Charles University in Prague

Faculty of Social Sciences
Institute of Economic Studies

MASTER'S THESIS

The least weighted squares and its asymptotics

Author: **Bc. Magdaléna Raušová**

Supervisor: **Prof. RNDr. Jan Ámos Víšek, CSc.**

Academic Year: **2015/2016**

Declaration of Authorship

The author hereby declares that she compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain a different or the same degree.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis document in whole or in part.

Prague, July 16, 2016

Signature

Acknowledgments

I would like to express my gratitude to Prof. RNDr. Jan Ámos Víšek, CSc. for his time, valuable advice, and for being in general very helpful during writing this thesis. I am also grateful for all the materials that he provided me with and that I could have based my thesis on.

Abstract

When there are some influential observations present in a data set (such as outliers or leverage points), the use of some robust method may be desirable for being able to draw relevant conclusions from an econometric analysis. In order to use these methods properly, we need some diagnostic tools. To be able to derive these tools theoretically, we first need to know the form of the asymptotic representation of corresponding estimator. This thesis derives the asymptotic representation of the estimator obtained by the method of least weighted squares under the assumption of heteroskedastic residuals. The tightness of the estimator and its asymptotic representation under several levels of contamination is also shown in a simulation study.

JEL Classification	C01, C15, C16, C46
Keywords	Regression model, asymptotic representation

Author's e-mail	`rausova.magdalena@gmail.com`
Supervisor's e-mail	`visek@fsv.cuni.cz`

Abstrakt

V případě, že se v datasetu vyskytují hodnoty, které se výrazně liší od většiny ostatních hodnot (jako například odlehlá a vlivná pozorování), může být vhodné použít některou z robustních metod, abychom mohli z ekonometrické analýzy vyvodit relevantní závěry. Pro správné použití těchto metod potřebujeme diagnostické nástroje. Aby bylo možné tyto nástroje teoreticky odvodit, je nutné znát formu asymptotické reprezentace příslušného odhadu. V této práci je odvozena asymptotická reprezentace pro odhad získaný metodou nejmenších vážených čtverců, a to za předpokladu heteroskedastických reziduí. Těsnost mezi tímto odhadem a jeho asymptotickou reprezentací při různých stupních kontaminace je ukázána také v simulační studii.

Klasifikace JEL	C01, C15, C16, C46
Klíčová slova	Regresní model, asymptotická reprezentace

E-mail autora	`rausova.magdalena@gmail.com`
E-mail vedoucího práce	`visek@fsv.cuni.cz`

Contents

List of Tables

List of Figures

Acronyms

ARE	asymptotic relative efficiency
CLT	central limit theorem
EDF	empirical distribution function
FE	fixed effects
FWE	fixed weighted effects
iid	independent and identically distributed
K-S test	Kolmogorov-Smirnov test
LMS	least median of squares
LTS	least trimmed squares
LWS	least weighted squares
MSE	mean squared error
OLS	ordinary least squares
RE	random effects
RWE	random weighted effects
WLOG	without loss of generality
WLS	weighted least squares

Master's Thesis Proposal

Author	Bc. Magdaléna Raušová
Supervisor	Prof. RNDr. Jan Ámos Víšek, CSc.
Proposed topic	The least weighted squares, its asymptotics and diagnostics

Motivation Most of the papers on economic topics include some econometric analysis. These analyses might be misleading when there are some influential observations – i.e. outliers or leverage points. As it is not so unusual that the data contain at least some such observations, it may be useful to employ the robust methods in order to find the problems and obtain relevant estimators. Because many of the research papers use panel data for the analysis, there has been some research also on robustification on the panel data methods. This can be achieved for the least weighted squares (Víšek, 2012).

Many previous results were derived under homoscedasticity because we have to start from the simple situation. On the other hand, the assumption of homoscedasticity is - namely in processing the economic data - a bit unrealistic assumption. Hence we need to "re-derive" (utilizing more sophisticated tools) the results in the heteroscedastic framework.

Moreover, to be able to work with the robust methods appropriately, we also need some of the diagnostic tools adjusted for the use of robust methods. For this purpose, the asymptotic representation of the LWS estimator is needed (to be able to "repeat" – in a generalized way – proofs of classical diagnostic tools, as e. g. Durbin-Watson or White tests). This thesis will try to derive this asymptotic representation of LWS under heteroscedasticity. This result then can be used to derive the specification test for LWS - needed e.g. to decide whether we should use fixed or random effects method for the analysis (or analogically to choose between their robust versions).

An alternative goal of our research could be simulations of asymptotic representation in order to find how far its representation works, for which sample sizes, under which "degree" of heteroscedasticity, how far it is affected by contamination, etc.

Hypotheses

Hypothesis #1: The asymptotic representation can be derived under standard (technical) conditions typically assumed to be fulfilled in robust identification of model, probably a bit strengthen as it was done e. g. in Víšek (2002).

Hypothesis #2: Comparison by simulations of the asymptotic representation and of the computed values of estimator reveals the tightness of approximation.

Methodology The first hypothesis will be confirmed by derivation of the asymptotic representation. As compared to using the asymptotic linearity of normal equations, this thesis will attempt to use the convergence of empirical distribution function (Víšek, 2011a), which should simplify the process of deriving the asymptotic representation of LWS. This result then can be employed to obtain the specification test, similarly as for M-estimators in Víšek (1998). The second one will be just carried out by simulations under various situations. The simulation study will be based on the Monte Carlo method and will be carried out in MATLAB.

Expected Contribution The asymptotic representation of the estimator is a valuable result as such. Similarly, any knowledge about the tightness of approximation of estimator by its asymptotic representation can be key information for deriving other diagnostic tools. Moreover, if we succeed in derivation of test specificity, it would be a significant contribution for robust data analysis as already indicated above.

Outline

1. Introduction
2. Overview of robust methods
3. Asymptotic representation of LWS
4. Specification test for LWS
5. Simulation study
6. Conclusion

Core bibliography

Hampel, F. R., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.A. (1986): Robust Statistics – The Approach Based on Influence Functions. New York: J.Wiley and Son.

Rousseeuw, P. J., Leroy, A.M. (1987): Robust Regression and Outlier Detection. New York: J.Wiley and Sons.

Víšek, J.Á. (1998): Robust specification test. Proceedings of Prague Stochastics 98 (eds. Hušková, M., Lachout, P.), Union of Czechoslovak Mathematicians and Physicists, p. 581 - 586.

Víšek, J.Á. (2002): The least weighted squares II. Consistency and asymptotic normality. Bulletin of the Czech Econometric Society, Vol. 9, no. 16, 1 - 28, 2002.

Víšek, J.Á. (2010): Weak $\sqrt{n}$-consistency of the Least Weighted Squares under Heteroskedasticity. Acta Universitatis Carolinae, 2/51, p. 71 - 82.

Víšek, J.Á. (2011a): Empirical Distribution Function under Heteroskedasticity. Statistics (ISSN: 0233-1888), Volume 45, Issue 5, p. 497-508.

Víšek J.Á. (2011b): Consistency of the Least Weighted Squares under Heteroskedasticity. Kybernetika 147, p. 179-206.

Víšek, J.Á. (2012): Robust estimation of model with the fixed and random effects. COMPSTAT 2012 Proceedings, The international statistical institute/ International association for statistical computing, p. 855-865.

Author

Supervisor

Chapter 1

Introduction

Econometric methods are a commonly used tool in economics. Although the most frequently used methods are the classical ones, it might be desirable to use also some of the robust methods for an econometric analysis, as the occurence of at least some influential observations (such as outliers or leverage points) is rather common in any data set. These influential data points can cause the results of the classical analysis to be completely misleading, and using some robust method as a complementary tool might help us to cope with it. To be able to decide between individual estimation methods and use them properly, we need some statistical tests and diagnostic tools. Although for the classical methods, such as ordinary least squares (OLS), the theoretical background and diagnostic tools are quite well developed, development of the diagnostic tools for robust methods is still in progress.

As the normal equations of the robust methods are usually not linear, the derivation of the theoretical tests is not as straightforward as e.g. for OLS. To be able to derive them, we need to know the asymptotic representation of corresponding robust estimator. This thesis derives the asymptotic representation of an estimator obtained by the method of least weighted squares (LWS) under the assumption of heteroskedastic residuals. The asymptotic representation of LWS estimator was previously derived under the assumption of homoskedasticity, see Víšek (2002b) and Víšek (2015). However, the assumption of homoskedasticity is often not satisfied and therefore it seems desirable to generalize this result for the heteroskedastic case. This thesis generalizes the result obtained in Víšek (2015).

Notice that as compared to the original proposal this thesis does not contain the specification test. This is because the attempt to derive the asymptotic representation using a combination of the asymptotic linearity of normal equations and the convergence of empirical distribution function was not successful yet under some reasonable assumptions. We did succeed in deriving the asymptotic representation under heteroskedasticity, however, the method used here is somewhat more complicated and extensive than expected. Therefore deriving the specification test based on the result in this thesis is left for future research.

Except for the Introduction, the thesis includes four other chapters. As robust methods are not a very commonly studied topic, Chapter 2 provides an overwiev of these methods, including motivation, and introducing the infinitesimal approach and the methods based on this approach - least median of squares, least trimmed squares and of course the method of LWS. In Chapter 3 there is the main result of the thesis, i.e. after summarizing the previous results and establishing some necessary assumptions and tools, the asymptotic representation of LWS estimator is derived under the assumption of heteroskedastic residuals. For clarity, the derivation is divided into proving several lemmas separately and the main result is stated in Theorem 3.1 in the end of the chapter. Chapter 4 provides results of a simulation study, where the tightness of the LWS estimator and the derived asymptotic representation is examined. The results are provided for several levels of contamination by both outliers and leverage points, and under the assumption of homoskedastic, as well as heteroskedastic residuals. Chapter 5 concludes.

All the figures in Chapter 2 were obtained using R. The numerical study in Chapter 4 is based on Monte Carlo method and all the results were obtained using MATLAB.

Chapter 2

Overview of robust methods

When doing any regression analysis, there are many assumptions that need to be satisfied in order for the estimation methods to work in the required way. One of the assumptions that is usually assumed to hold is normality of disturbances. However, it is quite common that this assumption is broken. Although the assumption of asymptotic normality is often sufficient due to the central limit theorem (CLT), there may be situations, where the deviations from normality in form of some outlying values cause the resulting estimates to be highly inefficient or completely misleading (it can indicate even e.g. wrong sign of the estimate of coefficient of an explanatory variable, different from an intuitively assumed or usually obtained one in similar models). The problem with these influential values can be overcome by robust methods.

This chapter is divided into four sections. The first section contains more on motivation for robust methods and introduces the first attempts to make robust estimators. The second section is devoted to the Hampel´s infinitesimal approach, which forms the basis for the method of least weighted squares introduced in the third section. The fourth section then presents, how the LWS can be generalized to panel data.

The core literature used in this chapter is Hampel *et al.* (1986), Rousseeuw & Leroy (1987) and Víšek (2000b). Throughout the thesis, knowledge of the basic econometrics concepts and terms is assumed. To read more about the basics see e.g. Greene (2012) or Wooldridge (2009).

2.1 Motivation and beginnings

Already in the 1920s, Fisher (1920) started the research on comparing estimation methods based on efficiency. Specifically, he showed that under the assumption of exact normality the standard deviation has higher asymptotic efficiency than the mean deviation (there is 12% efficiency gain).[1]

Even Fisher (1922) already considered data contamination and according to his paper the classical estimation methods are reliable only within a system of Pearsonian curves. He also showed that the efficiency decreases rather quickly when the data are contaminated by comparing the asymptotic efficiency of the arithmetic mean and variance estimators for normal distribution with corresponding estimators for t-distributions with various degrees of freedom. It appears that already for t_3 there is 50% loss of asymptotic efficiency for the estimator of the mean and 100% loss for the estimator of variance.[2] Although the data are usually not contaminated in a way that corresponds to t-distribution, further research supports also the idea of rather high efficiency losses when applying classical methods to contaminated real data - see Jeffreys (1961) or Tukey (1960).

The data contamination that requires the use of robust methods can have the form of outliers or leverage points. Outlier is an observation, which has the value of the explanatory variable inside a bulk of data, but the value of the response variable lies far away from the values of other observations. On the contrary, leverage point is an observation, which has the value of the response variable inside a bulk of data, but the value of the explanatory variable lies far away from the values of other observations. Moreover, we can distinguish between so called good and bad leverage points, where although the value of an explanatory variable of a good leverage point lies outside the bulk of data, it does not influence the regression line.

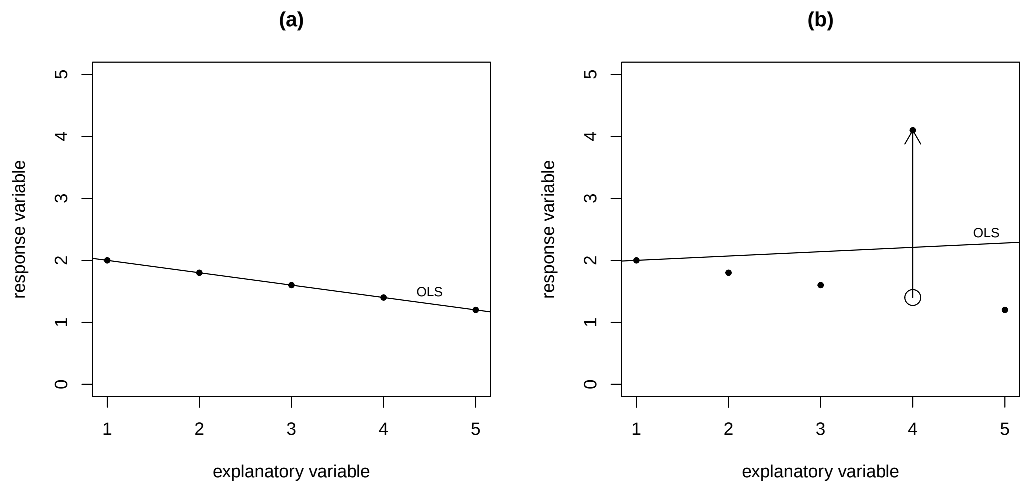
For a simple regression, it is not difficult to detect the influential observations (e.g. graphically), as can be seen in Figure 2.1 and Figure 2.2 (the shift to outlying values is indicated in parts (b) of both figures). The figures also show the effect of these influential observations on a classical estimation method, such as ordinary least squares. Although in a simple regression like this the

[1] Where standard (s_n) and mean (d_n) deviations can be computed by following formulas: $s_n = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$ and $d_n = \frac{\pi}{2n}\sum_{i=1}^{n}|x_i - \bar{x}|$.

[2] The asymptotic efficiency of corresponding estimators can be computed as $1 - \frac{6}{\nu(\nu+1)}$ for the mean and as $1 - \frac{12}{\nu(\nu+1)}$ for the variance (where ν is the number of degrees of freedom).

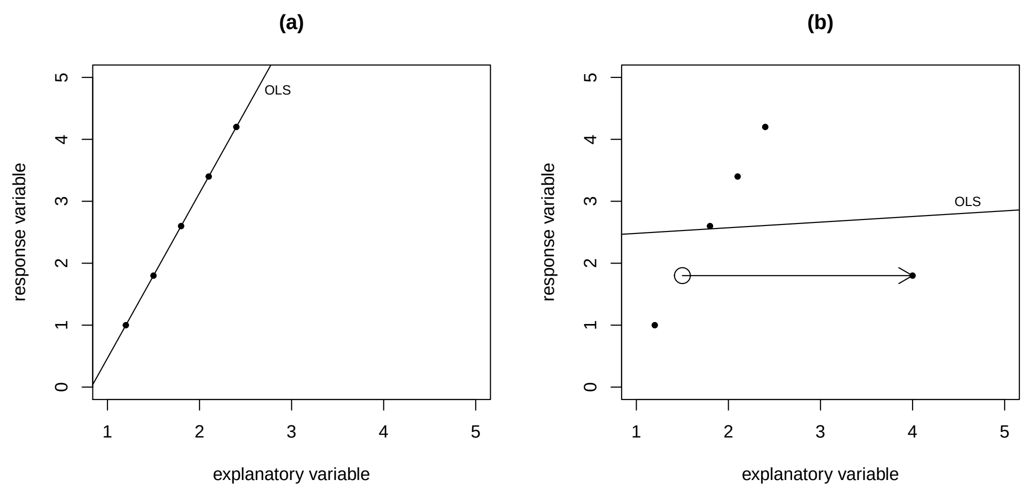
outlying values could be easily detected and treated, in a multiple regression the detection is much more complicated. That is where we need some theoretical background for the detection of influential values, such as the robust methods.

Figure 2.1: Outlier



Source: author's computations (based on Rousseeuw & Leroy (1987)).

Figure 2.2: Leverage point



Source: author's computations (based on Rousseeuw & Leroy (1987)).

For completeness and comparison to the robust estimators defined later, let us recall the definition of OLS estimator. We consider the standard regression model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u},$$

where $\mathbf{y} \in \mathbb{R}^n$ represents the response variable, $\mathbf{X} \in \mathbb{M}(n \times (k+1))$ represents the explanatory variables, $\mathbf{u} \in \mathbb{R}^n$ is the disturbance term and $\boldsymbol{\beta} \in \mathbb{R}^{k+1}$ is the vector of coefficients to be estimated.

Definition 2.1 (OLS estimator). The OLS estimator is defined as:

$$\hat{\boldsymbol{\beta}}^{OLS} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\ \mathbf{r}'\mathbf{r} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\ (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}),$$

where $\mathbf{r}$, $\mathbf{r}(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$, is the vector of residuals.

It is important to note that the robust methods should not necessarily be used instead of the classical methods, but rather as a complementary tool. In case that there are no outlying values, the classical methods are more efficient as compared to the robust ones. However, it is quite common that there are at least some outliers in a data set, and it is easier to detect them by using the robust methods than by the classical ones. Therefore to avoid the completely misleading results in case that some outliers are present, we need to use at least some robust method. Nevertheless, to achieve as high efficiency as possible, we need a "good" robust method. Since the 1960s, several researchers tried to find the best one.

The first person who considered contamination of data after Fisher and tried to develop some theoretical framework to deal with it was Tukey (1960), see also Huber (1981). He worked with parametric model - however, following Fisher´s idea he considered a certain neighbourhood of an ideal model (exact normal distribution), which contains also the outliers. More specifically, he allowed for an ε-proportion of "bad" observations, resulting in following underlying distribution:

$$F(x) = (1-\varepsilon)\Phi(x) + \varepsilon\Phi\left(\frac{x}{\sigma}\right),$$

where Φ stands for the standard normal distribution function and $\sigma > 0$ for the standard deviation. When allowing for deviations in this way, it follows from the asymptotic relative efficiency (ARE)[3] that already for very small proportion of outlying values ($\varepsilon = 0.0018$), the asymptotic efficiency of the mean deviation is higher relative to the standard deviation.

[3] $\text{ARE}(\varepsilon) = \lim_{n\to\infty} \frac{Var(s_n)}{Var(d_n)} \cdot \frac{E^2(d_n)}{E^2(s_n)}$.

Based on these results, there has been developed an extensive theory of robustness. This theory includes 3 main approaches - 2 approaches developed by Huber and Hampel´s infinitesimal approach based on influence functions. This theory can be applied also to non-parametric models or to parametric models with various underlying distributions - however, as was mentioned before, for the usual statistical inference the commonly used underlying distribution is the normal distribution.

The first approach proposed by Huber (1964) is the minmax approach, which attempts to minimize the maximal asymptotic variance (this may be useful when the number of observations is rather high and the average number of outliers is rather small). In this approach, the underlying distribution has following form:

$$F(x) = (1 - \varepsilon)\Phi(x) + \varepsilon H(x)$$

which is a generalized version of the model proposed by Tukey (ε is again the percentage of contamination and Φ is the standard normal distribution function). However, the unknown contaminating distribution H allows for more general form of outliers. Although the estimators from this model are not uniquely determined, this obstacle can be controlled for by imposing further restrictions - e.g. for uniquely determined parameter of location it is sufficient if H is symmetric about 0. In the same paper Huber proposed the M-estimator (maximum-likelihood-like estimator), which uses a suitable non-constant function ρ instead of the logarithm of density function (as it is for maximum-likelihood estimator).

Definition 2.2 (M-estimator). The M-estimator is defined as:

$$\hat{\boldsymbol{\beta}}^M = \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\ \rho(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

The M-estimators have quite pleasant properties, however, they lack the property of scale and regression equivariance. Therefore one needs to standardize the residuals in order to obtain an equivariant estimator. The robust estimators proposed later (based on the infinitessimal approach) are scale and regression equivariant even without standardization.

Although it is well mathematically developed, the second approach proposed by Huber (Huber & Strassen 1973) is not very commonly used in applications, since there is no general procedure how to use it. This approach is based on even more general neighbourhoods of normal distribution. To describe these neighbourhoods Huber used a special type of probabilities (called Choquet-capacities).

Before moving on to the infinitesimal approach, let us discuss some reasons, why outliers and leverage points appear in data. According to Hampel *et al.* (1986) there are four main reasons why the data deviate from strict parametric model - gross errors, rounding and grouping, approximate model, approximate fulfilment of the independence assumption.

The deviations in form of gross errors are usually considered the most dangerous for a regression analysis. They can be caused e.g. by measurement or typing errors and may have completely misleading values. This kind of deviation therefore brings more problems than deviations caused by random variability and can affect the resulting estimators the most, when untreated. Nevertheless, the presence of gross errors is quite a common phenomenon, which indicates the need of robust methods.

The deviations in form of rounding and grouping are considerably less problematic and it is often possible to neglect them without serious consequences. However, it can happen that also these deviations play an important role and therefore should not be completely ignored. They are important e.g. in case of very coarse classification (cannot be well approximated by a continuous distribution) or in case of superefficiency (infinite Fisher information). As we will see later, the problem with rounding and grouping can be detected by one of the properties of influence function (local-shift sensitivity).

When the model is only approximately normal, there still are some deviations, although they are not as obvious and dangerous as e.g. the gross errors. It was empirically found that these high-quality data are usually distributed according to a distribution that has longer tails than the normal distribution, see e.g. Romanowski (1970) for some examples. As the length of tails is correlated with the level of serial correlation (Jeffreys 1961), the distributional assumption might cause the statistical inference to be invalid. Similar consequences has also the approximate fulfilment of independence assumption (if the assumption is broken, we need to deal with serial correlation).

The aim of robust statistics then is to deal with these deviations - i.e. to find the outliers and leverage points, describe the structure of data and deal with unsuspected serial correlation.

2.2 Infinitesimal approach

The Hampel's (1968) approach links together 3 important robustness concepts: qualitative robustness, the influence function (utilizing the infinitesimal calculus) and the breakdown point. The qualitative robustness is a rather weak but necessary condition, which is not unique for this approach. Although the breakdown point is a general feature of robust methods as well, the influence function is the key notion of Hampel´s approach. The properties of robust estimators defined by means of influence function are very useful for treating contaminated data. Also other statisticians proposed several estimators based on Hampel´s approach.

The approach stands on the idea that many statistics are dependent only on the empirical distribution function (EDF). Prior to introducing the features of this approach itself, let us introduce the necessary notation. The EDF is defined as:

$$F_{n,\omega}(z) = \frac{1}{n} \sum_{i=1}^{n} I_{\{Z_i(\omega) \le z\}}(\omega),$$

where $\{Z_i(\omega)\}_{i=1}^{n}$, $\omega \in \Omega$, is a sequence of independent and identically distributed (iid) random variables defined on a probability space $(\Omega, \mathcal{A}, P)$ and $I_M(\omega)$ is an indicator of the set M defined as:

$$I_M(\omega) = \begin{cases} 1 & \text{if } \omega \in M \\ 0 & \text{if } \omega \notin M. \end{cases}$$

We can then consider many estimators as a functional of EDF (i.e. $\hat{\beta} = T(F_n)$, where $\hat{\beta}$ is the estimator of interest and $T(F_n)$ is the functional T of an EDF F_n), which inspires the idea to use derivatives in order to study properties of that estimator.

Now we can move to the definitions of influence function and properties of estimators derived from it that are crucial for the infinitesimal approach. These definitions can be found in Hampel *et al.* (1986).

Definition 2.3 (Influence function). The influence function of the functional T for the distribution F and at the point z is given by:

$$IF(z; T, F) = \lim_{t \to 0} \frac{T((1-t)F + t\Delta_z) - T(F)}{t}$$

in those z, where the limit exists.

The influence function measures the influence on corresponding estimate, when we add an observation at point z. Note that

$$T(F_n) = T(F) + \frac{1}{n}\sum_{i=1}^{n} IF(z_i; T, F).$$

Realizing this will be useful for the interpretation of the following properties of the influence function.

Alternatively, the influence function also has a heuristic interpretation. Specifically, it measures the asymptotic bias that arises due to the data contamination, i.e. it helps us to find the effect that an infinitesimal contamination at z has on the estimates.

To find the effect of contamination we use 3 following properties defined using the influence function. The most important one is the gross-error sensitivity, which is usually used as the first criterion when trying to robustify any estimator.

Definition 2.4 (Gross-error sensitivity). The gross-error sensitivity of the functional T for the distribution F is defined as:

$$\gamma^* = \sup_{z \in \mathbb{R}} |IF(z; T, F)|,$$

where the supremum is in fact taken over all z, where the influence function exists.

The gross-error sensitivity is defined as a supremum of an absolute value of the influence function. It follows that it describes the maximal effect on an estimator that can be caused by the infinitesimal contamination. The requirement for γ^* is to be finite. When we put an upper bound on γ^*, we assure that the asymptotic bias of the estimator is bounded.

Another requirement on the properties of robust estimators is as small local-shift sensitivity as possible.

Definition 2.5 (Local-shift sensitivity). The local-shift sensitivity is defined as:

$$\lambda^* = \sup_{z,y \in \mathbb{R}} \frac{|IF(z; T, F) - IF(y; T, F)|}{|z - y|}$$

for all $z \neq y$, where the influence functions exist.

The local-shift sensitivity is a measure of an effect of shifting one observation (from z to y). Therefore it measures the maximal effect of infinitesimal fluctuations within the data. These fluctuations (e.g. caused by rounding and grouping as mentioned before) will be present in any data that are not continuous - and by the means of λ^* it is possible to find, whether these fluctuations cause the estimates to be invalid. Such a discontinuity and consequently the fluctuations occur e.g. when data are measured by a digital device and we cannot (with probability equal to 1) measure the true value, so instead we measure the value nearest to it that is obtainable by the device.

Another desirable property is a finite rejection point.

Definition 2.6 (Rejection point). The rejection point is defined as:

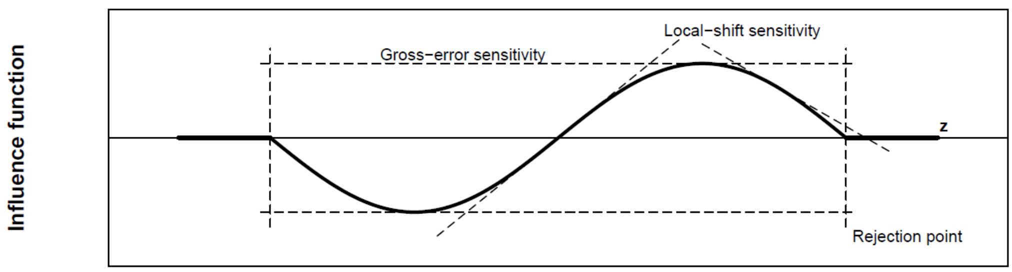
$$\rho^* = \inf_{r \in \mathbb{R}} \{r > 0; \quad IF(z; T, F) = 0 \quad when \quad |z| > r\}$$

for F symmetric about 0.[4]

The idea of rejection of outliers is one of the first robust methods. Even without the mathematical framework and only with subjective rejection, this can bring some efficiency gain when the data are contaminated. However, sometimes it is not possible to detect the outliers without the mathematical framework. This definition suggests that after reaching certain boundary (r), the influence function is constantly 0 and the points behind this boundary are therefore rejected. If r satisfying the stated condition does not exist, then $\rho^* = \infty$ by definition of infimum.

For illustration, the influence function and its properties can be depicted as in Figure 2.3.

Figure 2.3: Influence function



Source: author's computations (based on Hampel *et al.* (1986)).

[4]Notice that some weaker assumpotion about F might be sufficient here, such as $median(F) = 0$.

Additional to the influence function, Hampel´s approach uses another important robustness concept, which is called the breakdown point. It was first introduced by Hodges (1967) and later generalized by Hampel (1971). However, the general definition requires rather deep mathematical knowledge and a more straightforward finite sample version of this definition can be considered instead (Donoho & Huber 1983). Slightly adjusted version of this definition looks as follows (see Hampel *et al.* 1986).

Definition 2.7 (Breakdown point). The finite sample breakdown point ε_n^* of the estimator T_n at the sample $(x_1, ..., x_n)$ is given by:

$$\varepsilon_n^* = \frac{1}{n}\, max\{m;\ \max_{i_q} \sup_{y_q} |T_n(z_1, ..., z_n)| < \infty\}, \quad q = 1, ..., m,$$

where m is the number of outliers and $(z_1, ..., z_n)$ are the contaminated data obtained by replacing $(x_{i_1}, ..., x_{i_m})$ by arbitrary values $(y_1, ..., y_m)$.

This finite sample definition is equivalent to the general definition for parameters of location. For the parameters of scale we need to impose an additional assumption, so that we ensure that the estimator does not break down: $\min_{i_q} \inf_{y_q} |T_n(z_1, ..., z_n)| > 0$. It also may be desirable to consider the asymptotic breakdown point $\hat{\varepsilon}^* = \lim_{n\to\infty} \varepsilon_n^*$, since the resulting ε_n^* may depend on $(z_1, ..., z_n)$.

Intuitively, the breakdown point has a very straightforward meaning. It is the minimal percentage of contamination within the data that causes the estimator to break down. For example OLS cannot cope even with one outlier (if it is far enough from the bulk of data), therefore the breakdown point is $\frac{1}{n}$. With increasing sample size $n \to \infty$ we arrive at $\hat{\varepsilon}^* = 0\%$. The robust methods can cope with some level of contamination, the maximal possible breakdown point is 50%. However, the modern robust methods do not aim for the breakdown point as high as possible any more, since the modern algorithms allow to adjust the level of robustness to the level of contamination present in the particular data set.

In terms of breakdown point, the first method to achieve the desirable value, i.e. 50%, was the method of repeated median, where Siegel (1982) tried to use the fact that as compared to mean with 0%, median has breakdown point 50%. Although this estimator has the desirable properties, the algorithm is too complicated to be used in practice. However, it led also other researchers to employing the median in the robust estimation procedures and played an important role in development of other robust methods.

The later commonly used robust methods based on the infinitesimal approach were proposed by Rousseeuw (1984) and Hampel *et al.* (1986), further studied in Rousseeuw & Leroy (1987). These are the least median of squares (LMS) and the least trimmed squares (LTS).

The original LMS estimator was obtained by minimizing the median of squared residuals. The later definition uses the order statistics (although there is no median in the latter definition, the name remained the same). Let $r^2_{(i)}(\boldsymbol{\beta})$ be the order statistic of the i-th squared residual, i.e.

$$r^2_{(1)}(\boldsymbol{\beta}) \leq r^2_{(2)}(\boldsymbol{\beta}) \leq ... \leq r^2_{(n)}(\boldsymbol{\beta}).$$

Then the LMS estimator is defined as follows.

Definition 2.8 (LMS estimator). The LMS estimator is defined as:

$$\hat{\boldsymbol{\beta}}^{(LMS,n,h)} = \underset{\boldsymbol{\beta} \in \mathbb{R}^p}{\text{argmin}} \quad r^2_{(h)}(\boldsymbol{\beta}),$$

where $h \in \mathbb{R}$, $\frac{n}{2} \leq h \leq n$.

Asymptotically, this estimator has a 50% breakdown point, which can be computed as $\frac{\frac{n-p}{2}+1}{n}$, where p is the number of parameters to estimate (the maximal value of breakdown point is achieved for $h = \frac{n}{2} + \frac{p+1}{2}$). The LMS estimator also has other desirable properties, including scale and regression equivariance. However, there is one disadvantage in using this estimator in form of slow convergence. It is only $\sqrt[3]{n}$-consistent, i.e. we need more observations for the estimator to converge to the true value. As a consequence, the LMS estimator is less efficient than $\sqrt{n}$-consistent estimators.

To prevent this loss of efficiency, one can use a one-step M-estimator after obtaining the LMS estimator. Another way how to obtain a $\sqrt{n}$-consistent estimator is the LTS estimation procedure.

Definition 2.9 (LTS estimator). The LTS estimator is defined as:

$$\hat{\boldsymbol{\beta}}^{(LTS,n,h)} = \underset{\boldsymbol{\beta} \in \mathbb{R}^p}{\text{argmin}} \quad \sum_{i=1}^{h} r^2_{(i)}(\boldsymbol{\beta}),$$

where $h \in \mathbb{R}$, $\frac{n}{2} \leq h \leq n$.

The LTS estimator has all the desirable properties as LMS estimator, the breakdown point is also the same. But as compared to LMS, the LTS estimator converges quicker and therefore we can avoid the loss in efficiency (for proof of consistency and other properties of LMS and LTS see Rousseeuw & Leroy (1987)). It is also easy to note that an LTS estimator is in fact an OLS estimator, which takes into account only h observations with the smallest squared residuals. This means that the problematic observations furthest away from the regression line are not considered. As will be discussed in the next section, the complete omission of these observations might cause problems (in terms of local-shift sensitivity).

2.3 Least weighted squares

As was discussed in the last section, the robust methods based on the infinitesimal approach provide us with estimators that satisfy the expected requirements (specifically, LMS or LTS estimators). However, it can happen that the estimates resulting from these methods differ from each other, which is very unsettling considering that both estimators are consistent. This diversity of estimates can be illustrated on real data or on academic example, but it can also be formalized. See e.g. Víšek (2000a) to see that it is possible that LMS and LTS result in completely different estimators (the regression lines can even be orthogonal to each other).

Another issue, which one should be aware of, is the high local-shift sensitivity of LMS and LTS. This can be seen on an example in Figure 2.4, where shifting just one observation changes the sign of resulting estimate.

To deal with these issues, Víšek (2000b) proposed the least weighted squares estimation method.

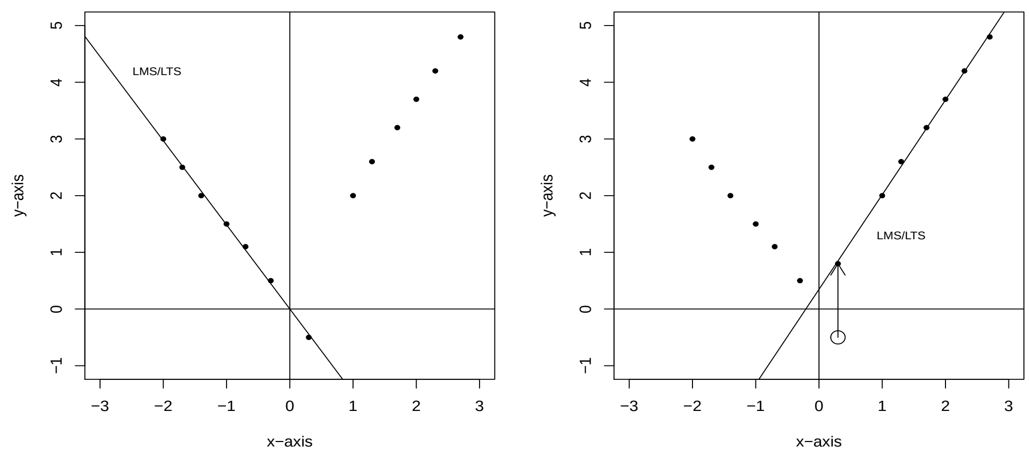
Definition 2.10 (LWS estimator). The LWS estimator is defined as:

$$\hat{\boldsymbol{\beta}}^{(LWS,n,w)} = \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\operatorname{argmin}} \quad \sum_{i=1}^{n} w\left(\frac{i-1}{n}\right) r_{(i)}^2(\boldsymbol{\beta}),$$

where $w(i)$ is the weight function.

We can see that the method of least weighted squares also uses the order statistics rather than the squared residuals directly. The main difference as compared to the other methods is the weight function. Due to this function LWS does not work only with the h observations as LTS but uses all of them

Figure 2.4: Local-shift sensitivity of robust estimators



Source: author's computations (based on Víšek (2000b)).

instead (which increases the efficiency if some of the not used data in LTS are not contaminated). Moreover, it allows us to assign also different weights than just 0/1 weights, which decreases the local shift sensitivity. It is also worth mentioning that as compared to the weighted least squares (WLS) estimator the weights in LWS are assigned implicitly by the method itself. On the contrary, to obtain the WLS estimator, we need to assign the weights explicitly (employing some external rule, frequently based on the "topology" of data). Before stating the definition of the weight function, let us for comparison and future reference recall the definition of a WLS estimator.

