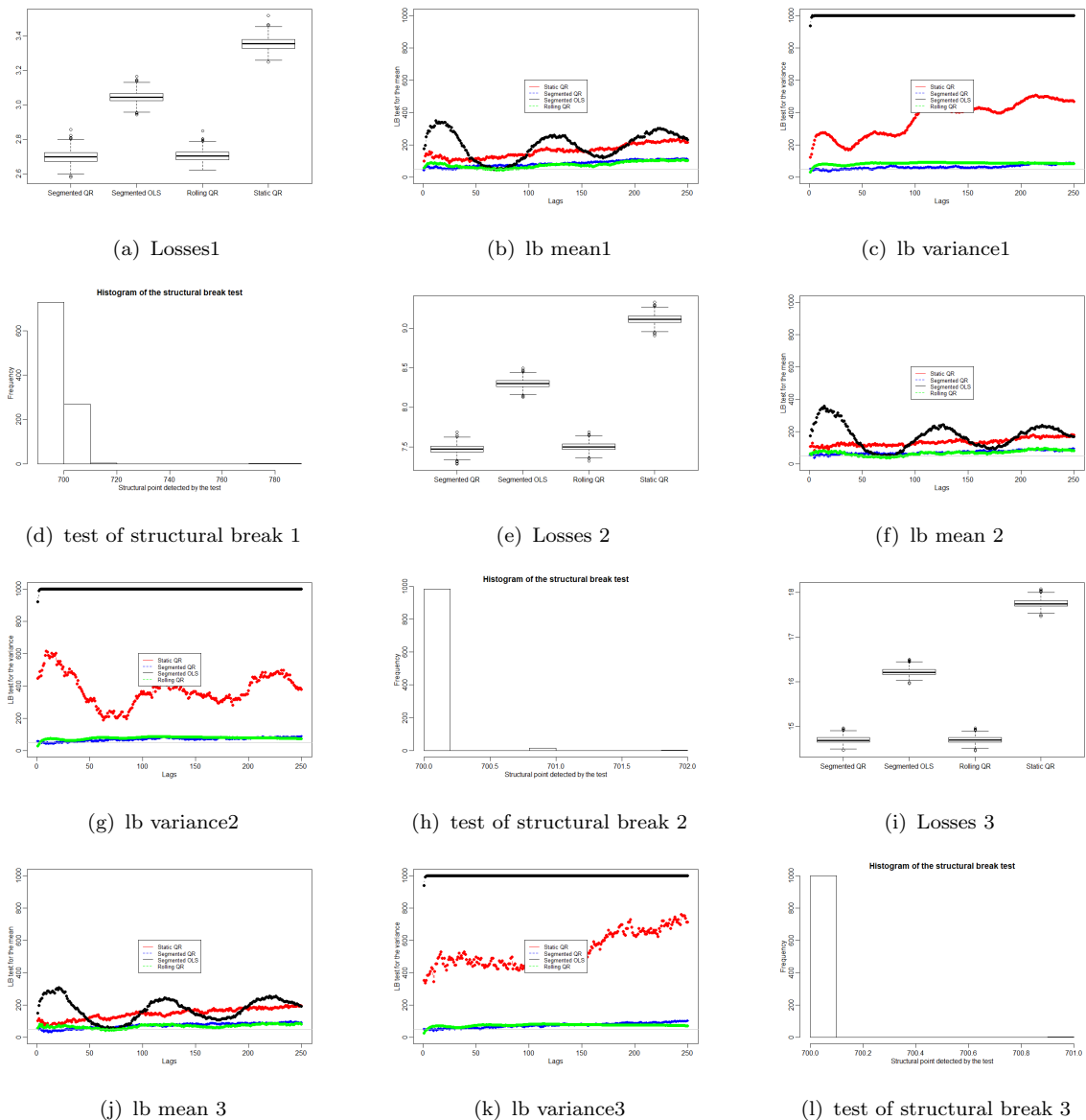

(k) lb variance3

(l) test of structural break 3

FIGURE 2.10: Ninth data-generating process: Structural break in the intercept of the data-generating process with a seasonal pattern changes across quantiles. Panels a, e and i refer to the losses with first, second and third specifications respectively. Plots b,f and j reports the frequency of rejection of the null for the Ljung-box test results with first,second and third specifications respectively and over lags from 1:250 on the seasonally adjusted series obtained from segmented least squares, segmented quantile regression, rolling quantile regression and static quantile regression adjustments. Panels c,g and k report the frequency of rejection of the null for the Ljung-Box test over lags from 1 to 250 on the squared seasonally adjusted series (to detect periodic components in the variances). Plots d,h and l report the histogram of the structural break test with first, second and third specifications respectively. Number of simulations: 1000. Series length: 1000

2.4   Application to real data

2.4.1   Industrial production Index

In this section, we apply our structural breaks-quantile regression approaches and compare them to the segmented least squares and static quantile regression adjustment methods. We work here on an industrial production index time series including 869 daily observations, from January 1947 to May 2019, these observations refer to the not seasonally adjusted industrial production (durable consumer goods) in the USA. Data has been recovered from the database of the Federal Reserve Economic Data (FRED).

