Three Essays in Econometrics

Inaugural-Dissertation
zur Erlangung des Grades eines Doktors
der Wirtschafts- und Gesellschaftswissenschaften
durch die
Rechts- und Staatswissenschaftliche Fakultät
der Rheinischen Friedrich-Wilhelms-Universität
Bonn

vorgelegt von
Thomas Nebeling
aus Siegen

Bonn 2015

Dekan:	Prof. Dr. Rainer Hüttemann
Erstreferent:	Prof. Dr. Jörg Breitung
Zweitreferent:	Prof. Dr. Christian Pigorsch

Tag der mündlichen Prüfung:	2. Februar 2016
Erscheinungsjahr:	2016

Diese Dissertation ist auf dem Hochschulschriftenserver der ULB Bonn (http://hss.ulb.uni-bonn.de/diss_online) elektronisch publiziert.

Acknowledgments

While I was writing this thesis, I received help and support of many people and I owe all of them an enormous amount of gratitude.

First and foremost, I would like to thank my main supervisor Jörg Breitung. He awakened my interest in the broad field of econometrics when I was a diploma student at the University of Bonn. I thank him for his steady support, patience, and constructive comments throughout the years.

Many thanks go to the former econometrics department at the University of Bonn. Especially I would like to thank Christian Pigorsch who accepted to act as my second advisor and to Matei Demetrescu whose invaluable comments helped to improve the first chapter.

I am indebted to my coauthors Nazarii Salish (Chapter 1) and Anna Bindler (Chapter 3). I really enjoyed working with you and I am grateful for many hours of fruitful discussions.

Furthermore, I would like to thank my fellow PhD students at the Bonn Graduate School of Economics, especially Max Conze and Uli Homm who enriched my academic as well as non-academic experiences during my time as a PhD student.

Last but not least, I would like to thank my family and my friends for their unconditional support. A very special thanks goes to Katharina. Thank you!

Contents

1 LM Tests for Shock Induced Asymmetries in Time Series **3**

1.1 Introduction 3

1.2 Preliminaries 5

1.3 The Lagrange Multiplier Test 7

1.4 Variants of the LM Test 9

1.4.1 Empirical Hessian and information matrix 9

1.4.2 OPG variant 11

1.4.3 Other regression based variants 12

1.5 Asymptotics 13

1.6 Monte Carlo Simulations 14

1.6.1 Normally distributed errors 14

1.6.2 Errors with skewed distribution 15

1.6.3 Conditional heteroskedasticity 17

1.6.4 Model Selection 18

1.7 Application 19

1.7.1 U.S. Unemployment Rate 20

1.7.2 U.S. Industrial Production 20

1.8 Conclusion 21

A Appendix: Proofs 22

B Appendix: Tables 28

C Appendix: Figures 34

2 Variance Bounds Tests and Forecast Evaluation **45**
2.1 Introduction 45
2.2 First Generation Tests: Shiller's Test 47
2.3 Second Generation Tests: Mankiw, Shapiro and Romer's Test 50
2.4 Variance Bounds Tests and Forecast Evaluation 52
2.5 Application 54
2.5.1 Simple Forecasts 55
2.5.2 Cointegration 56
2.5.3 Multivariate Forecasts 60
2.6 Conclusion 67

3 Social Networks and the Distribution of Wealth **69**
3.1 Introduction 69
3.2 Previous Literature 71
3.3 Data 73
3.4 The Institutional Background 76
3.5 Empirical Approach 78
3.6 Empirical Evidence 82
3.6.1 Descriptive Evidence 82
3.6.2 First Stage, Reduced Form and IV Estimation 85
3.6.3 Conditional and Unconditional QTE Estimation 87
3.6.4 IV Estimation by Wealth and Asset Type 92
3.7 Conclusion 93
Appendix 95

Bibliography **103**

List of Figures

1.1 Power of the LM_T Variants when $\varepsilon_t \sim N(0,1)$ . . . 35

1.2 Kernel Density Estimate of $\varepsilon_t \sim \mathcal{B}(0,1,0.8,3)$ . . . 36

1.3 Power of the $LM_T^{(IM)}$ Test when $\varepsilon_t \sim \mathcal{B}(0,1,0.8,3)$ . . . 37

1.4 Power of the $LM_T^{(OPG)}$ Test when $\varepsilon_t \sim \mathcal{B}(0,1,0.8,3)$ . . . 38

1.5 Power of the $LM_T^{(RB)}$ Test when $\varepsilon_t \sim \mathcal{B}(0,1,0.8,3)$ . . . 39

1.6 Power of the Robustified Version of the $LM_T^{(IM)}$ Test when $\varepsilon_t \sim \mathcal{B}(0,1,0.8,3)$. 40

1.7 Power of the $LM_T^{(OPG)}$ Test when $\varepsilon_t \sim GARCH(1,1)$ . . . 41

1.8 Power of the LM_T Variants when the Lag Length is Determined in a Preliminary Stage on the basis of the BIC . . . 42

1.9 Quarterly U.S. Unemployment Rate in % . . . 43

1.10 Monthly U.S. Industrial Production Index . . . 44

2.1 Real Stock Price and Real Dividend Series . . . 58

2.2 Differenced Real Stock Price and Real Dividend Series . . . 64

3.1 West and East German Territories . . . 77

3.2 Total Wealth . . . 83

3.3 Safe Assets . . . 83

3.4 Risky Assets . . . 83

3.5 Real Estate . . . 83

3.6 Non-Financial Assets . . . 83

3.7 Debt . . . 83

3.8 Conditional Quantile Treatment Effects . . . 89

3.9 Unconditional Quantile Treatment Effects 91

List of Tables

1.1 Rejection frequencies (in %) under the null of no asymmetric effects for AsMA processes (left panel) and AsAR processes (right panel); the nominal size is 5%. $\varepsilon_t \sim N(0,1)$. . . . 28

1.2 Rejection frequencies (in %) under the null of no asymmetric effects for AsMA processes (left panel) and AsAR processes (right panel); the nominal size is 5%. $\varepsilon_t \sim \mathcal{B}(0,1,0.8,3)$. . . . 29

1.3 Rejection frequencies (in %) under the null of no asymmetric effects for AsMA processes (left panel) and AsAR processes (right panel); the nominal size is 5%. $\varepsilon_t \sim GARCH(1,1)$. . . . 30

1.4 Size of the $maxLM_T(IM)$ test without and with Bonferroni correction in case of an $MA(1)$; the nominal size is 5% 31

1.5 Model Selection based decision frequencies (in %) under different AsMA DGPs . 32

1.6 Model Selection based decision frequencies (in %) under different AsAR DGPs . 33

2.1 Diebold-Mariano Test: Simple Forecasts 57

2.2 Augmented Dickey-Fuller Test: Price, Dividend, Ex Post Rational Price 59

2.3 Johansen and Engle-Granger Cointegration Test 60

2.4 Diebold-Mariano Test: Multivariate Forecasts 65

3.1 Wealth and Wealth Components 95

3.2 Weighted Descriptive Statistics 96

3.3 First Stage Estimation Results 97

3.4 Additional First Stage Estimation Results 98

3.5 Reduced Form, OLS and IV (2SLS) Results 99
3.6 Robustness Check - Functional Form (Log-Level) 100
3.7 IV (2SLS) Estimation - Wealth Decile Interactions 101

Introduction

This thesis consists of three self-contained chapters in applied and theoretical econometrics. The first chapter presents a Lagrange Multiplier test for detecting shock triggered asymmetries in time series and investigates its properties. Chapter 2 links variance bounds tests to forecast evaluation tests. Exploiting the cointegration relation between prices and dividends, we find rejections of the null hypothesis of market efficiency. The last chapter analyses the effect of social networks on wealth and wealth components on the basis of a novel dataset. Distributional effects are investigated in an instrumental variable quantile regression framework.

In particular, *Chapter 1*, joint work with Nazarii Salish, develops a Lagrange Multiplier test statistic and its variants to test for the null hypothesis of no asymmetric effects of innovations on time series. The test is built on asymmetric time series models that allow for different responses to positive and negative past shocks, and for which the likelihood functions, in general, are discontinuous. By making use of the theory of generalized functions, the Lagrange Multiplier type tests as well as the resulting asymptotics are derived. The suggested test statistics possess a standard asymptotic limiting behaviour under the null hypothesis. Monte Carlo experiments illustrate the accuracy of the asymptotic approximation. Moreover, it is shown that conventional model selection criteria can be used to estimate lag lengths required for the implementation of the suggested tests. Empirical applications to the quarterly U.S. unemployment rate and to the monthly U.S. industrial production index are provided.