Definition 2.11 (WLS estimator). The WLS estimator is defined as:

$$\hat{\boldsymbol{\beta}}^{(WLS,n,w)} = \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\operatorname{argmin}} \quad \sum_{i=1}^{n} w_i r_i^2(\boldsymbol{\beta}),$$

where w_i are the weights assigned to the squared residuals.

Definition 2.12 (Weight function). The weight function $w(i)$, $w : [0;1] \rightarrow [0;1]$, is absolutely continuous and non-increasing, $w(0) = 1$ and its derivative is bounded below by $L \in \mathbb{R}$.

In fact, the LWS estimator is a generalization of the methods mentioned above, which indicates that it has their desirable properties. We can obtain those estimators (LMS, LTS and even OLS) by appropriately selecting the weights (for details see e.g. Víšek (2011a)). It follows that an LWS estimator is scale and regression equivariant and can attain 50% breakdown point (however, the method of LWS allows for adjusting the level of robustness to the level of contamination of corresponding data set, as is required by the modern robust methods).

An important feature of any estimator is consistency (as was mentioned above, ideal is $\sqrt{n}$-consistency). As was shown in Víšek (2011a), under a certain set of assumptions the LWS estimator is weakly consistent. Another advantage of LWS estimator, important for our purposes, is its applicability to panel data (specifically, by the means of LWS, we can robustify the estimators resulting from a regression model with the fixed and random effects).

Prior to introducing the generalization of LWS to panel data, let us remind the algorithm that gives us the LWS estimator, and will be used later in the simulation study (Víšek 2012a).

(i) Let us have n observations. Select randomly $p+1$ out of these n observations and run a regression in order to find a regression plane.

(ii) Find the squared residuals of all the n observations with respect to the regression plane obtained in (i).

(iii) Order the squared residuals from (ii) by size, which gives us the order statistics of the squared residuals. Sum the weighted obtained order statistics as $\sum_{i=1}^{n} w(i) r_{(i)}^2(\beta)$ and denote this sum $S(\hat{\beta}_{current})$.

(iv) Compare $S(\hat{\beta}_{current})$ to the sum that was obtained in the previous cycle, $S(\hat{\beta}_{former})$[5].

- If $S(\hat{\beta}_{current}) < S(\hat{\beta}_{former})$, continue with (v).
- If $S(\hat{\beta}_{current}) \geq S(\hat{\beta}_{former})$, continue with (vi).

(v) Denote the current value of the sum of weighted ordered squared residuals as the former value, i.e. put $S(\hat{\beta}_{former}) = S(\hat{\beta}_{current})$ and compute the new value of $S(\hat{\beta}_{current})$ by finding $\hat{\beta}^{(WLS,n,w)}$ for the dataset that we obtained in (iii) by reordering the original data, evaluating the squared residuals and repeating step (iii); then continue again with step (iv).

(vi) Consider $t \in \mathbb{N}$.

[5]Set the initial value of $S(\hat{\beta}_{former})$ that we compare $S(\hat{\beta}_{current})$ to in the first cycle to $S(\hat{\beta}_{former}) = \infty$.

- If the same estimates were obtained t-times, continue with (viii).
- If not, continue with (vii).

(vii) Consider $v \in \mathbb{N}$.

- If the cycle was iterated v-times, continue with (viii).
- If not, repeat the whole procedure by going back to (i).

(viii) Take the last obtained $\hat{\beta}_{current}$ as $\hat{\beta}^{(LWS,n,w)}$.

Defining $v \in \mathbb{N}$ in the last step is important for the process to end at some point with certainty, as it may happen that we cannot find the same estimates t-times, even after many iterations. As the threshold numbers t and v, we should choose some reasonable values to make the estimator reliable, e.g. $t = 20$ and $v = 10000$.

2.4 LWS for panel data

Although there have been some attempts to robustify some of the panel data methods, robust estimation methods for data with panel structure have not been widely studied. Because of the advantages of LWS mentioned before and because of its applicability to panel data, this section explains, how the methods of fixed and random effects can be robustified utilizing the LWS (Víšek 2012b). To learn about some of the other attempts using other methods (in this case LTS), see e.g. Bramati & Croux (2007).

Before starting to explain the robust version, let us recall the panel data model and classical fixed and random effects estimation methods. The corresponding regression model looks as follows:

$$y_{it} = \beta_0 + \sum_{j=1}^{k} \beta_j x_{itj} + a_i + u_{it},$$

where $i = 1, ..., n$ represents the individual dimension and $t = 1, ..., T$ represents the time dimension. The matrix notation is no longer used for clarity (larger number of dimensions). Except for the time dimension and the unobserved heterogeneity a_i, the model is analogical to the one for cross-sections. The unobserved heterogeneity is the reason, why we need to treat the data with panel structure differently, using some special methods: if a_i is correlated with the explanatory variables, the OLS estimator is inconsistent; if there is no correlation, the OLS estimator is consistent, but we still need to deal with inefficiency caused by autocorrelation of the composite errors v_{it},$v_{it} = a_i + u_{it}$. The most commonly used methods to deal with panel data are fixed effects (FE) for the case of correlation ($Cov(x_{itj}, a_i) \neq 0$) and random effects (RE) for the case of no correlation ($Cov(x_{itj}, a_i) = 0$).

The method of FE demeans the data and estimates the demeaned regression equation (by the means of OLS), where a_i no longer appears since it is time invariant. In mathematical terms, we estimate:

$$y_{it} - \bar{y}_i = \sum_{j=1}^{k} \beta_j (x_{itj} - \bar{x}_{ij}) + u_{it} - \bar{u}_i,$$

where $\bar{y}_i = \frac{\sum_{t=1}^{T} y_{it}}{T}$, $\bar{x}_{ij} = \frac{\sum_{t=1}^{T} x_{itj}}{T}$ and $\bar{u}_i = \frac{\sum_{t=1}^{T} u_{it}}{T}$.

The method of RE subtracts only a proportion of the mean that causes the inefficiency, otherwise the procedure is the same. Specifically, a λ-proportion of the mean is subtracted, where $\lambda = 1 - \frac{\sigma_u}{\sqrt{\sigma_u^2 + T\sigma_a^2}}$ and σ_u^2 and σ_a^2 are the variances of u_{it} and a_i, respectively (these variances usually need to be estimated and therefore in practice the exact value of λ is mostly not known and we use the estimate $\hat{\lambda}$). Then we estimate:

$$y_{it} - \hat{\lambda}\bar{y}_i = \beta_0(1 - \hat{\lambda}) + \sum_{j=1}^{k} \beta_j (x_{itj} - \hat{\lambda}\bar{x}_{ij}) + v_{it} - \hat{\lambda}\bar{v}_i.$$

If a certain set of assumptions is satisfied, both FE and RE estimators are consistent and under the assumption of normality of disturbances also efficient. For details see again Greene (2012) or Wooldridge (2009).

To be able to generalize LWS estimator to the panel data, we first need to redefine the order statistics of the squared residuals, since the total number of observations is nT and not n as it was in the cross-section case. Therefore we have nT squared residuals and we obtain:

$$r^2_{(1)}(\boldsymbol{\beta}) \leq r^2_{(2)}(\boldsymbol{\beta}) \leq ... \leq r^2_{(nT)}(\boldsymbol{\beta}).$$

It follows that we have to slightly adjust the definition of the LWS estimator for the different number of observations.

Definition 2.13 (LWS estimator for panel data). The LWS estimator for panel data is defined as:

$$\hat{\boldsymbol{\beta}}^{(LWS,n,T,w)} = \underset{\boldsymbol{\beta} \in \mathbb{R}^p}{\operatorname{argmin}} \quad \sum_{i=1}^{nT} w\left(\frac{i-1}{nT}\right) r_{(i)}^2(\boldsymbol{\beta}),$$

where $w(i)$ is the weight function, defined in the same way as before.

Note that this generalized estimator inherits most of the desirable properties from the cross-sectional LWS estimator, such as equivariance, high efficiency or adjustable breakdown point. It can also be shown that this estimator is again weakly consistent (just by reformulation of the proof for cross-sectional case, see Víšek (2012b)). We can also arrive at OLS, LMS or LTS estimators by setting appropriate weights, just as before.

It remains to explain, how the LWS method can be used to robustify FE and RE estimators. The estimation procedure is similar to the classical one, but we need to use some robust methods in the process. For the fixed effects, we do the same demeaning transformation as was mentioned above, but we use a robust estimator of the subtracted mean, which is estimated by the means of LWS. After we obtain this demeaned equation, we estimate it by the means of LWS in order to obtain the fixed weighted effects (FWE) estimator.

In case of RE the procedure is similar, but we need to use a robust estimator of λ in the transformation. This robust estimator can be obtained by using LWS for estimation of σ_u^2 and σ_a^2 and applying the formula for λ stated above. When estimating the transformed equation by the means of LWS, we arrive at the random weighted effects (RWE) estimator.

Since the method of LWS, as well as the methods of fixed and random effects, result in a consistent estimator, it follows that the resulting FWE and RWE estimators are also consistent. Moreover, as follows from previous simulation studies, there is a significant gain in efficiency when using FWE and RWE estimators over their classical versions. Although when the contamination is high, the LWS method, which ignores the panel structure of the data, gives more reliable results than FWE or RWE (used appropriately according to the correlation between unobserved heterogeneity and explanatory variables), for the level of contamination up to around 10%, the FWE and RWE estimators are more efficient.

It can be concluded that in case of (even an infinitesimal) contamination, these robust methods for panel data using LWS give better results than the classical ones. However, one still needs to be careful about the choice of the weight function. If we have an idea about the level of contamination, then we can assign the weights appropriately (0 weights for the percentage of data that is contaminated with certainty). The problem can arise when the contaminated data are assigned a weight 1 (but there also are efficiency losses when too many of the non-contaminated data are assigned 0).

Chapter 3

Asymptotic representation of LWS

As the normal equations of the method of least weighted squares are not linear as e.g. the normal equations of OLS, performing some diagnostic tools is not as straightforward. Therefore to be able to derive some tests theoretically (e.g. the Hausman test), we first need to derive the asymptotic representation of the corresponding estimator. The main goal of this thesis is to derive this asymptotic representation for LWS estimator under the assumption of heteroskedasticity of residuals. The derivation is mainly based on Víšek (2015) and generalized for heteroskedastic residuals.

Note that the asymptotic representation is derived for cross-sections. However, the representation for pooled LWS would be done analogically and we would obtain the same result (where the total number of observations would be nT instead of n). It is also worth noticing that based on the simulation studies, the pooled LWS is the safest choice when we do not have an exact idea about the level of contamination. Nevertheless, the result derived in this thesis could be used to derive the asymptotic representation also for FWE and RWE in future research.

3.1 Previous research and necessary tools

The asymptotic representation of LWS estimator was previously derived under the assumption of homoskedastic residuals, for both non-random and random carriers. First derivation can be found in Víšek (2002a) and Víšek (2002b), where the case of non-random carriers is considered and the asymptotic representation is derived using the asymptotic linearity of normal equations.

Further research about the convergence of empirical distribution function

(Víšek 2011b) showed another way, how the asymptotic representation can be derived. The second derivation that uses this convergence of EDF can be found in Víšek (2015). As compared to the result in Víšek (2002b), where the estimator must lie in a compact set $\kappa \subset \mathbb{R}^p$, this second derivation generalizes the result to random carriers (which is a commonly considered framework nowadays) and allows to carry out the minimization problem over the whole space, i.e. for $\boldsymbol{\beta} \in \mathbb{R}^p$.

Although in Víšek (2015) the result is also derived under the assumption of homoskedastic residuals, the convergence of EDF, as shown in Víšek (2011b), holds also under the assumption of heteroskedasticity. Moreover, based on that result, we can also show that the consistency and $\sqrt{n}$-consistency of LWS estimator also holds for a model with heteroskedastic residuals as was shown in Víšek (2011a) and Víšek (2010). These reults therefore allow to generalize the derivation in Víšek (2015) for the case of heteroskedastic residuals, which is shown in the next section.

Note that combining the asymptotic linearity of normal equations and the convergence of EDF might offer an alternative (and possibly more straightforward) way to derive the asymptotic representation of LWS under heteroskedasticity. However, so far we were not able to derive the result without some rather restrictive assumptions.

Before moving on to the derivation itself, let us state the necessary conditions and recall the main results about the ($\sqrt{n}$-)consistency, as we will need them later. To be able to show the ($\sqrt{n}$-)consistency and derive the asymptotic representation we need following assumptions.

Assumptions 3.1. Let us assume that $\{(V_i', u_i)'\}_{i=1}^{\infty}$ is a sequence of independent p-dimensional random variables with absolutely continuous distribution functions $F_{V,u_i}(v,r) = F_V(v)F_{u_i}(r)$, $v \in \mathbb{R}^{p-1}$, $r \in \mathbb{R}$, where $F_{u_i}(r) = F_u(r\sigma_i^{-1})$, $E(u_i) = 0$, $Var(u_i) = \sigma_i^2$ and $0 < a = \liminf_{i\to\infty} \sigma_i \leq \limsup_{i\to\infty} \sigma_i = b < \infty$. Further, denote the densities of $F_V(v)$ and $F_{u_i}(r)$ by $f_V(v)$ and $f_{u_i}(r)$ respectively, where $f_{u_i}(r) = f_u(r\sigma_i^{-1})\sigma_i^{-1}$, and the densities $f_V(v)$ and $f_u(r)$ are bounded by some constants $B_V < \infty$ and $B_u < \infty$. Moreover, $E(V_1) = 0$, $E(V_1 \cdot V_1')$ is positive definite, and there is $q > 1$ such that $E(||V_1||^{2q}) < \infty$ (note that $\{V_i\}_{i=1}^{\infty}$ is a sequence of iid random variables since the distribution function $F_V(v)$ does not depend on i). Finally, consider $\{(X_i', u_i)'\}_{i=1}^{\infty}$ where we put $X_{i1} = 1$ and $X_{ij} = V_{i,j-1}$, $j = 2, 3, ..., p$ for all $i \in N$. Then we can denote $F_{X,u_i}(x,r) = F_X(x)F_{u_i}(r)$ the distribution function of $(X_1', u_i)'$.

Note that from the construction of the variables X and V follows that we can write the regression model as:

$$Y_i = X_i'\beta^0 + u_i = \sum_{j=1}^{p} X_{ij}\beta_j^0 + u_i = \beta_1^0 + \sum_{j=2}^{p} V_{ij-1}\beta_j^0 + u_i,$$

where we denote the true value of the β-coefficients by β^0. Throughout the derivation in the next section, we will assume without loss of generality (WLOG) that $\beta^0 = 0$ to simplify the procedure. However, at certain points we will write $\beta - \beta^0$ to obtain the results in their usual form.

Further notice that the assumption $E(V_1) = 0$ is also WLOG. Otherwise we could consider the demeaned value $\tilde{V}_i = V_i - E(V_i)$ together with the adjusted value of β^0. Moreover, let $\{\tilde{u}_i\}_{i=1}^{\infty}$ be a sequence of iid random variables. Then the sequence $\{u_i\}_{i=1}^{\infty}$, where $u_i = \tilde{u}_i \cdot \sigma_i$, can satisfy the Assumptions 3.1.

The estimator $\hat{\beta}^{(LWS,n,w)}$ can be found as a solution of the normal equation that can be derived as in Víšek (2011a). We can consider the definition of LWS estimator in Definition 2.10, where (following Hájek & Šidák (1967)) we can put $\pi(\beta, i) = j,\ j \in (1, ..., n) \Leftrightarrow r_i^2(\beta) = r_{(j)}^2(\beta)$. Then we can write the definition in following form:

$$\hat{\boldsymbol{\beta}}^{(LWS,n,w)} = \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\operatorname{argmin}} \quad \sum_{i=1}^{n} w\left(\frac{\pi(\beta,i)-1}{n}\right) r_i^2(\boldsymbol{\beta}).$$

Moreover, let us denote the EDF of the absolute value of the residuals by $F_\beta^{(n)}(r)$, so that we arrive at[1]

$$F_\beta^{(n)}(r) = \frac{1}{n}\sum_{j=1}^{n} I_{\{|r_j(\beta)|<r\}} = \frac{1}{n}\sum_{j=1}^{n} I_{\left\{|u_j - X_j'\beta|<r\right\}} =$$

$$= \frac{1}{n}\sum_{j=1}^{n} I_{\left\{\omega\in\Omega:\ |u_j(\omega) - X_j'(\omega)\beta|<r\right\}}. \tag{3.1}$$

It is important to realize that the order statistic of the absolute value of residuals assigns to an observation the same rank as the order statistic of squared residuals. Then we can write

$$\frac{\pi(\beta,i)-1}{n} = F_\beta^{(n)}\left(|r_i(\beta)|\right). \tag{3.2}$$

[1]It may seem somewhat unusual to speak about the EDF when we consider variables that are not iid. However, it can make sense as we will see later in the next section.

Considering also the equivalence of the WLS and LWS estimator for appropriate permutation, we can write the normal equations as follows:

$$\sum_{i=1}^{n} w\left(F_{\beta}^{(n)}\left(|r_i(\beta)|\right)\right) X_i(Y_i - X_i'\beta) = 0. \tag{3.3}$$

Analogically to the EDF of the absolute value of the residuals, let us denote the theoretical distribution function of the absolute value of the residuals by

$$F_{i,\beta}(v) = P(|r_i(\beta)| < v). \tag{3.4}$$

Then the mean distribution function can be defined as:

$$\bar{F}_{n,\beta}(v) = \frac{1}{n}\sum_{i=1}^{n} F_{i,\beta}(v). \tag{3.5}$$

In addition to the Definition 2.12 we need some more assumptions on the weight function.

Assumptions 3.2. The weight function w as defined in Definition 2.12 is continuous on $[0,1)$ with $w(1) = 0$. Moreover, it is Lipschitz in absolute value, i.e. there exists L_w such that for any $a, b \in [0,1]$ we have

$$|w(a) - w(b)| \leq L_w|a - b|,$$

and its derivative $w'(\alpha)$ is bounded in absolute value by a finite constant. Finally, let $E\left\{w(\bar{F}_{n,\beta^0}(|u|))X_1X_1'\right\}$ be positive definite.

If the weight function was not continuous, it still would be possible to prove the asymptotic properties, but we would need some more complicated techniques. The weight function is not continuous e.g. for the method of LTS, see Víšek (2006) for the proofs and derivations for this method with non-continuous weight function.

For the consistency of the LWS estimator we need one additional condition. Note that e.g. for OLS (where $w(i) = 1$ for all i) Equation 3.6 would be satisfied, as the normal equations have a unique solution.

Assumptions 3.3. There is the only solution of

$$\beta' E \sum_{i=1}^{n} \left[w\left(\bar{F}_{n,\beta}(|r_i(\beta)|)\right) X_i \left(u_i - X_i'\beta\right)\right] = 0. \tag{3.6}$$

Moreover,

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \sigma_i = 1. \tag{3.7}$$

Note that in the usual form we would again need to replace β by $(\beta - \beta^0)$ in Equation 3.6 and this form of Assumptions 3.3 is possible only due to the assumption that $\beta^0 = 0$ (WLOG). It is further not clear, if Equation 3.7 can be easily satisfied. However, when we let $\{\kappa_i\}_{i=1}^{\infty}$ be a sequence of iid random variables with mean value equal to 1, we obtain $\frac{1}{n} \sum_{i=1}^{n} \kappa_i \to 1$ (in probability).

Lemma 3.1. *Under Assumptions 3.1, 3.2 and 3.3 any sequence* $\left\{ \hat{\beta}^{(LWS,n,w)} \right\}_{n=1}^{\infty}$ *of the solutions of the normal equations given in Equation 3.3 is weakly consistent.*

For the proof of Lemma 3.1 see Víšek (2011a). To obtain the $\sqrt{n}$-consistency, we have to enlarge the conditions as follows.

Assumptions 3.4. The density $f_u(r)$ is uniformly with respect to x Lipschitz of the first order, i.e. there exists L_u such that for any $a, b \in [0, 1]$ we have

$$|f_u(a) - f_u(b)| \leq L_u |a - b|.$$

In addition, the derivative $f_u'(r)$ exists and is bounded in absolute value by $U' < \infty$.

Note that if Assumptions 3.4 hold for every r, they hold for every i when we put $r = \tilde{r} \sigma_i^{-1}$. We further need an additional assumption about the derivative of the weight function.

Assumptions 3.5. Let the derivative of the weight function $w'(\alpha)$ be Lipschitz of the first order, i.e. there exists J_w such that for any $a, b \in [0, 1]$ we have

$$|w'(a) - w'(b)| \leq J_w |a - b|.$$

Lemma 3.2. *Under the Assumptions 3.1, 3.2, 3.3, 3.4 and 3.5 any sequence* $\left\{ \hat{\beta}^{(LWS,n,w)} \right\}_{n=1}^{\infty}$ *of the solutions of the normal equations given in Equation 3.3 is weakly* $\sqrt{n}$*-consistent.*

For the proof of Lemma 3.2 see Víšek (2010). In order to derive the asymptotic representation we will need some additional assumptions (mainly specifying the character of heteroskedasticity). To be able to specify these assumptions we will also need distribution function of squared residuals. To be able to define it, let us recall Equation 3.4 and Equation 3.5. Analogically, we can write

$$G_i(z) = P(u_i^2 < z) = F_u\left(\sigma_i^{-\frac{1}{2}} z^{\frac{1}{2}}\right) - F_u\left(-\sigma_i^{-\frac{1}{2}} z^{\frac{1}{2}}\right)$$

and

$$G(z) = \frac{1}{n}\sum_{i=1}^{n} P(u_i^2 < z) = \frac{1}{n}\sum_{i=1}^{n}\left(F_u\left(\sigma_i^{-\frac{1}{2}} z^{\frac{1}{2}}\right) - F_u\left(-\sigma_i^{-\frac{1}{2}} z^{\frac{1}{2}}\right)\right). \tag{3.8}$$

Then the corresponding densities are $g_i(z)$ and $g(z)$, respectively, where

$$g_i(z) = \frac{\sigma_i^{-\frac{1}{2}}}{2z^{\frac{1}{2}}}\left(f_u\left(\sigma_i^{-\frac{1}{2}} z^{\frac{1}{2}}\right) + f_u\left(-\sigma_i^{-\frac{1}{2}} z^{\frac{1}{2}}\right)\right)$$

and

$$g(z) = \frac{1}{2nz^{\frac{1}{2}}}\sum_{i=1}^{n}\sigma_i^{-\frac{1}{2}}\left(f_u\left(\sigma_i^{-\frac{1}{2}} z^{\frac{1}{2}}\right) + f_u\left(-\sigma_i^{-\frac{1}{2}} z^{\frac{1}{2}}\right)\right).$$

Notice that $G(z) = \frac{1}{n}\sum_{i=1}^{n} G_i(z)$, $g(z) = \frac{1}{n}\sum_{i=1}^{n} g_i(z)$ and for $z < 0$ we have $G_i(z) = 0$, $G(z) = 0$, $g_i(z) = 0$ and $g(z) = 0$. Then let us assume the following.

Assumptions 3.6. Let $d \in \mathbb{R}^+$. Then there exists $\Delta(d) > 0$ such that

$$\inf_{z \in (0, d+\Delta(d))} G(z) > L_{g,d} > 0,$$

$$\sum_{i=1}^{n}\left(\sigma_i^{-1} - 1\right) = O(n^{\frac{1}{2}}) \tag{3.9}$$

and

$$\sup_{-\infty < z < \infty}\left|\left[\sup_{v \in (zb^{-1}, za^{-1})} f_u(v)\right] \cdot z\right| < L_\varepsilon < \infty. \tag{3.10}$$

Moreover, for any $\delta > 0$ there exists $n_\delta \in \mathbb{N}$ and $B_g > 0$ such that for all $n > n_\delta$ and any $0 < r \le s < \infty$ such that $G(s) - G(r) < \delta$, we have

$$G(a^{-1}s) - G(a^{-1}r) < B_g\left(G(s) - G(r)\right).$$

At this point, let us make several remarks justifying some of the expressions in Assumptions 3.6, as it is not very straightforward to see, if they are likely to be fulfilled. Let us first consider Equation 3.9 and put $\gamma_i = \sigma_i^{-1}$. Then we obtain $b^{-1} < \gamma_i < a^{-1}$ (see Assumptions 3.1). We can assume for simplicity that $(b^{-1}, a^{-1}) = (1-m, 1+m)$ for some $m \in (0,1)$, and that $\{\gamma_i\}_{i=1}^{\infty}$ is a sequence of iid random variables distributed uniformly on $(1-m, 1+m)$. Then we arrive at $E(\gamma_i) = 1$ and $Var(\gamma_i) = m^2$. Using CLT we obtain

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\gamma_i - 1) = O_p(1).$$

Equation 3.10 also deserves a comment. The form of this assumption might seem somewhat unusual. More usual form would be

$$\sup_{-\infty < v < \infty} |f_u(v) \cdot v| < L_\varepsilon < \infty.$$

It would be possible to derive Equation 3.10 from this more usual form under some additional (not very restrictive) assumptions on $f_u(v)$ (e.g. we would need $f_u(v)$ to be Lipschitz etc.). It seems preferable to assume directly Equation 3.10. In the next section we will need one more condition.

Assumptions 3.7. There exists $q' > 1$ such that $\sup_i E\left(|u_i|^{2q'}\right) < \infty$.

Let us further state some other definitions and previously proven lemmas that we will need in the next section for the derivation of the asymptotic representation under heteroskedasticity.

Definition 3.1 (Separability). Let $V = (V(s),\ s \in S) \subset \mathbb{R}^p$, where $S \subset \mathbb{R}^q$ and $p, q \in \mathbb{N}$, be a stochastic process. The process is called separable if there exists a countable dense set T, such that $T \subset S$, and for any pair $(\omega, t) \in \Omega \times S$ there exists a sequence $\{s_n\}_{n=1}^{\infty} \subset T$ for which

$$\lim_{n \to \infty} s_n = t \qquad \textit{and} \qquad \lim_{n \to \infty} V(\omega, s_n) = V(\omega, t).$$

The separability of stochastic processes is needed for the following lemma. Lemma 3.3 (Štěpán 1987) tells us that when two stochastic processes have the same distribution, they also have the same supremum over corresponding index set.

Lemma 3.3. *Let $V = (V(s),\ s \in S) \subset \mathbb{R}^\ell$, where $\ell \in \mathbb{N}$, be a separable stochastic process defined on probability space $(\Omega, \mathcal{A}, P)$. Let further $G \subset S$ be an open* *set and denote the set of all finite subsets of G by $k(G)$. Then we have for any closed set $K \subset \mathbb{R}^p$, where $p \in \mathbb{N}$*

$$\{\omega \in \Omega : V(s) \in K, s \in G\} \in \mathcal{A}$$

and

$$P(\{\omega \in \Omega : V(s) \in K, s \in G\}) = \inf_{J \in k(G)} P(\{\omega \in \Omega : V(s) \in K, s \in J\}).$$

Proof of Lemma 3.3 can be found in Štěpán (1987), see also Víšek (2015). Since we use the embedding into Wiener process (see Lemma 3.4 below) throughout the derivation in the next section, let us now define the Wiener process and state some of its useful properties.

Definition 3.2 (Wiener process). Let $W(s)$, where $s \in \mathbb{R}^+$, be a stochastic process. We say that $W(s)$ is a Wiener process (called also Brownian motion), if it satisfies following properties

(i) $W(0) = 0$.

(ii) The increments of $W(s)$ are stationary and independent.

(iii) The increments of $W(s)$ (i.e. $W(s+t) - W(t)$) are normally distributed with zero mean and variance s.

(iv) The function $s \mapsto W(s)$ is continuous on $\mathbb{R}^+$ with probability 1.

Further recall that if a random variable X has variance $Var(X) = \sigma^2$, then $Var(a \cdot X) = a^2 \cdot \sigma^2$. Moreover, if random variables X, Y are independent, then $Var(X+Y) = Var(X) + Var(Y)$. It follows that when we have $Var(W(s)) = s$, then

$$Var\left(n^{-\frac{1}{4}} W(s)\right) = n^{-\frac{1}{2}} \cdot s = Var\left(W\left(n^{-\frac{1}{2}} \cdot s\right)\right)$$

and consequently

$$n^{-\frac{1}{4}} \sum_{i=1}^{n} W(\tau_i) =_D n^{-\frac{1}{4}} W\left(\sum_{i=1}^{n} \tau_i\right) =_D W\left(n^{-\frac{1}{2}} \sum_{i=1}^{n} \tau_i\right),$$

where "$=_D$" denotes equivalence in distribution.

The embedding into Wiener process was proposed by Skorohod and "rediscovered" by Portnoy (1983). Let us now state some lemmas regarding the

Wiener processes that will be used later (note that all the processes used in the proofs in the next section are always defined only within the given proof). Lemma 3.4 and Lemma 3.5 are taken from Štěpán (1987), see also Víšek (2015).

Lemma 3.4. *Let $W(s)$ be a Wiener process, a and b some positive numbers, and ξ a random variable satisfying $P(\xi = -a) = \pi$ and $P(\xi = b) = 1 - \pi$, where $\pi \in (0, 1)$, and $E(\xi) = 0$. Further let τ be the time when $W(s)$ exits the interval $(-a, b)$. Then*

$$\xi =_D W(\tau).$$

In addition, $E(\tau) = a \cdot b = Var(\xi)$.

Lemma 3.5. *Let again $W(s)$ be a Wiener process and a and b some positive numbers. Then*

$$P\left(\max_{0 \le t \le b} |W(t)| > a\right) \le 2 \cdot P\left(|W(b)| > a\right).$$

The proofs of Lemma 3.4 and Lemma 3.5 can be found in Štěpán (1987). We will further need following lemma from Rao (1973). Notice that this lemma is formulated for iid random variables. However, as will be shown in the next section in Lemma 3.20, we can use it also for our purpose. Denote $[h]_{int}$ as the floor of h for any $h \in \mathbb{R}$.

Lemma 3.6. *Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of iid random variables distributed according to distribution function $F(x)$ with a continuous density $f(x)$. Moreover, for $\alpha \in (0, 1)$ let the upper α-quantile of $F(x)$, q_α, be given uniquely and let $f(q_\alpha) > 0$. Finally, put $\hat{q}_\alpha = X_{(\ell_n(\alpha))}$, where $\ell_n(\alpha) = [(1-\alpha)n]_{int}$. Then*

$$\sqrt{n}\left(\hat{q}_\alpha - q_\alpha\right) \underset{D}{\longrightarrow} \mathcal{N}\left(0, \frac{\alpha(1-\alpha)}{f^2(q_\alpha)}\right).$$

For the proof of Lemma 3.6 see Rao (1973). A sketch of the proof can be found also in Víšek (2015), where some of the expressions from the proof were used, specifically

$$F_{\hat{q}_\alpha}(x) = P\left(\hat{q}_\alpha < x\right) = P\left(\{\omega \in \Omega : \#\left\{X_i(\omega) < x, i = 1, 2, ..., n\right\} \ge \ell_n(\alpha)\}\right) =$$

$$= \sum_{\ell=\ell_n(\alpha)}^{n} \frac{n!}{\ell!(n-\ell)!} F^{\ell}(x) \left[1 - F(x)\right]^{n-\ell} \tag{3.11}$$

and the density of $\hat{q}_\alpha$ given by

$$\frac{n!}{(\ell_n(\alpha)-1)!(n-\ell_n(\alpha))!} F^{\ell_n(\alpha)-1}(x) \left[1-F(x)\right]^{n-\ell_n(\alpha)} f(x). \tag{3.12}$$

We will need to adjust this probability and corresponding density for the heteroskedastic case. However, these expressions are stated here as well to see the analogy between homoskedastic and heteroskedastic case.

Lemma 3.7. *Let $\alpha_0 \in (0,1)$. Then under the assumptions of Lemma 3.6 the density of $\hat{q}_\alpha = X_{(\ell_n(\alpha))}$ is for any $\alpha \in (\alpha_0, 1)$ given by*

$$h_{n,\alpha}(q) = h^*_{n,\alpha}(q) + \rho_{n,\alpha}(q),$$

where $h^*_{n,\alpha}(r)$ is a density symmetric around q_α and for any $K < \infty$ we have

$$\sup_{\alpha \in (\alpha_0,1)} \sup_{|q| \le n^{-\frac{1}{2}}K} |\rho_{n,\alpha}(q)| = O\left(n^{-\frac{1}{2}}\right).$$

The proof of Lemma 3.7 follows from Lemma 3.6 and can be again found in Rao (1973). For details see also the appendix of Víšek (2015). Before starting the derivation itself, let us recall one more lemma.

Lemma 3.8. *Let the Assumptions 3.1 hold. Then we have*

$$\sup_{v \in \mathbb{R}^+} \sup_{\beta \in \mathbb{R}^p} \sqrt{n} \left| F^{(n)}_\beta(v) - \bar{F}_{n,\beta}(v) \right| = O_p(1).$$

Lemma 3.8 along with corresponding proof can be found in Víšek (2011b). Now we have prepared all the necessary tools and we can move on to the next section.

3.2 Derivation under heteroskedasticity

As was mentioned above, the derivation of the asymptotic representation under heteroskedasticity in this section is mainly based on Víšek (2015). As compared to the previous paper, we need to prove some additional lemmas and make several adjustments to generalize the derivation for heteroskedastic case. Moreover, some of the proofs are more thorough for easier understanding.

For clarity of the text we will divide the derivation into proving several lemmas separately. Let us define $\tau_M = \{t \in \mathbb{R}^p, ||t|| < M\}$, where $M \in \mathbb{R}^+$. Then we can obtain the following.

Lemma 3.9. *Let the Assumptions 3.1 and 3.4 hold. Choose arbitrarily $\varepsilon > 0$ and $\tau \in (\frac{1}{2}, \frac{3}{4})$. Then there exists $K \in (0, \infty)$ and $n_{\varepsilon,M,\tau} \in \mathbb{N}$ such that for all $n > n_{\varepsilon,M,\tau}$*

$$P\left(\left\{\omega \in \Omega : \sup_{r\in\mathbb{R}} \sup_{t\in\tau_M} n^{\tau} \left|F^{(n)}_{\beta^0-n^{-\frac{1}{2}}t}(r) - F^{(n)}_{\beta^0}(r)\right| < K\right\}\right) > 1 - \varepsilon. \quad (3.13)$$

PROOF The proof is similar to the one in Víšek (2015), as the main idea of the proof is embedding into the Wiener process, which is not influenced by heteroskedasticity. However, some adjustments for heteroskedasticity are necessary. Moreover, some additional steps are included for more straightforward understanding.

As follows from Equation 3.1, we have

$$F^{(n)}_{\beta^0}(r) = \frac{1}{n}\sum_{i=1}^{n} I_{\{|u_i|<r\}} = \frac{1}{n}\sum_{i=1}^{n} I_{\{\omega\in\Omega\ :\ |u_i(\omega)|<r\}}$$

and

$$F^{(n)}_{\beta^0-n^{-\frac{1}{2}}t}(r) = \frac{1}{n}\sum_{i=1}^{n} I_{\left\{|u_i+n^{-\frac{1}{2}}X_i't|<r\right\}} = \frac{1}{n}\sum_{i=1}^{n} I_{\left\{\omega\in\Omega\ :\ |u_i(\omega)+n^{-\frac{1}{2}}X_i'(\omega)t|<r\right\}}.$$

Notice that both of these empirical distribution functions are zero for any $r \le 0$. Therefore we can consider only $r > 0$.