We present the time series in Figure 2.11 (a), the growth rate of the time series in Figure 2.11 (b), the auto-correlation function of the growth rate in Figure 2.11 (c), the partial auto-correlation function of the growth rate in 2.11 (d) and the periodogram in Figure 2.11 (e). Visual inspection of graph Figure 2.11 (a) shows a clear evidence of an existence of at least one change-point in the time series around the 688th observation. this will impact on the auto correlation function in Figure 2.11 (c) as the observed pattern is due to the existence of both the seasonal pattern, as evidenced in the periodogram Figure 2.11 (e), and the structural break effect. Since the seasonal variation might be due to a number of seasonal factors, we sort out of the essential frequencies from the periodogram. From these we we choose the five essential frequencies that, to the best of our understanding, are the most relevant for the seasonality in the industrial production index growth rate time series. More specifically these were the six months, two months, three months, twelve months and four months. The first two columns in Table 2.1 represent the chosen empirical periods and their corresponding frequencies, while the third and fourth columns include the theoretical periods and the corresponding frequencies. In our empirical analyses, we also verified that the inclusion of additional frequencies does not lead to an improvement in the model fit (both for segmented least squares and quantile regression approaches).

Period (days)	Frequency	Adopted period	Adopted frequency
5.986	0.167	6.000	0.167
2.397	0.417	2.000	0.500
3.000	0.332	3.000	0.033
12.055	0.082	12.000	0.083
4.000	0.250	4.000	0.250

TABLE 2.1: Essential frequencies of the industrial production index growth rate.

(a) Monthly Industrial production index time series

(b) Monthly Industrial production index growth rate

(c) Acf of the growth rate time series

(d) Pacf of the growth rate time series

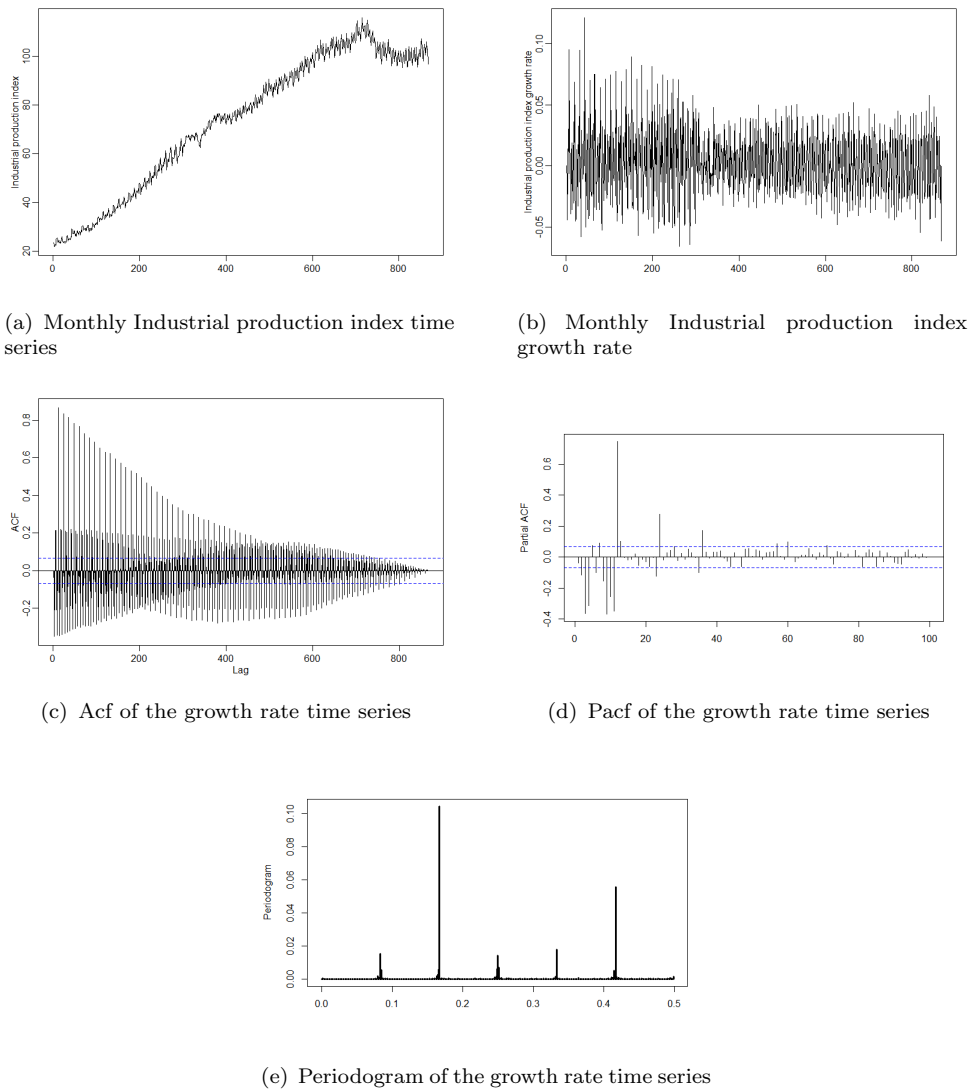(e) Periodogram of the growth rate time series

FIGURE 2.11: Monthly Industrial Production Index.