Chapter 2 contributes to the variance bounds literature. There is a long-standing debate whether stock prices are more volatile than traditional models imply. The chapter links variance bounds tests to the forecast evaluation literature. The Diebold Mariano test is introduced as a robust tool to overcome drawbacks of former testing procedures. Contrary to previous studies, we find that naive forecasts are not able to outperform the market price as a predictor for the ex post rational price. We further estimate a panoply of multivariate models to construct more sophisticated forecasts based on the cointegration relation between prices and dividends. Moreover, we distinguish between iterated and direct multi-step forecasts. We provide an application of our test procedure based on the Standard & Poor's 500 Composite Price Index. When we test the null hypothesis of market efficiency, we find rejections for longer holding periods of the underlying asset.

Chapter 3, joint work with Anna Louisa Bindler, uses data from a novel dataset to estimate the effect of social networks on the distribution of wealth in Germany. In particular, we study the relationship between wealth and church as an example for a strong social network. First, we exploit the natural experiment of Germany's reunification to identify the local average treatment effect on wealth. Second, we use an IV-quantile regression framework to show heterogenous effects along the distribution of wealth and estimate conditional as well as unconditional quantile treatment effect models. Our results show that church affiliation indeed has a positive impact on wealth that is driven by church attendance. We argue that this can be interpreted as evidence that the effect of church affiliation is driven by specific social network effects. We test for confounding factors such as risk aversion and other personality traits to single out the social network mechanism. Given the very rich dataset, we are able to disaggregate the results for different asset types and show heterogeneous effects for different asset classifications.

Chapter 1

LM Tests for Shock Induced Asymmetries in Time Series

1.1 Introduction

In the last decades there has been a significant increase in findings from empirical studies in economics and finance which indicate that the response of the economic processes is asymmetric with respect to positive or negative shocks (see, for instance, Elwood, 1998; Koutmos, 1999; Karras and Stokes, 1999; Kilian and Vigfusson, 2011; Brännäs et al., 2012, among others). In the univariate time series literature this led to an asymmetric time series paradigm as introduced by Wecker (1981). Wecker suggested asymmetric moving average models (henceforth AsMA) to model the asymmetries triggered by the sign of innovations.

In this chapter, we consider alternative generalizations of time series models in the context of asymmetries. In particular, an extension of (linear) autoregressive to asymmetric processes (henceforth AsAR) is suggested in Section 1.2. Here, the asymmetry is generated by a distinct influence of positive and negative past shocks on the underlying process. To the best of our knowledge, this type of model has not been considered in the literature before. Alternative extensions of classical time series models could be relevant, too, in which the potential presence of asymmetries triggered by shocks rises the question of (pre)testing for the correct model specification.

This testing problem has been discussed in the literature. Wecker (1981) suggested the use of the likelihood ratio test to test for the conventional moving average model against AsMA. Brännäs and De Gooijer (1994) constructed a Wald-type test to choose the correct model specification. Brännäs et al. (1998) considered a test statistic based on the artificial regression constructed from the Lagrange multiplier (henceforth LM) principle. Yet, these studies focus on presenting the idea of new asymmetric time series models and do not elaborate on the discontinuous nature of these models. More precisely, in the setting of AsMA models, the respective log-likelihood function is always discontinuous. Therefore, the standard approach of deriving the gradient and the Hessian from the log-likelihood function as well as the asymptotic behaviour of likelihood based statistics are not valid anymore.

In this chapter we contribute to the literature by suggesting new test statistics based on the Lagrange multiplier approach which account for discontinuities in the log-likelihood function. The tests are aimed for asymmetric time series models such as AsMA and AsAR. However, the suggested techniques can be used for other asymmetric time series models, too. We apply the treatment of non-differentiability as offered by Phillips (1991) for LAD estimators in order to deal with the non-smooth log-likelihood function. The main idea is to examine our problem in the mathematical space of generalized functions (distributions) whose derivatives do not exist in the classical sense, however can be formalized as (accommodated by) distributional derivatives. This solution allows us to operate with first order conditions and thereby derive LM type test statistics.

In fact, with this generalization of the classical approach, the asymptotic properties of the test statistics can be obtained. It is shown that the limiting distribution is a standard χ^2 distribution under the null of no asymmetric effects. Further, by means of Monte Carlo simulations the finite sample properties of the new test statistics are explored in different setups. Finally, in order to make testing procedures more applicable, we suggest a solution to select an appropriate model specification for the test implementation. We show in Monte Carlo experiments that the use of standard model selection criteria, such as the BIC or HQ, applied to a linear time series model provide a reliable estimate of the required lag length.

We apply these methods to the U.S. unemployment rate and the U.S. industrial production index. We find strong evidence that the growth of the unemployment rate as well as the U.S. industrial production index are affected by an asymmetric impact of positive/negative shocks.

The outline of this chapter is as follows. Section 1.2 introduces the notation as well as the modelling framework. The construction of the LM type tests is described in Section 1.3. Section 1.4 presents variants of the LM test. In Section 1.5 the asymptotic properties of the statistics are investigated. In Section 1.6 we present results from simulation studies. Two empirical examples are given in Section 1.7. The final section contains concluding remarks.

1.2 Preliminaries

This section introduces the asymmetric time series models as a counterpart to the conventional linear moving average and autoregressive models. The main characteristic of this model class which distinguishes it from other, well established, nonlinear models (such as threshold AR models for instance) is that two distinct filters, one for positive and one for negative innovations, are used. Wecker (1981) advocates the use of the asymmetric moving average model which takes the following form:

$$y_t = \varepsilon_t + \alpha_1 \varepsilon_{t-1} + ... + \alpha_p \varepsilon_{t-p} + \beta_1 \varepsilon_{t-1}^{+} + ... + \beta_p \varepsilon_{t-p}^{+}, \tag{1.1}$$

where $\varepsilon_t^{+} = \varepsilon_t 1 (\varepsilon_t \geq 0)$ and $1(\cdot)$ defines an indicator function. We extent Wecker's approach by considering the asymmetric autoregressive model defined as:

$$y_t = \varepsilon_t - \alpha_1 y_{t-1} - ... - \alpha_p y_{t-p} - \beta_1 y_{t-1}^{+} - ... - \beta_p y_{t-p}^{+}, \tag{1.2}$$

where $y_t^{+} = y_t 1 (\varepsilon_t \geq 0)$. In both models, it is assumed that $y_t = 0$ for $t \leq 0$ and that the random disturbance term ε_t is a real *i.i.d.* sequence with a $N(0, \sigma^2)$ distribution. Note that the normality assumption is necessary for the derivation of the LM statistics only. For the applications as well as for the derivations of asymptotic results this assumption can be relaxed. In general, it is necessary for the asymptotic analysis that the process y_t is stationary and invertible under the null hypothesis of no asymmetric effects. For this reason it is assumed that the roots of $\alpha(z) = 1 + \sum_{i=1}^{p} \alpha_i z^i$ lie outside the unit circle.[1]

To express models (1.1) and (1.2) in matrix notation, define $\mathbf{B}$ as a $T \times T$ backshift matrix with typical element $B_{ij} = 1$ if $i - j = 1$ and zero otherwise. $\mathbf{D}_{1(\varepsilon)} = diag\{1(\varepsilon_1 \geq 0), ...,$ $1(\varepsilon_T \geq 0)\}$ is a $T \times T$ diagonal matrix and $\boldsymbol{\alpha} \equiv (\alpha_1, ..., \alpha_p)'$, $\boldsymbol{\beta} \equiv (\beta_1, ..., \beta_p)'$ are parameter vectors. Therefore, models (1.1) and (1.2) can be rewritten as:

$$\mathbf{y} = \left(\mathbf{M}_{\boldsymbol{\alpha}} + \mathbf{M}_{\boldsymbol{\beta}} \mathbf{D}_{1(\boldsymbol{\varepsilon})}\right) \boldsymbol{\varepsilon}, \tag{1.3}$$

and

$$\left(\mathbf{M}_{\boldsymbol{\alpha}} + \mathbf{M}_{\boldsymbol{\beta}} \mathbf{D}_{1(\boldsymbol{\varepsilon})}\right) \mathbf{y} = \boldsymbol{\varepsilon}, \tag{1.4}$$

where $\mathbf{M}_{\boldsymbol{\alpha}} = \sum_{k=0}^{p} \alpha_k \mathbf{B}^k$ and $\mathbf{M}_{\boldsymbol{\beta}} = \sum_{k=1}^{p} \beta_k \mathbf{B}^k$, with $\alpha_0 = 1$ and $\mathbf{B}^0 = \mathbf{I}$ being the identity matrix. $\mathbf{y} = (y_1, ..., y_T)'$ denotes a $T \times 1$ vector of time series observations and $\boldsymbol{\varepsilon} = (\varepsilon_1, ..., \varepsilon_T)'$ is a $T \times 1$ vector of error terms.