Let further $\#A$ denote the number of elements of set A and let us define for all $n \in \mathbb{N}$, $r \in \mathbb{R}^+$ and $t \in \tau_M$

$$m^{(+)}_{n,U}(r,t) = \#\left\{i \in \{1,2,...,n\}\ :\ u_i \ge r \ \ and \ \ \left|u_i + n^{-\frac{1}{2}}X_i't\right| < r\right\} \quad (3.14)$$

$$m^{(-)}_{n,U}(r,t) = \#\left\{i \in \{1,2,...,n\}\ :\ |u_i| < r \ \ and \ \ u_i + n^{-\frac{1}{2}}X_i't \ge r\right\} \quad (3.15)$$

$$m^{(+)}_{n,L}(r,t) = \#\left\{i \in \{1,2,...,n\}\ :\ u_i \le -r \ \ and \ \ \left|u_i + n^{-\frac{1}{2}}X_i't\right| < r\right\} \quad (3.16)$$

$$m^{(-)}_{n,L}(r,t) = \#\left\{i \in \{1,2,...,n\}\ :\ |u_i| < r \ \ and \ \ u_i + n^{-\frac{1}{2}}X_i't \le -r\right\}. \quad (3.17)$$

Finally, let us define $m_n(r,t)$ as:

$$m_n(r,t) = m^{(+)}_{n,U}(r,t) - m^{(-)}_{n,U}(r,t) + m^{(+)}_{n,L}(r,t) - m^{(-)}_{n,L}(r,t). \tag{3.18}$$

It follows that the indices included in $m^{(+)}_{n,U}(r,t)$ are those for which we have simultaneously $\left|u_i + n^{-\frac{1}{2}}X_i't\right| < r$ and $u_i \geq r$. I.e. the indices that belong in $m^{(+)}_{n,U}(r,t)$ are considered for computing $F^{(n)}_{\beta^0 - n^{-\frac{1}{2}}t}(r)$, but not for computing $F^{(n)}_{\beta^0}(r)$. We can make analogical conclusions for $m^{(-)}_{n,U}(r,t)$, $m^{(+)}_{n,L}(r,t)$ and $m^{(-)}_{n,L}(r,t)$. Then we can conclude that

$$\left|F^{(n)}_{\beta^0 - n^{-\frac{1}{2}}t}(r) - F^{(n)}_{\beta^0}(r)\right| \leq \frac{1}{n}\left|m_n(r,t)\right|.$$

Therefore in order to prove Equation 3.13, it suffices to prove that for all $\varepsilon > 0$ and $\tau \in (\frac{1}{2}, \frac{3}{4})$ there exists $K \in (0,\infty)$ and $n_{\varepsilon,M,\tau} \in \mathbb{N}$ such that for all $n > n_{\varepsilon,M,\tau}$

$$P\left(\left\{\omega \in \Omega : n^{\tau-1} \sup_{r\in\mathbb{R}} \sup_{t\in\tau_M} \left|m_n(r,t)\right| < K\right\}\right) > 1 - \varepsilon.$$

In order to prove it, let us first consider Equation 3.14. The conditions $\left|u_i + n^{-\frac{1}{2}}X_i't\right| < r$ and $u_i \geq r$ imply that

$$r \leq u_i < r - n^{-\frac{1}{2}}X_i't.$$

Note that the only possibility, how both of these conditions can be satisfied simultaneously, is when $r < r - n^{-\frac{1}{2}}X_i't$, i.e. when $n^{-\frac{1}{2}}X_i't < 0$. It follows that when we define

$$b^{(+)}_i(r,t) = I_{\left\{r \leq u_i < r - n^{-\frac{1}{2}}X_i't\right\}}, \tag{3.19}$$

we obtain

$$m^{(+)}_{n,U}(r,t) \leq \sum_{i=1}^{n} b^{(+)}_i(r,t). \tag{3.20}$$

We can further denote

$$\pi_i(r,t) = E\left(b^{(+)}_i(r,t)\right)$$

and define following process

$$\xi^{(+)}_i(r,t) = b^{(+)}_i(r,t) - \pi_i(r,t). \tag{3.21}$$

Although $\left\{\xi_i^{(+)}(r,t)\right\}_{i=1}^{\infty}$ is not a sequence of iid processes, the stochastic processes are independent. Moreover, it can be shown that these processes are separable, which we will need later. Equation 3.21 hints that separability of the process $\left\{\xi_i^{(+)}(r,t)\right\}_{r\in\mathbb{R},t\in\mathbb{R}^p}$ follows from separability of the process $\left\{b_i^{(+)}(r,t)\right\}_{r\in\mathbb{R},t\in\mathbb{R}^p}$.

Let us first denote the set of rational numbers by $\mathbb{Q}$ and its p-th cartesian product by $\mathbb{Q}^p$. Looking at Definition 3.1 we can put $S = \mathbb{Q} \times (\tau_M \cap \mathbb{Q}^p)$ and $T = \mathbb{R} \times \tau_M$. Then fix some $\omega \in \Omega$ and select a sequence $\{r_k, t_k\}_{k=1}^{\infty} \subset \mathbb{R} \times \tau_M$ such that for all $k \in \mathbb{N}$ we obtain

$$[r_k, r_k - n^{-\frac{1}{2}} X_i' t_k) \subset [r, r - n^{-\frac{1}{2}} X_i' t).$$

We can see that the sequence $\{t_k\}_{k=1}^{\infty}$ is dependent on the sequence $\{r_k\}_{k=1}^{\infty}$ and also on the sign of $n^{-\frac{1}{2}} X_i' t$ (as was mentioned before, in the considered case we have $n^{-\frac{1}{2}} X_i' t < 0$ and as $n \in \mathbb{N}$, also $X_i' t = c < 0$). Let us for all $k \in \mathbb{N}$ put $r_k \in \mathbb{Q}$ such that $r_k \in (r, r - \frac{1}{k} \cdot c)$ and $t_k \in \mathbb{Q}^p$ such that for some appropriate $c_k \in \mathbb{R}$ and $t_k^* = t \cdot c_k$ we obtain $\|t_k - t_k^*\| < \frac{|c|}{k^2}$. Under these conditions we have $\lim_{k\to\infty} r_k = r$, $\lim_{k\to\infty} t_k = t$ and because $b_i^{(+)}(r,t)$ is continuous in r and t, also $\lim_{k\to\infty} b_i^{(+)}(r_k, t_k) = b_i^{(+)}(r,t)$. Therefore we can conclude that the process $\left\{b_i^{(+)}(r,t)\right\}_{r\in R,t\in R^p}$ (and consequently the process $\left\{\xi_i^{(+)}(r,t)\right\}_{r\in R,t\in R^p}$) is separable.

Due to the Assumptions 3.1 we have

$$\pi_i(r,t) = \int I_{\left\{r \le v < r - n^{-\frac{1}{2}} x' t\right\}} \mathrm{d}F_{X,u_i}(x,v) = \tag{3.22}$$

$$= \int \left[\int I_{\left\{r \le v < r - n^{-\frac{1}{2}} x' t\right\}} f_{u_i}(v) \mathrm{d}v\right] \mathrm{d}F_X(x) =$$

$$= \int \left[\int_r^{r - n^{-\frac{1}{2}} x' t} f_{u_i}(v) \mathrm{d}v\right] \mathrm{d}F_X(x) \le \int \left[\int_r^{r - n^{-\frac{1}{2}} x' t} B_u \sigma_i^{-1} \mathrm{d}v\right] \mathrm{d}F_X(x) =$$

$$= \int \left[B_u \sigma_i^{-1} v\right]_r^{r - n^{-\frac{1}{2}} x' t} \mathrm{d}F_X(x) = \int B_u \sigma_i^{-1} (-n^{-\frac{1}{2}} x' t) \mathrm{d}F_X(x) \le$$

$$\le n^{-\frac{1}{2}} \sigma_i^{-1} B_u \|t\| \int \|x\| \, \mathrm{d}F_X(x) \le \int \|x\| \, \mathrm{d}F_X(x) =$$

$$= n^{-\frac{1}{2}} a^{-1} B_u M \cdot E(||X_1||) = n^{-\frac{1}{2}} \Delta,$$

where $\Delta = a^{-1} B_u M \cdot E(||X_1||) < \infty$. We can further find $n_0 \in \mathbb{N}$ such that for all $n > n_0$ we get $n^{-\frac{1}{2}} \Delta \in (0, 1)$ and hence on $\mathbb{R} \times \tau_M$ we obtain for all $n > n_0$ $\pi_i(r, t) < n^{-\frac{1}{2}} \Delta$.

We will further consider only $n > n_0$ and use the embedding into a Wiener process. Following Portnoy (1983), Jurečková & Sen (1989) or e.g. Víšek (2011b), we can make use of Lemma 3.4.

Let us have for each i a probability space $(\Omega_i, \mathcal{A}_i, P_i)$ such that $(\Omega, \mathcal{A}, P)$ is the product space of these i spaces. On each of these probability spaces define the Wiener process $W_i(s)$ and denote a sequence of these independent Wiener processes as $\mathcal{W} = \{W_i(s)\}_{i=1}^{\infty}$. Let $\tau_i^{(+)}(r, t)$ (defined on the probability space $(\Omega_i, \mathcal{A}_i, P_i)$) be the first time when $W_i(s)$ exits the interval $(-\pi_i(r, t), 1 - \pi_i(r, t))$; in mathematical terms we have

$$\tau_i^{(+)}(r, t) = \inf \left\{ s \geq 0, W_i(s) \notin (-\pi_i(r, t, 1 - \pi_i(r, t)) \right\}.$$

From Lemma 3.4 follows that

$$\xi_i^{(+)}(r, t) =_D W_i(\tau_i^{(+)}(r, t))$$

and due to properties of a Wiener process (see Definition 3.2 and the properties of Wiener process following the definition) we arrive at

$$n^{-\frac{1}{4}} \sum_{i=1}^{n} \xi_i^{(+)}(r, t) =_D n^{-\frac{1}{4}} \sum_{i=1}^{n} W_i(\tau_i^{(+)}(r, t)) =_D W \left(n^{-\frac{1}{2}} \sum_{i=1}^{n} \tau_i^{(+)}(r, t) \right),$$

where $W(s)$ is again a Wiener process, such that $W(s)$ is independent from $\mathcal{W} = \{W_i(s)\}_{i=1}^{\infty}$.

As we have shown above, $\left\{ \xi_i^{(+)}(r, t) \right\}_{r \in R, t \in R^p}$ is a separable process. Therefore we can apply Lemma 3.3 to obtain

$$n^{-\frac{1}{4}} \sup_{r \in \mathbb{R}^+} \sup_{t \in \tau_M} \left| \sum_{i=1}^{n} \xi_i^{(+)}(r, t) \right| =_D \sup_{r \in \mathbb{R}^+} \sup_{t \in \tau_M} \left| W \left(n^{-\frac{1}{2}} \sum_{i=1}^{n} \tau_i^{(+)}(r, t) \right) \right|. \quad (3.23)$$

Let further U_i (defined again on the probability space $(\Omega_i, \mathcal{A}_i, P_i)$) be the first time when $W_i(s)$ exits the interval $(-n^{-\frac{1}{2}} \Delta, 1)$. Taking into consideration that for all $r \in \mathbb{R}^+$, $t \in \tau_M$ and $i = 1, ..., n$ we have

$$\pi_i(r,t) \leq n^{-\frac{1}{2}}\Delta \ \text{ and } \ 1 - \pi_i(r,t) \leq 1,$$

we can also conclude that for all $r \in \mathbb{R}^+$, $t \in \tau_M$ and $i = 1, ..., n$ it holds that

$$\tau_i^{(+)}(r,t) < U_i.$$

Note that U_i does not depend on r or t. Then it follows that the expression on the right hand side of Equation 3.23 is not larger than

$$\sup_{s \in \left(0, n^{-\frac{1}{2}} \sum_{i=1}^n U_i\right)} |W(s)| \,. \tag{3.24}$$

Moreover, applying the last part of Lemma 3.4, we have $E(U_i) = n^{-\frac{1}{2}}\Delta$.

We can further find $K_1 < \infty$ such that $\frac{\Delta}{K_1} < \frac{\varepsilon}{2}$. Then utilizing Markov inequality (i.e. Chebyshev inequality for positive variables) we obtain for every $n > n_0$

$$P\left(\left\{\omega \in \Omega \ : \ n^{-\frac{1}{2}} \sum_{i=1}^n U_i > K_1\right\}\right) \leq \frac{1}{K_1\sqrt{n}} \sum_{i=1}^n E(U_i) = \frac{\Delta}{K_1} < \frac{\varepsilon}{2}. \tag{3.25}$$

Moreover, let us find $K_2 > 0$ such that $\frac{2K_1}{K_2^2} \leq \frac{\varepsilon}{2}$. Then from Equation 3.23, Equation 3.24 and Equation 3.25 follows that

$$P\left(n^{-\frac{1}{4}} \sup_{r \in \mathbb{R}} \sup_{t \in \tau_M} \left|\sum_{i=1}^n \xi_i^{(+)}(r,t)\right| > K_2\right) =$$

$$= P\left(\sup_{r \in \mathbb{R}^+} \sup_{t \in \tau_M} \left|W\left(n^{-\frac{1}{2}} \sum_{i=1}^n \tau_i^{(+)}(r,t)\right)\right| > K_2\right) \leq$$

$$\leq P\left(\sup_{s \in \left(0, n^{-\frac{1}{2}} \sum_{i=1}^n U_i\right)} |W(s)| > K_2\right) =$$

$$= P\left(\left\{\sup_{s \in \left(0, n^{-\frac{1}{2}} \sum_{i=1}^n U_i\right)} |W(s)| > K_2\right\} \cap \left\{n^{-\frac{1}{2}} \sum_{i=1}^n U_i \leq K_1\right\}\right) +$$

$$+ P\left(\left\{\sup_{s \in \left(0, n^{-\frac{1}{2}} \sum_{i=1}^n U_i\right)} |W(s)| > K_2\right\} \cap \left\{n^{-\frac{1}{2}} \sum_{i=1}^n U_i > K_1\right\}\right) \leq$$

$$\leq P\left(\sup_{0\leq s\leq K_1}|W(s)|>K_2\right)+\frac{\varepsilon}{2}.$$

Applying Lemma 3.5, using Chebyshev inequality and recalling the property of a Wiener process that $Var(W(K_1)) = K_1$, we find that this is further bounded by

$$2P\left(|W(K_1)|>K_2\right)+\frac{\varepsilon}{2}\leq\frac{2K_1}{K_2^2}+\frac{\varepsilon}{2}\leq\frac{\varepsilon}{2}+\frac{\varepsilon}{2}=\varepsilon.$$

Therefore we have derived that

$$P\left(n^{-\frac{1}{4}}\sup_{r\in\mathbb{R}}\sup_{t\in\tau_M}\left|\sum_{i=1}^{n}\xi_i^{(+)}(r,t)\right|>K_2\right)\leq\varepsilon. \tag{3.26}$$

Similarly, we can derive analogical conclusions based on Equation 3.15, Equation 3.16 and Equation 3.17. Let us define

$$b_i^{(-)}(r,t)=I_{\left\{r-n^{-\frac{1}{2}}X_i't\leq u_i<r\right\}},\quad c_i^{(+)}(r,t)=I_{\left\{-r-n^{-\frac{1}{2}}X_i't<u_i\leq -r\right\}}$$

$$\textit{and}\quad c_i^{(-)}(r,t)=I_{\left\{-r<u_i\leq -r-n^{-\frac{1}{2}}X_i't\right\}}. \tag{3.27}$$

Then for

$$\xi_i^{(-)}(r,t)=b_i^{(-)}(r,t)-E\left(b_i^{(-)}(r,t)\right),\quad \zeta_i^{(+)}(r,t)=c_i^{(+)}(r,t)-E\left(c_i^{(+)}(r,t)\right)$$

$$\textit{and}\quad \zeta_i^{(-)}(r,t)=c_i^{(-)}(r,t)-E\left(c_i^{(-)}(r,t)\right)$$

hold conclusions that are analogical to Equation 3.26. From these conclusions, Equation 3.18 and Equation 3.20 we obtain[2]

$$n^{-\frac{1}{4}}\sup_{r\in\mathbb{R}}\sup_{t\in\tau_M}|m_n(r,t)|\leq$$

$$\leq n^{-\frac{1}{4}}\sup_{r\in\mathbb{R}}\sup_{t\in\tau_M}\left|\sum_{i=1}^{n}\left[b_i^{(+)}(r,t)-b_i^{(-)}(r,t)+c_i^{(+)}(r,t)-c_i^{(-)}(r,t)\right]\right|=$$

$$=n^{-\frac{1}{4}}\sup_{r\in\mathbb{R}}\sup_{t\in\tau_M}$$

$$\left|\sum_{i=1}^{n}\left[\left(b_i^{(+)}(r,t)-E\left(b_i^{(+)}(r,t)\right)\right)-\left(b_i^{(-)}(r,t)-E\left(b_i^{(-)}(r,t)\right)\right)+\right.\right.$$

$$+\left(c_i^{(+)}(r,t)-E\left(c_i^{(+)}(r,t)\right)\right)-\left(c_i^{(-)}(r,t)-E\left(c_i^{(-)}(r,t)\right)\right]+$$

[2]Note also that generally $|y|\leq|y-x|+|x|$ and $|x+y|\leq|x|+|y|$.

$$+\left[E\left(b_i^{(+)}(r,t)\right)-E\left(b_i^{(-)}(r,t)\right)+E\left(c_i^{(+)}(r,t)\right)-E\left(c_i^{(-)}(r,t)\right)\right]\Bigg|\leq$$

$$\leq n^{-\frac{1}{4}}\sup_{r\in\mathbb{R}}\sup_{t\in\tau_M}$$

$$\left|\sum_{i=1}^{n}\left[\left(b_i^{(+)}(r,t)-E\left(b_i^{(+)}(r,t)\right)\right)-\left(b_i^{(-)}(r,t)-E\left(b_i^{(-)}(r,t)\right)\right)+\right.\right.$$

$$\left.\left.+\left(c_i^{(+)}(r,t)-E\left(c_i^{(+)}(r,t)\right)\right)-\left(c_i^{(-)}(r,t)-E\left(c_i^{(-)}(r,t)\right)\right)\right]\right|+$$

$$+n^{-\frac{1}{4}}\sup_{r\in\mathbb{R}}\sup_{t\in\tau_M}$$

$$\left|\sum_{i=1}^{n}\left[E\left(b_i^{(+)}(r,t)\right)-E\left(b_i^{(-)}(r,t)\right)+E\left(c_i^{(+)}(r,t)\right)-E\left(c_i^{(-)}(r,t)\right)\right]\right|=$$

$$=n^{-\frac{1}{4}}\sup_{r\in\mathbb{R}}\sup_{t\in\tau_M}\left|\sum_{i=1}^{n}\left[\xi_i^{(+)}(r,t)-\xi_i^{(-)}(r,t)+\zeta_i^{(+)}(r,t)-\zeta_i^{(-)}(r,t)\right]\right|+ \tag{3.28}$$

$$+n^{-\frac{1}{4}}\sup_{r\in\mathbb{R}}\sup_{t\in\tau_M}$$

$$\left|\sum_{i=1}^{n}\left[E\left(b_i^{(+)}(r,t)\right)-E\left(b_i^{(-)}(r,t)\right)+E\left(c_i^{(+)}(r,t)\right)-E\left(c_i^{(-)}(r,t)\right)\right]\right| \tag{3.29}$$

where due to Equation 3.26 and the analogical results we conclude that Equation 3.28 is bounded in probability. To conclude the proof it remains to show that Equation 3.29 is also small in probability. For that purpose we need to estimate all the mean values.

Let us first realize that due to construction of $b_i^{(+)}(r,t)$, $b_i^{(-)}(r,t)$, $c_i^{(+)}(r,t)$ and $c_i^{(-)}(r,t)$ in Equation 3.19 and Equation 3.27 all the mean values in Equation 3.29 are nonnegative. Moreover, as was mentioned before, for $E\left(b_i^{(+)}(r,t)\right)$ to be nonnegative we need $x't<0$. Similarly we need $x't<0$ for $E\left(c_i^{(-)}(r,t)\right)$ to be nonnegative and analogicaly we obtain $x't>0$ for the two remaining mean values.

Let us consider $x't<0$. Recall the derivation in Equation 3.22 and the following lines and write

$$\pi_i(r,t)=E\left(b_i^{(+)}(r,t)\right)=\int\left[\int_{r}^{r-n^{-\frac{1}{2}}x't}f_{u_i}(v)\mathrm{d}v\right]\mathrm{d}F_X(x)=$$

$$= \int \left[\int_r^{r-n^{-\frac{1}{2}}x't} (f_{u_i}(v) - f_{u_i}(r))\,\mathrm{d}v \right] \mathrm{d}F_X(x) +$$

$$+ \int \left[f_{u_i}(r) \int_r^{r-n^{-\frac{1}{2}}x't} 1\mathrm{d}v \right] \mathrm{d}F_X(x) = R_b^{(+)}(r,t) - n^{-\frac{1}{2}} \int x't f_{u_i}(r) \mathrm{d}F_X(x),$$

where recalling Assumptions 3.1, Assumptions 3.4 and using the substitution $q = v - r$ we can write

$$\left| R_b^{(+)}(r,t) \right| = \left| \int \left[\int_r^{r-n^{-\frac{1}{2}}x't} (f_{u_i}(v) - f_{u_i}(r))\,\mathrm{d}v \right] \mathrm{d}F_X(x) \right| =$$

$$= \left| \int \left[\int_r^{r-n^{-\frac{1}{2}}x't} \sigma_i^{-1} \left(f_u(v\sigma_i^{-1}) - f_u(r\sigma_i^{-1}) \right) \mathrm{d}v \right] \mathrm{d}F_X(x) \right| \le$$

$$\le \int \left[\int_r^{r-n^{-\frac{1}{2}}x't} \sigma_i^{-2} L_u \left| v - r \right| \mathrm{d}v \right] \mathrm{d}F_X(x) =$$

$$= \sigma_i^{-2} L_u \int \left[\int_0^{-n^{-\frac{1}{2}}x't} q\mathrm{d}q \right] \mathrm{d}F_X(x) = \frac{1}{2n} \sigma_i^{-2} L_u \int (x't)^2 \,\mathrm{d}F_X(x) \le$$

$$\le \frac{1}{2n} a^{-2} L_u \int (x't)^2 \,\mathrm{d}F_X(x).$$

Analogically, we obtain for $E\left(c_i^{(-)}(r,t)\right)$:

$$E\left(c_i^{(-)}(r,t)\right) = \int I_{\left\{-r \le v < -r - n^{-\frac{1}{2}}x't\right\}} \mathrm{d}F_{X,u_i}(x,v) =$$

$$= R_c^{(-)}(r,t) - n^{-\frac{1}{2}} \int x't f_{u_i}(r) \mathrm{d}F_X(x)$$

with

$$\left| R_c^{(-)}(r,t) \right| \le \frac{1}{2n} a^{-2} L_u \int (x't)^2 \,\mathrm{d}F_X(x).$$

Therefore we can write for the estimate of the absolute value of the difference of these two mean values:

$$\left| E\left(b_i^{(+)}(r,t)\right) - E\left(c_i^{(-)}(r,t)\right) \right| \le \left| R_b^{(+)}(r,t) \right| + \left| R_c^{(-)}(r,t) \right| \le$$

$$\leq \frac{1}{n} a^{-2} L_u \int (x't)^2 \, \mathrm{d}F_X(x),$$

where the same result holds for $\left| E\left(c_i^{(+)}(r,t)\right) - E\left(b_i^{(-)}(r,t)\right) \right|$ with $x't > 0$. Hence we can bound Equation 3.29 as follows:

$$n^{-\frac{1}{4}} \sup_{r \in \mathbb{R}} \sup_{t \in \tau_M}$$

$$\left| \sum_{i=1}^{n} \left[E\left(b_i^{(+)}(r,t)\right) - E\left(b_i^{(-)}(r,t)\right) + E\left(c_i^{(+)}(r,t)\right) - E\left(c_i^{(-)}(r,t)\right) \right] \right| \leq$$

$$\leq 2n^{-\frac{1}{4}} \sup_{t \in \tau_M} \sum_{i=1}^{n} \frac{1}{n} a^{-2} L_u \int (x't)^2 \, \mathrm{d}F_X(x) =$$

$$= 2n^{-\frac{1}{4}} \sup_{t \in \tau_M} a^{-2} L_u \int (x't)^2 \, \mathrm{d}F_X(x) \leq$$

$$\leq 2n^{-\frac{1}{4}} a^{-2} L_u M^2 \int \|x\|^2 \, \mathrm{d}F_X(x) = O(n^{-\frac{1}{4}}).$$

As $n^{\tau-1} < n^{-\frac{1}{4}}$, this concludes the proof.

Q.E.D.

Based on this result, we can easily prove following lemmas.

Lemma 3.10. *Let the Assumptions 3.1, 3.2, 3.3, 3.4 and 3.5 hold. Choose arbitrarily* $\varepsilon > 0$ *and* $\tau \in (\frac{1}{2}, \frac{3}{4})$. *Then there exists* $K \in (0, \infty)$ *and* $n_{\varepsilon,M,\tau} \in \mathbb{N}$ *such that for all* $n > n_{\varepsilon,M,\tau}$

$$P\left(\left\{\omega \in \Omega : \sup_{r \in \mathbb{R}} n^{\tau} \left| F^{(n)}_{\hat{\beta}^{(LWS,n,w)}}(r) - F^{(n)}_{\beta^0}(r) \right| < K \right\}\right) > 1 - \varepsilon.$$

PROOF Put $t = -\sqrt{n}\left(\hat{\beta}^{(LWS,n,w)} - \beta^0\right)$. Then $\beta^0 - n^{-\frac{1}{2}} t = \hat{\beta}^{(LWS,n,w)}$ and according to Lemma 3.2 we have $t = O_p(1)$. Hence the proof follows from Lemma 3.9.

Q.E.D.

Lemma 3.11. *Let the Assumptions 3.1 and 3.4 hold. Choose arbitrarily* $\varepsilon > 0$ *and* $\tau \in (\frac{1}{2}, \frac{3}{4})$. *Then there exists* $K \in (0, \infty)$ *and* $n_{\varepsilon,M,\tau} \in \mathbb{N}$ *such that for all* $n > n_{\varepsilon,M,\tau}$

$$P\left(\left\{\omega \in \Omega : \sup_{r \in \mathbb{R}} \sup_{t \in \tau_M} n^{\tau} \left| F^{(n)}_{\beta^0}\left(\left|r - n^{-\frac{1}{2}} X_i' t\right|\right) - F^{(n)}_{\beta^0}(|r|) \right| < K \right\}\right) > 1 - \varepsilon.$$

PROOF Let $r > 0$ and realize that

$$F^{(n)}_{\beta^0}\left(\left|r - n^{-\frac{1}{2}}X_i't\right|\right) - F^{(n)}_{\beta^0}(|r|) = \frac{1}{n}\sum_{i=1}^{n}\left[I_{\left\{|u_i|<\left|r-n^{-\frac{1}{2}}X_i't\right|\right\}} - I_{\{|u_i|<|r|\}}\right].$$

It follows that the proof will be essentially the same as the proof of Lemma 3.9.

Q.E.D.

Lemma 3.12. *Let the Assumptions 3.1, 3.2, 3.3, 3.4 and 3.5 hold. Choose arbitrarily* $\varepsilon > 0$ *and* $\tau \in (\frac{1}{2}, \frac{3}{4})$. *Then there exists* $K \in (0, \infty)$ *and* $n_{\varepsilon,M,\tau} \in \mathbb{N}$ *such that for all* $n > n_{\varepsilon,M,\tau}$

$$P\left(\left\{\omega \in \Omega : \max_i n^{\tau}\left|F^{(n)}_{\beta^0}\left(\left|r_i(\hat{\beta}^{(LWS,n,w)})\right|\right) - F^{(n)}_{\beta^0}(|u_i|)\right| < K\right\}\right) > 1 - \varepsilon,$$

where $i = 1, 2, ..., n$.

PROOF Let again $t = -\sqrt{n}\left(\hat{\beta}^{(LWS,n,w)} - \beta^0\right)$. Then when we put $r = u_i$, we have $r - n^{-\frac{1}{2}}X_i't = r_i(\hat{\beta}^{(LWS,n,w)})$ and the proof follows from Lemma 3.11.

Q.E.D.

Lemma 3.10 and Lemma 3.12 enable us to prove the following.

Lemma 3.13. *Let the Assumptions 3.1, 3.2, 3.3, 3.4 and 3.5 hold. Then*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} w\left(F^{(n)}_{\beta^0}(|u_i|)\right) X_i\left(Y_i - X_i'\hat{\beta}^{(LWS,n,w)}\right) = o_p(1). \tag{3.30}$$

PROOF Recall that β^0 is (WLOG) assumed to be 0. Therefore we can write

$$\frac{1}{\sqrt{n}}\left\|\sum_{i=1}^{n} w\left(F^{(n)}_{\beta^0}(|u_i|)\right) X_i\left(Y_i - X_i'\hat{\beta}^{(LWS,n,w)}\right)\right\| =$$

$$= \frac{1}{\sqrt{n}}\left\|\sum_{i=1}^{n} w\left(F^{(n)}_{\beta^0}(|u_i|)\right) X_i\left(u_i - X_i'\hat{\beta}^{(LWS,n,w)}\right)\right\|,$$

which is bounded by

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left|w\left(F^{(n)}_{\beta^0}(|u_i|)\right) - w\left(F^{(n)}_{\beta^0}(|r_i(\hat{\beta}^{(LWS,n,w)})|)\right)\right| \times$$

$$\times \|X_i\| \cdot \left|u_i - X_i'\hat{\beta}^{(LWS,n,w)}\right| + \tag{3.31}$$

$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left|w\left(F_{\beta^0}^{(n)}(|r_i(\hat{\beta}^{(LWS,n,w)})|)\right)-w\left(F_{\hat{\beta}^{(LWS,n,w)}}^{(n)}(|r_i(\hat{\beta}^{(LWS,n,w)})|)\right)\right|\times$$

$$\times\|X_i\|\cdot\left|u_i-X_i'\hat{\beta}^{(LWS,n,w)}\right|+ \tag{3.32}$$

$$+\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}w\left(F_{\hat{\beta}^{(LWS,n,w)}}^{(n)}(|r_i(\hat{\beta}^{(LWS,n,w)})|)\right)X_i\left(u_i-X_i'\hat{\beta}^{(LWS,n,w)}\right)\right\|. \tag{3.33}$$

As $\hat{\beta}^{(LWS,n,w)}$ is one of the solutions of the normal equation in Equation 3.3, we can conclude that the expression in Equation 3.33 is equal to 0 (realizing again that $\beta^0 = 0$). Let us now show that the expression in Equation 3.31 converges in probability to 0. Equation 3.31 is bounded by

$$\frac{1}{n^{\tau+\frac{1}{2}}}\cdot n^{\tau}\max_i\left|w\left(F_{\beta^0}^{(n)}(|u_i|)\right)-w\left(F_{\beta^0}^{(n)}(|r_i(\hat{\beta}^{(LWS,n,w)})|)\right)\right|\cdot\sum_{i=1}^{n}\|X_i\|\cdot|u_i|+$$

$$+\frac{1}{n^{\tau+\frac{1}{2}}}\cdot n^{\tau}\max_i\left|w\left(F_{\beta^0}^{(n)}(|u_i|)\right)-w\left(F_{\beta^0}^{(n)}(|r_i(\hat{\beta}^{(LWS,n,w)})|)\right)\right|\times$$

$$\times\sum_{i=1}^{n}\|X_i\|^2\cdot\left\|\hat{\beta}^{(LWS,n,w)}-\beta^0\right\|.$$

From Assumptions 3.2 it follows that this is further bounded by

$$\frac{L_w}{n^{\tau+\frac{1}{2}}}\cdot n^{\tau}\max_i\left|F_{\beta^0}^{(n)}\left(\left|r_i(\hat{\beta}^{(LWS,n,w)})\right|\right)-F_{\beta^0}^{(n)}(|u_i|)\right|\cdot\sum_{i=1}^{n}\|X_i\|\cdot|u_i|+$$

$$+\frac{L_w}{n^{\tau+\frac{1}{2}}}\cdot n^{\tau}\max_i\left|F_{\beta^0}^{(n)}\left(\left|r_i(\hat{\beta}^{(LWS,n,w)})\right|\right)-F_{\beta^0}^{(n)}(|u_i|)\right|\times$$

$$\times\sum_{i=1}^{n}\|X_i\|^2\cdot\left\|\hat{\beta}^{(LWS,n,w)}-\beta^0\right\|.$$

As $\tau+\frac{1}{2}>1$, $E\left(\|X_i\|\right)<\infty$, $E\left(\|X_i\|^2\right)<\infty$, $E\left(|u_i|\right)<\infty$, we can use Lemma 3.1 and Lemma 3.12 to conclude that Equation 3.31 is $o_p(1)$.

Similarly, we can show that Equation 3.32 is bounded by

$$\frac{1}{n^{\tau+\frac{1}{2}}}\cdot n^{\tau}\sup_{r\in\mathbb{R}}\left|w\left(F_{\beta^0}^{(n)}(|r|)\right)-w\left(F_{\hat{\beta}^{(LWS,n,w)}}^{(n)}(|r|)\right)\right|\cdot\sum_{i=1}^{n}\|X_i\|\cdot|u_i|+$$

$$+\frac{1}{n^{\tau+\frac{1}{2}}}\cdot n^{\tau}\sup_{r\in\mathbb{R}}\left|w\left(F_{\beta^0}^{(n)}(|r|)\right)-w\left(F_{\hat{\beta}^{(LWS,n,w)}}^{(n)}(|r|)\right)\right|\times$$

$$\times \sum_{i=1}^{n} \|X_i\|^2 \cdot \left\| \hat{\beta}^{(LWS,n,w)} - \beta^0 \right\| .$$

Using again that the weight function is Lipschitz in absolute value, this can be bounded by

$$\frac{L_w}{n^{\tau+\frac{1}{2}}} \cdot n^{\tau} \sup_{r\in\mathbb{R}} \left| F^{(n)}_{\beta^0}(|r|) - F^{(n)}_{\hat{\beta}^{(LWS,n,w)}}(|r|) \right| \cdot \sum_{i=1}^{n} \|X_i\| \cdot |u_i| +$$

$$+\frac{L_w}{n^{\tau+\frac{1}{2}}} \cdot n^{\tau} \sup_{r\in\mathbb{R}} \left| F^{(n)}_{\beta^0}(|r|) - F^{(n)}_{\hat{\beta}^{(LWS,n,w)}}(|r|) \right| \cdot \sum_{i=1}^{n} \|X_i\|^2 \cdot \left\| \hat{\beta}^{(LWS,n,w)} - \beta^0 \right\| .$$

For the same reasons as before for Equation 3.31, just using Lemma 3.10 instead of Lemma 3.12, we can conclude that Equation 3.32 is $o_p(1)$. This concludes the proof.

Q.E.D.

Before moving on to the next lemma, let us prepare some additional necessary tools. Racalling Equation 3.1 and considering that $r_j(\beta^0) = u_j$, we obtain

$$F^{(n)}_{\beta^0}(|u_i|) = \frac{1}{n} \sum_{j=1}^{n} I_{\{|u_j|<|u_i|\}}.$$

When we further recall Equation 3.2, we get

$$F^{(n)}_{\beta^0}(|u_i|) = \frac{\pi(\beta^0, i) - 1}{n}. \tag{3.34}$$

Moreover, we can denote

$$w^*_k = w\left(\frac{k-1}{n}\right) - w\left(\frac{k}{n}\right) \tag{3.35}$$

and recalling Assumptions 3.2 (specifically $w(1) = 0$), by summation we obtain

$$w\left(\frac{k-1}{n}\right) = \sum_{j=k}^{n} w^*_j. \tag{3.36}$$

Then combining Equation 3.34 and Equation 3.36 we arrive at

$$w\left(F^{(n)}_{\beta^0}(|u_i|)\right) = w\left(\frac{\pi(\beta^0, i) - 1}{n}\right) = \sum_{\ell=\pi(\beta^0,i)}^{n} w^*_\ell = \sum_{\ell=1}^{n} w^*_\ell \cdot I_{\left\{u_i^2 \le u^2_{(\ell)}\right\}}. \tag{3.37}$$

Using this result, we can write Equation 3.30 as

$$\frac{1}{\sqrt{n}}\sum_{\ell=1}^{n} w_\ell^* \sum_{i=1}^{n} X_i \left(Y_i - X_i'\hat{\beta}^{(LWS,n,w)}\right) \cdot I_{\left\{u_i^2 \leq u_{(\ell)}^2\right\}} = o_p(1). \tag{3.38}$$

At this point, let us recall Equation 3.8, for any $\alpha \in (0,1)$ define q_α^2 as the upper α-quantile of $G(z)$ (i.e. we have $1 - G(q_\alpha^2) = \alpha$) and let us prove following lemmas. Note that Lemma 3.14 is new as compared to Víšek (2015) and we need it to be able to generalize for heteroskedasticity Lemma 3.15 and what follows.