We start modeling the industrial production index growth rate by identifying the structural change using the structural break test. The test detects the change points at observations 201, 300 and 578. The overall model we consider for the growth rate time series is presented in equation 2.20 :

$$
\begin{aligned}
x_t \;=\; & d_1\alpha_0 + d_2\alpha_0 + d_3\alpha_0 + d_4\alpha_0 + \sum_{j=1}^{5} d1\left[\delta_j cos\left(2\pi f_j\right) + \gamma_j sin\left(2\pi f_j\right)\right] \\
& + \sum_{j=1}^{5} d2\left[\delta_j cos\left(2\pi f_j\right) + \gamma_j sin\left(2\pi f_j\right)\right] + \sum_{j=1}^{5} d3\left[\delta_j cos\left(2\pi f_j\right) + \gamma_j sin\left(2\pi f_j\right)\right] \\
& + \sum_{j=1}^{5} d4\left[\delta_j cos\left(2\pi f_j\right) + \gamma_j sin\left(2\pi f_j\right)\right] + \varepsilon_t
\end{aligned}
\tag{2.20}
$$

d_i is a dummy variable that takes the value of 1 if t $\in$ sub-period i as determined by the break dates and 0 otherwise. The frequencies f_j reported in Table 2.1. Figure 2.12 shows the estimated parameters for the non-crossing static quantile regression curves over the considered 99 quantiles. On the other hand, Figures 2.13, 2.14, 2.15 and 2.16 present the estimated parameters for the non crossing segmented quantile regression. The coefficients of the static quantile regression show a relevant variation over quantiles which suggest using quantile regression over the least-squares approach. However, by focusing on the estimation of the four segments it is clear that the structural breaks affect the variation of the coefficients across quantiles for each segment in a different way. This supports the existence of the seasonal patterns that change across both the quantiles and the segments defined by the change-points. As a further confirmation of our choice, we compute the coefficient stability test on the original series using static quantile regression and the four-segments quantile regression, obtaining a p-value equal to zero in the five cases.

Figures 2.17 and 2.18 represent a comparison between the seasonal adjustment approaches using the auto-correlation function and the periodogram of the seasonally adjusted time series level and squared values. It is clear that using the static quantile regression approach leads to improper seasonally adjusted series due to the neglect of the effect of the structural breaks in the series. In contrast both the segmented approaches and the rolling analysis show almost similar and satisfactory results as the auto-correlation functions and the periodograms are free from the periodic behavior in both the series level and its squared level.
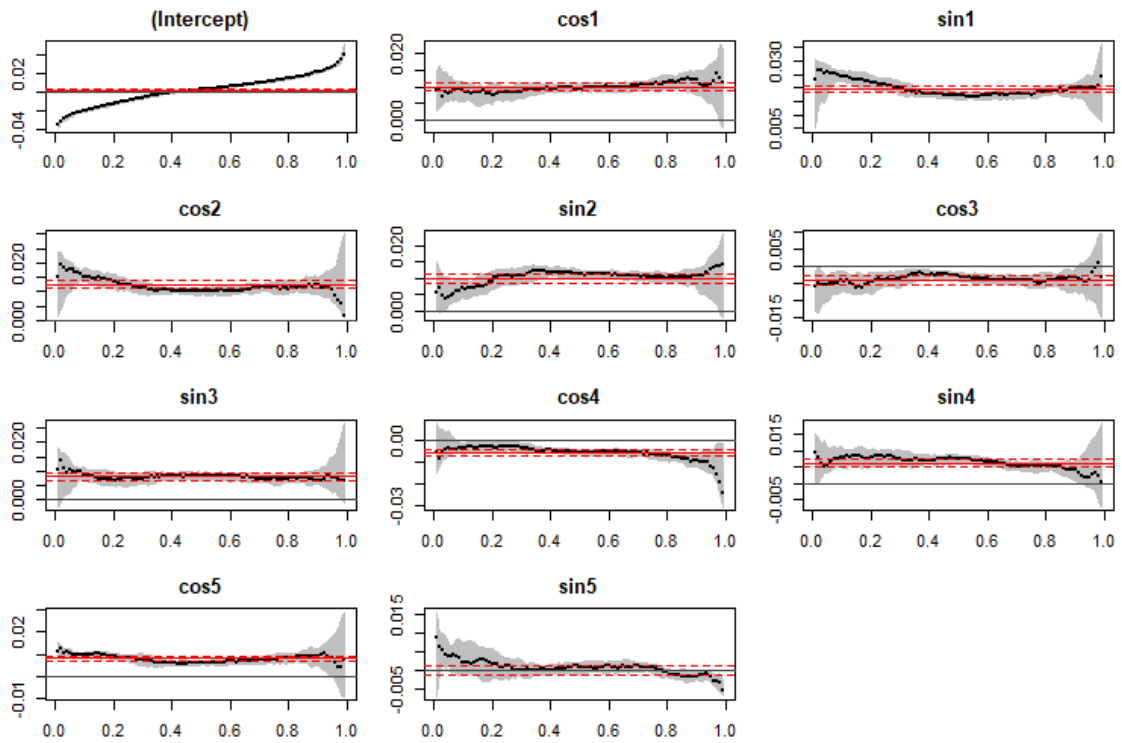
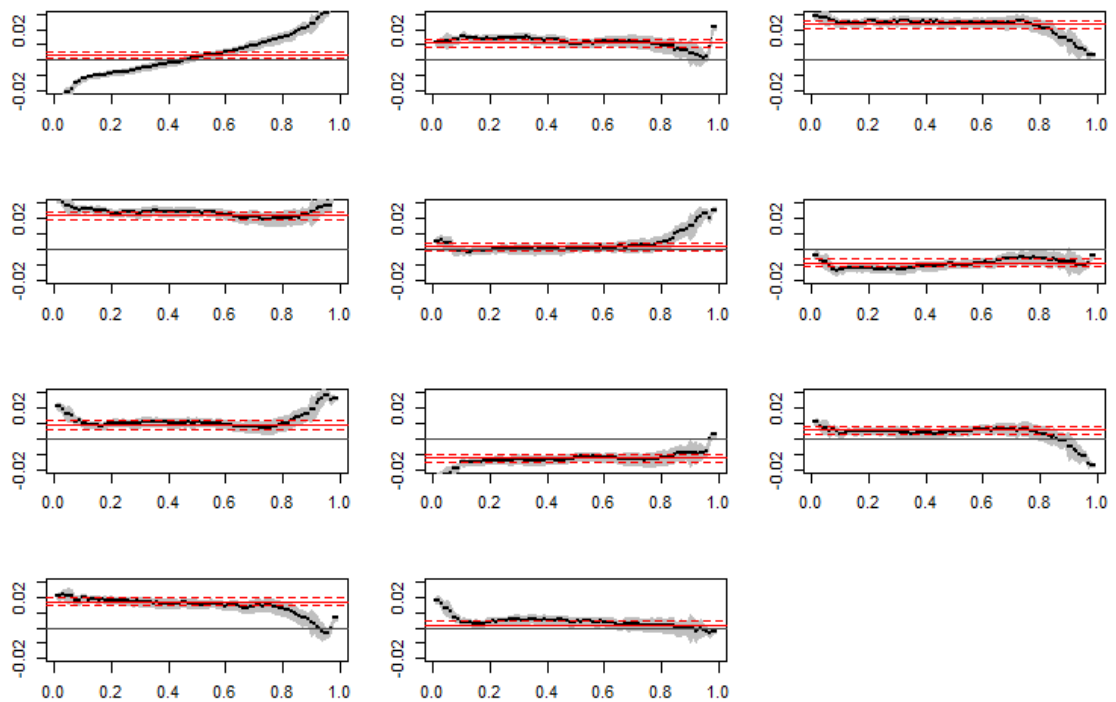FIGURE 2.12: Static quantile regression estimation..