The matrix representations (1.3) and (1.4) are convenient for our discussion, as deviations from the conventional symmetric MA(p) or AR(p) models are now represented in both cases by the matrix $\mathbf{M}_{\boldsymbol{\beta}}$. Therefore, the main hypothesis of interest can be formulated as follows:

$$\mathrm{H}_0: \; \mathbf{M}_{\boldsymbol{\beta}} = 0, \; (\text{or } \boldsymbol{\beta} = 0)$$

The null hypothesis is tested against the two alternatives that

$$\begin{aligned} \mathrm{H}_A &: \; \{y_t\} \text{ is generated by (1.3) or} \\ \mathrm{H}_B &: \; \{y_t\} \text{ is generated by (1.4).} \end{aligned}$$

[1] To complete the discussion we further outline implications of a violation of the stationarity assumption for the asymptotics. See Remark 1 of Section 1.5 for details.

1.3 The Lagrange Multiplier Test

In this section we derive the Lagrange multiplier test. The corresponding log-likelihood function for time series processes (1.3) and (1.4) is given by

$$\mathcal{L}\left(\boldsymbol{\alpha},\boldsymbol{\beta},\sigma^2\right) = const - \frac{T}{2}\ln\left(\sigma^2\right) - \frac{1}{2\sigma^2}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}, \tag{1.5}$$

where $\boldsymbol{\varepsilon} = \left(\mathbf{M}_{\boldsymbol{\alpha}} + \mathbf{M}_{\boldsymbol{\beta}}\mathbf{D}_{1(\boldsymbol{\varepsilon})}\right)^{-1}\mathbf{y}$ for the AsMA(p) and $\boldsymbol{\varepsilon} = \left(\mathbf{M}_{\boldsymbol{\alpha}} + \mathbf{M}_{\boldsymbol{\beta}}\mathbf{D}_{1(\boldsymbol{\varepsilon})}\right)\mathbf{y}$ for the AsAR(p) model. Let $\boldsymbol{\theta} = (\boldsymbol{\alpha}', \boldsymbol{\beta}')'$ be the parameter vector of interest and $\widehat{\boldsymbol{\theta}}_0 = (\widehat{\boldsymbol{\alpha}}', \mathbf{0})'$ the restricted ML estimator of $\boldsymbol{\theta}_0 = \left(\boldsymbol{\alpha}', \mathbf{0}\right)'$ for the LM test. The parameter σ^2 can be concentrated out. Furthermore, let $\mathbf{s}(\boldsymbol{\theta}) = \partial\mathcal{L}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ denote the score and $\mathcal{H}(\boldsymbol{\theta}) = -\operatorname{plim}_{T\to\infty} T^{-1}\partial^2\mathcal{L}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$ the asymptotic Hessian of the log-likelihood (1.5). It is convenient in this testing framework to use a partitioning of the score $\mathbf{s}(\boldsymbol{\theta}) = \left(\mathbf{s}_{\boldsymbol{\alpha}}(\boldsymbol{\theta})', \mathbf{s}_{\boldsymbol{\beta}}(\boldsymbol{\theta})'\right)'$, with $\mathbf{s}_{\boldsymbol{\alpha}}(\boldsymbol{\theta}) = \partial\mathcal{L}(\boldsymbol{\theta})/\partial\boldsymbol{\alpha}$ and $\mathbf{s}_{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \partial\mathcal{L}(\boldsymbol{\theta})/\partial\boldsymbol{\beta}$, respectively. The asymptotic Hessian matrix can then be expressed as:

$$\mathcal{H}(\boldsymbol{\theta}) = \begin{bmatrix} \mathcal{H}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}(\boldsymbol{\theta}) & \mathcal{H}_{\boldsymbol{\alpha}\boldsymbol{\beta}}(\boldsymbol{\theta}) \\ \mathcal{H}_{\boldsymbol{\beta}\boldsymbol{\alpha}}(\boldsymbol{\theta}) & \mathcal{H}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\theta}) \end{bmatrix}.$$

Here, $\mathcal{H}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}(\boldsymbol{\theta}) = -\operatorname{plim}_{T\to\infty} T^{-1}\partial^2\mathcal{L}(\boldsymbol{\theta})/\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}'$, $\mathcal{H}_{\boldsymbol{\alpha}\boldsymbol{\beta}}(\boldsymbol{\theta}) = -\operatorname{plim}_{T\to\infty} T^{-1}\partial^2\mathcal{L}(\boldsymbol{\theta})/\partial\boldsymbol{\alpha}\partial\boldsymbol{\beta}'$, etc.. Hence, the conventional LM test statistic for testing $\mathbf{H}_0$ against H_A or H_B can be written as:

$$\mathrm{LM}_T = \frac{1}{T}\mathbf{s}_{\boldsymbol{\beta}}\left(\widehat{\boldsymbol{\theta}}_0\right)'\mathbf{V}_{\boldsymbol{\beta}}^{-1}\left(\widehat{\boldsymbol{\theta}}_0\right)\mathbf{s}_{\boldsymbol{\beta}}\left(\widehat{\boldsymbol{\theta}}_0\right), \tag{1.6}$$

where $\mathbf{V}_{\boldsymbol{\beta}}(\boldsymbol{\theta})$ represents the variance of the score $\mathbf{s}_{\boldsymbol{\beta}}(\boldsymbol{\theta})$ and is taken from the respective diagonal block of the $\mathcal{H}(\boldsymbol{\theta})$ matrix, i.e. $\mathbf{V}_{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \mathcal{H}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\theta}) - \mathcal{H}_{\boldsymbol{\beta}\boldsymbol{\alpha}}(\boldsymbol{\theta})\,\mathcal{H}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}(\boldsymbol{\theta})^{-1}\,\mathcal{H}_{\boldsymbol{\alpha}\boldsymbol{\beta}}(\boldsymbol{\theta})$.

Note that the indicator function in the log-likelihood function (1.5) results in discontinuities of the function. Therefore, the standard framework for deriving the LM test and its asymptotics is, in general, not applicable. Phillips (1991) suggests a solution to non-regular problems like discontinuities in the criterion function as for the example of LAD estimator's. In particular, if derivatives do not exist in the usual sense, this may be accommodated directly by the use of generalized functions or distributions.[2] As presented in the following, this generalization does not only provide a justification for the LM test derivation, but it also helps to develop generalized Taylor series expansions of the first order conditions. These, in turn, are useful in order to derive asymptotic properties.

We start with the derivative of the indicator function which can be written as the Dirac delta (generalized) function:

$$\partial 1_{(x\geq 0)}/\partial x = \delta(x)$$

Details on the required properties of the $\delta(x)$ function are given in Appendix A and Lemma 3. In order to simplify the notation we define the matrix $\mathbf{M}_{\boldsymbol{\alpha\beta}} \equiv \mathbf{M}_{\boldsymbol{\alpha}} + \mathbf{M}_{\boldsymbol{\beta}}\mathbf{D}_{1(\boldsymbol{\varepsilon})}$ which essentially represents the filtering structure of processes (1.3) and (1.4). The derivative of $\mathbf{M}_{\boldsymbol{\alpha\beta}}$ with respect to $\boldsymbol{\theta}$ can be compactly written as:

$$\frac{\partial \mathbf{M}_{\boldsymbol{\alpha},\boldsymbol{\beta}}}{\partial \theta_i} = \begin{cases} \mathbf{B}^i + \mathbf{M}_{\boldsymbol{\beta}}\mathbf{D}_{\delta(\boldsymbol{\varepsilon})}\mathbf{D}_{\partial\boldsymbol{\varepsilon}/\partial\theta_i} & \text{for } \theta_i = \alpha_i \\ \mathbf{B}^i\mathbf{D}_{1(\boldsymbol{\varepsilon})} + \mathbf{M}_{\boldsymbol{\beta}}\mathbf{D}_{\delta(\boldsymbol{\varepsilon})}\mathbf{D}_{\partial\boldsymbol{\varepsilon}/\partial\theta_i} & \text{for } \theta_i = \beta_i \end{cases}, \tag{1.7}$$