Lemma 3.14. *Let the Assumptions 3.1 hold. Then we have*

$$G(z) - [F_u(z^{\frac{1}{2}}) - F_u(-z^{\frac{1}{2}})] = O(n^{-\frac{1}{2}}).$$

PROOF We can write

$$\left|G(z) - [F_u(z^{\frac{1}{2}}) - F_u(-z^{\frac{1}{2}})]\right| =$$

$$= \left|\frac{1}{n}\sum_{i=1}^{n} F_u\left(\sigma_i^{-\frac{1}{2}} z^{\frac{1}{2}}\right) - F_u\left(-\sigma_i^{-\frac{1}{2}} z^{\frac{1}{2}}\right) - F_u(z^{\frac{1}{2}}) + F_u(-z^{\frac{1}{2}})\right| =$$

$$= \left|\frac{1}{n}\sum_{i=1}^{n}\left\{\int_{-\sigma_i^{-\frac{1}{2}} z^{\frac{1}{2}}}^{-z^{\frac{1}{2}}} f_u(v)\mathrm{d}v + \int_{z^{\frac{1}{2}}}^{\sigma_i^{-\frac{1}{2}} z^{\frac{1}{2}}} f_u(v)\mathrm{d}v\right\}\right| \leq$$

$$\leq \left[\sup_{v\in(z^{\frac{1}{2}}b^{-\frac{1}{2}}, z^{\frac{1}{2}}a^{-\frac{1}{2}})} f_u(v)\right] \cdot \left|\frac{1}{n}\sum_{i=1}^{n}\left\{\int_{-\sigma_i^{-\frac{1}{2}} z^{\frac{1}{2}}}^{-z^{\frac{1}{2}}} 1\mathrm{d}v + \int_{z^{\frac{1}{2}}}^{\sigma_i^{-\frac{1}{2}} z^{\frac{1}{2}}} 1\mathrm{d}v\right\}\right| \leq$$

$$\leq 2 \cdot \left|\left[\sup_{v\in(z^{\frac{1}{2}}b^{-\frac{1}{2}}, z^{\frac{1}{2}}a^{-\frac{1}{2}})} f_u(v)\right] \cdot z^{\frac{1}{2}}\right| \cdot \left|\frac{1}{n}\sum_{i=1}^{n}(\sigma_i^{-\frac{1}{2}} - 1)\right| \leq$$

$$\leq L_\varepsilon \left|\frac{1}{n}\sum_{i=1}^{n}(\sigma_i^{-\frac{1}{2}} - 1)\right| = O(n^{-\frac{1}{2}}),$$

where except for the definition of $G(z)$ we have used Equation 3.9 and Equation 3.10.

Q.E.D.

Note that here we need somewhat weaker assumptions than are stated in Assumptions 3.6 (here we have everything to the power of $\frac{1}{2}$; if it is bounded for the expressions with the power of 1, it must be bounded also for these expressions with the power of $\frac{1}{2}$). However, the Assumptions 3.6 are stated in the stronger form, as that form is more usual. We can use this result to prove another lemma.

Lemma 3.15. *Let Assumptions 3.1 hold and put* $\ell_n(\alpha) = [(1-\alpha)n]_{int}$ *for any* $n \in \mathbb{N}$. *Then for any* $\varepsilon \in (0,1)$ *there are constants* $K^\varepsilon < \infty$ *and* $\tilde{K}^\varepsilon < \infty$, *and there is* $n_\varepsilon \in \mathbb{N}$ *such that for all* $n > n_\varepsilon$ *and any* $\alpha \in (0,1)$ *there exists an interval* $I^\varepsilon_{\alpha,n}$ *such that*

$$q^2_\alpha \in I^\varepsilon_{\alpha,n} \quad for\ any \quad \alpha \in (0,1), \tag{3.39}$$

$$P\left(\bigcap_{\alpha\in(0,1)} \left\{\omega \in \Omega \ : \ u^2_{(\ell_n(\alpha))} \in I^\varepsilon_{\alpha,n}\right\} \right) > 1-\varepsilon, \tag{3.40}$$

and for all $i = 1, 2, ..., n$ *we have*

$$\sup_{\alpha\in(0,1)} P\left(u^2_i \in I^\varepsilon_{\alpha,n}\right) \le n^{-\frac{1}{2}} K^\varepsilon \tag{3.41}$$

and

$$\sup_{\alpha\in(0,1)} E\left[|u_i| \cdot I^{(\varepsilon)}_{\alpha,n}\right] \le n^{-\frac{1}{2}} \tilde{K}^\varepsilon.$$

PROOF The proof is again mainly based on Víšek (2015). However, several adjustments for heteroskedasticity are necessary here. Before starting the proof let us shorten the notation and write ℓ_n instead of $\ell_n(\alpha)$. Then we can divide the proof in two parts. In the first part we will show that for any $\varepsilon \in (0,1)$ there are constants $K^U_\varepsilon < \infty$ and $\tilde{K}_\varepsilon < \infty$, and there is $n_\varepsilon \in \mathbb{N}$ such that for all $n > n_\varepsilon$ and any $\alpha \in (0,1)$ there exists $U^{(\varepsilon)}_{\alpha,n}$ such that

$$q^2_\alpha \le U^{(\varepsilon)}_{\alpha,n} \quad \text{for any} \quad \alpha \in (0,1), \tag{3.42}$$

$$P\left(\bigcap_{\alpha\in(0,1)} \left\{\omega \in \Omega \ : \ u^2_{(\ell_n)} \le U^{(\varepsilon)}_{\alpha,n}\right\} \right) > 1-\frac{1}{2}\varepsilon, \tag{3.43}$$

and for all $i = 1, 2, ..., n$ we have

$$\sup_{\alpha \in (0,1)} P\left(u_i^2 \in \left[q_\alpha^2, U_{\alpha,n}^{(\varepsilon)}\right]\right) < \frac{1}{2} n^{-\frac{1}{2}} K_\varepsilon^U \tag{3.44}$$

and

$$\sup_{\alpha \in (0,1)} E\left[|u_i| \cdot I_{\left\{q_\alpha^2 \le u_i^2 \le U_{\alpha,n}^{(\varepsilon)}\right\}}\right] \le \sup_{\alpha \in (0,1)} \int_{z^2 \in \left[q_\alpha^2, U_{\alpha,n}^{(\varepsilon)}\right]} |z| f_{u_i}(z) \mathrm{d}z <$$

$$< \frac{1}{2} n^{-\frac{1}{2}} \tilde{K}_\varepsilon. \tag{3.45}$$

Further notice that $U_{\alpha,n}^{(\varepsilon)}$ does not have to be finite and let it satisfy

$$G(U_{\alpha,n}^{(\varepsilon)}) - G(q_\alpha^2) = min\left\{n^{-\frac{1}{2}} K_\varepsilon, \alpha\right\}. \tag{3.46}$$

Choose and fix some $\varepsilon > 0$. Let $W(s)$ be a Wiener process, see again Definition 3.2. Then we can use Lemma 3.5 to find $K_\varepsilon < \infty$ such that

$$P\left(\sup_{0 \le s \le 2} |W(s)| > \frac{1}{2} K_\varepsilon\right) < \frac{1}{4}\varepsilon. \tag{3.47}$$

Similarly as before, denote a sequence of independent Wiener processes as $\mathcal{W} = \{W_i(s)\}_{i=1}^\infty$. Further, let U_i be the first time when $W_i(s)$ exits the interval $(-1, 1)$ and note that $\{U_i\}_{i=1}^\infty$ is a sequence of iid random variables. Then applying again the last part of Lemma 3.4, we obtain $E(U_i) = 1$. Moreover, put

$$B_n = \left\{\omega \in \Omega \;:\; n^{-1} \sum_{i=1}^n U_i > 2\right\}.$$

For the fixed $\varepsilon > 0$ we can then find $n_\varepsilon \in N$ such that for all $n > n_\varepsilon$

$$P(B_n) < \frac{1}{4}\varepsilon \tag{3.48}$$

and further consider only $n > n_\varepsilon$.

Returning to Equation 3.46, let us split it into two cases and define

$$A_n^{(\varepsilon)} = \left\{\alpha \in (0,1) \;:\; n_\varepsilon^{-\frac{1}{2}} K_\varepsilon < \alpha\right\}.$$

Then consider first $\alpha \in (0,1) \setminus A_n^{(\varepsilon)}$, i.e. the case when $G(U_{\alpha,n}^{(\varepsilon)}) - G(q_\alpha^2) = \alpha$. In this case we can put $U_{\alpha,n}^{(\varepsilon)} = \infty$. It follows that $q_\alpha^2 < U_{\alpha,n}^{(\varepsilon)} = \infty$ and Equation 3.42 holds. Moreover, if $U_{\alpha,n}^{(\varepsilon)} = \infty$, we have

$$P\left(\left\{u_{(\ell_n)}^2 < U_{\alpha,n}^{(\varepsilon)}\right\}\right) = 1. \tag{3.49}$$

To show that Equation 3.44 holds for any $\alpha \in (0,1) \setminus A_n^{(\varepsilon)}$, we can write

$$P\left(u_i^2 \in \left[q_\alpha^2, U_{\alpha,n}^{(\varepsilon)}\right]\right) = G_i(U_{\alpha,n}^{(\varepsilon)}) - G_i(q_\alpha^2) =$$

$$= F_u\left(\sigma_i^{-\frac{1}{2}}(U_{\alpha,n}^{(\varepsilon)})^{\frac{1}{2}}\right) - F_u\left(-\sigma_i^{-\frac{1}{2}}(U_{\alpha,n}^{(\varepsilon)})^{\frac{1}{2}}\right) - \left[F_u\left(\sigma_i^{-\frac{1}{2}}q_\alpha\right) - F_u\left(-\sigma_i^{-\frac{1}{2}}q_\alpha\right)\right] =$$

$$= G\left(\sigma_i^{-1}U_{\alpha,n}^{(\varepsilon)}\right) - G\left(\sigma_i^{-1}q_\alpha^2\right) + n^{-\frac{1}{2}}K^g \leq B_g \cdot \left(G(U_{\alpha,n}^{(\varepsilon)}) - G(q_\alpha^2)\right) + n^{-\frac{1}{2}}K^g =$$

$$= B_g \cdot \alpha + n^{-\frac{1}{2}}K^g \leq n^{-\frac{1}{2}}\left(B_g \cdot K_\varepsilon + K^g\right),$$

where we put $K_\varepsilon^U = 2\cdot(B_g \cdot K_\varepsilon + K^g)$ and we used Assumptions 3.6 and Lemma 3.14. When we denote

$$C_n = \bigcap_{\alpha \in (0,1)\setminus A_n^{(\varepsilon)}} \left\{\omega \in \Omega \ : \ u_{(\ell_n)}^2 < U_{\alpha,n}^{(\varepsilon)}\right\},$$

then from Equation 3.49 follows that $P(C_n) = 1$ and therefore also Equation 3.43 holds for $\alpha \in (0,1) \setminus A_n^{(\varepsilon)}$.

Let us now turn to the case when $\alpha \in A_n^{(\varepsilon)}$, i.e. $G(U_{\alpha,n}^{(\varepsilon)}) - G(q_\alpha^2) = n^{-\frac{1}{2}}K_\varepsilon$. We will denote

$$v_i^{(\alpha)} = I_{\left\{u_i^2 \leq U_{\alpha,n}^{(\varepsilon)}\right\}} - E\left(I_{\left\{u_i^2 \leq U_{\alpha,n}^{(\varepsilon)}\right\}}\right),$$

where due to the assumption that the distribution functions of u_i (and therefore also distribution functions of u_i^2) are absolutely continuous we can conclude that the processes $v_i^{(\alpha)}$, $i = 1, ..., n$, are separable (see again Definition 3.1). Moreover, we can write

$$a_{n,i} = E\left(I_{\left\{u_i^2 \leq U_{\alpha,n}^{(\varepsilon)}\right\}}\right) = P\left(u_i^2 \leq U_{\alpha,n}^{(\varepsilon)}\right) < 1. \tag{3.50}$$

Then we have $v_i^{(\alpha)} = 1 - a_{n,i} > 0$ if $u_i^2 \leq U_{\alpha,n}^{(\varepsilon)}$ and $v_i^{(\alpha)} = -a_{n,i} < 0$ otherwise. Let $\tau_{in}^{(\alpha)}$ be the first time when $W_i(s)$ exits the interval $(-a_{n,i}, 1 - a_{n,i})$ and apply Lemma 3.4 to obtain

$$v_i^{(\alpha)} =_D W_i(\tau_{in}^{(\alpha)}),$$

and similarly as in the proof of Lemma 3.9

$$n^{-\frac{1}{2}} \sum_{i=1}^{n} v_i^{(\alpha)} =_D n^{-\frac{1}{2}} \sum_{i=1}^{n} W_i(\tau_{in}^{(\alpha)}) =_D W \left(n^{-1} \sum_{i=1}^{n} \tau_{in}^{(\alpha)} \right), \tag{3.51}$$

where $W(s)$ is again a Wiener process, such that $W(s)$ is independent from $\mathcal{W} = \{W_i(s)\}_{i=1}^{\infty}$. Since $(-a_{n,i}, 1 - a_{n,i}) \subset (-1, 1)$, we have further for every $\alpha \in A_n^{(\varepsilon)}$ and $n > n_\varepsilon$

$$n^{-1} \sum_{i=1}^{n} \tau_{in}^{(\alpha)} \leq n^{-1} \sum_{i=1}^{n} U_i \tag{3.52}$$

and we can use Equation 3.47, Equation 3.48, Equation 3.51, Equation 3.52 and Lemma 3.3 (remember that the processes $v_i^{(\alpha)}$ are separable) to obtain

$$P\left(\left\{ n^{-\frac{1}{2}} \sup_{\alpha \in A_n^{(\varepsilon)}} \left| \sum_{i=1}^{n} v_i^{(\alpha)} \right| > \frac{1}{2} K_\varepsilon \right\} \right) \leq$$

$$\leq P\left(\left\{ n^{-\frac{1}{2}} \sup_{\alpha \in A_n^{(\varepsilon)}} \left| \sum_{i=1}^{n} v_i^{(\alpha)} \right| > \frac{1}{2} K_\varepsilon \right\} \cap B_n^c \right) + P(B_n) \leq$$

$$\leq P\left(\left\{ n^{-\frac{1}{2}} \sup_{\alpha \in A_n^{(\varepsilon)}} \left| \sum_{i=1}^{n} W_i(\tau_{in}^{(\alpha)}) \right| > \frac{1}{2} K_\varepsilon \right\} \cap B_n^c \right) + \frac{1}{4}\varepsilon =$$

$$= P\left(\left\{ \sup_{\alpha \in A_n^{(\varepsilon)}} \left| W\left(n^{-1} \sum_{i=1}^{n} \tau_{in}^{(\alpha)} \right) \right| > \frac{1}{2} K_\varepsilon \right\} \cap B_n^c \right) + \frac{1}{4}\varepsilon \leq$$

$$\leq P\left(\left\{ \sup_{0 \leq s \leq 2} |W(s)| > \frac{1}{2} K_\varepsilon \right\} \cap B_n^c \right) + \frac{1}{4}\varepsilon < \frac{1}{2}\varepsilon.$$

Define the set $D_n = \left\{ \omega \in \Omega \ : \ n^{-\frac{1}{2}} \sup_{\alpha \in A_n^{(\varepsilon)}} \left| \sum_{i=1}^{n} v_i^{(\alpha)} \right| < \frac{1}{2} K_\varepsilon \right\}$. We have just derived that $P(D_n) \geq 1 - \frac{1}{2}\varepsilon$ and therefore we have for all $\alpha \in A_n^{(\varepsilon)}$ with probability at least $1 - \frac{1}{2}\varepsilon$

$$\sum_{i=1}^{n} I_{\left\{u_i^2 \leq U_{\alpha,n}^{(\varepsilon)}\right\}} > \sum_{i=1}^{n} E\left(I_{\left\{u_i^2 \leq U_{\alpha,n}^{(\varepsilon)}\right\}}\right) - \frac{1}{2} n^{\frac{1}{2}} K_\varepsilon.$$

From Equation 3.50 further follows that

$$\sum_{i=1}^{n} E\left(I_{\left\{u_i^2 \leq U_{\alpha,n}^{(\varepsilon)}\right\}}\right) = \sum_{i=1}^{n} P(u_i^2 \leq U_{\alpha,n}^{(\varepsilon)}) = \sum_{i=1}^{n} G_i(U_{\alpha,n}^{(\varepsilon)}) =$$

$$= n \cdot \frac{\sum_{i=1}^{n} G_i(U_{\alpha,n}^{(\varepsilon)})}{n} = n \cdot G(U_{\alpha,n}^{(\varepsilon)}).$$

Then we can write

$$\sum_{i=1}^{n} I_{\left\{u_i^2 \leq U_{\alpha,n}^{(\varepsilon)}\right\}} > n \cdot G(U_{\alpha,n}^{(\varepsilon)}) - \frac{1}{2} n^{\frac{1}{2}} K_\varepsilon =$$

$$= n \cdot \left[G(U_{\alpha,n}^{(\varepsilon)}) - G(q_\alpha^2)\right] + n \cdot G(q_\alpha^2) - \frac{1}{2} n^{\frac{1}{2}} K_\varepsilon =$$

$$= n^{\frac{1}{2}} K_\varepsilon + n(1-\alpha) - \frac{1}{2} n^{\frac{1}{2}} K_\varepsilon = n(1-\alpha) + \frac{1}{2} n^{\frac{1}{2}} K_\varepsilon > \ell_n,$$

where the last inequality follows from the way, how ℓ_n was defined. Recall that we have put $\ell_n = [(1-\alpha)n]_{int}$.

Therefore there are at least ℓ_n squared error terms for which it holds with probability at least $1-\frac{1}{2}\varepsilon$ that $u_i^2 \leq U_{\alpha,n}^{(\varepsilon)}$. It follows that the ℓ_n-th order statistic is lower than $U_{\alpha,n}^{(\varepsilon)}$ with probability at least $1 - \frac{1}{2}\varepsilon$ (uniformly for all $\alpha \in A_n^{(\varepsilon)}$). This means that we have shown that Equation 3.43 holds also for $\alpha \in A_n^{(\varepsilon)}$. Equation 3.42 for $\alpha \in A_n^{(\varepsilon)}$ follows from definition of $U_{\alpha,n}^{(\varepsilon)}$, Equation 3.44 for $\alpha \in A_n^{(\varepsilon)}$ follows from definition of $U_{\alpha,n}^{(\varepsilon)}$, Assumptions 3.6 and Lemma 3.14.

To conclude the first part of the proof it remains to show that Equation 3.45 also holds. Let us recall Assumptions 3.1, specifically that $E(u_i) = 0$ and $Var(u_i) = \int z^2 f_{u_i}(z)\mathrm{d}z = \sigma_i^2 < b^2 < \infty$. Therefore the function $z^2 f_{u_i}(z)$ is bounded, say by C. Then we can employ Hölder´s inequality to obtain

$$\int_{z^2 \in \left[q_\alpha^2, U_{\alpha,n}^{(\varepsilon)}\right]} |z| f_{u_i}(z)\mathrm{d}z = \int_{z^2 \in \left[q_\alpha^2, U_{\alpha,n}^{(\varepsilon)}\right]} \left\{|z|\sqrt{f_{u_i}(z)}\right\} \cdot \left\{\sqrt{f_{u_i}(z)}\right\} \mathrm{d}z \leq$$

$$\leq \left\{\int_{z^2 \in \left[q_\alpha^2, U_{\alpha,n}^{(\varepsilon)}\right]} z^2 f_{u_i}(z)\mathrm{d}z \cdot \int_{z^2 \in \left[q_\alpha^2, U_{\alpha,n}^{(\varepsilon)}\right]} f_{u_i}(z)\mathrm{d}z\right\}^{\frac{1}{2}} \leq$$

$$\leq \left\{ C \cdot \int_{z^2 \in \left[q_\alpha^2, U_{\alpha,n}^{(\varepsilon)}\right]} f_{u_i}(z) \mathrm{d}z \cdot P\left(u_i^2 \in \left[q_\alpha^2, U_{\alpha,n}^{(\varepsilon)}\right]\right) \right\}^{\frac{1}{2}} =$$

$$= C^{\frac{1}{2}} \cdot P\left(u_i^2 \in \left[q_\alpha^2, U_{\alpha,n}^{(\varepsilon)}\right]\right) < \frac{1}{2} n^{-\frac{1}{2}} \tilde{K}_\varepsilon.$$

The second part of the proof is similar to the first one. We can analogically show that for any $\varepsilon \in (0,1)$ there are constants $K'_\varepsilon < \infty$ and $\tilde{K}'_\varepsilon < \infty$ such that for all $n > n_\varepsilon$ and any $\alpha \in (0,1)$ there exists $L_{\alpha,n}^{(\varepsilon)}$ such that

$$q_\alpha^2 \geq L_{\alpha,n}^{(\varepsilon)} \quad \text{for any} \quad \alpha \in (0,1),$$

$$P\left(\bigcap_{\alpha \in (0,1)} \left\{ \omega \in \Omega \ : \ u_{(\ell_n)}^2 \geq L_{\alpha,n}^{(\varepsilon)} \right\} \right) > 1 - \frac{1}{2}\varepsilon,$$

and for all $i = 1, 2, ..., n$ we have

$$\sup_{\alpha \in (0,1)} P\left(u_i^2 \in \left[L_{\alpha,n}^{(\varepsilon)}, q_\alpha^2\right]\right) < \frac{1}{2} n^{-\frac{1}{2}} K'_\varepsilon$$

and

$$\sup_{\alpha \in (0,1)} E\left[|u_i| \cdot I_{\left\{L_{\alpha,n}^{(\varepsilon)} \leq u_i^2 \leq q_\alpha^2\right\}} \right] \leq \sup_{\alpha \in (0,1)} \int_{z^2 \in \left[L_{\alpha,n}^{(\varepsilon)}, q_\alpha^2\right]} |z| f_{u_i}(z) \mathrm{d}z < \frac{1}{2} n^{-\frac{1}{2}} \tilde{K}'_\varepsilon.$$

Then we can put $I_{\alpha,n}^{(\varepsilon)} = \left(L_{\alpha,n}^{(\varepsilon)}, U_{\alpha,n}^{(\varepsilon)}\right)$, which concludes the proof.

Q.E.D.

Let us further denote $[a,b]_{ord} = [\min\{a,b\}, \max\{a,b\}]$ for any $a, b \in \mathbb{R}$. Then we can use Lemma 3.15 to show the following.

Lemma 3.16. *Let the assumptions of Lemma 3.15 hold. Then for any $\varepsilon > 0$ there is a constant $K_\varepsilon < \infty$ and $n_\varepsilon \in \mathbb{N}$ such that for all $n > n_\varepsilon$ and any $\alpha \in (0,1)$ there exists some set B_n such that $P(B_n) > 1 - \varepsilon$,*

$$\left\{ \left[u_{(\ell_n)}^2, q_\alpha^2\right]_{ord} \cap B_n \right\} \subset \left\{ I_{\alpha,n}^{(\varepsilon)} \cap B_n \right\}$$

and

$$\sup_{\alpha \in (0,1)} P\left(u_i^2 \in \left\{ \left[u_{(\ell_n)}^2, q_\alpha^2\right]_{ord} \cap B_n \right\}\right) \leq n^{-\frac{1}{2}} K_\varepsilon.$$

PROOF After adjusting the proof of Lemma 3.15, this lemma can be proven in the same way as in Víšek (2015). Let us state the proof also here for completeness. Fix some $\varepsilon \in (0,1)$ and use Lemma 3.15 to find a constant $K_\varepsilon < \infty$ and $n_\varepsilon \in \mathbb{N}$ such that for all $n > n_\varepsilon$ and any $\alpha \in (0,1)$ we have $q_\alpha^2 \in I_{\alpha,n}^\varepsilon$, the set

$$B_n = \bigcap_{\alpha\in(0,1)} \left\{\omega \in \Omega \ : \ u^2_{(\ell_n)} \in I^\varepsilon_{\alpha,n}\right\}$$

satisfies $P(B_n) > 1 - \varepsilon$ and

$$\sup_{\alpha\in(0,1)} P\left(u_i^2 \in \left\{I^\varepsilon_{\alpha,n} \cap B_n\right\}\right) \leq n^{-\frac{1}{2}} K_\varepsilon,$$

see Equation 3.39, Equation 3.40 and Equation 3.41. Then realize that we have $\left\{\left[u^2_{(\ell_n)}, q^2_\alpha\right]_{ord} \cap B_n\right\} \subset \left\{I^{(\varepsilon)}_{\alpha,n} \cap B_n\right\}$ and therefore

$$\sup_{\alpha\in(0,1)} P\left(u_i^2 \in \left\{\left[u^2_{(\ell_n)}, q^2_\alpha\right]_{ord} \cap B_n\right\}\right) \leq \sup_{\alpha\in(0,1)} P\left(u_i^2 \in \left\{I^{(\varepsilon)}_{\alpha,n} \cap B_n\right\}\right) \leq n^{-\frac{1}{2}} K_\varepsilon$$

which concludes the proof.

Q.E.D.

Lemma 3.17. *Let $\{u_i\}_{i=1}^\infty$, where $u_i \in R$, be a sequence of independent random variables with absolutely continuous distribution functions $F_{u_i}(z) = F_u(z\sigma_i^{-1})$ as specified in the Assumptions 3.1. Then fix some $\delta \in (0,1)$. Finally, for some $\Delta = \Delta(q_\delta^2) \in (0,\infty)$ let*

$$\inf_{z\in(0,u_\delta^2+\Delta)} g(z) > L_g > 0. \tag{3.53}$$

Then for all $\varepsilon \in (0,1)$ there exists a constant $K^{(\varepsilon,\delta)} < \infty$ and $n_{\varepsilon,\delta} \in N$ such that for all $n > n_{\varepsilon,\delta}$ we obtain

$$P\left(\sup_{\alpha\in(\delta,1)} \left|u^2_{(\ell_n(\alpha))} - q^2_\alpha\right| < n^{-\frac{1}{2}} K^{(\varepsilon,\delta)}\right) > 1 - \varepsilon.$$

PROOF To prove this lemma, some adjustments for heteroskedasticity are again necessary as compared to Víšek (2015). We can divide the proof into two parts, split the absolute value, and show that for all $\varepsilon \in (0,1)$ there exists a constant $K_\varepsilon < \infty$ such that we have

$$P\left(\sup_{\alpha\in(\delta,1)} u^2_{(\ell_n)} - q^2_\alpha < n^{-\frac{1}{2}} K_\varepsilon\right) > 1-\varepsilon, \tag{3.54}$$

where we again write ℓ_n instead of $\ell_n(\alpha)$ for simplicity. Then similarly as before choose and fix some $\varepsilon \in (0,1)$, let $W(s)$ be a Wiener process (see again Definition 3.2) and find $K_1 < \infty$ such that

$$P\left(\sup_{0\le s\le 2} |W(s)| > \frac{1}{2}K_1\right) < \frac{1}{2}\varepsilon. \tag{3.55}$$

We can put $K^{(\varepsilon)} = K_1 L_g^{-1}$, $n_\Delta = 2\cdot[K^{(\varepsilon)}\Delta^{-1}]_{int} + 1$ and further consider only $n > n_\Delta$. Then when we denote

$$v_i^{(\alpha)} = I_{\left\{u_i^2 \le q_\alpha^2 + n^{-\frac{1}{2}}K^{(\varepsilon)}\right\}} - E\left(I_{\left\{u_i^2 \le q_\alpha^2 + n^{-\frac{1}{2}}K^{(\varepsilon)}\right\}}\right),$$

we can write

$$a_{n,i} = E\left(I_{\left\{u_i^2 \le q_\alpha^2 + n^{-\frac{1}{2}}K^{(\varepsilon)}\right\}}\right) = P\left(u_i^2 \le q_\alpha^2 + n^{-\frac{1}{2}}K^{(\varepsilon)}\right),$$

and we have $v_i^{(\alpha)} = 1 - a_{n,i} > 0$ if $u_i^2 \le q_\alpha^2 + n^{-\frac{1}{2}}K^{(\varepsilon)}$ and $v_i^{(\alpha)} = -a_{n,i} < 0$ otherwise. Let $\tau_{in}^{(\alpha)}$ be the first time when $W_i(s)$ exits the interval $(-a_{n,i}, 1-a_{n,i})$ and apply Lemma 3.4 to obtain

$$v_i^{(\alpha)} =_D W_i(\tau_{in}^{(\alpha)}),$$

and similarly as in the proofs of previous lemmas

$$n^{-\frac{1}{2}}\sum_{i=1}^n v_i^{(\alpha)} =_D n^{-\frac{1}{2}}\sum_{i=1}^n W_i(\tau_{in}^{(\alpha)}) =_D W\left(n^{-1}\sum_{i=1}^n \tau_{in}^{(\alpha)}\right), \tag{3.56}$$

where $W(s)$ is again a Wiener process, such that $W(s)$ is independent from $\mathcal{W} = \{W_i(s)\}_{i=1}^\infty$. Further, let U_i be the first time when $W_i(s)$ exits the interval $(-1,1)$ and note that $\{U_i\}_{i=1}^\infty$ is a sequence of iid random variables. Then applying again the last part of Lemma 3.4, we obtain $E(U_i) = 1$. Since

$(-a_{n,i}, 1 - a_{n,i}) \subset (-1, 1)$, we have

$$n^{-1} \sum_{i=1}^{n} \tau_{in}^{(\alpha)} \leq n^{-1} \sum_{i=1}^{n} U_i \tag{3.57}$$

for every $\alpha \in (\delta, 1)$. Moreover, let us put

$$B_n = \left\{ \omega \in \Omega \ : \ n^{-1} \sum_{i=1}^{n} U_i > 2 \right\}.$$

Then for the ε that we have fixed earlier we can find $n_\varepsilon > n_\Delta$ such that for any $n > n_\varepsilon$ we have

$$P(B_n) < \frac{1}{2}\varepsilon, \tag{3.58}$$

and similarly as in the proof of Lemma 3.15 we can use Equation 3.55, Equation 3.56, Equation 3.57, Equation 3.58 and Lemma 3.3 (realizing that the processes $v_i^{(\alpha)}$ are again separable) to obtain

$$P\left(\left\{ n^{-\frac{1}{2}} \sup_{\alpha \in (\delta,1)} \left| \sum_{i=1}^{n} v_i^{(\alpha)} \right| > \frac{1}{2} K_1 \right\}\right) \leq$$

$$\leq P\left(\left\{ n^{-\frac{1}{2}} \sup_{\alpha \in (\delta,1)} \left| \sum_{i=1}^{n} v_i^{(\alpha)} \right| > \frac{1}{2} K_1 \right\} \cap B_n^c \right) + P\left(B_n\right) \leq$$

$$\leq P\left(\left\{ n^{-\frac{1}{2}} \sup_{\alpha \in (\delta,1)} \left| \sum_{i=1}^{n} W_i(\tau_{in}^{(\alpha)}) \right| > \frac{1}{2} K_1 \right\} \cap B_n^c \right) + \frac{1}{2}\varepsilon =$$

$$= P\left(\left\{ \sup_{\alpha \in (\delta,1)} \left| W\left(n^{-1} \sum_{i=1}^{n} \tau_{in}^{(\alpha)} \right) \right| > \frac{1}{2} K_1 \right\} \cap B_n^c \right) + \frac{1}{2}\varepsilon \leq$$

$$\leq P\left(\left\{ \sup_{0 \leq s \leq 2} |W(s)| > \frac{1}{2} K_1 \right\} \cap B_n^c \right) + \frac{1}{2}\varepsilon < \varepsilon.$$

Define the set $D_n = \left\{ \omega \in \Omega \ : \ n^{-\frac{1}{2}} \sup_{\alpha \in (\delta,1)} \left| \sum_{i=1}^{n} v_i^{(\alpha)} \right| < \frac{1}{2} K_1 \right\}$. We have just derived that $P(D_n) \geq 1 - \varepsilon$ and therefore we have for all $\alpha \in (\delta, 1)$ with probability at least $1-\varepsilon$

$$\sum_{i=1}^{n} I_{\left\{u_i^2 \leq q_\alpha^2 + n^{-\frac{1}{2}} K^{(\varepsilon)}\right\}} > \sum_{i=1}^{n} E\left(I_{\left\{u_i^2 \leq q_\alpha^2 + n^{-\frac{1}{2}} K^{(\varepsilon)}\right\}}\right) - \frac{1}{2} n^{\frac{1}{2}} K_1.$$

Moreover, we can write

$$\sum_{i=1}^{n} E\left(I_{\left\{u_i^2 \leq q_\alpha^2 + n^{-\frac{1}{2}} K^{(\varepsilon)}\right\}}\right) = \sum_{i=1}^{n} P\left(u_i^2 \leq q_\alpha^2 + n^{-\frac{1}{2}} K^{(\varepsilon)}\right) =$$

$$= \sum_{i=1}^{n} G_i\left(q_\alpha^2 + n^{-\frac{1}{2}} K^{(\varepsilon)}\right) = n \cdot G\left(q_\alpha^2 + n^{-\frac{1}{2}} K^{(\varepsilon)}\right) \geq n(1-\alpha) + n^{\frac{1}{2}} K^{(\varepsilon)} L_g =$$

$$= n(1-\alpha) + n^{\frac{1}{2}} K_1$$

and therefore we arrive at

$$\sum_{i=1}^{n} I_{\left\{u_i^2 \leq q_\alpha^2 + n^{-\frac{1}{2}} K^{(\varepsilon)}\right\}} > n(1-\alpha) + n^{\frac{1}{2}} K_1 - \frac{1}{2} n^{\frac{1}{2}} K_1 = n(1-\alpha) + \frac{1}{2} n^{\frac{1}{2}} K_1 > \ell_n,$$

where the last inequality follows again from the way, how ℓ_n was defined. Recall that we have put $\ell_n = [(1-\alpha)n]_{int}$.

Therefore there are at least ℓ_n squared error terms for which it holds with probability at least $1-\varepsilon$ that $u_i^2 \leq q_\alpha^2 + n^{-\frac{1}{2}} K^{(\varepsilon)}$. It follows that the ℓ_n-th order statistic is lower than $q_\alpha^2 + n^{-\frac{1}{2}} K^{(\varepsilon)}$ with probability at least $1-\varepsilon$ (uniformly for all $\alpha \in (\delta, 1)$), which proves Equation 3.54.

The other part of the proof would be done analogically by considering

$$\tilde{v}_i^{(\alpha)} = I_{\left\{u_i^2 \geq q_\alpha^2 + n^{-\frac{1}{2}} K^{(\varepsilon)}\right\}} - E\left(I_{\left\{u_i^2 \geq q_\alpha^2 + n^{-\frac{1}{2}} K^{(\varepsilon)}\right\}}\right)$$

instead of $v_i^{(\alpha)}$.

Q.E.D.

Now recall Equation 3.5 and let us denote the density of $\bar{F}_{n,\beta^0}(r)$ by $\bar{f}_n(r)$, i.e. we have

$$\bar{F}_{n,\beta^0}(r) = \frac{1}{n} \sum_{i=1}^{n} P(|u_i| < r) = \frac{1}{n} \sum_{i=1}^{n} \left(F_u\left(\sigma_i^{-1} r\right) - F_u\left(-\sigma_i^{-1} r\right)\right)$$

and

$$\bar{f}_n(r) = \frac{1}{n}\sum_{i=1}^{n} \sigma_i^{-1}\left(f_u\left(\sigma_i^{-1}r\right) + f_u\left(-\sigma_i^{-1}r\right)\right).$$

Then analogically to Lemma 3.17 we can prove Lemma 3.18.

Lemma 3.18. *Let $\{u_i\}_{i=1}^{\infty}$, where $u_i \in R$, be a sequence of independent random variables with absolutely continuous distribution functions $F_{u_i}(z) = F_u(z\sigma_i^{-1})$ as specified in the Assumptions 3.1. Then fix some $\delta \in (0,1)$. Finally, for some $\Delta = \Delta(q_\delta^2) \in (0,\infty)$ and any $n \in \mathbb{N}$ let*

$$\inf_{r\in(0,\sqrt{u_\delta^2+\Delta})} \bar{f}_n(r) > L_f > 0. \tag{3.59}$$

Then for all $\varepsilon \in (0,1)$ there exists a constant $K^{(\varepsilon,\delta)} < \infty$ and $n_{\varepsilon,\delta} \in \mathbb{N}$ such that for all $n > n_{\varepsilon,\delta}$ we obtain

$$P\left(\sup_{\alpha\in(\delta,1)} \left|\sqrt{u^2_{(\ell_n(\alpha))}} - q_\alpha\right| < n^{-\frac{1}{2}}K^{(\varepsilon,\delta)}\right) > 1-\varepsilon.$$

PROOF The proof is analogical to the proof of Lemma 3.17.

Q.E.D.

To be able to derive the asymptotic representation, we further need to know the probability that the i-th variable in a sequence of n independent variables is equal to the ℓ-th order statistic. For homoskedastic case this probability is equal to $\frac{1}{n}$, as was shown in previous publications. The proof for iid variables can be found also in Víšek (2015). In what follows, it is sufficient to know the upper bound of this probability for the case where the variables are not identically distributed. Let us find it in the following lemma.

Lemma 3.19. *Let $\{u_i\}_{i=1}^{\infty}$, where $u_i \in \mathbb{R}$, be a sequence of independent random variables with absolutely continuous distribution functions $H_i(z)$. Then for all $n \in \mathbb{N}$, $i = 1, ..., n$ and $\ell = 1, ..., n$ we have*

$$P\left(u_i^2 = u_{(\ell)}^2\right) \le \frac{c}{n},$$

where c is some constant satisfying $c \ge 1$.