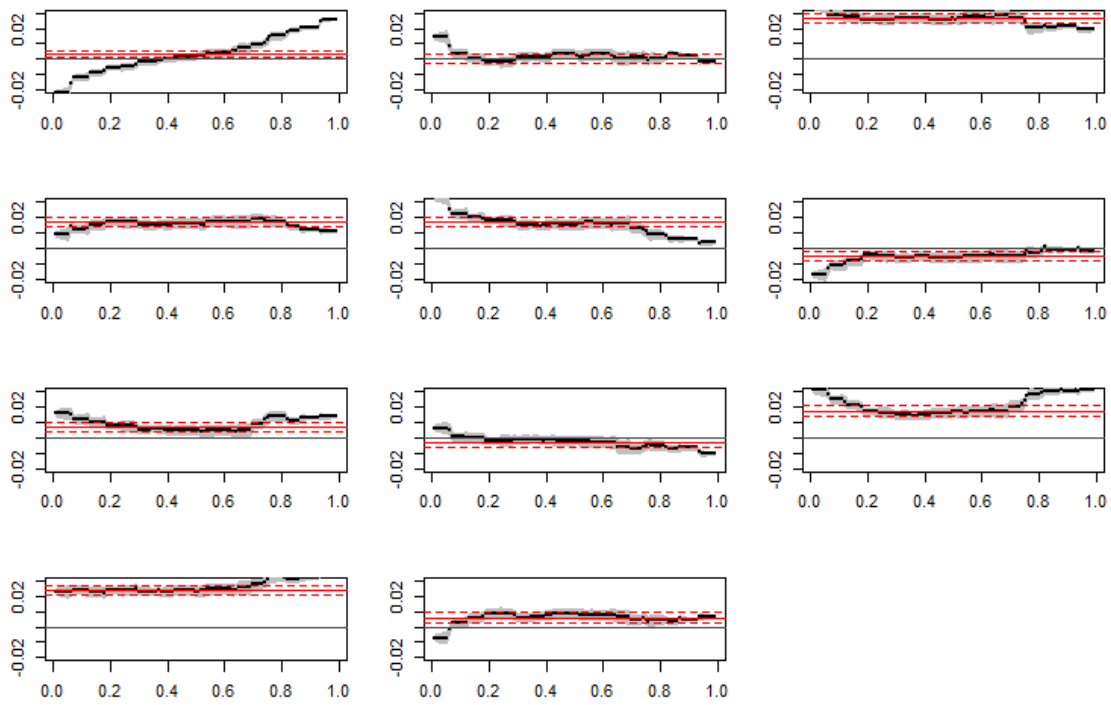FIGURE 2.13: Quantile regression estimation segment 1.