where $\mathbf{D}_{\delta(\boldsymbol{\varepsilon})}$ is a $T \times T$ diagonal matrix defined as $diag\{\delta(\varepsilon_1), ..., \delta(\varepsilon_T)\}$ and $\mathbf{D}_{\partial\boldsymbol{\varepsilon}/\partial\theta_i} = diag\{\partial\varepsilon_1/\partial\theta_i, ..., \partial\varepsilon_T/\partial\theta_i\}$. Further, under the null hypothesis $\mathbf{M}_{\boldsymbol{\beta}} = 0$ and $\partial\mathbf{M}_{\boldsymbol{\alpha},\boldsymbol{\beta}}/\partial\theta_i$ takes a simple matrix form $\mathbf{B}^i$ or $\mathbf{B}^i\mathbf{D}_{1(\boldsymbol{\varepsilon})}$. Finally, using standard results for matrix derivatives (see e.g. Lütkepohl, 1996), the elements of the score vector $\mathbf{s}_{\boldsymbol{\beta}}\left(\widehat{\boldsymbol{\theta}}_0\right)$ under the null hypothesis can be presented in a quadratic form as:

$$s_{\beta,i}\left(\widehat{\boldsymbol{\theta}}_0\right) = \frac{1}{\widehat{\sigma}^2}\widehat{\boldsymbol{\varepsilon}}'\left(\widehat{\mathbf{M}}_{\boldsymbol{\alpha}}^{-1}\mathbf{B}^i\widehat{\mathbf{D}}_{1(\boldsymbol{\varepsilon})}\right)'\widehat{\boldsymbol{\varepsilon}}, \tag{1.8}$$

for process (1.3) and as:

$$s_{\beta,i}\left(\widehat{\boldsymbol{\theta}}_0\right) = -\frac{1}{\widehat{\sigma}^2}\widehat{\boldsymbol{\varepsilon}}'\left(\mathbf{B}^i\widehat{\mathbf{D}}_{1(\boldsymbol{\varepsilon})}\widehat{\mathbf{M}}_{\boldsymbol{\alpha}}^{-1}\right)'\widehat{\boldsymbol{\varepsilon}}, \tag{1.9}$$

for process (1.4), where $i = 1, ..., p$, $\widehat{\boldsymbol{\varepsilon}}$ is the ML estimate of $\boldsymbol{\varepsilon}$ under H_0 and $\widehat{\mathbf{D}}_{1(\boldsymbol{\varepsilon})} = diag\{1_{(\widehat{\varepsilon}_1\geq 0)}, ..., 1_{(\widehat{\varepsilon}_T\geq 0)}\}$. The vector $\widehat{\boldsymbol{\varepsilon}}$ is estimated by an MA process as $\widehat{\mathbf{M}}_{\boldsymbol{\alpha}}^{-1}\boldsymbol{y}$ or by an AR process as $\widehat{\mathbf{M}}_{\boldsymbol{\alpha}}\boldsymbol{y}$, respectively, where $\widehat{\mathbf{M}}_{\boldsymbol{\alpha}} = \sum_{k=0}^{p}\widehat{\alpha}_k\mathbf{B}^k$.

[2]See e.g. Gelfand and Shilov (1964) for a more detailed overview of the theory of generalized functions.

1.4 Variants of the LM Test

There are a number of methods to compute the LM statistic (1.6), as there are different asymptotically valid ways to estimate the covariance matrix $\mathbf{V}_{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)$. So far, we have assumed that $\mathbf{V}_{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)$ is derived from the asymptotic Hessian matrix evaluated under the null hypothesis. However, any method that allows to consistently estimate $\mathbf{V}_{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)$ is valid. In what follows, several approaches that are widely used in the literature are discussed.

1.4.1 Empirical Hessian and information matrix

The most straightforward method, based on (1.6), is to compute the negative of the Hessian evaluated at the restricted vector of ML estimates $\widehat{\boldsymbol{\theta}}_0$. This is referred to as the empirical Hessian estimator:

$$\mathbf{V}_{\boldsymbol{\beta}}^{(H)}\left(\widehat{\boldsymbol{\theta}}_0\right)=\frac{1}{T}\left(\mathbf{H}_{\boldsymbol{\beta}\boldsymbol{\beta}}\left(\widehat{\boldsymbol{\theta}}_0\right)-\mathbf{H}_{\boldsymbol{\beta}\boldsymbol{\alpha}}\left(\widehat{\boldsymbol{\theta}}_0\right)\mathbf{H}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}\left(\widehat{\boldsymbol{\theta}}_0\right)^{-1}\mathbf{H}_{\boldsymbol{\alpha}\boldsymbol{\beta}}\left(\widehat{\boldsymbol{\theta}}_0\right)\right),$$

where $\mathbf{H}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}\left(\widehat{\boldsymbol{\theta}}_0\right)=-\partial^2\mathcal{L}\left(\widehat{\boldsymbol{\theta}}_0\right)/\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}'$, $\mathbf{H}_{\boldsymbol{\alpha}\boldsymbol{\beta}}\left(\widehat{\boldsymbol{\theta}}_0\right)=\partial^2\mathcal{L}\left(\widehat{\boldsymbol{\theta}}_0\right)/\partial\boldsymbol{\alpha}\partial\boldsymbol{\beta}'$, etc.. However, this estimator is complex to estimate in practice due to the Dirac delta functions and its derivatives even under the null.

Yet, it can be shown that taking expectations eliminates terms that include delta functions in the expression of $\mathbf{V}_{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)$. This comes from the definition of the function itself and the so called sifting property of the delta functions (see Lemma 3). That means that the information matrix approach can be applied instead of the empirical Hessian in order to obtain an efficient and applicable estimator of $\mathbf{V}_{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)$. Hence, in what follows the estimator $\mathbf{V}_{\boldsymbol{\beta}}^{(IM)}\left(\widehat{\boldsymbol{\theta}}_0\right)$ is defined as:

$$\mathbf{V}_{\boldsymbol{\beta}}^{(IM)}\left(\widehat{\boldsymbol{\theta}}_0\right)=\frac{1}{T}\left(\mathbf{J}_{\boldsymbol{\beta}\boldsymbol{\beta}}\left(\widehat{\boldsymbol{\theta}}_0\right)-\mathbf{J}_{\boldsymbol{\beta}\boldsymbol{\alpha}}\left(\widehat{\boldsymbol{\theta}}_0\right)\mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}\left(\widehat{\boldsymbol{\theta}}_0\right)^{-1}\mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\beta}}\left(\widehat{\boldsymbol{\theta}}_0\right)\right), \tag{1.10}$$

where $\mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}(\boldsymbol{\theta})=\mathbb{E}\left[\mathbf{s}_{\boldsymbol{\alpha}}(\boldsymbol{\theta})\,\mathbf{s}_{\boldsymbol{\alpha}}(\boldsymbol{\theta})'\right]$, $\mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\beta}}(\boldsymbol{\theta})=\mathbb{E}\left[\mathbf{s}_{\boldsymbol{\alpha}}(\boldsymbol{\theta})\,\mathbf{s}_{\boldsymbol{\beta}}(\boldsymbol{\theta})'\right]$, etc..

Finally, in order to derive an analytical expression for $\mathbf{V}_{\boldsymbol{\beta}}^{(IM)}$ we relax the Gaussian distributional assumption of ε_t for more specific restrictions on the existence of higher-order moments.

This allows to robustify the estimator $\mathbf{V}_{\boldsymbol{\beta}}^{(IM)}\left(\widehat{\boldsymbol{\theta}}_0\right)$ to non-normal disturbances.

Assumption 1 *(i)* $\{\varepsilon_t\}$ *is an iid sequence with zero mean and* $\mathbb{E}\left[\varepsilon_t^2\right]=\sigma^2>0$.
(ii) There is a positive constant $B>0$ *such that* $\mathbb{E}\left|\varepsilon_t\right|^{4+r}\leq B<\infty$ *for some* $r>0$ *and all* t.
(iii) The density function of ε_t*, defined as* $f_\varepsilon(\cdot)$*, is continuous and differentiable at zero.*

Assumption 1 is sufficient to fulfill all conditions required for the asymptotic properties derived in this chapter. While part (i) and (ii) are standard identification assumptions in the time series literature, part (iii) restricts the analysis to innovations with a smooth density function at point zero.