PROOF Let us realize that we can write

$$P\left(u_i^2 = u_{(\ell)}^2\right) = E_{u_i}\left(P\left(u_{(\ell)}^2 = z\right)|u_i^2 = z\right).$$

Then we want to find the probability that ℓ of the variables are smaller than z. Let us denote for any k, where $0 \leq k \leq n$ the set of indices $1 \leq i_1, i_2, ..., i_k \leq n$ by $I\{i_1, i_2, .., i_k\}$. Then the required probability is given by

$$\sum_{I\{i_1,i_2,..,i_k\}} H_{i_1}(z) \cdot H_{i_2}(z) \cdot ... \cdot H_{i_k}(z) \times$$

$$\times \left[1 - H_{i_{k+1}}(z)\right] \cdot \left[1 - H_{i_{k+2}}(z)\right] \cdot ... \cdot \left[1 - H_{i_n}(z)\right],$$

where $\sum_{I\{i_1,i_2,..,i_k\}}$ is a sum over all k-tuples ($\ell \leq k \leq n$) of indices from $\{1, 2, ..., n\}$ and we can notice the analogy to the probability for homoskedastic case in Equation 3.11 .Therefore we want to bound the following expression

$$\int_0^\infty \sum_{I\{i_1,i_2,..,i_k\}} H_{i_1}(z) \cdot H_{i_2}(z) \cdot ... \cdot H_{i_k}(z) \times$$

$$\times \left[1 - H_{i_{k+1}}(z)\right] \cdot \left[1 - H_{i_{k+2}}(z)\right] \cdot ... \cdot \left[1 - H_{i_n}(z)\right] h_{i_j}(z) \mathrm{d}z.$$

We can notice that since we do not need the exact value, only the upper bound, we can consider the maximal and minimal distributions, multiplied by some constant $c \geq 1$. Moreover, when considering these extreme distributions, the sum in the integral becomes the same binomial coefficient as we would have for the homoskedastic case. We can further bound the density $h_{i_j}(z)$ by a density corresponding to the maximal distribution function. Then we obtain following expression

$$c \cdot \int_0^\infty \frac{(n-1)!}{(\ell-1)!(n-\ell)!} H_{max}^{\ell-1}(z)(1 - H_{min}(z))^{n-\ell} \cdot h_{max}(z) \mathrm{d}z \qquad (3.60)$$

and we can integrate it by parts. Recall that to integrate by parts we can use the formula $\int f'(z)g(z)\mathrm{d}z = f(z)g(z) - \int f(z)g'(z)\mathrm{d}z$. Here we have

$$f'(z) = \frac{(n-1)!}{(\ell-1)!(n-\ell)!} H_{max}^{\ell-1}(z) \cdot h_{max}(z)$$

$$f(z) = \frac{(n-1)!}{\ell!(n-\ell)!} H_{max}^{\ell}(z)$$

$$g(z) = (1 - H_{min}(z))^{n-\ell}$$

$$g'(z) = -(n-\ell)(1 - H_{min}(z))^{n-\ell-1} h_{min}(z).$$

Therefore when we integrate Equation 3.60 by parts, we arrive at

$$c \cdot \left[\frac{(n-1)!}{\ell!(n-\ell)!} H^{\ell}_{max}(z) \cdot (1 - H_{min}(z))^{n-\ell} \right]_0^{\infty} +$$

$$+ c \cdot \int_0^{\infty} \frac{(n-1)!}{\ell!(n-\ell)!} H^{\ell}_{max}(z) \cdot (n-\ell)(1 - H_{min}(z))^{n-\ell-1} h_{min}(z) \mathrm{d}z, \qquad (3.61)$$

where the first summand becomes 0 and to bound the second summand we can replace $h_{min}(z)$ by $h_{max}(z)$. After some rearranging, we can conclude that Equation 3.61 can be bounded by

$$c \cdot \int_0^{\infty} \frac{(n-1)!}{\ell!(n-\ell-1)!} H^{\ell}_{max}(z) \cdot (1 - H_{min}(z))^{n-\ell-1} h_{max}(z) \mathrm{d}z.$$

We can integrate it in the same manner $(n-\ell)$-times and we arrive at

$$P\left(u_i^2 = u_{(\ell)}^2\right) \leq c \cdot \int_0^{\infty} \frac{(n-1)!}{(n-1)!} H^{n-1}_{max}(z) h_{max}(z) \mathrm{d}z = c \cdot \left[\frac{1}{n} H^{n}_{max}(z) \right]_0^{\infty} = \frac{c}{n},$$

which concludes the proof. Note that computations for a concrete distribution (specifically, results for exponential distribution were obtained) also support this result.

Q.E.D.

In what follows, we will also need to use Lemma 3.7 and therefore also Lemma 3.6. These lemmas and corresponding proofs are formulated for iid variables. However, we can show that we are not able to distinguish the EDF of the homoskedastic residuals from the EDF of heteroskedastic residuals in probability. This result follows from Lemma 3.20. Then we will be able to use Lemma 3.6 and Lemma 3.7 also for the heteroskedastic case in the proof of Lemma 3.21.

Lemma 3.20. *Let $\{u_i\}_{i=1}^{\infty}$ be a sequence of independent random variables with EDF $F^{(n)}_{u_i}(v)$ and $\{u_i^*\}_{i=1}^{\infty}$ a sequence of iid random variables with EDF $F^{(n)}_{u_i^*}(v)$. Then we have*

$$\sup_{v \in \mathbb{R}} \sqrt{n} \left| F^{(n)}_{u_i}(v) - F^{(n)}_{u_i^*}(v) \right| = O_p(1).$$

PROOF To conduct this proof we can use the result from Lemma 3.8. Notice that Lemma 3.8 is proven under the assumption of heteroskedasticity. It implies that it also holds under homoskedasticity, i.e. the EDF of the homoskedastic as well as heteroskedastic residuals are close in probability to the mean theoretical distribution function. In mathematical terms we have

$$\sup_{v\in\mathbb{R}} \sqrt{n} \left| F_{u_i}^{(n)}(v) - \bar{F}_n(v) \right| = O_p(1)$$

and

$$\sup_{v\in\mathbb{R}} \sqrt{n} \left| F_{u_i^*}^{(n)}(v) - \bar{F}_n(v) \right| = O_p(1).$$

Therefore we can write

$$\sup_{v\in\mathbb{R}} \sqrt{n} \left| F_{u_i}^{(n)}(v) - F_{u_i^*}^{(n)}(v) \right| = \sup_{v\in\mathbb{R}} \sqrt{n} \left| F_{u_i}^{(n)}(v) - \bar{F}_n(v) + \bar{F}_n(v) - F_{u_i^*}^{(n)}(v) \right| \leq$$

$$\leq \sup_{v\in\mathbb{R}} \sqrt{n} \left| F_{u_i}^{(n)}(v) - \bar{F}_n(v) \right| + \sup_{v\in\mathbb{R}} \sqrt{n} \left| F_{u_i^*}^{(n)}(v) - \bar{F}_n(v) \right| = O_p(1).$$

Q.E.D.

Taking into account Lemma 3.19 and Lemma 3.20, we can generalize for heteroskedasticity another lemma from Víšek (2015). The adjusted lemma and corresponding proof then look as follows.

Lemma 3.21. *Let $\{u_i\}_{i=1}^{\infty}$ be a sequence of independent random variables with absolutely continuous distribution functions $F_{u_i}(x) = F_u(x\sigma_i^{-1})$ with corresponding densities $f_{u_i}(x) = f_u(x\sigma_i^{-1})\sigma_i^{-1}$, as specified in Assumptions 3.1. Further, let $f_u(z)$ be bounded by some finite constant (say by U_f). Moreover, let it be uniformly locally Lipschitz of the first order in z, i.e. there exists a constant $K_f < \infty$ and $\tau > 0$ such that for any $z_1, z_2 \in \mathbb{R}$, $|z_1 - z_2| < \tau$, we have*

$$|f_u(z_1) - f_u(z_2)| \leq K_f \cdot |z_1 - z_2| \, .$$

Finally, let us fix some $\delta \in (0,1)$ and let for $\Delta = \Delta(q_\delta^2) \in (0,\infty)$

$$\inf_{z\in(0,q_\delta^2+\Delta)} g(z) > L_g > 0$$

and

$$\inf_{r\in(0,\sqrt{q_\delta^2+\Delta})} \bar{f}_n(r) > L_f > 0$$

as in Equation 3.53 and Equation 3.59. Then for all $\varepsilon > 0$ *there exists a constant* $K^{(\varepsilon,\delta)} < \infty$ *and* $n_{\varepsilon,\delta} \in \mathbb{N}$ *such that for all* $n > n_{\varepsilon,\delta}$ *there exists a set* D_n *such that* $P(D_n) > 1 - \varepsilon$,

$$\max_{1\le i\le n} \sup_{\alpha\in(\delta,1)} \left| E\left\{ I_{\{u_i>0\}} \cdot \left[I_{\left\{u_i^2 \le u^2_{(\ell_n(\alpha))}\right\}} - I_{\left\{u_i^2 \le q^2_{1-\frac{\ell_n(\alpha)}{n}}\right\}} \right] \cdot I_{\{D_n\}} \right\} \right| < \\ < n^{-1} K^{(\varepsilon,\delta)} \tag{3.62}$$

and

$$\max_{1\le i\le n} \sup_{\alpha\in(\delta,1)} \left| E\left\{ I_{\{u_i<0\}} \cdot \left[I_{\left\{u_i^2 \le u^2_{(\ell_n(\alpha))}\right\}} - I_{\left\{u_i^2 \le q^2_{1-\frac{\ell_n(\alpha)}{n}}\right\}} \right] \cdot I_{\{D_n\}} \right\} \right| < \\ < n^{-1} K^{(\varepsilon,\delta)}. \tag{3.63}$$

PROOF First of all note that if $u_i = 0$, the expectation would be equal to 0. Then let us consider Equation 3.62 and fix some $\varepsilon > 0$. Moreover, denote

$$B_n^{(1)} = \left\{ \sup_{\alpha\in(\delta,1)} \left| \sqrt{u^2_{(\ell_n(\alpha))}} - q_\alpha \right| < n^{-\frac{1}{2}} \cdot K_{\varepsilon,\delta} \right\} \tag{3.64}$$

and

$$B_n^{(2)} = \left\{ \sup_{\alpha\in(\delta,1)} \left| u^2_{(\ell_n(\alpha))} - q^2_\alpha \right| < n^{-\frac{1}{2}} \cdot K_{\varepsilon,\delta} \right\}$$

and put $C_n = B_n^{(1)} \cap B_n^{(2)}$. We can employ Lemma 3.17 and Lemma 3.18 to find $K_{\varepsilon,\delta} < \infty$ and $n_{\varepsilon,\delta} \in \mathbb{N}$ such that for all $n > n_{\varepsilon,\delta}$ we have $P(C_n) > 1 - \varepsilon$.

In what follows, let us write in short ℓ instead of $\ell_n(\alpha)$ for clarity and denote $D_{i,\ell,n} = \left\{\omega \in \Omega : u_i(\omega) = u_{(\ell)}(\omega)\right\}$. Lemma 3.19 implies that $P(D_{i,\ell,n})$ is bounded by $\frac{c}{n}$, where $c \ge 1$ is some constant. When we realize that

$$\left| I_{\{u_i>0\}} \cdot \left[I_{\left\{u_i^2 \le u^2_{(\ell)}\right\}} - I_{\left\{u_i^2 \le q^2_{1-\frac{\ell}{n}}\right\}} \right] \cdot I_{\{C_n\}} \right| \le 1,$$

we obtain

$$E\left(\left| I_{\{u_i>0\}} \cdot \left[I_{\left\{u_i^2 \le u^2_{(\ell)}\right\}} - I_{\left\{u_i^2 \le q^2_{1-\frac{\ell}{n}}\right\}} \right] \cdot I_{\{C_n\}} \cdot I_{\left\{D_{i,\ell,n}\right\}} \right| \right) \le E\left(I_{\left\{D_{i,\ell,n}\right\}} \right) \le \frac{c}{n}$$

and therefore

$$\left| E\left\{ I_{\{u_i>0\}} \cdot \left[I_{\left\{u_i^2 \le u_{(\ell)}^2\right\}} - I_{\left\{u_i^2 \le q_{1-\frac{\ell}{n}}^2\right\}} \right] \cdot I_{\{C_n\}} \right\} \right| =$$

$$= \left| E\left\{ I_{\{u_i>0\}} \cdot \left[I_{\left\{u_i^2 \le u_{(\ell)}^2\right\}} - I_{\left\{u_i^2 \le q_{1-\frac{\ell}{n}}^2\right\}} \right] \cdot I_{\{C_n\}} \cdot I_{\{D_{i,\ell,n}^c\}} \right\} + \right.$$

$$\left. + E\left\{ I_{\{u_i>0\}} \cdot \left[I_{\left\{u_i^2 \le u_{(\ell)}^2\right\}} - I_{\left\{u_i^2 \le q_{1-\frac{\ell}{n}}^2\right\}} \right] \cdot I_{\{C_n\}} \cdot I_{\{D_{i,\ell,n}\}} \right\} \right| \le$$

$$\le \left| E\left\{ I_{\{u_i>0\}} \cdot \left[I_{\left\{u_i^2 \le u_{(\ell)}^2\right\}} - I_{\left\{u_i^2 \le q_{1-\frac{\ell}{n}}^2\right\}} \right] \cdot I_{\{C_n\}} \cdot I_{\{D_{i,\ell,n}^c\}} \right\} \right| + \frac{c}{n}. \qquad (3.65)$$

Let us further recall that $\sqrt{u_i^2} = |u_i|$. In what follows we will denote the squared root of the ℓ-th order statistic of the squared disturbances a little non-tradionally by $|u|_\ell$, i.e. we have $\sqrt{u_\ell^2} = |u|_\ell$. Then we can write

$$\left| E\left\{ I_{\{u_i>0\}} \cdot \left[I_{\left\{u_i^2 \le u_{(\ell)}^2\right\}} - I_{\left\{u_i^2 \le q_{1-\frac{\ell}{n}}^2\right\}} \right] \cdot I_{\{C_n\}} \cdot I_{\{D_{i,\ell,n}^c\}} \right\} \right| =$$

$$= \left| E_{|u|_\ell} \left(E\left\{ I_{\{u_i>0\}} \cdot \left[I_{\left\{u_i^2 \le u_{(\ell)}^2\right\}} - I_{\left\{u_i^2 \le q_{1-\frac{\ell}{n}}^2\right\}} \right] \cdot I_{\{C_n\}} \cdot I_{\{D_{i,\ell,n}^c\}} \,\middle|\, |u|_\ell = z \right\} \right) \right|.$$

Let us now find the upper bound of

$$E\left\{ I_{\{u_i>0\}} \cdot \left[I_{\left\{u_i^2 \le u_{(\ell)}^2\right\}} - I_{\left\{u_i^2 \le q_{1-\frac{\ell}{n}}^2\right\}} \right] \cdot I_{\{C_n\}} \cdot I_{\{D_{i,\ell,n}^c\}} \,\middle|\, |u|_\ell = z \right\}. \qquad (3.66)$$

First of all, let us realize that we can consider only $z \in \left(0, q_\delta + n^{-\frac{1}{2}} \cdot K_{\varepsilon,\delta}\right)$ when we evaluate this expected value, as there is $I_{\{C_n\}}$ present in the integral. To show that, notice that we have for all $n > n_{\varepsilon,\delta}$ and $\omega \in C_n$

$$\left| \, |u|_{(\ell)} - q_{1-\frac{\ell}{n}} \right| = \left| \sqrt{e_{(\ell)}^2} - u_{1-\frac{\ell}{n}} \right| < n^{-\frac{1}{2}} \cdot K_{\varepsilon,\delta}$$

and therefore

$$\left(|u|_{(\ell)}, q_{1-\frac{\ell}{n}} \right)_{ord} = \left(\sqrt{u_{(\ell)}^2}, q_{1-\frac{\ell}{n}} \right)_{ord} \subset \left(0, q_{1-\frac{\ell}{n}} + n^{-\frac{1}{2}} \cdot K_{\varepsilon,\delta} \right).$$

As q_α^2 is defined as the upper α-quantile of $G(z)$, we have $q_\alpha < q_\delta$ for any $\alpha \in (\delta, 1)$. Recall the assumptions of this lemma and define $b = q_\delta + \Delta(q_\delta^2)$.

Then we can find $\tilde{n}_{\varepsilon,\delta}$ such that for all $n > \tilde{n}_{\varepsilon,\delta}$ and all $\alpha \in (\delta, 1)$we have

$$\left(\sqrt{u^2_{(\ell)}}, q_{1-\frac{\ell}{n}}\right)_{ord} \subset (0, b).$$

Moreover, we can see that

$$I_{\{u_i>0\}} \cdot \left[I_{\left\{u_i^2 \le u^2_{(\ell)}\right\}} - I_{\left\{u_i^2 \le q^2_{1-\frac{\ell}{n}}\right\}} \right] \cdot I_{\{C_n\}} \cdot I\{D^c_{i,\ell,n}\} = 1 \tag{3.67}$$

if and only if $q_{1-\frac{\ell}{n}} < u_i < \sqrt{u^2_{(\ell)}}$ and

$$I_{\{u_i>0\}} \cdot \left[I_{\left\{u_i^2 \le u^2_{(\ell)}\right\}} - I_{\left\{u_i^2 \le q^2_{1-\frac{\ell}{n}}\right\}} \right] \cdot I_{\{C_n\}} \cdot I\{D^c_{i,\ell,n}\} = -1 \tag{3.68}$$

if and only if $\sqrt{u^2_{(\ell)}} < u_i \le q_{1-\frac{\ell}{n}}$.

To be able to find the upper bound of Equation 3.66, we need to know the conditional density of u_i given $\sqrt{u^2_{(\ell)}}$. Let us denote it $f_{u_i \big| |u|_{(\ell)}}(v \mid |u|_{(\ell)} = z)$. Note that to derive this conditional density we need to use a different method than was used in Víšek (2015) for homoskedasticity, where the derivation was based mainly on the proof of Lemma 3.6. Although we can again see some similarities to Equation 3.11 and Equation 3.12 when deriving the density of $|u|_{(\ell)}$, the idea here is somewhat different.

We can realize that as the expectation in Equation 3.66 includes $I_{\{D^c_{i,\ell,n}\}}$, we are restricted to the case where $\left\{\omega \in \Omega : u_i(\omega) \neq u_{(\ell)}(\omega)\right\}$. It means that u_i and $u_{(\ell)}$ are independent and therefore $f_{u_i \big| |u|_{(\ell)}}(v \mid |u|_{(\ell)} = z) = f_{u_i}(v)$. This implies that if $f_{u_i}(v)$ is Lipschitz, then the conditional density is Lipschitz as well, which is what we need in what follows. However, we can show the independence of u_i and $u_{(\ell)}$ also formally.

By definition, the conditional density is a ratio of the joint density of the two variables in question and the marginal density of $|u|_{(\ell)}$. Denote $H_i(z)$ the distribution function of $|u_i|$ and $h_i(z)$ the corresponding density. Let us first find the marginal density of $|u|_{(\ell)}$.

Similarly as in the proof of Lemma 3.19, we need to find the probability that at least ℓ of the variables are smaller than z. Let us denote for any k, where $0 \le k \le n$ the set of indices $1 \le i_1, i_2, ..., i_k \le n$ by $I\{i_1, i_2, .., i_k\}$. Then the required probability is given by

$$\sum_{k=\ell}^{n} \sum_{I\{i_1,i_2,..,i_k\}} H_{i_1}(z) \cdot H_{i_2}(z) \cdot ... \cdot H_{i_k}(z) \times$$

$$\times \left[1 - H_{i_{k+1}}(z)\right] \cdot \left[1 - H_{i_{k+2}}(z)\right] \cdot ... \cdot \left[1 - H_{i_n}(z)\right],$$

where $\sum_{I\{i_1,i_2,..,i_k\}}$ is a sum over all k-tuples ($\ell \leq k \leq n$) of indices from $\{1, 2, ..., n\}$. Let us enlarge the notation in a way that $I_{i_j}\{i_1, i_2, .., i_k\}$ represents the set of indices $1 \leq i_1, i_2, ..., i_{j-1}, i_{j+1}, ..., i_k \leq n$ for $\ell \leq k \leq n$, where i_j is the index of the random variable whose distribution function $H_{i_j}(z)$ is derivated to obtain $h_{i_j}(z)$. Then the density of $|u|_{(\ell)}$ is given by

$$\sum_{k=\ell}^{n} \sum_{I_{i_j}\{i_1,i_2,..,i_k\}} H_{i_1}(z) \cdot H_{i_2}(z) \cdot ... \cdot H_{i_{j-1}}(z) \cdot H_{i_{j+1}}(z) \cdot ... \cdot H_{i_k}(z) \times$$

$$\times \left[1 - H_{i_{k+1}}(z)\right] \cdot \left[1 - H_{i_{k+2}}(z)\right] \cdot ... \cdot \left[1 - H_{i_n}(z)\right] \cdot h_{i_j}(z), \quad (3.69)$$

where $\sum_{I_{i_j}\{i_1,i_2,..,i_k\}}$ stays for the sum over all $(k-1)$-tuples ($\ell \leq k \leq n$) of indices from $\{1, 2, ..., n\}$ and all j, $1 \leq j \leq k$.

We can further find the joint density of u_i and $\sqrt{u^2_{(\ell)}}$. Recall that in this part of the proof we consider only the case when $u_i > 0$ (due to the presence of $I_{\{u_i>0\}}$ in Equation 3.66). Therefore we will find the probability that $0 < u_i < z'$ and at the same time $\sqrt{u^2_{(\ell)}} < z$. We can consider the cases when $z < z'$ and when $z' < z$ separately. Let us find the joint density for $z' < z$, the density for the other case would be derived analogically.

Let us realize that $0 < u_i < z'$ has probability $F_{u_i}(z') - F_{u_i}(0)$. Moreover, since $u_i < z' < z$, we already have one u_i such that $|u_i|$ is smaller than z. So that we obtain $\sqrt{u^2_{(\ell)}} < z$, we must find at least $\ell - 1$ other such u_i's. Similarly as before, we need to find at least $k \geq \ell - 1$ out of the remaining $n-1$ variables that will be smaller than z. The probability of this event is given by

$$\sum_{k=\ell-1}^{n-1} \sum_{I\{i_1,i_2,..,i_{k-1}\}} H_{i_1}(z) \cdot H_{i_2}(z) \cdot ... \cdot H_{i_{k-1}}(z) \times$$

$$\times \left[1 - H_{i_{k+1}}(z)\right] \cdot \left[1 - H_{i_{k+2}}(z)\right] \cdot ... \cdot \left[1 - H_{i_{n-1}}(z)\right],$$

where we have assumed that $i_k = i$, i.e. the observation with the index i_k is the one that we already know is smaller than z. Then the joint distribution function of interest is given by

$$[F_{u_i}(z') - F_{u_i}(0)] \cdot \sum_{k=\ell-1}^{n-1} \sum_{I\{i_1,i_2,..,i_{k-1}\}} H_{i_1}(z) \cdot H_{i_2}(z) \cdot \ldots \cdot H_{i_{k-1}}(z) \times$$

$$\times \left[1 - H_{i_{k+1}}(z)\right] \cdot \left[1 - H_{i_{k+2}}(z)\right] \cdot \ldots \cdot \left[1 - H_{i_{n-1}}(z)\right]$$

and the corresponding joint density is given by

$$\sum_{k=\ell-1}^{n-1} \sum_{I_{i_j}\{i_1,i_2,..,i_{k-1}\}} H_{i_1}(z) \cdot H_{i_2}(z) \cdot \ldots \cdot H_{i_{j-1}}(z) \cdot H_{i_{j+1}}(z) \cdot \ldots \cdot H_{i_{k-1}}(z) \times$$

$$\times \left[1 - H_{i_{k+1}}(z)\right] \cdot \left[1 - H_{i_{k+2}}(z)\right] \cdot \ldots \cdot \left[1 - H_{i_{n-1}}(z)\right] \cdot h_{i_j}(z) \cdot f_{u_i}(z'). \quad (3.70)$$

It follows that the conditional density that we are looking for is given as a ratio of Equation 3.70 and Equation 3.69. Note that as $H_i(z) \in (0,1)$ for every i and we can assume $h_i(z)$ to be bounded and non-zero, we can conclude that $f_{u_i\big| |u|_{(\ell)}}(v \mid |u|_{(\ell)} = z) = C \cdot f_{u_i}(v)$ with $C < \infty$ being some normalizing constant. This would be sufficient for the conditional density to be Lipschitz. However, it can be shown that $C = 1$ and therefore the variables are independent. It suffices to realize that $f_{u_i}(v)$ is a density and hence $\int_{-\infty}^{\infty} f_{u_i}(v)\mathrm{d}v = 1$.

Then recall Equation 3.68 and we can evaluate Equation 3.66 for the case when $\sqrt{u_{(\ell)}^2} = z < q_{1-\frac{\ell}{n}}$ as

$$E\left\{ I_{\{u_i>0\}} \cdot \left[I_{\left\{u_i^2 \le u_{(\ell)}^2\right\}} - I_{\left\{u_i^2 \le q_{1-\frac{\ell}{n}}^2\right\}} \right] \cdot I_{\{C_n\}} \cdot I_{\{D_{i,\ell,n}^c\}} \,\Big|\, |u|_\ell = z \right\} =$$

$$= -\int_z^{q_{1-\frac{\ell}{n}}} f_{u_i\big| |u|_{(\ell)}}(z' \mid |e|_{(\ell)} = z)\mathrm{d}z' = -\int_z^{q_{1-\frac{\ell}{n}}} f_{u_i\big| |u|_{(\ell)}}(q_{1-\frac{\ell}{n}} \mid |u|_{(\ell)} = z)\mathrm{d}z' -$$

$$-\int_z^{q_{1-\frac{\ell}{n}}} \left\{ f_{u_i\big| |u|_{(\ell)}}(z' \mid |u|_{(\ell)} = z) - f_{u_i\big| |u|_{(\ell)}}(q_{1-\frac{\ell}{n}} \mid |u|_{(\ell)} = z) \right\} \mathrm{d}z'$$

$$= -f_{u_i\big| |u|_{(\ell)}}(q_{1-\frac{\ell}{n}} \mid |u|_{(\ell)} = z)\left(q_{1-\frac{\ell}{n}} - z\right) + R_{n_1}(z). \quad (3.71)$$

As we derived that the conditional density is Lipschitz of the first order, we have

$$\left| f_{u_i\big| |u|_{(\ell)}}(z' \mid |u|_{(\ell)} = z) - f_{u_i\big| |u|_{(\ell)}}(q_{1-\frac{\ell}{n}} \mid |u|_{(\ell)} = z) \right| \le$$

$$\le \tilde{K}_f \cdot \left| z' - q_{1-\frac{\ell}{n}} \right| \le \tilde{K}_f \cdot \left| z - q_{1-\frac{\ell}{n}} \right|,$$

where the last inequality follows from the fact that $z' \in \left(z, q_{1-\frac{\ell}{n}}\right)$. Then we obtain

$$|R_{n_1}(z)| \leq \tilde{K}_f \cdot \left|z - q_{1-\frac{\ell}{n}}\right| \int_z^{q_{1-\frac{\ell}{n}}} 1 \mathrm{d}z' \leq \tilde{K}_f \cdot \left[q_{1-\frac{\ell}{n}} - z\right]^2.$$

When we further recall Equation 3.64, we find that for all $n > n_{\varepsilon,\delta}$ and for all $\omega \in C_n$

$$|R_{n_1}(z)| \leq n^{-1} \cdot \tilde{K}_f \cdot K^2_{\varepsilon,\delta}. \tag{3.72}$$

Analogically we can recall Equation 3.67 and evaluate Equation 3.66 for the case when $\sqrt{u^2_{(\ell)}} = z > q_{1-\frac{\ell}{n}}$ as

$$E\left\{ I_{\{u_i>0\}} \cdot \left[I_{\left\{u_i^2 \leq u^2_{(\ell)}\right\}} - I_{\left\{u_i^2 \leq q^2_{1-\frac{\ell}{n}}\right\}} \right] \cdot I_{\{C_n\}} \cdot I\{D^c_{i,\ell,n}\} \,\Big|\, |u|_\ell = z \right\} =$$

$$= f_{u_i \big| |u|_{(\ell)}}(q_{1-\frac{\ell}{n}} \mid |u|_{(\ell)} = z) \left(q_{1-\frac{\ell}{n}} - z\right) + R_{n_2}(z). \tag{3.73}$$

Similarly as before we obtain for all $n > n_{\varepsilon,\delta}$ and for all $\omega \in C_n$

$$|R_{n_2}(z)| \leq n^{-1} \cdot \tilde{K}_f \cdot K^2_{\varepsilon,\delta}. \tag{3.74}$$

Then we can employ Lemma 3.7 to find a constant $K^*_{\varepsilon,\delta} < \infty$ and $n'_{\varepsilon,\delta} > n_{\varepsilon,\delta}$ such that for all $n > n'_{\varepsilon,\delta}$ the density of $\hat{q}_{1-\frac{\ell}{n}} = \sqrt{u^2_{(\ell)}}$ can be written as

$$h_{n,\alpha}(q) = h^*_{n,\alpha}(q) + \rho_{n,\alpha}(q),$$

where

$$\sup_{\alpha \in (\delta,1)} \sup_{|q| \leq K_{\varepsilon,\delta}} |\rho_{n,\alpha}(q)| \leq n^{-\frac{1}{2}} \cdot K^*_{\varepsilon,\delta},$$

where recalling the expressions in Lemma 3.7 we put $K = K^*_{\varepsilon,\delta}$ and we can notice that the density $h^*_{n,\alpha}(q)$ is (uniformly in $n \in \mathbb{N}$, $q \in \mathbb{R}$ and in $\alpha \in (0,1)$) bounded by some constant $U_h < \infty$. Then we can write

$$E_{|u|_\ell}\left(E\left\{ I_{\{u_i>0\}} \cdot \left[I_{\left\{u_i^2 \leq u^2_{(\ell)}\right\}} - I_{\left\{u_i^2 \leq q^2_{1-\frac{\ell}{n}}\right\}} \right] \cdot I_{\{C_n\}} \cdot I\{D^c_{i,\ell,n}\} \,\Big|\, |u|_\ell = z \right\}\right) =$$

$$= -\int_{q_{1-\frac{\ell}{n}} - n^{-\frac{1}{2}} \cdot K_{\varepsilon,\delta}}^{q_{1-\frac{\ell}{n}}} \left[f_{u_i \big| |u|_{(\ell)}}(q_{1-\frac{\ell}{n}} \mid |u|_{(\ell)} = z) \left(q_{1-\frac{\ell}{n}} - z\right) + R_{n_1}(z) \right] \times$$

$$\times \left[h^*_{n,\alpha}(z) + \rho_{n,\alpha}(z) \right] \mathrm{d}z +$$

$$+\int_{q_{1-\frac{\ell}{n}}}^{q_{1-\frac{\ell}{n}}+n^{-\frac{1}{2}}\cdot K_{\varepsilon,\delta}} \left[f_{u_i\big| |u|_{(\ell)}}(q_{1-\frac{\ell}{n}} \mid |u|_{(\ell)}=z)\left(z-q_{1-\frac{\ell}{n}}\right)+R_{n_2}(z)\right]\times$$

$$\times\left[h^*_{n,\alpha}(z)+\rho_{n,\alpha}(z)\right]\mathrm{d}z.$$

It follows that for any pair of z^*, z^{**}, where $z^* \in \left[q_{1-\frac{\ell}{n}} - n^{-\frac{1}{2}} \cdot K_{\varepsilon,\delta}, q_{1-\frac{\ell}{n}}\right]$ and $z^{**} \in \left[q_{1-\frac{\ell}{n}}, q_{1-\frac{\ell}{n}} + n^{-\frac{1}{2}} \cdot K_{\varepsilon,\delta}\right]$, satisfying $q_{1-\frac{\ell}{n}} - z^* = z^{**} - q_{1-\frac{\ell}{n}}$, we obtain

$$f_{u_i\big| |u|_{(\ell)}}(q_{1-\frac{\ell}{n}} \mid |u|_{(\ell)}=z)\left(q_{1-\frac{\ell}{n}}-z^*\right)\cdot h^*_{n,\alpha}(z^*)=$$

$$=f_{u_i\big| |u|_{(\ell)}}(q_{1-\frac{\ell}{n}} \mid |u|_{(\ell)}=z)\left(z^{**}-q_{1-\frac{\ell}{n}}\right)\cdot h^*_{n,\alpha}(z^{**}).$$

Then we can combine Equation 3.71, Equation 3.72, Equation 3.73 and Equation 3.74 to obtain

$$\left|E_{|u|_\ell}\left(E\left\{I_{\{u_i>0\}}\cdot\left[I_{\left\{u_i^2\le u^2_{(\ell)}\right\}}-I_{\left\{u_i^2\le q^2_{1-\frac{\ell}{n}}\right\}}\right]\cdot I_{\{C_n\}}\cdot I\{D^c_{i,\ell,n}\}\;\Big|\;|u|_\ell=z\right\}\right)\right|\le$$

$$\le\left|\int_{q_{1-\frac{\ell}{n}}-n^{-\frac{1}{2}}K_{\varepsilon,\delta}}^{q_{1-\frac{\ell}{n}}} f_{u_i\big| |u|_{(\ell)}}(q_{1-\frac{\ell}{n}} \mid |u|_{(\ell)}=z)\left(q_{1-\frac{\ell}{n}}-z\right)\cdot\rho_{n,\alpha}(z)\mathrm{d}z\right|+$$

$$+\left|\int_{q_{1-\frac{\ell}{n}}-n^{-\frac{1}{2}}K_{\varepsilon,\delta}}^{q_{1-\frac{\ell}{n}}} R_{n_1}(z)\cdot\left[U_h+\rho_{n,\alpha}(z)\right]\mathrm{d}z\right|+$$

$$+\left|\int_{q_{1-\frac{\ell}{n}}}^{q_{1-\frac{\ell}{n}}+n^{-\frac{1}{2}}K_{\varepsilon,\delta}} f_{u_i\big| |u|_{(\ell)}}(q_{1-\frac{\ell}{n}} \mid |u|_{(\ell)}=z)\left(q_{1-\frac{\ell}{n}}-z\right)\cdot\rho_{n,\alpha}(z)\mathrm{d}z\right|+$$

$$+\left|\int_{q_{1-\frac{\ell}{n}}}^{q_{1-\frac{\ell}{n}}+n^{-\frac{1}{2}}K_{\varepsilon,\delta}} R_{n_2}(z)\cdot\left[U_h+\rho_{n,\alpha}(z)\right]\mathrm{d}z\right|\le$$

$$\le\left\{U_f\cdot n^{-\frac{1}{2}}\cdot K_{\varepsilon,\delta}\cdot n^{-\frac{1}{2}}\cdot K^*_{\varepsilon,\delta}+n^{-1}\cdot\tilde{K}_f\cdot K^2_{\varepsilon,\delta}\cdot\left[U_h+n^{-\frac{1}{2}}\cdot K^*_{\varepsilon,\delta}\right]\right\}\times$$

$$\times\left\{\int_{q_{1-\frac{\ell}{n}}-n^{-\frac{1}{2}}K_{\varepsilon,\delta}}^{q_{1-\frac{\ell}{n}}} 1\mathrm{d}z+\int_{q_{1-\frac{\ell}{n}}}^{q_{1-\frac{\ell}{n}}+n^{-\frac{1}{2}}K_{\varepsilon,\delta}} 1\mathrm{d}z\right\}=$$

$$=\left\{U_f\cdot n^{-1}\cdot K_{\varepsilon,\delta}\cdot K^*_{\varepsilon,\delta}+n^{-1}\cdot\tilde{K}_f\cdot K^2_{\varepsilon,\delta}\cdot\left[U_h+n^{-\frac{1}{2}}\cdot K^*_{\varepsilon,\delta}\right]\right\}\times$$

$$\times\left\{2\cdot n^{-\frac{1}{2}}K_{\varepsilon,\delta}\right\}\le n^{-\frac{3}{2}}\cdot C',$$

where C' is a finite constant. Therefore we can conclude that

$$\sup_{\alpha\in(\delta,1)} \left| E_{|u|_\ell} \left(E \left\{ I_{\{u_i>0\}} \cdot \left[I_{\left\{u_i^2 \le u_{(\ell)}^2\right\}} - I_{\left\{u_i^2 \le q^2_{1-\frac{\ell}{n}}\right\}} \right] \times \right.\right.\right.$$

$$\left.\left.\left. \times I_{\{C_n\}} \cdot I\{D^c_{i,\ell,n}\} \,\Big|\, |u|_\ell = z \right\} \right) \right| \le n^{-\frac{3}{2}} \cdot C'$$

and together with Equation 3.65 this proves Equation 3.62. Equation 3.63 can be proven analogically. This concludes the proof.