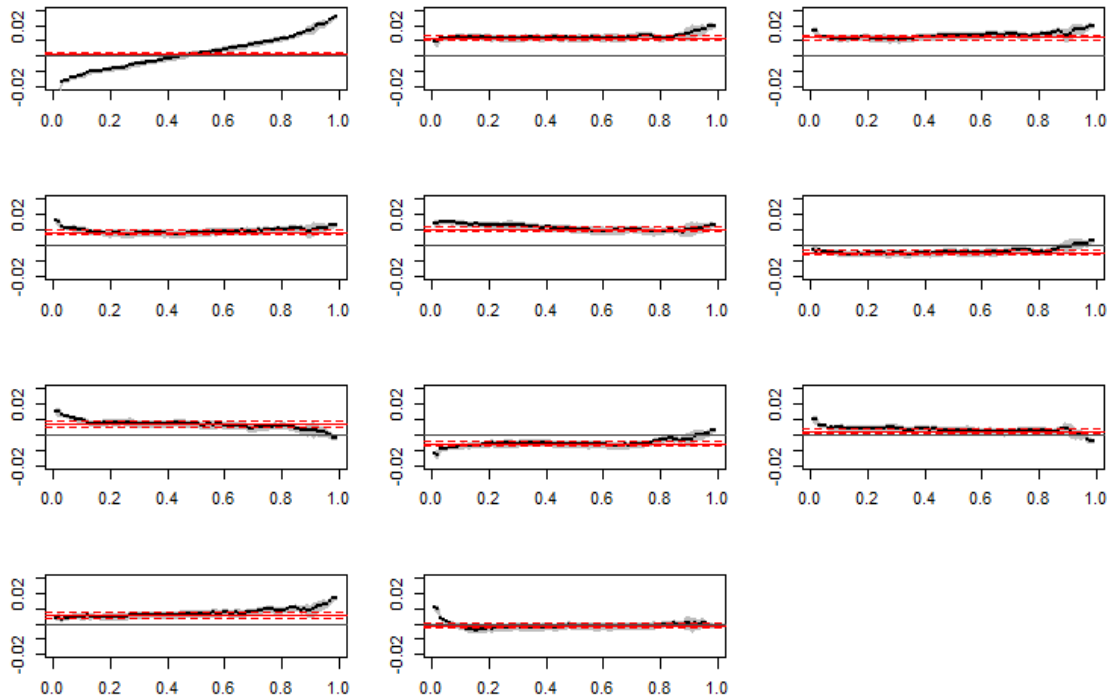FIGURE 2.14: Quantile regression estimation segment 2.

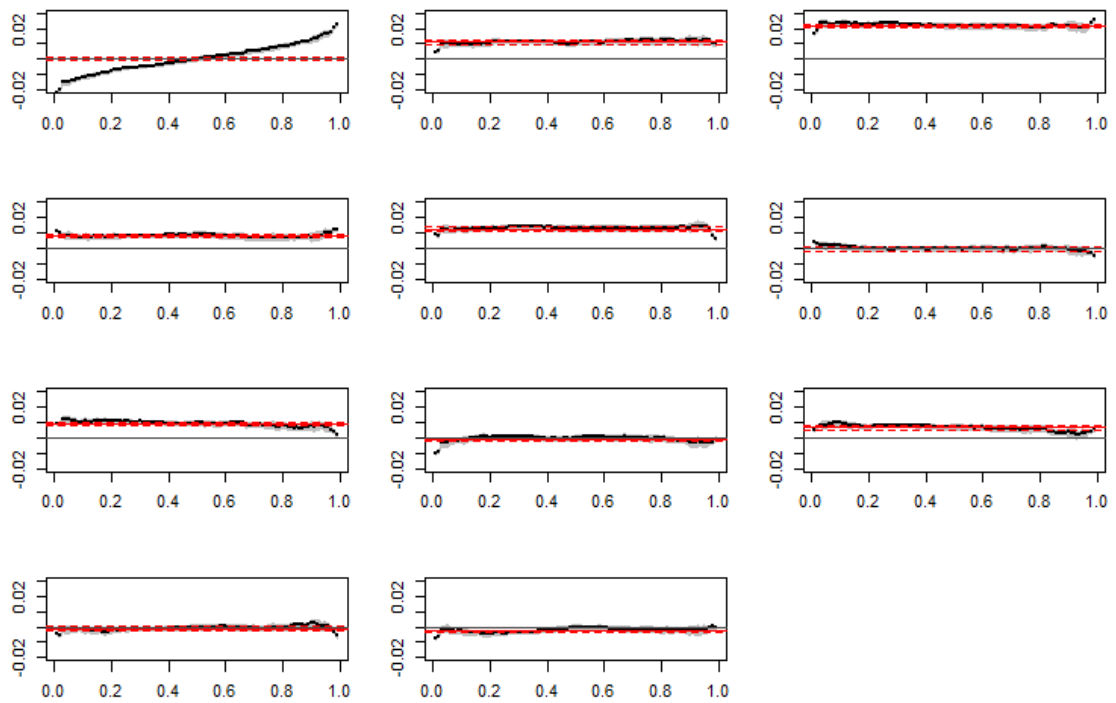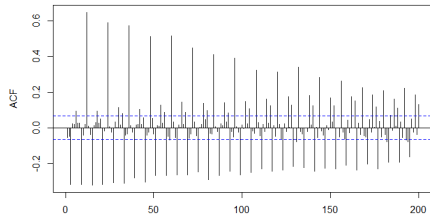FIGURE 2.15: Quantile regression estimation segment 3. .
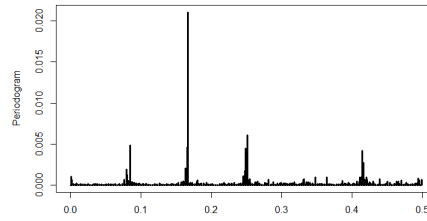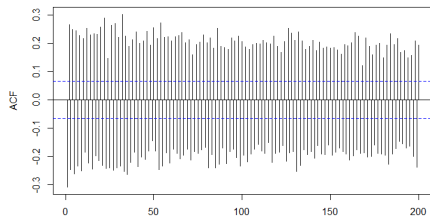
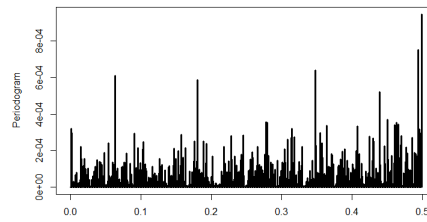FIGURE 2.16: Quantile regression estimation segment 4. .