The matrix $\mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}(\boldsymbol{\theta}_0)$ can be calculated by using standard results for quadratic forms (see e.g. Ullah, 2004, Appendix A.5) and has the same form for both H_A and H_B alternatives, with typical element $J_{i,j}(\boldsymbol{\theta}_0)$:

$$J_{i,j}(\boldsymbol{\theta}_0)\equiv\mathbb{E}\left[s_{\alpha,i}(\boldsymbol{\theta}_0)\,s_{\alpha,j}(\boldsymbol{\theta}_0)\right]=\operatorname{tr}\left[(\mathbf{M}_{\boldsymbol{\alpha}}^{-1}\mathbf{B}^i)(\mathbf{M}_{\boldsymbol{\alpha}}^{-1}\mathbf{B}^j)'\right], \tag{1.11}$$

for $i,j=1,...,p$. However, the results for the other components of the matrix differ depending on the modeling framework as outlined below. Note, that in the following Lemmas we omit the argument $\boldsymbol{\theta}_0$ in $J_{i,j}$, $s_{\alpha,i}$ and $s_{\beta,i}$ in order to ease notation.

Lemma 1 *Let* $\phi_k=\mathbb{E}\left(\varepsilon_t^+\right)^k$ *for* $k=1,2$*. Then, under data generating process (1.3), assumption 1 and the null hypothesis,*

$$\mathbb{E}\left[s_{\alpha,i}s_{\beta,j}\right] = \gamma_1 J_{i,j}, \tag{1.12}$$

$$\mathbb{E}\left[s_{\beta,i}s_{\beta,j}\right] = (\gamma_1-\gamma_2)\,J_{i,j}+\gamma_2 W_{i,j} \tag{1.13}$$

where $1\leq i,j\leq p$, $\gamma_1=\phi_2/\sigma^2$, $\gamma_2=(\phi_1)^2/\sigma^2$ *and* $W_{i,j}=\boldsymbol{l}'(\mathbf{M}_{\boldsymbol{\alpha}}^{-1}\mathbf{B}^i)(\mathbf{M}_{\boldsymbol{\alpha}}^{-1}\mathbf{B}^j)'\boldsymbol{l}$ *with* $\boldsymbol{l}$ *being a* $T\times 1$ *vector of ones.*

The invertibility of the process y_t ensures that the inverse of $M_{\boldsymbol{\alpha}}$ exists under the null. Hence, we can write

$$\mathbf{M}_{\boldsymbol{\alpha}}^{-1}=\left(\sum_{k=0}^{p}\alpha_k\mathbf{B}^k\right)^{-1}=\sum_{k=0}^{\infty}\psi_k\mathbf{B}^k, \tag{1.14}$$

where $\psi_0 = 1$ and $\sum_{k=0}^{\infty} |\psi_k| < \infty$.

Lemma 2 *Let $\phi_k = \mathbb{E}\left(\varepsilon_t^+\right)^k$ for $k = 1, 2$. Then under data generating process (1.4), assumption 1 and the null hypothesis,*

$$\mathbb{E}\left[s_{\alpha,i}s_{\beta,j}\right] = \begin{cases} \mathcal{F}_0 J_{i,j} & \text{for } i > j \\ \mathcal{F}_0 J_{i,j} + \gamma_1 \psi_{|i-j|} & \text{for} i \leq j \end{cases}, \tag{1.15}$$

$$\mathbb{E}\left[s_{\beta,i}s_{\beta,j}\right] = \begin{cases} \mathcal{F}_0 J_{i,j} + \gamma_1 & \text{for } i = j \\ \mathcal{F}_0^2 J_{i,j} + \gamma_2 & \text{for } i \neq j \end{cases}, \tag{1.16}$$

where $1 \leq i, j \leq p$, $\mathcal{F}_0 = (1 - F_\varepsilon(0))$ and $F_\varepsilon(\cdot)$ denote the distribution function of ε; $\gamma_1 = (T - i)(\phi_2 - \sigma^2 \mathcal{F}_0)/\sigma^2$, $\gamma_2 = \gamma_1 \mathcal{F}_0 \psi_{|i-j|} + \phi_1^2/\sigma^2\,(T - \max(i,j))$.

Therefore, to test for the null of no asymmetric effects of innovations, it is sufficient to estimate parameter vector $\boldsymbol{\alpha}$ and error vector $\boldsymbol{\varepsilon}$ under the null and use the estimates to construct the components of the LM test (1.6), i.e.,

$$\mathrm{LM}_T^{(IM)} = \mathbf{s}\left(\widehat{\boldsymbol{\theta}}_0\right)' \left[\mathbf{V}_{\boldsymbol{\beta}}^{(IM)}\left(\widehat{\boldsymbol{\theta}}_0\right)\right]^{-1} \mathbf{s}\left(\widehat{\boldsymbol{\theta}}_0\right), \tag{1.17}$$

where $\mathbf{s}\left(\widehat{\boldsymbol{\theta}}_0\right)$ is given by (1.8) or (1.9) and $\mathbf{V}_{\boldsymbol{\beta}}^{(IM)}\left(\widehat{\boldsymbol{\theta}}_0\right)$ is derived as in (1.11) and Lemma 1 or Lemma 2 for the null hypothesis of interest, respectively.

1.4.2 OPG variant

The second method is the most straightforward option. It is based on the outer product of the gradient and is referred to as the OPG estimator. First, recall that the inverse of $\mathbf{M}_{\boldsymbol{\alpha}}$ under the null is given as $\mathbf{M}_{\boldsymbol{\alpha}}^{-1} = \sum_{k=0}^{\infty} \psi_k \mathbf{B}^k = \sum_{k=0}^{T-1} \psi_k L^k$, where $\psi_0 = 1$ and $\sum_{k=0}^{\infty} |\psi_k| < \infty$. Then write the score vector $\mathbf{s}\left(\widehat{\boldsymbol{\theta}}_{\mathbf{0}}\right)$ as the sum of T contributions

$$s_{\theta,i}\left(\widehat{\boldsymbol{\theta}}_{\mathbf{0}}\right) = \sum_{t=1}^{T} g_{t,i}\left(\widehat{\boldsymbol{\theta}}_{\mathbf{0}}\right), \tag{1.18}$$

where $g_{t,i}\left(\widehat{\boldsymbol{\theta}}_{\mathbf{0}}\right) = \sum_{s=1}^{t-i} \varepsilon_t \varepsilon_s \widehat{\psi}_{t-s-i}$ for $\theta_i = \alpha_i$. If $\theta_i = \beta_i$ then $g_{t,i}\left(\widehat{\boldsymbol{\theta}}_{\mathbf{0}}\right) = \sum_{s=1}^{t-i} \varepsilon_t \varepsilon_s^+ \widehat{\psi}_{t-s-i}$ for the AsMA model and $g_{t,i}\left(\widehat{\boldsymbol{\theta}}_{\mathbf{0}}\right) = \sum_{s=1}^{t-i} \varepsilon_t \varepsilon_s 1(\varepsilon_{t-1} \geq 0)\widehat{\psi}_{t-s-i}$ for the AsAR model. Define the

$T \times 2p$ matrix $\mathbf{G}\left(\widehat{\boldsymbol{\theta}}_{\mathbf{0}}\right)$ with typical element $g_{t,i}\left(\widehat{\boldsymbol{\theta}}_{\mathbf{0}}\right)$. Then, if the OPG estimator is used in (1.6) the test statistic becomes

$$\mathrm{LM}_T^{(OPG)} = \mathbf{s}\left(\widehat{\boldsymbol{\theta}}_0\right)' \left[\mathbf{G}\left(\widehat{\boldsymbol{\theta}}_{\mathbf{0}}\right)' \mathbf{G}\left(\widehat{\boldsymbol{\theta}}_{\mathbf{0}}\right)\right]^{-1} \mathbf{s}\left(\widehat{\boldsymbol{\theta}}_0\right). \tag{1.19}$$

This statistic can be computed using an artificial regression which has the form

$$\boldsymbol{l} = \mathbf{G}\left(\widehat{\boldsymbol{\theta}}_{\mathbf{0}}\right)\boldsymbol{c} + \boldsymbol{u}, \tag{1.20}$$

where $\boldsymbol{l}$ is the unity vector, $\boldsymbol{c}$ is a parameter vector and $\boldsymbol{u}$ is a residual vector. The explained sum of squares obtained from (1.20) is numerically equal to the OPG variant of the LM statistic (1.19).