Q.E.D.

Using all the previous results, we can now prove the last lemma needed to find the asymptotic representation of the LWS estimator. Notice that in what follows, we write $\left(\hat{\beta}^{(LWS,n,w)} - \beta^0\right)$, although we assumed $\beta^0 = 0$. This is to obtain the usual form of the asymptotic representation.

Lemma 3.22. *Let Assumptions 3.1, 3.2, 3.3, 3.4, 3.5, 3.6 and 3.7 hold. Then we can show that*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} w\left(\bar{F}_{n,\beta^0}(|u_i|)\right) \cdot X_i u_i =$$

$$= \frac{1}{n} \sum_{i=1}^{n} w\left(\bar{F}_{n,\beta^0}(|u_i|)\right) \cdot X_i X_i' \cdot \left\{ \sqrt{n} \left(\hat{\beta}^{(LWS,n,w)} - \beta^0 \right) \right\} + o_p(1).$$

PROOF After adjusting all the previoius lemmas for heteroskedasticity, the proof of this lemma can be done almost in the same way as it was done in Víšek (2015). Only a few more adjustements are necessary. First of all, recall Equation 3.38. Then we can write

$$\frac{1}{\sqrt{n}} \sum_{\ell=1}^{n} w_\ell^* \sum_{i=1}^{n} X_i \left(Y_i - X_i'\beta^0\right) I_{\left\{u_i^2 \le u_{(\ell)}^2\right\}} = \tag{3.75}$$

$$= \frac{1}{\sqrt{n}} \sum_{\ell=1}^{n} w_\ell^* \sum_{i=1}^{n} X_i X_i' \left(\hat{\beta}^{(LWS,n,w)} - \beta^0 \right) \cdot I_{\left\{u_i^2 \le u_{(\ell)}^2\right\}} + o_p(1), \tag{3.76}$$

where Equation 3.75 can be rewritten as

$$\frac{1}{\sqrt{n}} \sum_{\ell=1}^{n} w_\ell^* \sum_{i=1}^{n} X_i u_i \cdot \left[I_{\left\{u_i^2 \le u_{(\ell)}^2\right\}} - I_{\left\{u_i^2 \le q^2_{1-\frac{\ell}{n}}\right\}} \right] + \tag{3.77}$$

$$+ \frac{1}{\sqrt{n}} \sum_{\ell=1}^{n} w_\ell^* \sum_{i=1}^{n} X_i u_i \cdot I_{\left\{u_i^2 \le q^2_{1-\frac{\ell}{n}}\right\}}.$$

We will first consider the case when $j = 2, ..., p$, i.e. for now we do not take the intercept into account. Then using Chebyshev inequality we obtain for the expression in Equation 3.77 and any $\varepsilon > 0$

$$P\left(\frac{1}{\sqrt{n}}\left|\sum_{\ell=1}^{n} w_\ell^* \sum_{i=1}^{n} X_{ij}u_i \cdot \left[I_{\left\{u_i^2 \le u_{(\ell)}^2\right\}} - I_{\left\{u_i^2 \le q^2_{1-\frac{\ell}{n}}\right\}}\right]\right| > \varepsilon\right) \le$$

$$\le \frac{1}{\varepsilon^2 n} E\left(\left\{\sum_{\ell=1}^{n} w_\ell^* \sum_{i=1}^{n} X_{ij}u_i \cdot \left[I_{\left\{u_i^2 \le u_{(\ell)}^2\right\}} - I_{\left\{u_i^2 \le q^2_{1-\frac{\ell}{n}}\right\}}\right]\right\}^2\right) =$$

$$= \frac{1}{\varepsilon^2 n} E\left(\left\{\sum_{i=1}^{n} X_{ij}u_i \cdot \sum_{\ell=1}^{n} w_\ell^* \left[I_{\left\{u_i^2 \le u_{(\ell)}^2\right\}} - I_{\left\{u_i^2 \le q^2_{1-\frac{\ell}{n}}\right\}}\right]\right\}^2\right). \tag{3.78}$$

Further realize that $E(X_{ij}) = 0$ for all $j = 2, ..., p$ and that X_i is independent from u_i (see Assumptions 3.1). Therefore for any $j = 2, ..., p$ and for any pair i, k, such that $i = 1, ..., n$, $k = 1, ..., n$, $i \neq k$ we obtain

$$E\left\{X_{ij}u_i \cdot \sum_{\ell=1}^{n} w_\ell^* \left[I_{\left\{u_i^2 \le u_{(\ell)}^2\right\}} - I_{\left\{u_i^2 \le q^2_{1-\frac{\ell}{n}}\right\}}\right] \times \right.$$

$$\left. \times X_{kj}u_k \cdot \sum_{\ell=1}^{n} w_\ell^* \left[I_{\left\{u_k^2 \le u_{(\ell)}^2\right\}} - I_{\left\{u_k^2 \le q^2_{1-\frac{\ell}{n}}\right\}}\right]\right\} =$$

$$= E\left(X_{ij}\right) \cdot E\left(X_{kj}\right) \cdot E\left\{u_i \cdot \sum_{\ell=1}^{n} w_\ell^* \left[I_{\left\{u_i^2 \le u_{(\ell)}^2\right\}} - I_{\left\{u_i^2 \le q^2_{1-\frac{\ell}{n}}\right\}}\right] \times \right.$$

$$\left. \times u_k \cdot \sum_{\ell=1}^{n} w_\ell^* \left[I_{\left\{u_k^2 \le u_{(\ell)}^2\right\}} - I_{\left\{u_k^2 \le q^2_{1-\frac{\ell}{n}}\right\}}\right]\right\} = 0.$$

It follows that Equation 3.78 is further equal to

$$\frac{1}{\varepsilon^2 n} \sum_{i=1}^{n} E(X_{ij}^2) \cdot E\left(\left\{u_i \cdot \sum_{\ell=1}^{n} w_\ell^* \left[I_{\left\{u_i^2 \le u_{(\ell)}^2\right\}} - I_{\left\{u_i^2 \le q^2_{1-\frac{\ell}{n}}\right\}}\right]\right\}^2\right) =$$

$$= \frac{1}{\varepsilon^2 n} \sum_{i=1}^{n} E(X_{ij}^2) \cdot E\left\{u_i^2 \cdot \sum_{\ell=1}^{n} w_\ell^* \cdot \left[I_{\left\{u_i^2 \le u_{(\ell)}^2\right\}} - I_{\left\{u_i^2 \le q^2_{1-\frac{\ell}{n}}\right\}}\right] \times \right.$$

$$\left. \times \sum_{m=1}^{n} w_m^* \cdot \left[I_{\left\{u_i^2 \le u_{(m)}^2\right\}} - I_{\left\{u_i^2 \le q^2_{1-\frac{m}{n}}\right\}}\right]\right\} \le$$

$$\leq \frac{1}{\varepsilon^2 n} \sum_{i=1}^{n} E(X_{ij}^2) \cdot E \left\{ u_i^2 \cdot \sum_{\ell=1}^{n} w_\ell^* \cdot \left| I_{\left\{u_i^2 \leq u_{(\ell)}^2\right\}} - I_{\left\{u_i^2 \leq q_{1-\frac{\ell}{n}}^2\right\}} \right| \times \right.$$

$$\left. \times \sum_{m=1}^{n} w_m^* \cdot \left| I_{\left\{u_i^2 \leq u_{(m)}^2\right\}} - I_{\left\{u_i^2 \leq q_{1-\frac{m}{n}}^2\right\}} \right| \right\}. \tag{3.79}$$

Moreover, let us realize that $\left| I_{\left\{u_i^2 \leq u_{(m)}^2\right\}} - I_{\left\{u_i^2 \leq q_{1-\frac{m}{n}}^2\right\}} \right| \leq 1$ and simultaneously $\sum_{m=1}^{n} w_m^* = 1$. Therefore we have $\sum_{m=1}^{n} w_m^* \cdot \left| I_{\left\{u_i^2 \leq u_{(m)}^2\right\}} - I_{\left\{u_i^2 \leq q_{1-\frac{m}{n}}^2\right\}} \right| \leq 1$ and the expression after the last inequality in Equation 3.79 can be bounded by

$$\frac{1}{\varepsilon^2 n} \sum_{i=1}^{n} E(X_{ij}^2) \cdot E \left\{ u_i^2 \cdot \sum_{\ell=1}^{n} w_\ell^* \cdot \left| I_{\left\{u_i^2 \leq u_{(\ell)}^2\right\}} - I_{\left\{u_i^2 \leq q_{1-\frac{\ell}{n}}^2\right\}} \right| \right\} =$$

$$= \frac{1}{\varepsilon^2 n} \sum_{\ell=1}^{n} w_\ell^* \sum_{i=1}^{n} E(X_{ij}^2) \cdot E \left\{ u_i^2 \cdot \left| I_{\left\{u_i^2 \leq u_{(\ell)}^2\right\}} - I_{\left\{u_i^2 \leq q_{1-\frac{\ell}{n}}^2\right\}} \right| \right\}. \tag{3.80}$$

Now we can notice that as u_i^2 is always positive, we can use Hölder´s inequality. Therefore the expression inEquation 3.80 can be further bounded by

$$\frac{1}{\varepsilon^2 n} \sum_{\ell=1}^{n} w_\ell^* \sum_{i=1}^{n} E(X_{ij}^2) \left\{ E\left(|u_i|^{2q'} \right) \right\}^{\frac{1}{q'}} \left\{ E\left(\left| I_{\left\{u_i^2 \leq u_{(\ell)}^2\right\}} - I_{\left\{u_i^2 \leq q_{1-\frac{\ell}{n}}^2\right\}} \right|^{q''} \right) \right\}^{\frac{1}{q''}} \leq$$

$$\leq \sup_i \frac{1}{\varepsilon^2} \sum_{\ell=1}^{n} w_\ell^* E(X_{ij}^2) \cdot \left\{ E\left(|u_i|^{2q'} \right) \right\}^{\frac{1}{q'}} \times$$

$$\times \left\{ E \left| I_{\left\{u_i^2 \leq u_{(\ell)}^2\right\}} - I_{\left\{u_i^2 \leq q_{1-\frac{\ell}{n}}^2\right\}} \right| \right\}^{\frac{1}{q''}}. \tag{3.81}$$

For q' see the Assumptions 3.7 and note that if $\frac{1}{q'} + \frac{1}{q''} = 1$ and $q' > 1$, then also $q'' > 1$. Moreover, realize that

$$\left| I_{\left\{u_i^2 \leq u_{(\ell)}^2\right\}} - I_{\left\{u_i^2 \leq q_{1-\frac{\ell}{n}}^2\right\}} \right|^{q''} = \left| I_{\left\{u_i^2 \leq u_{(\ell)}^2\right\}} - I_{\left\{u_i^2 \leq q_{1-\frac{\ell}{n}}^2\right\}} \right|.$$

We can further employ Lemma 3.15 to find a constant $K < \infty$ and $n' \in \mathbb{N}$

such that for all $n > n'$ we have

$$\sup_i E\left|I_{\left\{u_i^2 \le u_{(\ell)}^2\right\}} - I_{\left\{u_i^2 \le q^2_{1-\frac{\ell}{n}}\right\}}\right| < n^{-\frac{1}{2}}K.$$

This result, together with the Assumptions 3.7, implies that Equation 3.81 can be bounded by

$$\sup_i K^{\frac{1}{q''}} \varepsilon^{-2} n^{-\frac{1}{2q''}} \sum_{\ell=1}^{n} w_\ell^* E(X_{1j}^2) \cdot \left\{E\left(|u_i|^{2q'}\right)\right\}^{\frac{1}{q'}} = O_p(n^{-\frac{1}{2q''}}).$$

It remains to consider the case when $j = 1$, i.e. we consider the intercept. As we have $X_{i1} = 1$ for all i, we can rewrite the expression in Equation 3.77 as

$$\frac{1}{\sqrt{n}} \sum_{\ell=1}^{n} w_\ell^* \sum_{i=1}^{n} u_i \cdot \left[I_{\left\{u_i^2 \le u_{(\ell)}^2\right\}} - I_{\left\{u_i^2 \le q^2_{1-\frac{\ell}{n}}\right\}}\right]. \tag{3.82}$$

Let us show that it also is bounded in probability. We can employ Lemma 3.15 to find a constant $\tilde{K}^{(\varepsilon)} < \infty$ and $n' \in \mathbb{N}$ such that for all $n > n'$ and all $\ell = 1, ..., n$ there exists an interval $I^{\varepsilon}_{1-\frac{\ell}{n},n}$ such that

$$q^2_{1-\frac{\ell}{n}} \in I^{\varepsilon}_{1-\frac{\ell}{n},n},$$

$$P\left(\bigcap_{\ell=1,\dots,n} \left\{\omega \in \Omega \ : \ u_{(\ell)}^2 \in I^{\varepsilon}_{1-\frac{\ell}{n},n}\right\}\right) > 1 - \frac{\varepsilon}{4},$$

and for all $i = 1, 2, ..., n$ we have

$$E\left[|u_i| \cdot I^{(\varepsilon)}_{1-\frac{\ell}{n},n}\right] \le n^{-\frac{1}{2}} \tilde{K}^{(\varepsilon)}.$$

Denote the set $\left\{\bigcap_{\ell=1,\dots,n} \left\{\omega \in \Omega \ : \ u_{(\ell)}^2 \in I^{\varepsilon}_{1-\frac{\ell}{n},n}\right\}\right\}$ by B_n. Then we have for any $\omega \in B_n$

$$\left[u_{(\ell)}^2, q^2_{1-\frac{\ell}{n},n}\right]_{ord} \subset I^{\varepsilon}_{1-\frac{\ell}{n},n}. \tag{3.83}$$

Let us further choose and fix some $\varepsilon > 0$ and some $\theta > 0$. Then we can find $\alpha_0 \in (0, 1)$ and $n_0 \in N$ such that $w(\frac{n_0-1}{n_0} - \alpha_0) \le \frac{\varepsilon \cdot \theta}{8 \cdot \tilde{K}^{(\varepsilon)}}$. Moreover, for all $n > n_0$ we have $\frac{n_0-1}{n_0} < \frac{n-1}{n}$ and hence $w(\frac{n_0-1}{n_0} - \alpha_0) > w(\frac{n-1}{n} - \alpha_0)$, and therefore

$$\sum_{\ell=\ell_n(\alpha_0)}^{n} w_\ell^* = w\left(\frac{[(1-\alpha_0)n]_{int}-1}{n}\right) \leq w\left(\frac{n-1}{n}-\alpha_0\right) \leq \frac{\varepsilon\cdot\theta}{8\cdot\tilde{K}^{(\varepsilon)}}$$

and we can split Equation 3.82 in two parts, so that we arrive at

$$\frac{1}{\sqrt{n}}\sum_{\ell=1}^{\ell_n(\alpha_0)} w_\ell^* \sum_{i=1}^{n} u_i\cdot\left[I_{\left\{u_i^2\leq u_{(\ell)}^2\right\}} - I_{\left\{u_i^2\leq q^2_{1-\frac{\ell}{n}}\right\}}\right]+ \tag{3.84}$$

$$+\frac{1}{\sqrt{n}}\sum_{\ell=\ell_n(\alpha_0)}^{n} w_\ell^* \sum_{i=1}^{n} u_i\cdot\left[I_{\left\{u_i^2\leq u_{(\ell)}^2\right\}} - I_{\left\{u_i^2\leq q^2_{1-\frac{\ell}{n}}\right\}}\right]. \tag{3.85}$$

We can show that both of these summands are small in probability. Let us consider the summand in Equation 3.85 first. We can write

$$P\left(\frac{1}{\sqrt{n}}\left|\sum_{\ell=\ell_n(\alpha_0)}^{n} w_\ell^* \sum_{i=1}^{n} u_i\cdot\left[I_{\left\{u_i^2\leq u_{(\ell)}^2\right\}} - I_{\left\{u_i^2\leq q^2_{1-\frac{\ell}{n}}\right\}}\right]\right| > \theta\right) =$$

$$= P\left(\frac{1}{\sqrt{n}}\left|\sum_{\ell=\ell_n(\alpha_0)}^{n} w_\ell^* \sum_{i=1}^{n} u_i\cdot\left[I_{\left\{u_i^2\leq u_{(\ell)}^2\right\}} - I_{\left\{u_i^2\leq q^2_{1-\frac{\ell}{n}}\right\}}\right]\cdot\left[I_{\{B_n\}}+I_{\{B_n^c\}}\right]\right| > \theta\right) \leq$$

$$\leq P\left(\frac{1}{\sqrt{n}}\left|\sum_{\ell=\ell_n(\alpha_0)}^{n} w_\ell^* \sum_{i=1}^{n} u_i\cdot\left[I_{\left\{u_i^2\leq u_{(\ell)}^2\right\}} - I_{\left\{u_i^2\leq q^2_{1-\frac{\ell}{n}}\right\}}\right]\cdot I_{\{B_n\}}\right| > \frac{\theta}{2}\right)+$$

$$+P\left(\frac{1}{\sqrt{n}}\left|\sum_{\ell=\ell_n(\alpha_0)}^{n} w_\ell^* \sum_{i=1}^{n} u_i\cdot\left[I_{\left\{u_i^2\leq u_{(\ell)}^2\right\}} - I_{\left\{u_i^2\leq q^2_{1-\frac{\ell}{n}}\right\}}\right]\cdot I_{\{B_n^c\}}\right| > \frac{\theta}{2}\right) \leq$$

$$\leq P\left(\frac{1}{\sqrt{n}}\left|\sum_{\ell=\ell_n(\alpha_0)}^{n} w_\ell^* \sum_{i=1}^{n} u_i\cdot\left[I_{\left\{u_i^2\leq u_{(\ell)}^2\right\}} - I_{\left\{u_i^2\leq q^2_{1-\frac{\ell}{n}}\right\}}\right]\cdot I_{\{B_n\}}\right| > \frac{\theta}{2}\right)+\frac{\varepsilon}{4} \leq$$

$$\leq P\left(\frac{1}{\sqrt{n}}\sum_{\ell=\ell_n(\alpha_0)}^{n} w_\ell^* \sum_{i=1}^{n} |u_i|\cdot\left|I_{\left\{u_i^2\leq u_{(\ell)}^2\right\}} - I_{\left\{u_i^2\leq q^2_{1-\frac{\ell}{n}}\right\}}\right|\cdot I_{\{B_n\}} > \frac{\theta}{2}\right)+\frac{\varepsilon}{4}.$$

As $\left|I_{\left\{u_i^2\leq u_{(\ell)}^2\right\}} - I_{\left\{u_i^2\leq q^2_{1-\frac{\ell}{n}}\right\}}\right| = 1$ if and only if $u_i^2 \in \left[u_{(\ell)}^2, q^2_{1-\frac{\ell}{n},n}\right]_{ord}$, due to

Equation 3.83 we can conclude that

$$P\left(\frac{1}{\sqrt{n}}\sum_{\ell=\ell_n(\alpha_0)}^{n} w_\ell^* \sum_{i=1}^{n} |u_i| \cdot \left| I_{\left\{u_i^2 \le u_{(\ell)}^2\right\}} - I_{\left\{u_i^2 \le q_{1-\frac{\ell}{n}}^2\right\}} \right| \cdot I_{\{B_n\}} > \frac{\theta}{2}\right) \le$$

$$\le P\left(\frac{1}{\sqrt{n}}\sum_{\ell=\ell_n(\alpha_0)}^{n} w_\ell^* \sum_{i=1}^{n} |u_i| \cdot I_{1-\frac{\ell}{n},n}^{\varepsilon} \cdot I_{\{B_n\}} > \frac{\theta}{2}\right).$$

Moreover, when we use Markov inequality, we find that this can be further bounded by

$$E\left(\frac{2}{\theta\sqrt{n}}\sum_{\ell=\ell_n(\alpha_0)}^{n} w_\ell^* \sum_{i=1}^{n} |u_i| \cdot I_{1-\frac{\ell}{n},n}^{\varepsilon} \cdot I_{\{B_n\}}\right) \le$$

$$\le \frac{2}{\theta\sqrt{n}}\sum_{\ell=\ell_n(\alpha_0)}^{n} w_\ell^* \sum_{i=1}^{n} E\left(|u_i| \cdot I_{1-\frac{\ell}{n},n}^{\varepsilon}\right) \le \frac{2}{\theta n}\sum_{\ell=\ell_n(\alpha_0)}^{n} w_\ell^* \sum_{i=1}^{n} \tilde{K}^{(\varepsilon)} =$$

$$= \frac{2n\tilde{K}^{(\varepsilon)}}{\theta n}\sum_{\ell=\ell_n(\alpha_0)}^{n} w_\ell^* \le \frac{\varepsilon}{4}$$

and we can conclude that

$$P\left(\frac{1}{\sqrt{n}}\left|\sum_{\ell=\ell_n(\alpha_0)}^{n} w_\ell^* \sum_{i=1}^{n} u_i \cdot \left[I_{\left\{u_i^2 \le u_{(\ell)}^2\right\}} - I_{\left\{u_i^2 \le q_{1-\frac{\ell}{n}}^2\right\}}\right]\right| > \theta\right) \le \frac{\varepsilon}{2}.$$

Now we can turn to the summand in Equation 3.84 and write it as

$$\frac{1}{\sqrt{n}}\sum_{\ell=1}^{\ell_n(\alpha_0)} w_\ell^* \sum_{i=1}^{n} \left[u_i - q_{1-\frac{\ell}{n}}\right] \cdot I_{\{u_i>0\}} \cdot \left[I_{\left\{u_i^2 \le u_{(\ell)}^2\right\}} - I_{\left\{u_i^2 \le q_{1-\frac{\ell}{n}}^2\right\}}\right] + \tag{3.86}$$

$$+ \frac{1}{\sqrt{n}}\sum_{\ell=1}^{\ell_n(\alpha_0)} w_\ell^* \sum_{i=1}^{n} q_{1-\frac{\ell}{n}} \cdot I_{\{u_i>0\}} \cdot \left[I_{\left\{u_i^2 \le u_{(\ell)}^2\right\}} - I_{\left\{u_i^2 \le q_{1-\frac{\ell}{n}}^2\right\}}\right] + \tag{3.87}$$

$$+ \frac{1}{\sqrt{n}}\sum_{\ell=1}^{\ell_n(\alpha_0)} w_\ell^* \sum_{i=1}^{n} \left[u_i - q_{1-\frac{\ell}{n}}\right] \cdot I_{\{u_i<0\}} \cdot \left[I_{\left\{u_i^2 \le u_{(\ell)}^2\right\}} - I_{\left\{u_i^2 \le q_{1-\frac{\ell}{n}}^2\right\}}\right] + \tag{3.88}$$

$$+ \frac{1}{\sqrt{n}}\sum_{\ell=1}^{\ell_n(\alpha_0)} w_\ell^* \sum_{i=1}^{n} q_{1-\frac{\ell}{n}} \cdot I_{\{u_i<0\}} \cdot \left[I_{\left\{u_i^2 \le u_{(\ell)}^2\right\}} - I_{\left\{u_i^2 \le q_{1-\frac{\ell}{n}}^2\right\}}\right]. \tag{3.89}$$

Then these four parts can be treated separately. We want to show that each of them is in probability smaller than $\frac{\varepsilon}{8}$. Let us first consider the expression in Equation 3.86. We can realize that due to the definition of $\ell_n(\alpha)$ we have for all $\alpha \in (\alpha_0, 1)$

$$1 - \frac{\ell_n(\alpha)}{n} = 1 - \frac{[(1-\alpha)n]_{int}}{n} \geq 1 - \frac{(1-\alpha)n}{n} = \alpha \geq \alpha_0$$

and consequently for all $n \in \mathbb{N}$ and all $\ell = 1, ..., \ell_n(\alpha_0)$

$$q^2_{1-\frac{\ell}{n}} \leq q^2_{\alpha_0}. \tag{3.90}$$

Further put $\delta = \alpha_0$ (i.e. $q^2_\delta = q^2_{\alpha_0}$) and define

$$B^1_n = \sup_{\alpha \in (\alpha_0, 1)} \left| \sqrt{u^2_{(\ell_n(\alpha))}} - q_\alpha \right| < n^{-\frac{1}{2}} K^*_{\varepsilon,\delta} \tag{3.91}$$

and

$$B^2_n = \sup_{\alpha \in (\alpha_0, 1)} \left| u^2_{(\ell_n(\alpha))} - q^2_\alpha \right| < n^{-\frac{1}{2}} K^*_{\varepsilon,\delta}.$$

Then recall Lemma 3.17 and Lemma 3.18. Employing these lemmas we can find (for ε that we have fixed before) $K^*_{\varepsilon,\delta} < \infty$ and $n^*_{\varepsilon,\delta} \in \mathbb{N}$ such that for all $n > n^*_{\varepsilon,\delta}$, when we put $\tilde{B}_n = B^1_n \cap B^2_n$, we obtain

$$P\left(\tilde{B}_n\right) > 1 - \frac{\varepsilon}{16}.$$

Moreover, we can employ Lemma 3.16 and find $K^{**}_{\varepsilon,\delta} < \infty$ and $n^{**}_{\varepsilon,\delta} \in \mathbb{N}$ such that for all $n > n^{**}_{\varepsilon,\delta}$ there exists a set C_n with probability at least $1 - \frac{\varepsilon}{16}$ such that

$$P\left(\left\{\left| I_{\left\{u_i^2 \leq u^2_{(\ell_n(\alpha))}\right\}} - I_{\left\{u_i^2 \leq q^2_\alpha\right\}} \right| = 1\right\} \cap I_{\{C_n\}}\right) < n^{-\frac{1}{2}} \cdot K^{**}_{\varepsilon,\delta}. \tag{3.92}$$

Put $K_{\varepsilon,\delta} = \max\{K^*_{\varepsilon,\delta}, K^{**}_{\varepsilon,\delta}\}$, $n_{\varepsilon,\delta} = \max\{n^*_{\varepsilon,\delta}, n^{**}_{\varepsilon,\delta}, 32^2\theta^{-2}\varepsilon^{-2}K^4_{\varepsilon,\delta}\}$, denote $D_n = \tilde{B}_n \cap C_n$, and in what follows, consider only $n > n_{\varepsilon,\delta}$. Notice that for the set D_n we obtain $P(D_n) > 1 - \frac{\varepsilon}{16}$. Then we get for Equation 3.86

$$P\left(\left\{\omega \in \Omega : \frac{1}{\sqrt{n}} \left| \sum_{\ell=1}^{\ell_n(\alpha_0)} w^*_\ell \sum_{i=1}^{n} \left[u_i - q_{1-\frac{\ell}{n}}\right] \cdot I_{\{u_i > 0\}} \times \right.\right.\right.$$

$$\times\left[I_{\left\{u_i^2\le u_{(\ell)}^2\right\}}-I_{\left\{u_i^2\le q^2_{1-\frac{\ell}{n}}\right\}}\right]\Bigg|<\theta\Bigg\}\Bigg)\le$$

$$\le P\left(\left\{\omega\in\Omega:\frac{1}{\sqrt{n}}\left|\sum_{\ell=1}^{\ell_n(\alpha_0)}w_\ell^*\sum_{i=1}^n\left[u_i-q_{1-\frac{\ell}{n}}\right]\cdot I_{\{u_i>0\}}\times\right.\right.\right.$$

$$\left.\left.\left.\times\left[I_{\left\{u_i^2\le u_{(\ell)}^2\right\}}-I_{\left\{u_i^2\le q^2_{1-\frac{\ell}{n}}\right\}}\right]\right|<\frac{\theta}{2}\right\}\cap D_n\right)+$$

$$+P\left(\left\{\omega\in\Omega:\frac{1}{\sqrt{n}}\left|\sum_{\ell=1}^{\ell_n(\alpha_0)}w_\ell^*\sum_{i=1}^n\left[u_i-q_{1-\frac{\ell}{n}}\right]\cdot I_{\{u_i>0\}}\times\right.\right.\right.$$

$$\left.\left.\left.\times\left[I_{\left\{u_i^2\le u_{(\ell)}^2\right\}}-I_{\left\{u_i^2\le q^2_{1-\frac{\ell}{n}}\right\}}\right]\right|<\frac{\theta}{2}\right\}\cap D_n^c\right)\le$$

$$\le\frac{2}{\theta n}\sum_{\ell=1}^{\ell_n(\alpha_0)}w_\ell^*\sum_{i=1}^n K_{\varepsilon,\delta}\cdot E\left(\left|I_{\left\{u_i^2\le u_{(\ell)}^2\right\}}-I_{\left\{u_i^2\le q^2_{1-\frac{\ell}{n}}\right\}}\right|\right)+\frac{\varepsilon}{16}\le$$

$$\le\frac{2}{\theta n^{\frac{3}{2}}}\sum_{\ell=1}^{\ell_n(\alpha_0)}w_\ell^*\sum_{i=1}^n K^2_{\varepsilon,\delta}+\frac{\varepsilon}{16}=\frac{2}{\theta n^{\frac{1}{2}}}K^2_{\varepsilon,\delta}\sum_{\ell=1}^{\ell_n(\alpha_0)}w_\ell^*+\frac{\varepsilon}{16}\le\frac{\varepsilon}{8},$$

where we have again used the Markov inequality, Equation 3.91 and Equation 3.92. Equation 3.88 can be treated similarly and we obtain an analogical result. For Equation 3.87 we obtain

$$P\left(\left\{\omega\in\Omega:\frac{1}{\sqrt{n}}\left|\sum_{\ell=1}^{\ell_n(\alpha_0)}w_\ell^*\sum_{i=1}^n q_{1-\frac{\ell}{n}}\cdot I_{\{u_i>0\}}\times\right.\right.\right.$$

$$\times\left[I_{\left\{u_i^2\le u_{(\ell)}^2\right\}}-I_{\left\{u_i^2\le q^2_{1-\frac{\ell}{n}}\right\}}\right]\Bigg|<\theta\Bigg\}\Bigg)\le$$

$$\le P\left(\left\{\omega\in\Omega:\frac{1}{\sqrt{n}}\left|\sum_{\ell=1}^{\ell_n(\alpha_0)}w_\ell^*\sum_{i=1}^n q_{1-\frac{\ell}{n}}\cdot I_{\{u_i>0\}}\times\right.\right.\right.$$

$$\left.\left.\left.\times\left[I_{\left\{u_i^2\le u_{(\ell)}^2\right\}}-I_{\left\{u_i^2\le q^2_{1-\frac{\ell}{n}}\right\}}\right]\right|<\frac{\theta}{2}\right\}\cap D_n\right)+$$

$$+P\left(\left\{\omega\in\Omega:\frac{1}{\sqrt{n}}\left|\sum_{\ell=1}^{\ell_n(\alpha_0)}w_\ell^*\sum_{i=1}^n q_{1-\frac{\ell}{n}}\cdot I_{\{u_i>0\}}\times\right.\right.\right.$$

$$\times \left[I_{\left\{u_i^2 \le u_{(\ell)}^2\right\}} - I_{\left\{u_i^2 \le q_{1-\frac{\ell}{n}}^2\right\}} \right] \Bigg| < \frac{\theta}{2} \Bigg\} \cap D_n^c \Bigg) \le$$

$$\le P \left(\left\{ \omega \in \Omega : \frac{1}{\sqrt{n}} \left| \sum_{\ell=1}^{\ell_n(\alpha_0)} w_\ell^* \sum_{i=1}^n q_{1-\frac{\ell}{n}} \cdot I_{\{u_i > 0\}} \times \right.\right.\right.$$

$$\left.\left.\left. \times \left[I_{\left\{u_i^2 \le u_{(\ell)}^2\right\}} - I_{\left\{u_i^2 \le q_{1-\frac{\ell}{n}}^2\right\}} \right] \right| < \frac{\theta}{2} \right\} \cap D_n \right) + \frac{\varepsilon}{16}.$$

To find the upper bound of this expression let us employ Lemma 3.21 and find a constant $\tilde{K}_{\varepsilon,\delta} < \infty$ and $\tilde{n}_{\varepsilon,\delta} \in \mathbb{N}$ such that for all $n > \tilde{n}_{\varepsilon,\delta}$ there exists a set A_n for which we have $P(A_n) > 1 - \frac{\varepsilon}{32}$ and

$$\max_{1 \le i \le n} \sup_{\alpha \in (\theta,1)} \left| E \left\{ I_{\{u_i > 0\}} \cdot \left[I_{\left\{u_i^2 \le u_{(\ell_n(\alpha))}^2\right\}} - I_{\left\{u_i^2 \le q_{1-\frac{\ell_n(\alpha)}{n}}^2\right\}} \right] \cdot I_{\{A_n\}} \right\} \right| < n^{-1} \cdot \tilde{K}_{(\varepsilon,\delta)}.$$

When we put $\tilde{n}_{\varepsilon,\delta}^* = \max\left\{ \tilde{n}_{\varepsilon,\delta}, \left[\frac{128}{\theta \cdot \varepsilon} q_{\alpha_0} \cdot \tilde{K}_{\varepsilon,\delta} \right]^2 \right\}$ and from now on consider only $n > \tilde{n}_{\varepsilon,\delta}^*$, then recalling Equation 3.90, we have

$$P \left(\left\{ \omega \in \Omega : \frac{1}{\sqrt{n}} \left| \sum_{\ell=1}^{\ell_n(\alpha_0)} w_\ell^* \sum_{i=1}^n q_{1-\frac{\ell}{n}} \cdot I_{\{u_i > 0\}} \times \right.\right.\right.$$

$$\left.\left.\left. \times \left[I_{\left\{u_i^2 \le u_{(\ell)}^2\right\}} - I_{\left\{u_i^2 \le q_{1-\frac{\ell}{n}}^2\right\}} \right] \right| < \frac{\theta}{2} \right\} \cap D_n \right) \le$$

$$\le P \left(\left\{ \omega \in \Omega : \frac{1}{\sqrt{n}} \left| \sum_{\ell=1}^{\ell_n(\alpha_0)} w_\ell^* \sum_{i=1}^n q_{1-\frac{\ell}{n}} \cdot I_{\{u_i > 0\}} \times \right.\right.\right.$$

$$\left.\left.\left. \times \left[I_{\left\{u_i^2 \le u_{(\ell)}^2\right\}} - I_{\left\{u_i^2 \le q_{1-\frac{\ell}{n}}^2\right\}} \right] \right| < \frac{\theta}{4} \right\} \cap D_n \cap A_n \right) +$$

$$+ P \left(\left\{ \omega \in \Omega : \frac{1}{\sqrt{n}} \left| \sum_{\ell=1}^{\ell_n(\alpha_0)} w_\ell^* \sum_{i=1}^n q_{1-\frac{\ell}{n}} \cdot I_{\{u_i > 0\}} \times \right.\right.\right.$$

$$\left.\left.\left. \times \left[I_{\left\{u_i^2 \le u_{(\ell)}^2\right\}} - I_{\left\{u_i^2 \le q_{1-\frac{\ell}{n}}^2\right\}} \right] \right| < \frac{\theta}{4} \right\} \cap D_n \cap A_n^c \right) \le$$

$$\leq P\left(\left\{\omega\in\Omega:\frac{1}{\sqrt{n}}\left|\sum_{\ell=1}^{\ell_n(\alpha_0)}w_\ell^*\sum_{i=1}^{n}q_{1-\frac{\ell}{n}}\cdot I_{\{u_i>0\}}\times\right.\right.\right.$$

$$\left.\left.\left.\times\left[I_{\left\{u_i^2\leq u_{(\ell)}^2\right\}}-I_{\left\{u_i^2\leq q^2_{1-\frac{\ell}{n}}\right\}}\right]\right|<\frac{\theta}{4}\right\}\cap D_n\cap A_n\right)+\frac{\varepsilon}{32}\leq$$

$$\leq\frac{4}{\theta\sqrt{n}}\sum_{\ell=1}^{\ell_n(\alpha_0)}w_\ell^*\sum_{i=1}^{n}q_{1-\frac{\ell}{n}}\cdot E\left(I_{\{u_i>0\}}\cdot\left[I_{\left\{u_i^2\leq u_{(\ell)}^2\right\}}-I_{\left\{u_i^2\leq q^2_{1-\frac{\ell}{n}}\right\}}\right]\times\right.$$

$$\left.\times I_{\{D_n\}}\cdot I_{\{A_n\}}\right)+\frac{\varepsilon}{32}\leq$$

$$\leq\frac{4}{\theta\sqrt{n}}\sum_{\ell=1}^{\ell_n(\alpha_0)}w_\ell^*q_{\alpha_0}\sum_{i=1}^{n}E\left(I_{\{u_i>0\}}\cdot\left[I_{\left\{u_i^2\leq u_{(\ell)}^2\right\}}-I_{\left\{u_i^2\leq q^2_{1-\frac{\ell}{n}}\right\}}\right]\cdot I_{\{A_n\}}\right)+\frac{\varepsilon}{32}\leq$$

$$\leq\frac{4}{\theta\sqrt{n}}\sum_{\ell=1}^{\ell_n(\alpha_0)}w_\ell^*q_{\alpha_0}\sum_{i=1}^{n}n^{-1}\cdot\tilde{K}_{(\varepsilon,\delta)}+\frac{\varepsilon}{32}\leq\frac{4}{\theta\sqrt{n}}q_{\alpha_0}\tilde{K}_{(\varepsilon,\delta)}+\frac{\varepsilon}{32}\leq\frac{\varepsilon}{16}.$$