(a) ACF of the seasonally adjusted time series using static quantile regression.
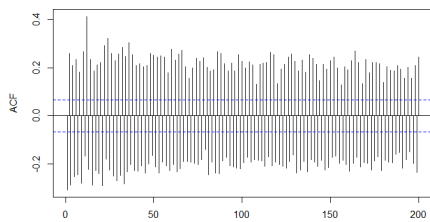
(b) Periodogram of the seasonally adjusted time series using static quantile regression.
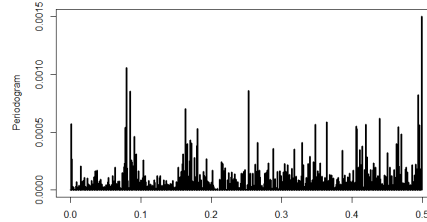
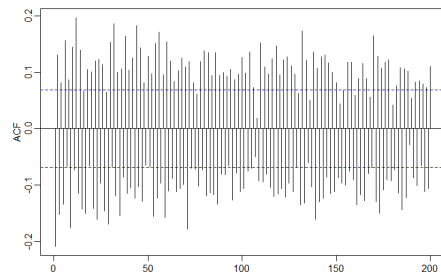(c) ACF of the seasonally adjusted time series using segmented quantile regression.

(d) Periodogram of the seasonally adjusted time series using segmented quantile regression.

(e) ACF of the seasonally adjusted time series using segmented least squares.

(f) Periodogram of the seasonally adjusted time series using segmented least squares.

(g) ACF of the seasonally adjusted time series using rolling quantile regression.

(h) Periodogram of the seasonally adjusted time series using rolling quantile regression.

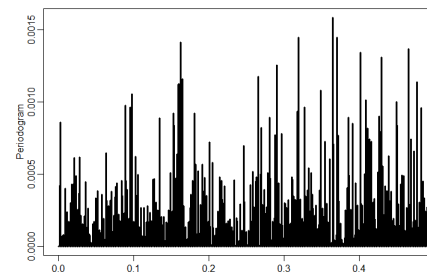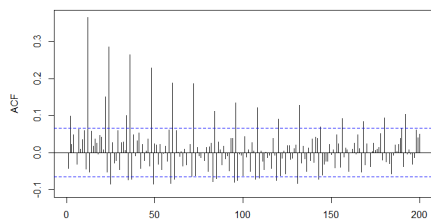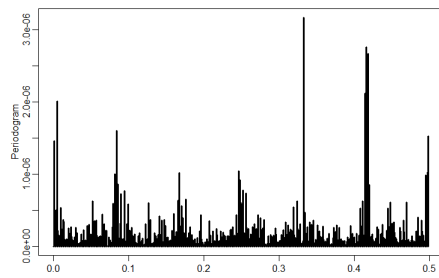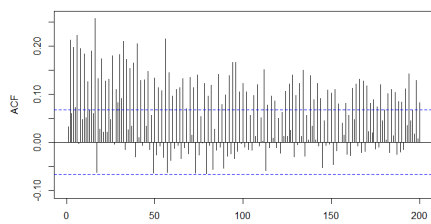FIGURE 2.17: The auto-correlation function and the periodogram of the seasonally adjusted time series

(a) ACF of the squared values of the seasonally adjusted time series using static quantile regression.

(b) Periodogram of the squared values of the seasonally adjusted time series using static quantile regression.

(c) ACF of the squared values of the seasonally adjusted time series using segmented quantile regression.

(d) Periodogram of the squared values of the seasonally adjusted time series using segmented quantile regression.

(e) ACF of the squared values of the seasonally adjusted time series using segmented least squares.

(f) Periodogram of the squared values of the seasonally adjusted time series using segmented least squares.
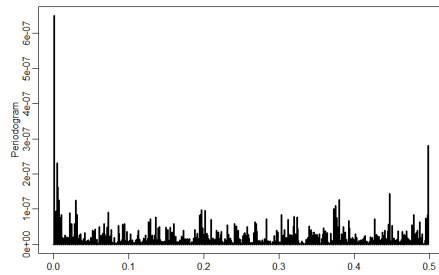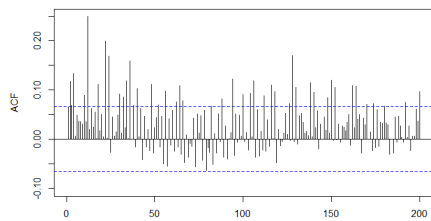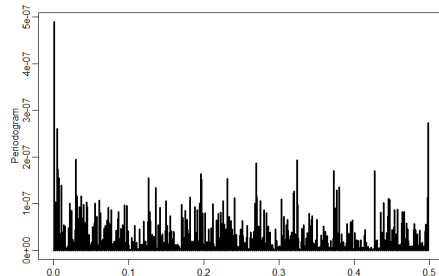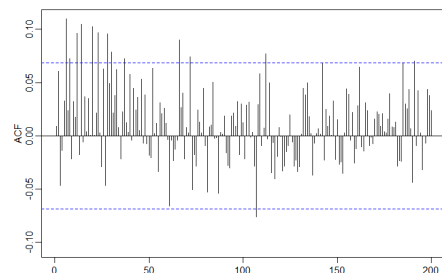
(g) ACF of the squared values of the seasonally adjusted time series using rolling quantile regression.