This OPG variant has an advantage of being relatively easy to compute and also is known to provide a heteroskedasticity robust version of the LM test (1.6). However, it should be used with caution since there is evidence suggesting that this form tends to be less reliable for finite samples (see e.g. Davidson and MacKinnon, 1983 among many others). Section 1.6 provides a further discussion on this issue.

1.4.3 Other regression based variants

Alternatives of the LM test presented in the form of artificial regressions can be used for our tests. In this section we discuss one of the best known artificial regression forms of the LM test which is based on the Gauss-Newton regressions.[3] This approach simply involves regressing the disturbances from the restricted model on the derivatives of the criterion function with respect to all parameters of the unrestricted model.

More precisely, consider the following auxiliary regression

$$\widehat{\boldsymbol{\varepsilon}} = \mathbf{X}_{\boldsymbol{\alpha}}\left(\widehat{\boldsymbol{\theta}}_{\mathbf{0}}\right)\boldsymbol{\rho}_{\alpha} + \mathbf{X}_{\boldsymbol{\beta}}\left(\widehat{\boldsymbol{\theta}}_{\mathbf{0}}\right)\boldsymbol{\rho}_{\beta} + \boldsymbol{v}, \tag{1.21}$$

[3]For a review of other available regression based procedures see for instance Davidson and MacKinnon (2001).

where $\mathbf{X}_{\boldsymbol{\alpha}}\left(\widehat{\boldsymbol{\theta}}_{\mathbf{0}}\right) = \left[\frac{\partial\varepsilon}{\partial\alpha_1}\left(\widehat{\theta}_0\right),\ldots,\frac{\partial\varepsilon}{\partial\alpha_p}\left(\widehat{\theta}_0\right)\right]$ and $\mathbf{X}_{\boldsymbol{\beta}}\left(\widehat{\boldsymbol{\theta}}_{\mathbf{0}}\right) = \left[\frac{\partial\varepsilon}{\partial\beta_1}\left(\widehat{\theta}_0\right),\ldots,\frac{\partial\varepsilon}{\partial\beta_p}\left(\widehat{\theta}_0\right)\right]$. Both regression matrices $\mathbf{X}_{\boldsymbol{\alpha}}\left(\widehat{\boldsymbol{\theta}}_{\mathbf{0}}\right)$ and $\mathbf{X}_{\boldsymbol{\beta}}\left(\widehat{\boldsymbol{\theta}}_{\mathbf{0}}\right)$ can be easily computed using the expressions for $\frac{\partial\varepsilon}{\partial\theta_i}$ derived in items (ii) and (ii') of Lemma 3 (see Appendix A). Testing the null hypothesis $\mathrm{H}_0 : \boldsymbol{\beta} = 0$ is asymptotically equivalent to test whether $\boldsymbol{\rho}_\beta = 0$ in the regression (1.21). Therefore, the test statistic can be readily computed as the standard Wald test from the Gauss-Newton regressions (1.21). In what follows we will refer to this variant of the LM test as regression based and denote it as $\mathrm{LM}_T^{(RB)}$.

Further, a careful inspection shows that this form of the statistic for the H_A alternative closely resembles the test suggested by Brännäs et al. (1998). Therefore, the arguments and the results obtained in this chapter can be used to justify the derivation of the statistics in Brännäs et al. (1998) and to establish its asymptotics.

1.5 Asymptotics

The difference between the different LM-type test statistics described above lies in the estimation of V_β. Since all considered approaches provide consistent estimators for the covariance matrix of the score vector under the null, $\mathrm{LM}_T^{(IM)}$, $\mathrm{LM}_T^{(OPG)}$ and $\mathrm{LM}_T^{(RB)}$ are asymptotically equivalent and follow a χ^2 distribution with p degrees of freedom. This result is summarized in the following theorem.

Theorem 1 *For both processes (1.3) and (1.4), under assumption 1 and the null hypothesis*

$$\mathrm{LM}_T \to \chi_p^2,$$

as $T \to \infty$.

Remark 1 Note that if the stationarity assumption is violated under the null hypothesis, the underlying asymptotics differ from the ones obtained in Theorem 1. For instance, consider the underlying process y_t to be near integrated, i.e.,

$$y_t = (1 + \frac{c}{T})y_{t-1} + \varepsilon_t, \tag{1.22}$$

then the LM test in order to test for an AsAR(1) process behaves asymptotically with

$$\mathrm{LM}_T \xrightarrow{p} \frac{\left(\int_0^1 J_c(r)\mathrm{d}W(r)\right)^2}{\int_0^1 J_c^2(r)\mathrm{d}r}, \tag{1.23}$$

where $J_c(r)$ is an Ornstein-Uhlenbeck process and $W(r)$ is a Brownian motion.[4] However, at this point it is not evident how to distinguish non-stationarity from asymmetry. Therefore, pretesting for a unit root before applying the LM test for asymmetries might provide invalid results. We do not pursue this problem in this chapter but would like to point out that this is an interesting line of research.

1.6 Monte Carlo Simulations

After deriving LM-type tests for testing shock induced asymmetries in time series and their asymptotics, we now turn to study the small sample properties of the test and its variants. The main aim of this section is to evaluate the performance of the tests in terms of their size and power in several setups. Moreover, for completeness of analysis we compare the LM variants to tests available in the literature for AsMA setups, namely the LR test (cf. Wecker, 1981) and the Wald test (cf. Brännäs and De Gooijer, 1994).

1.6.1 Normally distributed errors

As a benchmark specification we consider two types of time series processes given as

$$y_t = \varepsilon_t + \alpha\varepsilon_{t-1}^- + \beta\varepsilon_{t-1}^+ \tag{1.24}$$

$$y_t = \varepsilon_t + \alpha y_{t-1}^- + \beta y_{t-1}^+, \tag{1.25}$$

$$\text{with} \quad \varepsilon_t \sim N(0,1), \tag{1.26}$$

where (1.24) corresponds to an AsMA(1) and (1.25) to an AsAR(1) model. We examine different combinations of α and β selected from the set {0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9} and three sample sizes $T = 50$, $T = 100$ and $T = 200$. All Monte Carlo simulations are based on

[4]The proof of this fact is almost identical to the proof presented in Phillips (1997).

$N = 2000$ replications and are executed for tests of a nominal size of 10%, 5% and 1%, but only the results for the size of 5% are reported, since no qualitative differences were observed.

The left panel of Table 1.1 (see Appendix B) shows rejection frequencies under the null hypothesis when the underlying process is an MA(1) (i.e., $\alpha = \beta$ in (1.24)) with a lag coefficient α. For this specification we are able to compare the variants of the LM test with the LR test and the Wald test. We observe that the LR test tends to overreject in small samples, when α is close to unity. Our study also supports the finding of Brännäs and De Gooijer (1994) who observe serious size distortions of the Wald test for most of the values of α. Turning to the LM type tests we see clear improvements in the size performance compared to the LR and Wald approaches. However, in the case of $T = 50$ we observe moderate deviations from the nominal size for the $\mathrm{LM}_T^{(OPG)}$ and the $\mathrm{LM}_T^{(RB)}$ test when the parameter α is close to unity, which disappear fast as T increases. The right panel of Table 1.1 shows rejection frequencies under the null of an AR(1) process. The obtained results show that the $\mathrm{LM}_T^{(OPG)}$ and the $\mathrm{LM}_T^{(RB)}$ perform equally well, while $\mathrm{LM}_T^{(IM)}$ slightly underrejects, especially when α is close to one.

Figure 1.1 (see Appendix C) illustrates the corresponding rejection frequencies under the alternative. In particular, parameter β in (1.24) and (1.25) is fixed to zero, while α takes values from the interval $[0, 1)$ as described above. At this point we point out that fixing α and allowing β to change produces symmetric results but is omitted from the discussion here. The left panel of Figure 1.1 shows the results for the AsMA alternative and the right one for the AsAR alternative. All three tests perform comparably well except for the case of $T = 50$ where the $\mathrm{LM}_T^{(IM)}$ test has marginally higher power than the other variants for the AsMA alternative and suffers slightly from a power loss relative to the other tests in the case of the AsAR alternative.