Equation 3.89 can be again treated analogically. Then we can conclude that

$$P\left(\frac{1}{\sqrt{n}}\left|\sum_{\ell=1}^{\ell_n(\alpha_0)}w_\ell^*\sum_{i=1}^{n}u_i\cdot\left[I_{\left\{u_i^2\leq u_{(\ell)}^2\right\}}-I_{\left\{u_i^2\leq q^2_{1-\frac{\ell}{n}}\right\}}\right]\right|>\theta\right)\leq\frac{\varepsilon}{2}$$

and hence

$$P\left(\frac{1}{\sqrt{n}}\left|\sum_{\ell=1}^{n}w_\ell^*\sum_{i=1}^{n}u_i\cdot\left[I_{\left\{u_i^2\leq u_{(\ell)}^2\right\}}-I_{\left\{u_i^2\leq q^2_{1-\frac{\ell}{n}}\right\}}\right]\right|>\theta\right)\leq\varepsilon.$$

From what we have derived it follows that we can write Equation 3.75 as

$$\frac{1}{\sqrt{n}}\sum_{\ell=1}^{n}w_\ell^*\sum_{i=1}^{n}X_i\left(Y_i-X_i'\beta^0\right)I_{\left\{u_i^2\leq u_{(\ell)}^2\right\}}=$$

$$=\frac{1}{\sqrt{n}}\sum_{\ell=1}^{n}w_\ell^*\sum_{i=1}^{n}X_iu_i\cdot I_{\left\{u_i^2\leq q^2_{1-\frac{\ell}{n}}\right\}}+o_p(1)=$$

$$=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}X_iu_i\sum_{\ell=1}^{n}w_\ell^*\cdot I_{\left\{u_i^2\leq q^2_{1-\frac{\ell}{n}}\right\}}+o_p(1).$$

When we further recall Equation 3.35 and Equation 3.36, we obtain

$$\sum_{\ell=1}^{n}w_\ell^*\cdot I_{\left\{u_i^2\leq q^2_{1-\frac{\ell}{n}}\right\}}=\sum_{\ell=\ell_n^{(i)}}^{n}\left[w\left(\frac{\ell-1}{n}\right)-w\left(\frac{\ell}{n}\right)\right]=w\left(\frac{\ell_n^{(i)}-1}{n}\right),$$

where we denote by $\ell_n^{(i)}$ the smallest ℓ for which $u_i^2 < q^2_{1-\frac{\ell}{n}}$. This implies that $u_i^2 \geq q^2_{1-\frac{\ell_n^{(i)}-1}{n}}$. Therefore we have

$$q^2_{1-\frac{\ell_n^{(i)}-1}{n}} \leq u_i^2 < q^2_{1-\frac{\ell_n^{(i)}}{n}}$$

and consequently

$$q_{1-\frac{\ell_n^{(i)}-1}{n}} \leq |u_i| < q_{1-\frac{\ell_n^{(i)}}{n}}.$$

As q_α is the upper α-qualtile of the function $\bar{F}_{n,\beta^0}(z)$ and this function is continuous from the left, we have

$$1 - \frac{\ell_n^{(i)} - 1}{n} = \bar{F}_{n,\beta^0}\left(q_{1-\frac{\ell_n^{(i)}-1}{n}}\right) \leq \bar{F}_{n,\beta^0}\left(|u_i|\right) \leq \bar{F}_{n,\beta^0}\left(q_{1-\frac{\ell_n^{(i)}}{n}}\right) = 1 - \frac{\ell_n^{(i)}}{n}$$

and considering that the weight function w is monotone, we have also

$$w\left(1 - \frac{\ell_n^{(i)}}{n}\right) \leq w\left(\bar{F}_{n,\beta^0}\left(|u_i|\right)\right) \leq w\left(1 - \frac{\ell_n^{(i)} - 1}{n}\right).$$

Taking into account all these results and the boundedness of $w'(\alpha)$, we arrive at

$$\sum_{\ell=1}^{n} w_\ell^* \cdot I_{\left\{u_i^2 \leq q^2_{1-\frac{\ell}{n}}\right\}} = w\left(\bar{F}_{n,\beta^0}\left(|u_i|\right)\right) + o\left(\frac{1}{n}\right)$$

and therefore Equation 3.75 can be written as

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i u_i \sum_{\ell=1}^{n} w_\ell^* \cdot I_{\left\{u_i^2 \leq q^2_{1-\frac{\ell}{n}}\right\}} + o_p(1) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} w\left(\bar{F}_{n,\beta^0}\left(|u_i|\right)\right) X_i u_i + o_p(1).$$

Let us now return to Equation 3.76 and consider the first term. When we take into account Equation 3.37, we can write

$$\frac{1}{\sqrt{n}} \sum_{\ell=1}^{n} w_\ell^* \sum_{i=1}^{n} X_i X_i' \left(\hat{\beta}^{(LWS,n,w)} - \beta^0\right) \cdot I_{\left\{u_i^2 \leq u^2_{(\ell)}\right\}} =$$

$$= \frac{1}{n} \sum_{i=1}^{n} X_i X_i' \sum_{\ell=1}^{n} w_\ell^* I_{\left\{u_i^2 \leq u^2_{(\ell)}\right\}} \cdot \left\{\sqrt{n}\left(\hat{\beta}^{(LWS,n,w)} - \beta^0\right)\right\} =$$

$$= \frac{1}{n} \sum_{i=1}^{n} w\left(F^{(n)}_{\beta^0}\left(|u_i|\right)\right) X_i X_i' \cdot \left\{\sqrt{n}\left(\hat{\beta}^{(LWS,n,w)} - \beta^0\right)\right\} =$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left[w\left(F_{\beta^0}^{(n)}\left(|u_i|\right)\right) - w\left(\bar{F}_{n,\beta^0}\left(|u_i|\right)\right)\right] X_i X_i' \cdot \left\{\sqrt{n}\left(\hat{\beta}^{(LWS,n,w)} - \beta^0\right)\right\} +$$

$$+\frac{1}{n}\sum_{i=1}^{n} w\left(\bar{F}_{n,\beta^0}\left(|u_i|\right)\right) X_i X_i' \cdot \left\{\sqrt{n}\left(\hat{\beta}^{(LWS,n,w)} - \beta^0\right)\right\}.$$

Then we can employ Lemma 3.8 and the assumption that the weight function is Lipschitz and we get

$$\left\|\frac{1}{n^{\frac{3}{2}}}\sum_{i=1}^{n}\sqrt{n}\left[w\left(F_{\beta^0}^{(n)}\left(|u_i|\right)\right) - w\left(\bar{F}_{n,\beta^0}\left(|u_i|\right)\right)\right] X_i X_i' \times \right.$$

$$\left. \times \left\{\sqrt{n}\left(\hat{\beta}^{(LWS,n,w)} - \beta^0\right)\right\}\right\| \leq$$

$$\leq \frac{1}{n^{\frac{1}{2}}}\sup_{r\in\mathbb{R}}\sqrt{n}\left|w\left(F_{\beta^0}^{(n)}(r)\right) - w\left(\bar{F}_{n,\beta^0}(r)\right)\right| \cdot n^{-1}\sum_{i=1}^{n}\|X_i\|\cdot\|X_i\| \times$$

$$\times \left\{\sqrt{n}\left(\hat{\beta}^{(LWS,n,w)} - \beta^0\right)\right\} \leq$$

$$\leq \frac{1}{n^{\frac{1}{2}}} L_w \cdot n^{-1}\sum_{i=1}^{n}\|X_i\|\cdot\|X_i\|\cdot\left\{\sqrt{n}\left(\hat{\beta}^{(LWS,n,w)} - \beta^0\right)\right\}\cdot O_p(1) = O_p(n^{-\frac{1}{2}}) = o_p(1),$$

where we used the $\sqrt{n}$-consistency of the LWS estimator from Lemma 3.2. This concludes the proof.

Q.E.D.

Finally, using this result, we can find the asymptotic representation of LWS estimator under the assumption of heteroskedasticity. The formula for the asymptotic representation is given in Theorem 3.1.

Theorem 3.1. *Let the Assumptions 3.1, 3.2, 3.3, 3.4, 3.5, 3.6 and 3.7 hold. Moreover, let* $Q = E\left\{w(\bar{F}_{\beta^0}(|u|))X_1X_1'\right\}$. *Then*

$$\sqrt{n}\left(\hat{\beta}^{(LWS,n,w)} - \beta^0\right) = Q^{-1} \cdot \frac{1}{\sqrt{n}}\sum_{i=1}^{n} w\left(\bar{F}_{n,\beta^0}\left(|u_i|\right)\right) X_i u_i + o_p(1). \quad (3.93)$$

PROOF Let us realize that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} w\left(\bar{F}_{n,\beta^0}\left(|u_i|\right)\right) X_i X_i' = Q \quad a.s.$$

and that Q is positive definite and therefore regular. Then the proof follows directly from Lemma 3.22.

Q.E.D.

Chapter 4

Simulation study

This chapter provides a numerical study that illustrates the results derived in previous chapter. The study is based on Monte Carlo method and was implemented in MATLAB. The chapter is divided in three sections. The first section explains the setup of the simulations. The second section provides results for homoskedastic residuals and is included only for complexity, therefore the results are not discussed in detail. The results of simulations of the main result of this thesis, i.e. the asymptotic representation of LWS estimator under heteroskedasticity, are provided in the third section.

4.1 Setup of the simulation study

To be able to conduct the simulation study, we first generate the matrix of explanatory variables $\mathbf{X}$ and the vector of error terms. These data are generated according to normal distribution. The errors are first generated with constant variance, which is used in the second section. To obtain heteroskedastic error terms used in the third section, we further multiply it by a random number for every observation i.

The second and third section are further divided in two subsections, based on the way how the data were contaminated (outliers or leverage points). To obtain the contaminated data, we define a threshold and contaminate all the observations behind this threshold. Therefore in this study we consider the case where the contamination influences the observations that are in the tails, not in the centre of the bulk of data. This way of contamination is also the reason for the percentage of contaminated data not to be an integer (e.g. in the second section we have 15.09% instead of exactly 15%). The outliers are obtained by multiplying the response variable by 15 (for the observations that are behind the threshold). The leverage points are obtained by multiplying the explanatory variable by 10 and then taking negative value of the resulting response variable to ensure that the leverage points are bad (of course again for the observations that are behind the threshold).

Each of the subsections contains results for several levels of contamination. The considered levels of contamination are 0%, 0.5%, 1%, 2%, 5%, 10% and 15%. Note that to obtain good results for significantly higher percentage of contamination with the method of LWS we would need to adjust the weight function.

The model used in the next two sections looks as follows

$$Y_i = \beta_j^0 X_{ij} + u_i$$

where $i = 1, ..., n$, $j = 1, ..., p$. The number of observations is set to $n = 500$ and the number of variables is $p = 5$, with $j = 1$ representing the intercept. As the number of repetitions to get the mean estimate stated in the table we choose $m = 500$. The vector of true coefficients is set to $\beta^0 = (2, 4, 5, -3, -6)'$. For every considered case (homoskedasticity/heteroskedasticity, outliers/leverage points, percentage of contamination) we estimate the coefficients using the methods of OLS and LWS for comparison of classical and robust methods, and provide the results for the asymptotic representation to illustrate how well it represents the LWS estimator. This procedure (along with estimating the mean squared error (MSE)) is repeated for every $k = 1, ..., m$ and then we take the mean values of these m repetitions to obtain the desired values. The estimators for each k are then used to obtain the empirical distributions of the estimators, which can be seen in the figures in Section 4.3.

The resulting estimators in the tables in the next sections are denoted $\hat{\beta}^{(OLS,n)}$ for the estimator obtained by the method of OLS, $\hat{\beta}^{(LWS,n,w)}$ for the estimator obtained by the method of LWS (see the algorithm described in Section 2.3), and $\hat{\beta}^{(AR,n,w)}$ for the estimator obtained from the asymptotic representation (see the formula derived in Section 3.2). The number in the parentheses stated in the lower index next to each of the estimates is the MSE.[1]

Overall, we can see in the following sections that the asymptotic representation has lower MSE than the LWS estimator. This is due to the construction

[1]Note that we could alternatively use the variance instead of MSE. However, the MSE is more appropriate here, as it accounts also for the bias.

of the asymptotic representation - in the simulations we omit the $o_p(1)$ term that goes in probability to 0 (see Equation 3.93). This does not influence the estimator, but it causes the MSE to be somewhat lower.

As was shown in previous studies, an important feature of a study including LWS estimator is the choice of weight function. In what follows, we use a weight function based on Tukey´s ρ, similarly as in Campbell & Rousseeuw (1998). The Tukey´s weight function can be also found in Hampel *et al.* (1986) and was also used e.g. in Gervini & Yohai (2002). We define the weight function in a way that 5% of the data with largest squared residuals are assigned $w = 0$ and therefore are omitted. 80% of the data with smallest squared residuals are assigned $w = 1$ and the remaining 15% of observations are assigned decreasing weights. Then we can continue with the results.

4.2 Homoskedastic errors

Let us first consider homoskedastic error terms. As was mentioned above, these results are provided just for completeness, therefore the details, such as the distribution of the resulting estimators, are studied only in the next section containing results for heteroskedastic residuals. We can divide this section according to the type of contamination as follows.

4.2.1 Outliers

As the titles hint, this subsection provides results for homoskedastic residuals, where the data were contaminated by outliers (in a way that was described in the first section). In Table 4.1 we can see that with homoskedastic errors and no contamination both the estimation methods, as well as the asymptotic representation, give reasonable results. We can notice that the method of OLS results in a more efficient estimator than the method of LWS, as the MSE is lower.

Table 4.2 suggests that even for a very small percentage of contamination the method of OLS gives misleading results. On the contrary, the results for both the LWS estimator and the asymptotic representation stay basically the same.

Similiar results follow from Table 4.3, Table 4.4 and Table 4.5, where we can see that the OLS estimator worsens with increasing level of contamination, as compared to the LWS estimator and the asymptotic representation.

Table 4.1: Homoskedastic errors, 0% of contamination

Level of contamination= 0.00%					
True β^0	2.000	4.000	5.000	-3.000	-6.000
$\hat{\beta}^{(OLS,n)}$	$2.000_{(0.0006)}$	$4.000_{(0.0006)}$	$4.999_{(0.0005)}$	$-2.999_{(0.0005)}$	$-6.000_{(0.0005)}$
$\hat{\beta}^{(LWS,n,w)}$	$2.001_{(0.0010)}$	$4.000_{(0.0009)}$	$4.999_{(0.0009)}$	$-2.998_{(0.0009)}$	$-6.000_{(0.0010)}$
$\hat{\beta}^{(AR,n,w)}$	$2.000_{(0.0003)}$	$4.000_{(0.0003)}$	$5.000_{(0.0003)}$	$-2.999_{(0.0003)}$	$-6.000_{(0.0004)}$

Source: author's computations.

Table 4.2: Homoskedastic errors, cca 0.5% of outliers

Level of contamination= 0.48%					
True β^0	2.000	4.000	5.000	-3.000	-6.000
$\hat{\beta}^{(OLS,n)}$	$2.562_{(2.2148)}$	$4.977_{(3.2634)}$	$6.229_{(4.2790)}$	$-3.789_{(2.8953)}$	$-7.595_{(5.7385)}$
$\hat{\beta}^{(LWS,n,w)}$	$2.000_{(0.0009)}$	$3.999_{(0.0009)}$	$5.000_{(0.0010)}$	$-3.000_{(0.0009)}$	$-6.000_{(0.0010)}$
$\hat{\beta}^{(AR,n,w)}$	$2.000_{(0.0003)}$	$3.999_{(0.0003)}$	$5.000_{(0.0003)}$	$-3.000_{(0.0003)}$	$-6.000_{(0.0003)}$

Source: author's computations.

Table 4.3: Homoskedastic errors, cca 1% of outliers

Level of contamination= 1.00%					
True β^0	2.000	4.000	5.000	-3.000	-6.000
$\hat{\beta}^{(OLS,n)}$	$2.855_{(3.6808)}$	$5.859_{(7.6102)}$	$7.453_{(10.3405)}$	$-4.417_{(5.2507)}$	$-8.954_{(14.4261)}$
$\hat{\beta}^{(LWS,n,w)}$	$1.999_{(0.0008)}$	$4.000_{(0.0010)}$	$4.999_{(0.0009)}$	$-3.002_{(0.0009)}$	$-6.000_{(0.0009)}$
$\hat{\beta}^{(AR,n,w)}$	$2.000_{(0.0003)}$	$4.000_{(0.0004)}$	$4.999_{(0.0004)}$	$-3.001_{(0.0003)}$	$-6.000_{(0.0003)}$

Source: author's computations.

Table 4.4: Homoskedastic errors, cca 2% of outliers

Level of contamination= 1.97%					
True β^0	2.000	4.000	5.000	-3.000	-6.000
$\hat{\beta}^{(OLS,n)}$	$3.720_{(7.8628)}$	$7.490_{(18.8485)}$	$9.272_{(25.1469)}$	$-5.400_{(10.3779)}$	$-11.110_{(34.1601)}$
$\hat{\beta}^{(LWS,n,w)}$	$2.000_{(0.0009)}$	$4.000_{(0.0009)}$	$5.001_{(0.0009)}$	$-3.001_{(0.0009)}$	$-6.000_{(0.0009)}$
$\hat{\beta}^{(AR,n,w)}$	$2.000_{(0.0003)}$	$4.000_{(0.0004)}$	$5.000_{(0.0004)}$	$-3.001_{(0.0004)}$	$-6.000_{(0.0004)}$

Source: author's computations.

Table 4.5: Homoskedastic errors, cca 5% of outliers

Level of contamination= 4.94%					
True β^0	2.000	4.000	5.000	-3.000	-6.000
$\hat{\beta}^{(OLS,n)}$	$5.768_{(22.7668)}$	$10.963_{(58.4090)}$	$14.098_{(93.8538)}$	$-8.616_{(40.2529)}$	$-17.209_{(137.1784)}$
$\hat{\beta}^{(LWS,n,w)}$	$1.999_{(0.0008)}$	$4.000_{(0.0008)}$	$5.001_{(0.0008)}$	$-3.001_{(0.0008)}$	$-6.001_{(0.0009)}$
$\hat{\beta}^{(AR,n,w)}$	$1.999_{(0.0004)}$	$4.000_{(0.0004)}$	$5.001_{(0.0004)}$	$-3.001_{(0.0004)}$	$-6.000_{(0.0004)}$

Source: author's computations.

As the percentage of contamination exceeds the 5% boundary and we get to the decreasing part of the weight function, the LWS estimator starts to worsen as well (as some of the contaminated observations influence the estimator, even though with weight lower than $w = 1$). This phenomenon can be seen in Table 4.6 and Table 4.7. Note that although the MSE of $\hat{\beta}^{(AR,n,w)}$ slightly increases, the asymptotic representation is not influenced as much as the LWS estimator.

Table 4.6: Homoskedastic errors, cca 10% of outliers

Level of contamination= 10.02%					
True β^0	2.000	4.000	5.000	-3.000	-6.000
$\hat{\beta}^{(OLS,n)}$	$8.626_{(53.5158)}$	$17.338_{(190.5969)}$	$21.502_{(287.3661)}$	$-12.950_{(109.9162)}$	$-26.020_{(418.9173)}$
$\hat{\beta}^{(LWS,n,w)}$	$2.064_{(0.0654)}$	$4.134_{(0.0737)}$	$5.113_{(0.0618)}$	$-3.082_{(0.0699)}$	$-6.157_{(0.0851)}$
$\hat{\beta}^{(AR,n,w)}$	$2.000_{(0.0005)}$	$4.001_{(0.0005)}$	$5.001_{(0.0005)}$	$-2.998_{(0.0005)}$	$-5.999_{(0.0006)}$

Source: author's computations.

Table 4.7: Homoskedastic errors, cca 15% of outliers

Level of contamination= 15.09%					
True β^0	2.000	4.000	5.000	-3.000	-6.000
$\hat{\beta}^{(OLS,n)}$	$11.106_{(94.3531)}$	$22.032_{(339.0322)}$	$27.453_{(521.8490)}$	$-16.476_{(193.6685)}$	$-32.824_{(736.9791)}$
$\hat{\beta}^{(LWS,n,w)}$	$2.451_{(0.9313)}$	$4.907_{(1.5958)}$	$6.129_{(2.1961)}$	$-3.555_{(0.9486)}$	$-7.277_{(2.5727)}$
$\hat{\beta}^{(AR,n,w)}$	$2.000_{(0.0006)}$	$4.001_{(0.0006)}$	$4.999_{(0.0007)}$	$-3.001_{(0.0006)}$	$-5.998_{(0.0007)}$

Source: author's computations.

4.2.2 Leverage points

This subsection contains the same analysis as the previous one with the difference that the contamination is caused by leverage points instead of outliers. If the data are not contaminated, the result is of course the same, therefore we can start with the level of contamination being 0.5% in Table 4.8. We can see that the results for all levels of contamination are very similar to the previous subsection. The only significant difference as compared to the previous subsection is that leverage points influence the OLS estimator more than outliers (this phenomenon was found in previous studies and is confirmed also here). In Table 4.11 and following two tables we can see that for the contamination level of 5% (or higher) the OLS estimates have even wrong sign.

Table 4.8: Homoskedastic errors, cca 0.5% of leverage points

Level of contamination= 0.46%					
True β^0	2.000	4.000	5.000	-3.000	-6.000
$\hat{\beta}^{(OLS,n)}$	$1.632_{(1.2370)}$	$3.218_{(1.8055)}$	$4.094_{(2.0562)}$	$-2.479_{(1.3157)}$	$-4.869_{(3.3078)}$
$\hat{\beta}^{(LWS,n,w)}$	$1.999_{(0.0009)}$	$3.998_{(0.0010)}$	$4.999_{(0.0008)}$	$-3.001_{(0.0010)}$	$-6.000_{(0.0008)}$
$\hat{\beta}^{(AR,n,w)}$	$1.999_{(0.0003)}$	$3.999_{(0.0004)}$	$5.000_{(0.0003)}$	$-3.000_{(0.0004)}$	$-5.999_{(0.0003)}$

Source: author's computations.

Table 4.9: Homoskedastic errors, cca 1% of leverage points

Level of contamination= 1.01%					
True β^0	2.000	4.000	5.000	-3.000	-6.000
$\hat{\beta}^{(OLS,n)}$	$1.288_{(2.3376)}$	$2.532_{(4.3847)}$	$3.026_{(6.9302)}$	$-1.835_{(3.4029)}$	$-3.776_{(7.8261)}$
$\hat{\beta}^{(LWS,n,w)}$	$1.999_{(0.0010)}$	$4.000_{(0.0010)}$	$5.003_{(0.0010)}$	$-2.999_{(0.0009)}$	$-5.997_{(0.0009)}$
$\hat{\beta}^{(AR,n,w)}$	$1.999_{(0.0003)}$	$4.000_{(0.0004)}$	$5.002_{(0.0004)}$	$-2.999_{(0.0003)}$	$-5.999_{(0.0003)}$

Source: author's computations.

Table 4.10: Homoskedastic errors, cca 2% of leverage points

Level of contamination= 1.93%					
True β^0	2.000	4.000	5.000	-3.000	-6.000
$\hat{\beta}^{(OLS,n)}$	$0.789_{(4.2693)}$	$1.315_{(10.7700)}$	$1.805_{(14.5045)}$	$-1.025_{(7.1307)}$	$-2.083_{(20.2994)}$
$\hat{\beta}^{(LWS,n,w)}$	$1.999_{(0.0009)}$	$4.002_{(0.0009)}$	$5.001_{(0.0010)}$	$-3.000_{(0.0008)}$	$-6.001_{(0.0009)}$
$\hat{\beta}^{(AR,n,w)}$	$2.000_{(0.0003)}$	$4.001_{(0.0004)}$	$5.000_{(0.0004)}$	$-3.000_{(0.0003)}$	$-6.001_{(0.0004)}$

Source: author's computations.

Table 4.11: Homoskedastic errors, cca 5% of leverage points

Level of contamination= 4.96%					
True β^0	2.000	4.000	5.000	-3.000	-6.000
$\hat{\beta}^{(OLS,n)}$	$-0.808_{(12.5530)}$	$-1.953_{(42.1990)}$	$-2.249_{(59.7705)}$	$1.338_{(24.6382)}$	$2.732_{(84.0307)}$
$\hat{\beta}^{(LWS,n,w)}$	$1.998_{(0.0008)}$	$4.001_{(0.0007)}$	$5.000_{(0.0008)}$	$-2.999_{(0.0008)}$	$-5.999_{(0.0008)}$
$\hat{\beta}^{(AR,n,w)}$	$1.998_{(0.0004)}$	$4.001_{(0.0004)}$	$5.000_{(0.0004)}$	$-3.000_{(0.0004)}$	$-6.000_{(0.0004)}$

Source: author's computations.

Table 4.12: Homoskedastic errors, cca 10% of leverage points

Level of contamination= 9.99%					
True β^0	2.000	4.000	5.000	-3.000	-6.000
$\hat{\beta}^{(OLS,n)}$	$-3.196_{(33.9782)}$	$-6.338_{(114.7301)}$	$-8.002_{(178.5700)}$	$4.769_{(67.8557)}$	$9.469_{(250.9593)}$
$\hat{\beta}^{(LWS,n,w)}$	$1.949_{(0.0311)}$	$3.917_{(0.0397)}$	$4.895_{(0.0465)}$	$-2.941_{(0.0400)}$	$-5.875_{(0.0610)}$
$\hat{\beta}^{(AR,n,w)}$	$2.000_{(0.0006)}$	$4.001_{(0.0005)}$	$5.001_{(0.0006)}$	$-2.999_{(0.0005)}$	$-6.000_{(0.0006)}$

Source: author's computations.

Table 4.13: Homoskedastic errors, cca 15% of leverage points

Level of contamination= 15.06%					
True β^0	2.000	4.000	5.000	-3.000	-6.000
$\hat{\beta}^{(OLS,n)}$	$-5.010_{(55.4246)}$	$-10.198_{(209.8713)}$	$-12.809_{(326.8115)}$	$7.528_{(117.6219)}$	$15.135_{(457.1284)}$
$\hat{\beta}^{(LWS,n,w)}$	$1.697_{(0.5156)}$	$3.308_{(0.9526)}$	$4.167_{(1.2896)}$	$-2.475_{(0.6916)}$	$-5.009_{(1.5066)}$
$\hat{\beta}^{(AR,n,w)}$	$2.002_{(0.0006)}$	$4.001_{(0.0006)}$	$5.001_{(0.0006)}$	$-3.000_{(0.0006)}$	$-6.001_{(0.0006)}$

Source: author's computations.

4.3 Heteroskedastic errors

This section contains the results obtained for heteroskedastic residuals. As this illustrates the main result of the thesis, we include somewhat more detailed analysis. In addition to the tables with resulting estimators, we include also the plots of the empirical distribution functions of $\hat{\beta}^{(LWS,n,w)}$ and $\hat{\beta}^{(AR,n,w)}$ to show that the distributions do not differ significantly from each other.

The figures of the two empirical distribution functions are accompanied by the resulting p-value of the two-sample Kolmogorov-Smirnov test (K-S test). The null hypothesis of the test is that the two functions come from the same distribution. Therefore in order to be able to conclude that the distributions do not significantly differ from each other, we should not reject the null hypothesis. Notice that the decision about rejection of the null hypothesis of the K-S test is dependent on the size of the sample. As we have quite large sample for this purpose, we are more likely to reject the null (i.e. even if the p-value is rather small, we might be able to conclude that the estimators come from the same distribution).

4.3.1 Outliers

In the first subsection we have again results for the case where the data were contaminated by outliers. First of all, let us consider the results for data with no contamination. We can see in Table 4.14 that with no contamination the OLS estimator gives reasonable results even for the heteroskedastic case. This of course makes sense as we do not need the assumption of homoskedasticity for the OLS estimator to be unbiased. However, heteroskedasticity does influence the efficiency of the OLS estimator and we can see that even with no contamination the LWS estimator outperforms the OLS estimator in this respect.

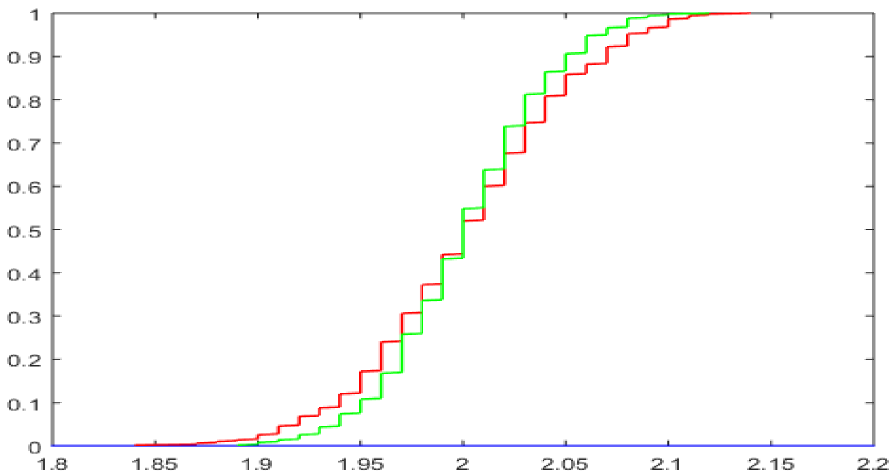
Table 4.14: Heteroskedastic errors, 0% of contamination

Level of contamination= 0.00%					
True β^0	2.000	4.000	5.000	-3.000	-6.000
$\hat{\beta}^{(OLS,n)}$	$2.002_{(0.0054)}$	$4.002_{(0.0051)}$	$5.000_{(0.0048)}$	$-3.004_{(0.0053)}$	$-6.004_{(0.0052)}$
$\hat{\beta}^{(LWS,n,w)}$	$2.002_{(0.0025)}$	$3.999_{(0.0027)}$	$5.002_{(0.0029)}$	$-3.002_{(0.0026)}$	$-6.002_{(0.0028)}$
$\hat{\beta}^{(AR,n,w)}$	$2.001_{(0.0015)}$	$4.000_{(0.0017)}$	$5.001_{(0.0016)}$	$-3.001_{(0.0015)}$	$-6.002_{(0.0017)}$

Source: author's computations.

As was mentioned before, the MSE of the asymptotic representation is lower than the MSE of the LWS estimator. This also causes the EDF of $\hat{\beta}^{(AR,n,w)}$ to be steeper than the EDF of $\hat{\beta}^{(LWS,n,w)}$, as we can see in Figure 4.1, where the steeper (green) function represents the asymptotic representation and the flatter (red) function represents the method of LWS. Specifically, we consider the estimator of the first coefficient β_1 (i.e. the intercept).[2] The p-value of the K-S test is 0.15, therefore even on 10% significance level we cannot reject the null that these two functions come from the same distribution.

Figure 4.1: EDF of $\hat{\beta}_1^{(LWS,n,w)}$ and $\hat{\beta}_1^{(AR,n,w)}$, 0% of contamination



Source: author's computations.

Table 4.15: Heteroskedastic errors, cca 0.5% of outliers

Level of contamination= 0.47%					
True β^0	2.000	4.000	5.000	−3.000	−6.000
$\hat{\beta}^{(OLS,n)}$	$2.532_{(2.0393)}$	$5.007_{(3.3806)}$	$6.255_{(4.3575)}$	$-3.608_{(2.1605)}$	$-7.500_{(5.0982)}$
$\hat{\beta}^{(LWS,n,w)}$	$2.002_{(0.0028)}$	$3.999_{(0.0026)}$	$5.000_{(0.0027)}$	$-3.004_{(0.0029)}$	$-6.001_{(0.0025)}$
$\hat{\beta}^{(AR,n,w)}$	$2.002_{(0.0017)}$	$3.999_{(0.0016)}$	$5.000_{(0.0016)}$	$-3.003_{(0.0017)}$	$-6.001_{(0.0015)}$

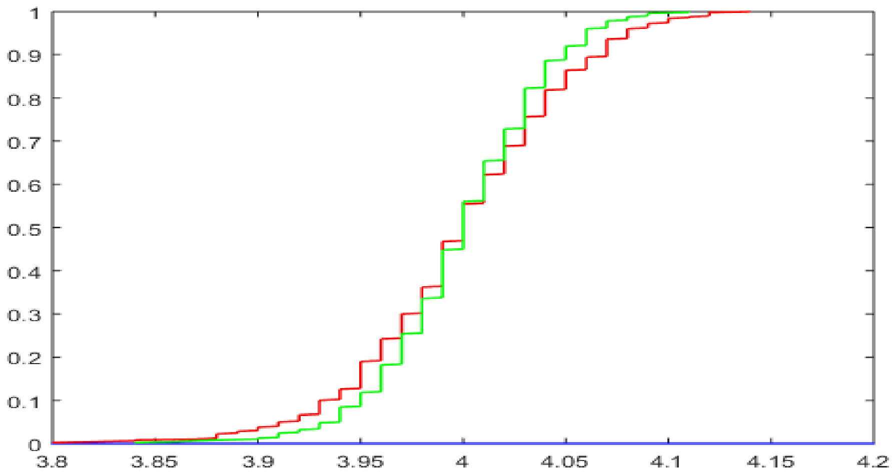
Source: author's computations.

In Table 4.15 we can see that even very small percentage of contamination affects the OLS estimator, similarly as in homoskedastic case. On the contrary, the LWS estimator and the asymptotic representation of LWS estimator are essentially the same as in the case with no contamination.

[2] For each level of contamination we always show the figure only for one of the five estimators, but note that the other four look very similarly in all cases.

It follows that also the empirical distribution functions in Figure 4.2 look similarly as for the case with no contaminated data. The steeper (green) function again stands for the asymptotic representation and the flatter (red) function represents the LWS estimator. We consider the second coefficient β_2 and the resulting p-value of the two sample K-S test is again 0.15. Therefore we can conclude that these two functions also come from the same distribution.

Figure 4.2: EDF of $\hat{\beta}_2^{(LWS,n,w)}$ and $\hat{\beta}_2^{(AR,n,w)}$, cca 0.5% of outliers



Source: author's computations.

In Table 4.16, Table 4.17 and Table 4.18 we can see that the results for LWS estimator and for the asymptotic representation remain essentially the same up to 5% of outliers, i.e. as long as the contaminated observations are not considered in the estimation process at all.

Table 4.16: Heteroskedastic errors, cca 1% of outliers

Level of contamination= 1.04%					
True β^0	2.000	4.000	5.000	-3.000	-6.000
$\hat{\beta}^{(OLS,n)}$	$2.979_{(4.1257)}$	$5.944_{(8.0633)}$	$7.451_{(10.2171)}$	$-4.442_{(5.3253)}$	$-9.012_{(14.9117)}$
$\hat{\beta}^{(LWS,n,w)}$	$2.006_{(0.0033)}$	$3.995_{(0.0032)}$	$4.999_{(0.0030)}$	$-2.999_{(0.0028)}$	$-5.999_{(0.0029)}$
$\hat{\beta}^{(AR,n,w)}$	$2.005_{(0.0019)}$	$3.996_{(0.0019)}$	$4.999_{(0.0018)}$	$-2.999_{(0.0017)}$	$-6.000_{(0.0018)}$

Source: author's computations.

The plots of the EDFs look very similarly as well, as we can see in Figure 4.3, Figure 4.4 and Figure 4.5. In all the figures the steeper (green) function represents again the asymptotic representation. The p-values of the K-S test are

Table 4.17: Heteroskedastic errors, cca 2% of outliers

Level of contamination= 2.01%					
True β^0	2.000	4.000	5.000	-3.000	-6.000
$\hat{\beta}^{(OLS,n)}$	$3.781_{(8.6033)}$	$7.449_{(18.3162)}$	$9.282_{(25.0117)}$	$-5.749_{(13.1579)}$	$-11.125_{(35.0836)}$
$\hat{\beta}^{(LWS,n,w)}$	$2.002_{(0.0030)}$	$4.002_{(0.0030)}$	$4.997_{(0.0028)}$	$-2.998_{(0.0029)}$	$-6.002_{(0.0036)}$
$\hat{\beta}^{(AR,n,w)}$	$2.001_{(0.0018)}$	$4.001_{(0.0019)}$	$4.998_{(0.0017)}$	$-2.999_{(0.0018)}$	$-6.001_{(0.0022)}$

Source: author's computations.