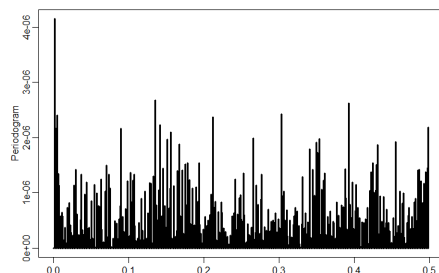(h) Periodogram of the squared values of the seasonally adjusted time series using rolling quantile regression.

FIGURE 2.18: The auto-correlation function and the periodogram of the squared values of the seasonally adjusted time series.

2.5   Conclusion

We extend the introduced seasonal adjustment approach in the first chapter to ensure a proper adjustment of the seasonal pattern when the time series is characterized by a periodic stochastic component. The proposals, in this situation, account for the case when the seasonal patterns might have varying impacts on the conditional density of a variable of interest. We compare our proposals, namely the segmented quantile regression and the rolling analysis, to the segmented least squares and the proposed approach in the first chapter for performing a seasonal adjustment. We evaluate the models using simulations of different data generating processes that present various seasonal patterns and structural breaks. The evaluation has been extended to a real time series namely the industrial production index. The findings of the first, second, third, fifth, sixth and seventh data generating processes, when the data-generating process includes additive and/or multiplicative seasonal components (invariant across quantiles) with different patterns of the structural breaks, show that the proposed approaches correctly perform the seasonal adjustment. The results of the validation criteria using the segmented approaches and the rolling analysis show small values of the loss function. The rejection frequencies of the Ljung-box test are near the nominal level using these approaches. The static quantile regression approach leads to undesirable results. However, the findings of the fourth, eighth and ninth data generating processes, when the periodic component changes across quantiles, show that our proposals outperform the other methods including the segmented least squares. This is clear from the results of the loss function which has very small values using the proposed approaches in comparison with other approaches. The rejection frequencies of the Ljung-box test are near the nominal level and at the same time very close using only the proposed approaches. We also evaluate the structural break test with different intensities of the introduced break-point. The results show a correct detection of the breakpoint in all data generating processes. The empirical example confirms the flexibility of the proposed approaches for economics data. Further extensions of this study could be possible by taking into account the methods of the time varying quantiles. The conditional autoregressive value at risk (CAViaR) model Engle and Manganelli (2004) or conditional score driven models Harvey (2013), Patton *et al.* (2019) can be considered for this purpose. Also the comparison of forecasts made on the seasonally adjusted series from both linear and quantile regression approaches is another extension.

Bibliography

Alexandridis, A. K. and Zapranis, A. D. (2014) Modeling financial wind derivatives. *Wavelet Neural Networks: With Applications in Financial Engineering, Chaos, and Classification* pp. 197–218.

Aretz, K. and Arisoy, E. (2016) Do stock markets price expected stock skewness? new evidence from quantile regression skewness forecasts. *Available at SSRN* .

Aue, A., Cheung, R. C., Lee, T. C., Zhong, M. *et al.* (2017) Piecewise quantile autoregressive modeling for nonstationary time series. *Bernoulli* **23**(1), 1–22.

Auger, I. E. and Lawrence, C. E. (1989) Algorithms for the optimal identification of segment neighborhoods. *Bulletin of mathematical biology* **51**(1), 39–54.

Bai, J. and Perron, P. (1998) Estimating and testing linear models with multiple structural changes. *Econometrica* pp. 47–78.

Bonaccolto, G. (2016) Quantile regression methods in economics and finance .

Bondell, H. D., Reich, B. J. and Wang, H. (2010) Noncrossing quantile regression curve estimation. *Biometrika* **97**(4), 825–838.

Brockwell, P. J., Davis, R. A. and Calder, M. V. (2002) *Introduction to time series and forecasting.* Volume 2. Springer.

Caporin, M. and Preś, J. (2012) Modelling and forecasting wind speed intensity for weather risk management. *Computational Statistics & Data Analysis* **56**(11), 3459–3476.

Caporin, M. and Preś, J. (2013) Forecasting temperature indices density with time-varying long-memory models. *Journal of Forecasting* **32**(4), 339–352.

Cenesizoglu, T. and Timmermann, A. (2007) Predictability of stock returns: a quantile regression approach. Technical report, Technical report, Technical Report, Cirano.

Chen, C. W., Khamthong, K. and Lee, S. (2017) Structural breaks of capm-type market model with heteroskedasticity and quantile regression. In *Robustness in Econometrics*, pp. 111–134. Springer.

Chen, J. and Gupta, A. K. (2011) *Parametric statistical change point analysis: with applications to genetics, medicine, and finance.* Springer Science & Business Media.

Chen, P.-F., Lee, C.-C., Lee, C.-C. and Huang, J.-X. (2016) A dynamic analysis of exchange rate exposure: The impact of china's renminbi. *The World Economy* **39**(1), 132–157.

Cho, J. S., Kim, T.-h. and Shin, Y. (2015) Quantile cointegration in the autoregressive distributed-lag modeling framework. *Journal of econometrics* **188**(1), 281–300.

Dette, H., Hallin, M., Kley, T., Volgushev, S. *et al.* (2015) Of copulas, quantiles, ranks and spectra: An $l_-\{1\}$-approach to spectral analysis. *Bernoulli* **21**(2), 781–831.

Eckley, I. A., Fearnhead, P. and Killick, R. (2011) Analysis of changepoint models. *Bayesian Time Series Models* pp. 205–224.

Engle, R. F. and Manganelli, S. (2004) Caviar: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics* **22**(4), 367–381.