1.6.2 Errors with skewed distribution

In the following, we investigate the behaviour of the LM type tests when the errors are not normally distributed. Since we construct test statistics that are aimed to distinguish the contribution of positive and negative errors, it is of particular interest to study if the obtained tests are robust to skewed distributions of the underlying errors. Therefore, we allow the errors in (1.24) and (1.25) to be generated from a beta distribution, i.e.,

$$\varepsilon_t \sim \mathcal{B}(\mu, \sigma, \xi, \kappa), \tag{1.27}$$

where the parameters $(\mu, \sigma, \xi, \kappa)$ are fixed to the values such that assumption (1) is satisfied. In particular, $\mu = 0$ refers to the mean of the distribution, $\sigma = 1$ refers to the standard deviation, $\xi = 0.8$ and $\kappa = 3$ refer to the skewness and to the kurtosis respectively.[5] All other specifications of the MC design remain as in section 1.6.1.

Table 1.2 (see Appendix B) shows the rejection frequencies under the null hypothesis for setups (1.24), (1.25) with (1.27). The reported results show only marginal changes to those obtained in Table 1.1. This indicates that all three tests are successful in controlling for the type I error in the setups where innovations are drawn from a skewed distribution.

Turning to the power analysis, an interesting observation is made. Figures 1.3, 1.4 and 1.5 (see Appendix C) illustrate the obtained power of the tests. As a deviation point from the benchmark design in Section 1.6.1 each panel reports two setups, one with fixed α and $\beta \in [0, 1)$ and one with fixed β and $\alpha \in [0, 1)$. It is clear from Figure 1.3 to 1.5 that while the power properties of the $\mathrm{LM}_T^{(OPG)}$ and $\mathrm{LM}_T^{(RB)}$ do not qualitatively change compared to the scenario with normal errors, a practical weakness of $\mathrm{LM}_T^{(IM)}$ is revealed. In particular, the power results are asymmetric with respect to the fixed α and fixed β. The problem vanishes with growing T. Yet, the $\mathrm{LM}_T^{(IM)}$ test seems to be less robust for small samples with skewed error distributions.

A simple solution to this issue is to use a maximum statistic that is based on the $\mathrm{LM}_{T+}^{(IM)} \equiv \mathrm{LM}_T^{(IM)}$ test and the $\mathrm{LM}_{T-}^{(IM)}$ test. The latter is constructed against alternatives where negative errors enter our modeling framework (1.3) or (1.4) instead of positive errors, i.e.,

$$\max\mathrm{LM}_T^{(IM)} = \max\left\{\mathrm{LM}_{T+}^{(IM)}, \mathrm{LM}_{T-}^{(IM)}\right\}. \tag{1.28}$$

As suggested by the standard theory on multiple comparison problems we use a Bonferroni

[5]Figure 3 in Appendix C illustrates the density function of this distribution.

correction to control for possible size distortions.[6] An additional Monte Carlo experiment (see Figure 1.6 in Appendix C) shows that this approach can successfully resolve the issue of asymmetric power loss previously detected for the case of AsMA models. As an alternative solution one could construct a test statistic that is jointly built on positive and negative residuals.

1.6.3 Conditional heteroskedasticity

To investigate the effect of conditional heteroskedasticity on the performance of the proposed LM type tests, instead of (1.26) we use a GARCH(1,1) specification to generate errors for the processes (1.24) and (1.25), i.e.,

$$\varepsilon_t = \sqrt{h_t}\nu_t, \qquad \text{where} \tag{1.29}$$

$$h_t = \kappa + \delta h_{t-1} + \theta \varepsilon_{t-1}^2, \text{and} \tag{1.30}$$

$$\nu_t \sim N(0,1) \tag{1.31}$$

with $\kappa = 0.01$, $\delta = 0.08$ and $\theta = 0.9$. We do not show other results for alternative parameter combinations as this would exceed the scope of this study. However, estimating a $GARCH(1,1)$ on daily stock market returns usually yields estimates close to the ones we have chosen here (see e.g., Pelagatti and Lisis, 2009).

Table 1.3 presents type I errors for this setup. For the case of an underlying MA(1), the $\text{LM}_T^{(IM)}$ and the $\text{LM}_T^{(RB)}$ tests are oversized for all sample sizes. The OPG variant of the LM test shows a good size control. However, when α is close to unity it overrejects for T=50 and T=100. When the underlying process is an AR(1), the $\text{LM}_T^{(OPG)}$ test performs equally well for all sample sizes. The regression based version of the LM test is again oversized. The $\text{LM}_T^{(IM)}$ test is oversized for T=200, but performs nearly as well as the OPG variant for T=50 and T=100.

Based on our findings, $\text{LM}_T^{(OPG)}$ is the best alternative in case of conditional heteroskedasticity. Therefore, we report the power of $\text{LM}_T^{(OPG)}$ in Figure 1.7 only. In comparison to our

[6] See Table 1.4 for additional insights on the size properties of the $\max\text{LM}_T^{(IM)}$ with and without Bonferroni correction.

benchmark specification 1.6.1 we observe a small drop in power.

1.6.4 Model Selection

In practice the knowledge of the lag length is required prior to the implementation of the LM test. Hence, in this section, we study the estimation of the true order, which shall be called p_0, and its impact on the test statistics. Our primary aim is to establish the small sample behaviour of $\widehat{p}$ estimated using a standard model selection approach within a linear time series model when the true underlying model is in fact an AsMA(p_0) or an AsAR(p_0) model. Specifically, the lag length is estimated from a linear MA(p) or AR(p) model with $1 \leq p \leq P_{\max}$ where $P_{\max}$ is known a priori. The model selection criteria such as the AIC, BIC or HQ are used for the estimation of p_0. The second aim of this section is to investigate the influence of the estimated lag length on the size-power properties of the LM test.

Therefore, in a first step we investigate the performance of the three mentioned model selection criteria in two model setups, each with two different parameterizations. In particular, we use the following specifications:

$$y_t = \varepsilon_t + \alpha_1 \varepsilon_{t-1}^- + \alpha_2 \varepsilon_{t-2}^- + \beta_1 \varepsilon_{t-1}^+ + \beta_2 \varepsilon_{t-2}^+ \tag{1.32}$$

$$y_t = \varepsilon_t + \alpha_1 y_{t-1}^- + \alpha_2 y_{t-2}^- + \beta_1 y_{t-1}^+ + \beta_2 y_{t-2}^+ \tag{1.33}$$

where the first corresponds to an AsMA(2) and the latter to an AsAR(2). We use the parameter combinations $\alpha_1 = 0.5$, $\alpha_2 = 0.4$, $\beta_1 = 0.3$, $\beta_2 = 0.2$ and $\alpha_1 = 0.5$, $\alpha_2 = 0.3$, $\beta_1 = 0.1$, $\beta_2 = 0.1$. Further, we calculate the selected lag length frequencies up to a lag of six periods (i.e., $P_{\max} = 6$) for sample sizes $T = 100$, $T = 200$, $T = 400$ and $T = 600$ using $N = 2000$ replications.

The results are presented in Table 1.5 and 1.6 and are qualitatively similar for both model specifications. For $T = 100$ the BIC has a clear tendency to under-select the lag length for both parameterizations. However, this improves rapidly with increasing T. Furthermore the BIC shows the highest percentage of correct lag selection (above 94%). Similar observations are made for the HQ criterion. As for the linear time series models, the AIC has a tendency to over-select the lag length for all sample sizes. As before, for the 'less linear' parameterization (the second specification) and $T = 100$ the correct lag decision frequency is only about 50%. Overall the correct decision frequencies are in the range from $50-75\%$. Interestingly, we observe an improvement of the performance from $T = 100$ to $T = 400$ but a decrease for $T = 600$. The performance of the HQ criterion is kind of a mixture between the aforementioned criteria.

Which criterion is preferable is nevertheless always context specific. For our purposes it is important to note that the standard criteria can be used to determine the lag length in finite samples, although one should be aware of a potential over-selection of the AIC criterion.

Eventually, we investigate the influence of a preliminary model selection stage on the power of the LM test. To be able to compare the performance with the results from our benchmark model in (1.6.1) we use the BIC in our baseline setup with normally distributed errors. BIC values are calculated up to a lag of six periods. The results are shown in Figure 1.8. In this setup we only observe a minor power loss compared to the case with a known time series process.