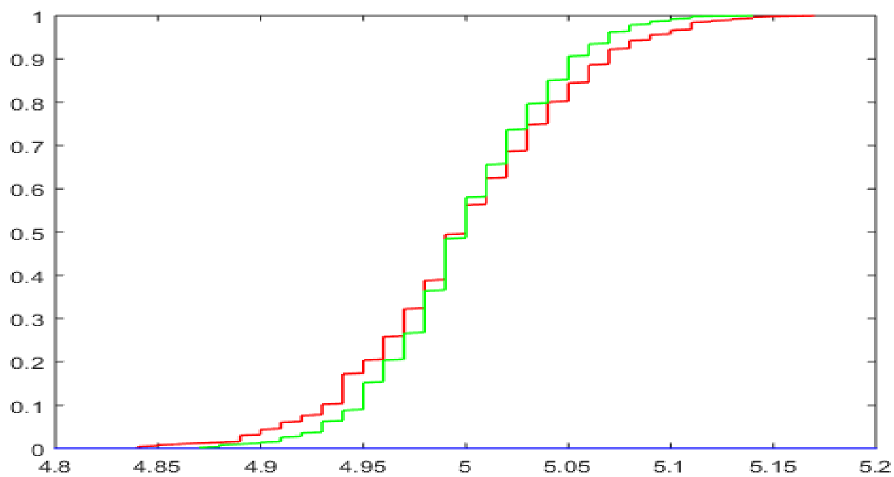
Table 4.18: Heteroskedastic errors, cca 5% of outliers

Level of contamination= 4.97%					
True β^0	2.000	4.000	5.000	-3.000	-6.000
$\hat{\beta}^{(OLS,n)}$	$5.643_{(21.4504)}$	$11.516_{(67.3940)}$	$14.311_{(99.5893)}$	$-8.611_{(41.0469)}$	$-17.033_{(136.6985)}$
$\hat{\beta}^{(LWS,n,w)}$	$2.004_{(0.0036)}$	$4.000_{(0.0031)}$	$4.998_{(0.0035)}$	$-3.001_{(0.0035)}$	$-5.997_{(0.0034)}$
$\hat{\beta}^{(AR,n,w)}$	$2.003_{(0.0025)}$	$4.000_{(0.0022)}$	$4.998_{(0.0024)}$	$-3.001_{(0.0024)}$	$-5.997_{(0.0024)}$

Source: author's computations.

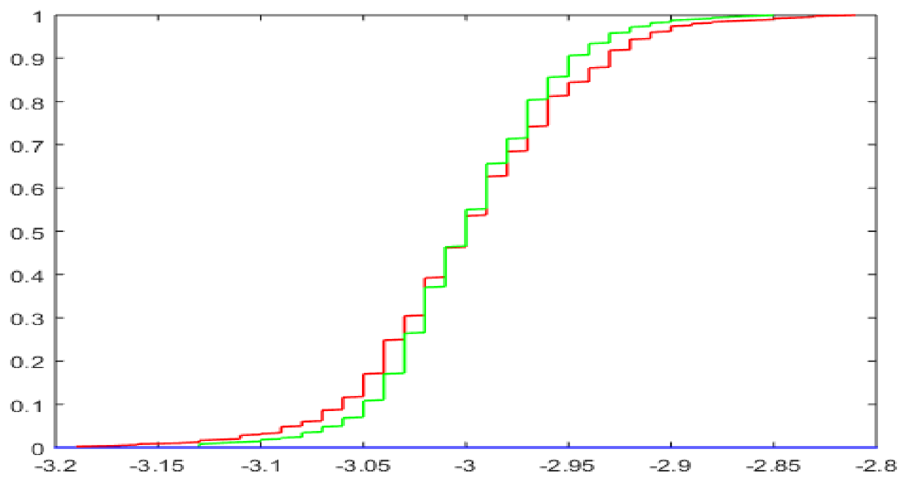
0.06, 0.09 and 0.5, respectively. Although the p-values for 1% and 2% of outliers are rather small, this can be caused by the high number of observations, as was mentioned before. Moreover, we still cannot reject the null for any of these cases on the commonly considered 5% significance level. From the overall results it seems reasonable to conclude that up to the contamination level of 5% $\hat{\beta}^{(LWS,n,w)}$ and $\hat{\beta}^{(AR,n,w)}$ come from the same distribution.

Figure 4.3: EDF of $\hat{\beta}_3^{(LWS,n,w)}$ and $\hat{\beta}_3^{(AR,n,w)}$, cca 1% of outliers



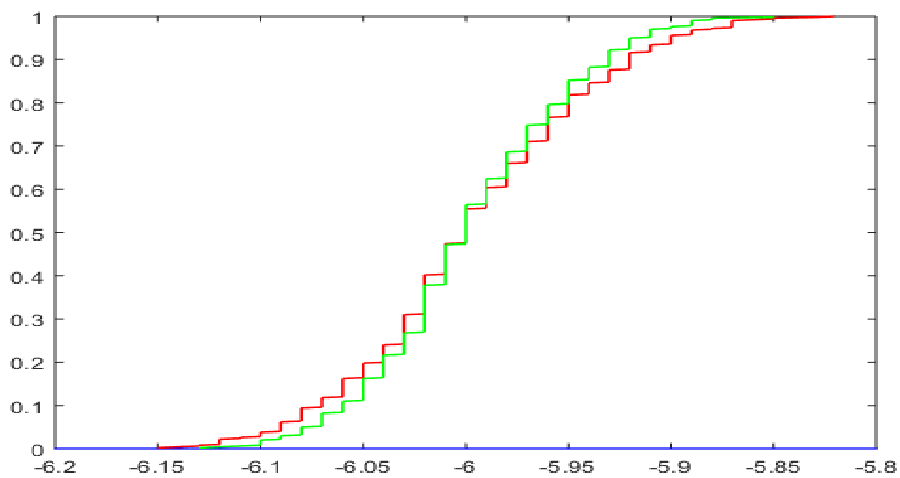
Source: author's computations.

Figure 4.4: EDF of $\hat{\beta}_4^{(LWS,n,w)}$ and $\hat{\beta}_4^{(AR,n,w)}$, cca 2% of outliers



Source: author's computations.

Figure 4.5: EDF of $\hat{\beta}_5^{(LWS,n,w)}$ and $\hat{\beta}_5^{(AR,n,w)}$, cca 5% of outliers



Source: author's computations.

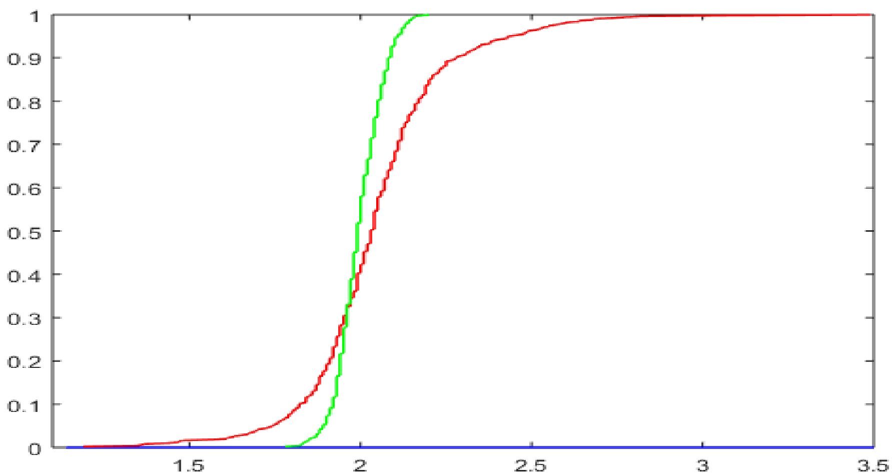
Table 4.19 and Table 4.20 suggest that when we exceed the 5% level of contamination (i.e. we start to take some of the contaminated data into account), the LWS estimator worsens faster than the asymptotic representation of this estimator. The same result follows also from Figure 4.6 and Figure 4.7, where the EDF of LWS estimator starts to spread and the EDF of the asymptotic representation remains very steep. Also the p-values of the K-S tests are very low.

Table 4.19: Heteroskedastic errors, cca 10% of outliers

Level of contamination= 10.11%					
True β^0	2.000	4.000	5.000	-3.000	-6.000
$\hat{\beta}^{(OLS,n)}$	$8.636_{(54.9530)}$	$17.230_{(188.4013)}$	$21.551_{(291.1745)}$	$-12.903_{(110.1769)}$	$-25.672_{(406.1717)}$
$\hat{\beta}^{(LWS,n,w)}$	$2.047_{(0.0508)}$	$4.104_{(0.0788)}$	$5.141_{(0.0786)}$	$-3.069_{(0.0422)}$	$-6.146_{(0.0683)}$
$\hat{\beta}^{(AR,n,w)}$	$1.997_{(0.0044)}$	$3.994_{(0.0042)}$	$4.997_{(0.0042)}$	$-2.998_{(0.0040)}$	$-5.996_{(0.0041)}$

Source: author's computations.

Figure 4.6: EDF of $\hat{\beta}_1^{(LWS,n,w)}$ and $\hat{\beta}_1^{(AR,n,w)}$, cca 10% of outliers



Source: author's computations.

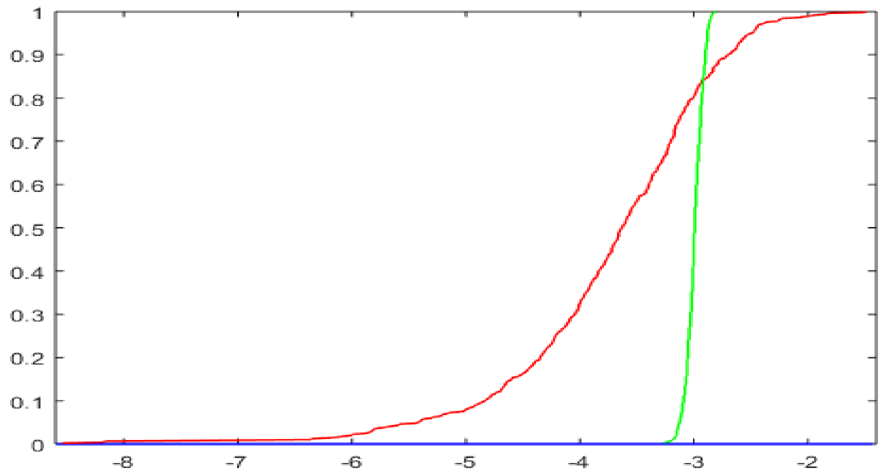
To conclude, the simulations showed that for the model with heteroskedastic errors, where the contamination is caused by outliers, the asymptotic representation represents the LWS estimator really well up to the contamination level of 5%. When we cross the contamination level, where all of the contaminated data are omitted, the results start to move away from each other. Of course, the results here also confirm that the method of OLS is not able to cope with even very small percentage of data contamination.

Table 4.20: Heteroskedastic errors, cca 15% of outliers

Level of contamination= 15.17%					
True β^0	2.000	4.000	5.000	-3.000	-6.000
$\hat{\beta}^{(OLS,n)}$	$11.021_{(93.0467)}$	$22.140_{(343.5782)}$	$27.536_{(523.6570)}$	$-16.657_{(198.2240)}$	$-33.009_{(747.1523)}$
$\hat{\beta}^{(LWS,n,w)}$	$2.387_{(0.7600)}$	$4.865_{(1.5158)}$	$6.101_{(2.1426)}$	$-3.729_{(1.3888)}$	$-7.309_{(2.6416)}$
$\hat{\beta}^{(AR,n,w)}$	$1.997_{(0.0058)}$	$3.994_{(0.0055)}$	$4.991_{(0.0061)}$	$-2.994_{(0.0055)}$	$-5.994_{(0.0053)}$

Source: author's computations.

Figure 4.7: EDF of $\hat{\beta}_4^{(LWS,n,w)}$ and $\hat{\beta}_4^{(AR,n,w)}$, cca 15% of outliers



Source: author's computations.

4.3.2 Leverage points

Let us move on to the second method of contamination, i.e. we will study, whether the results obtained in previous subsection will differ, when we contaminate the data by leverage points instead of outliers. The case with no contamination is again of course the same, therefore we can start with the level of contamination 0.5% in Table 4.21. Similarly as in the case of outliers, the method of OLS is not able to cope with even this small level of contamination. The LWS estimator and its asymptotic representation again seem to give very similar results.

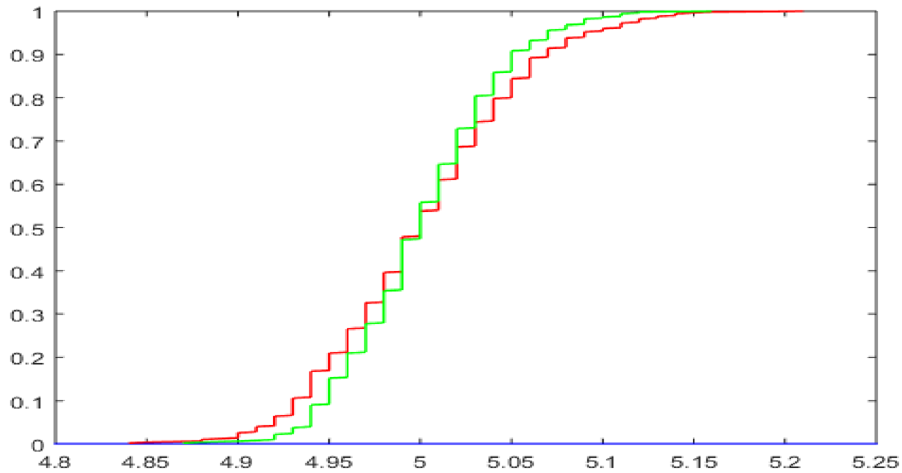
This claim is also supported by the plots of the EDFs as can be seen in Figure 4.8. The flatter (red) function represents again the LWS estimator and the steeper (green) one stands for the asymptotic representation. The p-value of the K-S test for these two samples is 0.1. Therefore we can again conclude that the functions come from the same distribution.

Table 4.21: Heteroskedastic errors, cca 0.5% of leverage points

Level of contamination= 0.48%					
True β^0	2.000	4.000	5.000	-3.000	-6.000
$\hat{\beta}^{(OLS,n)}$	$1.640_{(1.1085)}$	$3.224_{(1.8941)}$	$4.110_{(2.5185)}$	$-2.446_{(1.4876)}$	$-4.763_{(3.3721)}$
$\hat{\beta}^{(LWS,n,w)}$	$1.999_{(0.0031)}$	$4.000_{(0.0032)}$	$5.001_{(0.0030)}$	$-3.000_{(0.0029)}$	$-5.996_{(0.0033)}$
$\hat{\beta}^{(AR,n,w)}$	$1.999_{(0.0018)}$	$4.001_{(0.0019)}$	$5.001_{(0.0018)}$	$-3.000_{(0.0017)}$	$-5.997_{(0.0019)}$

Source: author's computations.

Figure 4.8: EDF of $\hat{\beta}_3^{(LWS,n,w)}$ and $\hat{\beta}_3^{(AR,n,w)}$, cca 0.5% of leverage points



Source: author's computations.

Also the rest of the analysis appears to lead to similar conclusions that we made for ouliers. The only significant difference is again that leverage points have somewhat larger influence on the estimates than outliers, similarly as in the case of homoskedastic errors. The results for data with 1%, 2% and 5% leverage points can be found in Table 4.22, Table 4.23 and Table 4.24, respectively.

Corresponding plots of the EDFs can be found in Figure 4.9, Figure 4.10 and Figure 4.11. The steeper (green) function in all these figures illustrates again the asymptotic representation. The resulting p-values from the two-sample K-S tests are 0.17 for 1% of leverage points, 0.14 for 2% of leverage points and 0.29 for 5% of leverage points. Therefore in neither of these cases we can reject the null hypothesis that the functions have the same underlying distribution (on 10% significance level).

Table 4.22: Heteroskedastic errors, cca 1% of leverage points

Level of contamination= 1.03%					
True β^0	2.000	4.000	5.000	-3.000	-6.000
$\hat{\beta}^{(OLS,n)}$	$1.228_{(2.7614)}$	$2.404_{(4.6888)}$	$3.100_{(6.4333)}$	$-1.850_{(3.3792)}$	$-3.617_{(9.5116)}$
$\hat{\beta}^{(LWS,n,w)}$	$1.997_{(0.0030)}$	$3.997_{(0.0031)}$	$5.003_{(0.0026)}$	$-3.004_{(0.0028)}$	$-6.004_{(0.0026)}$
$\hat{\beta}^{(AR,n,w)}$	$1.998_{(0.0018)}$	$3.998_{(0.0019)}$	$5.002_{(0.0016)}$	$-3.003_{(0.0017)}$	$-6.003_{(0.0016)}$

Source: author's computations.

Table 4.23: Heteroskedastic errors, cca 2% of leverage points

Level of contamination= 1.99%					
True β^0	2.000	4.000	5.000	-3.000	-6.000
$\hat{\beta}^{(OLS,n)}$	$0.739_{(4.5330)}$	$1.262_{(11.2618)}$	$1.689_{(15.3432)}$	$-0.914_{(7.4647)}$	$-1.972_{(20.7739)}$
$\hat{\beta}^{(LWS,n,w)}$	$2.001_{(0.0027)}$	$4.004_{(0.0030)}$	$5.000_{(0.0032)}$	$-3.004_{(0.0034)}$	$-6.002_{(0.0030)}$
$\hat{\beta}^{(AR,n,w)}$	$2.001_{(0.0017)}$	$4.003_{(0.0019)}$	$5.000_{(0.0020)}$	$-3.003_{(0.0021)}$	$-6.002_{(0.0019)}$

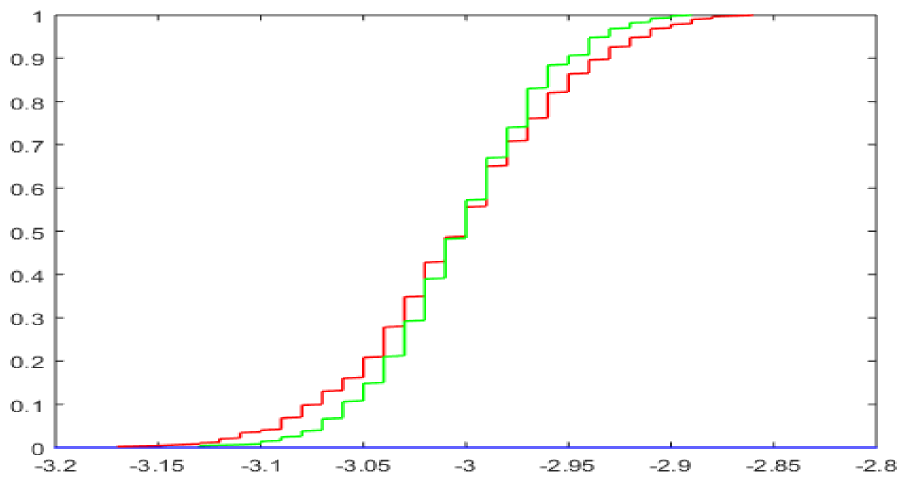
Source: author's computations.

Table 4.24: Heteroskedastic errors, cca 5% of leverage points

Level of contamination= 4.86%					
True β^0	2.000	4.000	5.000	-3.000	-6.000
$\hat{\beta}^{(OLS,n)}$	$-0.861_{(13.3486)}$	$-1.740_{(39.4102)}$	$-2.104_{(57.6227)}$	$1.251_{(22.9531)}$	$2.637_{(82.1514)}$
$\hat{\beta}^{(LWS,n,w)}$	$2.002_{(0.0033)}$	$3.999_{(0.0036)}$	$5.002_{(0.0036)}$	$-3.001_{(0.0032)}$	$-5.999_{(0.0034)}$
$\hat{\beta}^{(AR,n,w)}$	$2.002_{(0.0023)}$	$4.000_{(0.0024)}$	$5.001_{(0.0025)}$	$-3.001_{(0.0022)}$	$-6.000_{(0.0024)}$

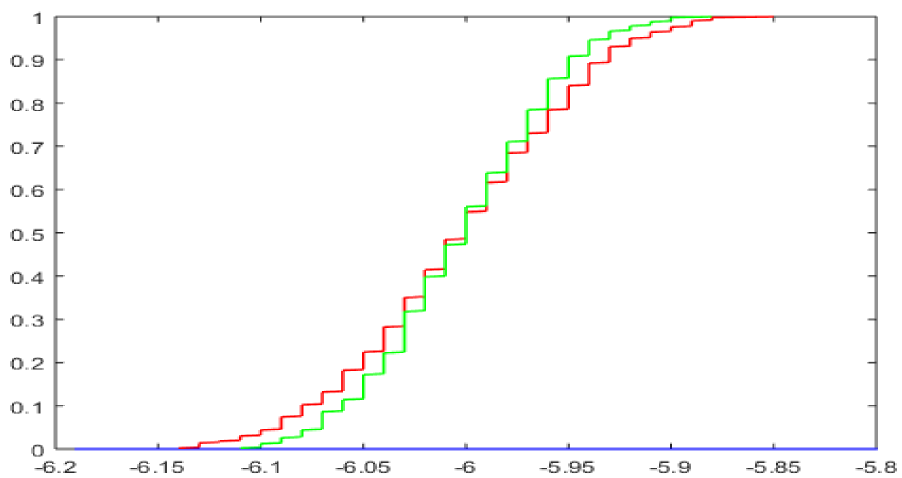
Source: author's computations.

Figure 4.9: EDF of $\hat{\beta}_4^{(LWS,n,w)}$ and $\hat{\beta}_4^{(AR,n,w)}$, cca 1% of leverage points



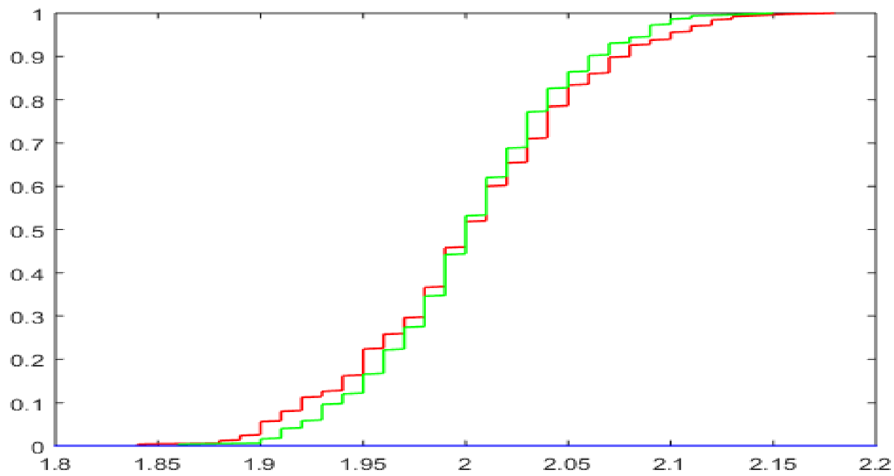
Source: author's computations.

Figure 4.10: EDF of $\hat{\beta}_5^{(LWS,n,w)}$ and $\hat{\beta}_5^{(AR,n,w)}$, cca 2% of leverage points



Source: author's computations.

Figure 4.11: EDF of $\hat{\beta}_1^{(LWS,n,w)}$ and $\hat{\beta}_1^{(AR,n,w)}$, cca 5% of leverage points



Source: author's computations.

Similarly as for outliers, the situation changes after crossing the boundary of 5% level of contamination. The results for the contamination level of 10% and 15% can be found in Table 4.25 and Table 4.26, respectively. Corresponding plots of the EDFs can be found in Figure 4.12 and Figure 4.13. The p-values of the K-S tests are very low in both cases and hence for higher level of contamination we can reject the null hypothesis.

Table 4.25: Heteroskedastic errors, cca 10% of leverage points

Level of contamination= 9.94%					
True β^0	2.000	4.000	5.000	-3.000	-6.000
$\hat{\beta}^{(OLS,n)}$	$-3.144_{(32.5133)}$	$-6.436_{(117.2655)}$	$-7.874_{(176.0813)}$	$4.436_{(63.4384)}$	$9.363_{(246.1871)}$
$\hat{\beta}^{(LWS,n,w)}$	$1.962_{(0.0259)}$	$3.919_{(0.0364)}$	$4.898_{(0.0463)}$	$-2.941_{(0.0367)}$	$-5.892_{(0.0394)}$
$\hat{\beta}^{(AR,n,w)}$	$1.996_{(0.0043)}$	$4.001_{(0.0043)}$	$4.996_{(0.0038)}$	$-3.000_{(0.0039)}$	$-5.996_{(0.0041)}$

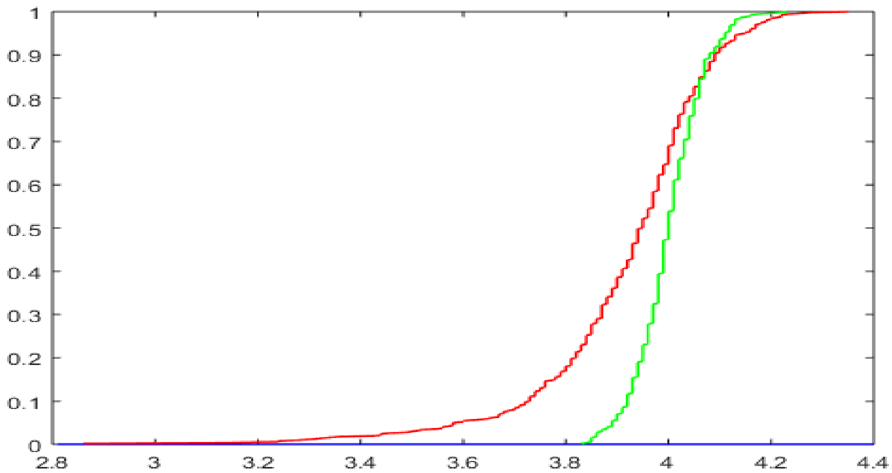
Source: author's computations.

Table 4.26: Heteroskedastic errors, cca 15% of leverage points

Level of contamination= 15.12%					
True β^0	2.000	4.000	5.000	-3.000	-6.000
$\hat{\beta}^{(OLS,n)}$	$-5.249_{(59.1848)}$	$-10.224_{(210.9355)}$	$-12.867_{(329.2009)}$	$7.613_{(120.2425)}$	$15.262_{(464.4025)}$
$\hat{\beta}^{(LWS,n,w)}$	$1.662_{(0.5733)}$	$3.337_{(0.9299)}$	$4.109_{(1.3942)}$	$-2.457_{(0.7643)}$	$-4.947_{(1.8217)}$
$\hat{\beta}^{(AR,n,w)}$	$2.000_{(0.0055)}$	$3.992_{(0.0054)}$	$4.985_{(0.0056)}$	$-2.998_{(0.0057)}$	$-5.993_{(0.0058)}$

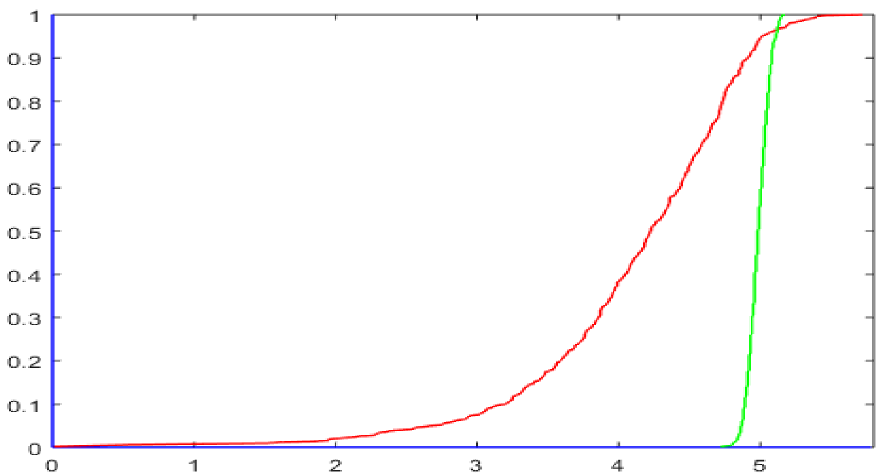
Source: author's computations.

Figure 4.12: EDF of $\hat{\beta}_2^{(LWS,n,w)}$ and $\hat{\beta}_2^{(AR,n,w)}$, cca 10% of leverage points



Source: author's computations.

Figure 4.13: EDF of $\hat{\beta}_3^{(LWS,n,w)}$ and $\hat{\beta}_3^{(AR,n,w)}$, cca 15% of leverage points



Source: author's computations.

It follows that when considering a regression model with heteroskedastic errors, we can draw similar conclusions for the case when the data are contaminated by leverage points as we did for outliers. I.e. the LWS estimator seems to be really well represented by its asymptotic representation up to the contamination level of 5%. The results for leverage points may be slightly more persuasive than the results for outliers, since in none of the considered cases we get too close to the 5% significance level when conducting the K-S test.

The results for higher contamination levels suggest that the underlying distributions of $\hat{\beta}^{(LWS,n,w)}$ and $\hat{\beta}^{(AR,n,w)}$ differ when there are too many contaminated observations. It follows that the optimal choice of weight function is even more crucial when considering the asymptotic representation than when we just want to obtain a reliable LWS estimator.

Chapter 5

Conclusion

The main purpose of this thesis was to derive the asymptotic representation of the least weighted squares estimator under the assumption of heteroskedasticity. After introducing the robust methods in general, and the method of LWS more in detail in Chapter 2, the derivation of the asymptotic representation is provided in Chapter 3. Moreover, Chapter 4 provides results of a simulation study illustrating the tightness of the LWS estimator and the derived asymptotic representation.

The derivation under the assumption of heteroskedastic residuals provided in Chapter 3 is a generalization of the homoskedastic case derived in Víšek (2015). To be able to generalize this result for heteroskedasticity, we had to impose some more assumptions, mainly specifying the form of heteroskedasticity. However, none of these assumptions seems to be too restrictive for a commonly considered regression framework.

Then using the adjusted assumptions, it was necesssary to rederive most of the proofs of the lemmas used in Víšek (2015). Moreover, some additional lemmas had to be formulated and proved in order to be able to derive the result under heteroskedasticity. This is the main contribution of the thesis.

Another original contribution are the results of the simulation study, which illustrate the theoretical result derived in Chapter 3. In the simulations in Chapter 4, we considered also the situation with homoskedastic residuals for completeness. However, the main result of this thesis is illustrated in Section 4.3, where heteroskedastic residuals were considered.

These results suggest that the LWS estimator is well represented by its asymptotic representation, as long as most of the contaminated observations are completely eliminated by the weight function. In that case, the empirical distribution functions of the LWS estimator and of the derived representation are statistically the same (for both outliers and leverage points). The situation changes when some of the contaminated observations are assigned non-zero weights - then the results start to move away from each other. It follows that when considering the asymptotic representation, the optimal choice of the weight function is even more crucial than when we just want to obtain a reliable LWS estimator.

The result derived in this thesis might be used for future research, e.g. for development of some of the diagnostic tools for the estimators based on the method of LWS, such as the specification test. Another possibility of future research might be to simplify the procedure of deriving the asymptotic representation. A possible (but so far unsuccessful) way to do this seems to be to combine the asymptotic linearity of normal equations with the convergence of empirical distribution function.

Bibliography

BRAMATI, M. C. & C. CROUX (2007): "Robust estimators for the fixed effects panel data model." *The Econometrics Journal* **10**: pp. 521–540.

CAMPBELL, N. & P. J. ROUSSEEUW (1998): "On the calculation of a robust s-estimator of a covariance matrix." *Statistics in Medicine* **(17)**: pp. 2685–2695.

DONOHO, D. L. & P. J. HUBER (1983): "The notion of breakdown point." *A Festschrift for Erich Lehmann* .

FISHER, R. A. (1920): "A mathematical examination of the methods of determining the accuracy of an observation by the mean error and by the mean square error." *Monthly Notices of the Royal Astronomical Society* **80**. Reprinted in Collected Papers of R. A. Fisher, ed. J. H. Bennett, Vol.1, 188–201, University of Adelaide.

FISHER, R. A. (1922): "On the mathematical foundations of theoretical statistics." *Philosophical Transactions of the Royal Society of London* **222**: pp. 309–368.

GERVINI, D. & V. YOHAI (2002): "A class of fully efficient regression estimators." *The Annals of Statistics* **30(2)**: pp. 583–616.

GREENE, W. H. (2012): *Econometric analysis.* Pearson, 7th edition.

HAMPEL, F. R. (1968): *Contributions to the theory of robust estimation.* Ph.D. thesis, University of California, Berkeley.

HAMPEL, F. R. (1971): "A general qualitative definition of robustness." *The Annals of Mathematical Statistics* **42**: pp. 1887–1896.

HAMPEL, F. R., E. M. RONCHETTI, P. J. ROUSSEEUW, & W. A. STAHEL (1986): *Robust statistics – the approach based on influence functions.* New York: Wiley.

HÁJEK, J. & Z. ŠIDÁK (1967): "Theory of rank tests." *Academia, Prague and Academic Press, New York* .

HODGES, Jr., J. L. (1967): "Efficiency in normal samples and tolerance of extreme values for some estimates of location." *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* **1**: pp. 163–168.

HUBER, P. J. (1964): "Robust estimation of a location parameter." *The Annals of Mathematical Statistics* **35**: pp. 73–101.

HUBER, P. J. (1981): *Robust statistics*. New York: Wiley.

HUBER, P. J. & V. STRASSEN (1973): "Minimax tests and the neyman–pearson lemma for capacities." *The Annals of Statistics* **1**: pp. 251–263.

JEFFREYS, H. (1961): *Theory of probability*. New York: Oxford University Press, 3rd edition. 1st eddition 1939.

JUREČKOVÁ, J. & P. K. SEN (1989): "Uniform second order asymptotic linearity of m-statistics in linear models." *Statistics and Decisions* **7**: pp. 263–276.

PORTNOY, S. (1983): "Tightness of the sequence of empiric c.d.f. processes defined from regression fractiles." *Robust and Nonlinear Times-Series Analysis, New York: Springer-Verlag* pp. 231–246.

RAO, R. C. (1973): *Linear Statistical Inference and Its Applications*. New York: J.Wiley & Sons.

ROMANOWSKI, M. (1970): *The theory of random errors based on the concept of modulated normal distributions*. National Research Council of Canada, Division of Physics.

ROUSSEEUW, P. J. (1984): "Least median of squares regression." *Journal of the American Statistical Association* **79**: pp. 871–880.

ROUSSEEUW, P. J. & A. M. LEROY (1987): *Robust regression and outlier detection*. New York: Wiley.

SIEGEL, A. F. (1982): "Robust regression using repeated medians." *Biometrika* **69**: pp. 242–244.

Štěpán, J. (1987): *Teorie pravděpodobnosti (Theory of probability)*. Academia, Praha.

Tukey, J. W. (1960): "A survey of sampling from contaminated distributions." *Contributions to probability and statistics* pp. 448–485. Stanford University Press.

Víšek, J. A. (2000a): "On the diversity of estimates." *Computational Statistics and Data Analysis* **34**: pp. 67–89.

Víšek, J. A. (2000b): "Regression with high breakdown point." *Proc. ROBUST 2000* pp. 324–356.

Víšek, J. A. (2002a): "The least weighted squares i. the asymptotic linearity of normal equations." *Bulletin of the Czech Econometric Society* **9(15)**: pp. 31–58.

Víšek, J. A. (2002b): "The least weighted squares ii. consistency and asymptotic normality." *Bulletin of the Czech Econometric Society* **9(16)**: pp. 1–28.

Víšek, J. A. (2006): "The least trimmed squares. part i - consistency. part ii - sqrt(n)-consistency. part iii - asymptotic normality and bahadur representation." *Kybernetika* **42**: pp. 1–36,181–202,203–224.

Víšek, J. A. (2010): "Weak sqrt(n)-consistency of the least weighted squares under heteroskedasticity." *Acta Universitatis Carolinae* **2/51**: pp. 71–82.

Víšek, J. A. (2011a): "Consistency of the least weighted squares under heteroskedasticity." *Kybernetika* **47**: pp. 179–206.

Víšek, J. A. (2011b): "Empirical distribution function under heteroskedasticity." *Statistics* **45**.

Víšek, J. A. (2012a): "Advantages and disadvantages, challenges and threads of robust methods." Faculty of Social Sciences, Charles University, Prague, personal correspondence.

Víšek, J. A. (2012b): "Robust estimation of model with the fixed and random effects." *COMPSTAT 2012 Proceedings* pp. 855–865. The international statistical institute/ International association for statistical computing.

Víšek, J. A. (2015): "Representation of the least weighted squares." *Advances and Applications in Statistics* **47**: pp. 91–144.

Wooldridge, J. M. (2009): *Introductory econometrics: a modern approach*. South-Western, Cengage Learning, 4th edition.