Fisher, R. A. (1929) Tests of significance in harmonic analysis. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* **125**(796), 54–59.

Guler, K., Ng, P. T. and Xiao, Z. (2017) Mincer–zarnowitz quantile and expectile regressions for forecast evaluations under aysmmetric loss functions. *Journal of Forecasting* **36**(6), 651–679.

Hällman, L. (2017) The rolling window method: Precisions of financial forecasting.

Harvey, A. C. (2013) *Dynamic models for volatility and heavy tails: with applications to financial and economic time series.* Volume 52. Cambridge University Press.

Hawkins, D. M., Qiu, P. and Kang, C. W. (2003) The changepoint model for statistical process control. *Journal of quality technology* **35**(4), 355–366.

Hsu, D. (1982) A bayesian robust detection of shift in the risk structure of stock market returns. *Journal of the American Statistical Association* **77**(377), 29–39.

Killick, R., Fearnhead, P. and Eckley, I. A. (2012) Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association* **107**(500), 1590–1598.

Koenker, R. and Bassett, J. G. (1978) Regression quantiles. *Econometrica* pp. 33–50.

Koenker, R. and Xiao, Z. (2002) Inference on the quantile regression process. *Econometrica* **70**(4), 1583–1612.

Lahiani, A. (2019) Exploring the inflationary effect of oil price in the us: A quantile regression approach over 1876-2014. *International Journal of Energy Sector Management* **13**(1), 60–76.

Lehmann, E. L. and Romano, J. P. (2006) *Testing statistical hypotheses.* Springer Science & Business Media.

Lindsey, D. T., Holzman, P. S., Haberman, S. and Yasillo, N. J. (1978) Smooth-pursuit eye movements: A comparison of two measurement techniques for studying schizophrenia. *Journal of Abnormal Psychology* **87**(5), 491.

Linton, O. and Whang, Y.-J. (2007) The quantilogram: With an application to evaluating directional predictability. *Journal of Econometrics* **141**(1), 250–282.

Neuenschwander, A. L. and Crews, K. A. (2008) Disturbance, management, and landscape dynamics. *Photogrammetric Engineering & Remote Sensing* **74**(6), 753–764.

Pascual, A. G. (2003) Assessing european stock markets (co) integration. *Economics Letters* **78**(2), 197–203.

Patton, A. J., Ziegel, J. F. and Chen, R. (2019) Dynamic semiparametric models for expected shortfall (and value-at-risk). *Journal of Econometrics* **211**(2), 388–413.

Scott, A. J. and Knott, M. (1974) A cluster analysis method for grouping means in the analysis of variance. *Biometrics* pp. 507–512.

Sen, A. and Srivastava, M. S. (1975) On tests for detecting change in mean. *The Annals of statistics* pp. 98–108.

Xu, Q. and Childs, T. (2013) Evaluating forecast performances of the quantile autoregression models in the present global crisis in international equity markets. *Applied Financial Economics* **23**(2), 105–117.

Young, P. C., Pedregal, D. J. and Tych, W. (1999) Dynamic harmonic regression. *Journal of forecasting* **18**(6), 369–394.

Zanin, L. and Marra, G. (2012) Rolling regression versus time-varying coefficient modelling: An empirical investigation of the okun's law in some euro area countries. *Bulletin of Economic Research* **64**(1), 91–108.

Zivot, E. and Wang, J. (2007) *Modeling financial time series with S-Plus®*. Volume 191. Springer Science & Business Media.

Mohammed Elseidi

CURRICULUM VITAE

Contact Information

University of Padova
Department of Statistics
via Cesare Battisti, 241-243
35121 Padova. Italy.

Tel. +39 049 827 4111
e-mail: elseidi@stat.unipd.it

Current Position

Since October 2016; (expected completion: September 2019)
PhD Student in Statistical Sciences, University of Padova.
Thesis title: Quantile regression-based seasonal adjustment
Supervisor: Prof. Massimiliano Caporin

Research interests

- Time series analysis.
- Seasonal adjustment.

Education

October 2012 – October 2015
Master (*laurea specialistica/magistrale*) degree in Applied statistics .
University of Mansoura, Faculty of commerce
Title of dissertation: "Forecasting By Using Fuzzy Time Series Models Acomparative study With Box-Jenkins: / Applied study. "

October 2008 – July 2012
Bachelor degree (*laurea triennale*) in Commerce.
University of Mansoura, Faculty of Commerce

Awards and Scholarship

October 2016
PhD scholarship in department of statistical sciences university of Padova.

Computer skills

- R
- LATEX
- SPSS
- Microsoft Office

Language skills

Arabic: native; English: good; Italian:basic

Publications

Working papers

Caporin, Massimiliano., Elseidi, Mohammed., Quantile regression-based seasonal adjustment (2019).
Caporin, Massimiliano., Elseidi, Mohammed., Seasonal adjustment by quantile regression in the presence of structural break (2019).

Conference presentations

Elseidi, Mohammed., Caporin, Massimiliano.,(2018).(Quantile Regression-based seasonal adjustment.). (invited) *The 11th International Conference of the ERCIM WG on Computational and Methodological Statistics (CMStatistics 2018*, 14/12/2018-18/12/2018 – University of Pisa, Italy.

References

Prof.Massimiliano Caporin
Full Professor of Economic Statistic, Department of Statistical Sciences, Padua University
Address: Via Cesare Battisti, 241/243 35121 Padua
e-mail: massimiliano.caporin@unipd.it