1.7 Application

The question whether macroeconomic variables exhibit asymmetries over the business cycle has a long tradition in macroeconomic research. In his General Theory Keynes (1936, p. 314) notes that "the substitution of a downward for an upward tendency often takes place suddenly and violently, whereas there is, as a rule, no such sharp turning-point when an upward is substituted for a downward tendency." Although this asymmetric pattern was recognized very early, the majority of business cycle models rely on a linear specification. However, standard linear models are not capable to explain this asymmetric behaviour. As a result, the development of tools to detect and to understand the nature of these asymmetries was of vast interest during the last decades (see e.g., Neftci, 1984; Beaudry and Koop, 1993; Elwood, 1998; Hansen and Prescott, 2005).

In the following, we provide two examples. We apply the LM test to two time series which are directly related to business cycles: The quarterly U.S. unemployment rate as well as the monthly U.S. industrial production index.

1.7.1 U.S. Unemployment Rate

It is a well known stylized fact that shocks are transmitted asymmetrically to the labour market (e.g. Blanchard and Summers, 1987).

A typical observation from unemployment time series data is shown in Figure 1.9. The figure shows the quarterly seasonally adjusted U.S. unemployment rate (all persons) for the period $1955Q1$ to $2013Q3$.[7] The unemployment rate is increasing sharply and fast during a contraction phase (grey shaded area) while it is moving downward slowly during an expansion phase of the business cycle. As a result the unemployment rate is often considered as a countercylical indicator for business cycles.[8]

Therefore, we test for asymmetries in the quarterly U.S. unemployment rate. Unit root tests, like the augmented Dickey-Fuller test, suggest that the unemployment rate has a unit root and hence we work with the first difference of the series to ensure stationarity. As our linear specification of the series, we use an AR(4) (without a constant), as this appears to be the necessary minimum to describe the short run dynamics. For this specification, the Breusch-Godfrey LM test shows no remaining serial correlation at the 5% significance level. The areas in the graph of the unemployment rate where we observe positive shocks for the AR(4) model are marked red in Figure 1.9. Positive shocks mostly coincide with recessions. This finding supports the presence of asymmetries of the form presented in this chapter. Hence, we run the $\mathrm{LM}_T^{(IM)}$ test for the given specification. With an empirical value of 15.01 and a critical value of 13.28 (at the 1% significance level) we strongly reject the null hypothesis of no asymmetric effects.

1.7.2 U.S. Industrial Production

Industrial production is one of the main indicators for the actual state of the economy. It is considered as a procyclical and coincident indicator for business cycles.

The monthly seasonally adjusted U.S. Industrial Production Index for the period January

[7] We have taken the data from the website of the OECD `http://stats.oecd.org/index.aspx?queryid=36324`.

[8] Dates for the recessions are taken from the NBER website `http://www.nber.org/cycles.html`.

1919 to September 2014 is shown in Figure 1.10 (the basis year of the index is 2007).[9] As in the first application, the grey shaded areas in the graph mark recessions during the sample period. Contractions often coincide with sharp decreases in industrial production. During times of economic growth, industrial production increases slowly.

We transform the index into growth rates by taking log differences in order to achieve stationarity. A suitable linear specification appears to be an AR(8) with a constant. After demeaning the series, we apply the $\mathrm{LM}_T^{(IM)}$ test. The empirical value of our test is 19.11 while the critical value is 15.51 at the 5% level. Hence, we again have a clear indication for asymmetries in the U.S. business cycle.

1.8 Conclusion

In this article, we derive different variants of an LM test to detect asymmetries which are triggered by a different persistence of positive and negative past shocks. Further, we investigate the asymptotic properties of the test. In an extensive simulation study, we examine the small sample properties of the LM_T test under different model specifications. The test has favourable small sample properties compared to already existing tests. Furthermore, it is easy to implement, as it only requires estimation of the model under the null hypothesis. Moreover, we show by means of Monte Carlo simulations that standard model selection criteria are still applicable for the implementation of the test.

In an application to business cycle related macroeconomic time series, we demonstrate the relevance of our testing procedure.

[9] We have taken the data from the website of the Federal Reserve Bank of St. Louis `https://research.stlouisfed.org/fred2/series/INDPRO/`.

A Appendix: Proofs

First, some auxiliary results are collected in the following Lemma to simplify the exposition of the following proofs.

Lemma 3

(i) *Sifting property of delta functions*

$$\int_{\Omega} \delta(x) f(x)\, dx = f(0) \quad \text{and} \quad \int_{\Omega} \dot{\delta}(x) f(x)\, dx = -\dot{f}(0),$$

where $\dot{\delta}(x)$ *defines the derivative of the delta function and* $\dot{f}(x)$ *is the derivative of* $f(x)$*;*

(ii) *For process (1.3) it holds under the alternative that*

$$\frac{\partial \boldsymbol{\varepsilon}}{\partial \beta_i} = -\widetilde{\mathbf{M}}_{\boldsymbol{\alpha},\boldsymbol{\beta}}^{-1} \mathbf{B}^i \mathbf{D}_{1(\boldsymbol{\varepsilon})} \boldsymbol{\varepsilon},$$

where $\widetilde{\mathbf{M}}_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \mathbf{M}_{\boldsymbol{\alpha}} + \mathbf{M}_{\boldsymbol{\beta}} \widetilde{\mathbf{D}}$*, with* $\widetilde{\mathbf{D}} = \mathbf{D}_{1(\boldsymbol{\varepsilon})} + \mathbf{D}_{\delta(\boldsymbol{\varepsilon})} \mathbf{D}_{\boldsymbol{\varepsilon}}$*;*

(iii) *Given a stochastic sequence* $\{x_{t,T}\}_{t=1}^{T}$ *such that* $\operatorname{plim}_{T\to\infty} x_{t,T} = 0$, $\|x_{t,T}\| \leq \|x_{t-1,T}\|$ *and* $x_{1,T} = o_p(T^{-1})$ *then*

$$\operatorname{plim}_{T\to\infty} \left\| \sum_{t=1}^{T} x_{t,T} \right\| = 0.$$

Proof. Sifting property **(i)** summarizes some of the features of delta functions (see e.g. Gelfand and Shilov, 1964).

Property **(ii)** comes directly from differentiation of (1.3), i.e.,

$$\begin{aligned} \frac{\partial \boldsymbol{\varepsilon}}{\partial \theta_i} &= -\mathbf{M}_{\boldsymbol{\alpha},\boldsymbol{\beta}}^{-1} \left[\mathbf{B}^i \mathbf{D}_{1(\boldsymbol{\varepsilon})} + \mathbf{M}_{\boldsymbol{\beta}} \mathbf{D}_{\delta(\boldsymbol{\varepsilon})} \mathbf{D}_{\partial \boldsymbol{\varepsilon}/\partial \theta_i} \right] \boldsymbol{\varepsilon} \\ &= -\mathbf{M}_{\boldsymbol{\alpha},\boldsymbol{\beta}}^{-1} \mathbf{B}^i \mathbf{D}_{1(\boldsymbol{\varepsilon})} \boldsymbol{\varepsilon} + \mathbf{M}_{\boldsymbol{\beta}} \mathbf{D}_{\delta(\boldsymbol{\varepsilon})} \mathbf{D}_{\boldsymbol{\varepsilon}} \partial \boldsymbol{\varepsilon} / \partial \theta_i. \end{aligned}$$

solving the last equality for $\frac{\partial \boldsymbol{\varepsilon}}{\partial \theta_i}$ will yield the required result.

Finally, the last property **(iii)** follows from the triangle inequality, i.e.,

$$\left\| \sum_{t=1}^{T} x_{t,T} \right\| \leq \sum_{t=1}^{T} \|x_{t,T}\| \leq T \|x_{1,T}\|,$$

where $T\left\|x_{1,T}\right\|$ behaves as $o_p(1)$. ■

Proof of Lemma 1

Recall that invertibility of the process y_t ensures the existence of the inverse of $\mathbf{M}_{\boldsymbol{\alpha}}$ under the null, i.e.,

$$\mathbf{M}_{\boldsymbol{\alpha}}^{-1}=\left(\sum_{k=0}^{p}\alpha_k\mathbf{B}^k\right)^{-1}=\sum_{l=0}^{\infty}\psi_l\mathbf{B}^l=\sum_{i=0}^{T-1}\psi_l\mathbf{B}^l,$$

where $\psi_0=1$ and $\sum_{k=0}^{\infty}|\psi_k|<\infty$.

(i) We have that

$$\begin{aligned}
s_{\alpha,i} &= \frac{1}{\sigma^2}\boldsymbol{\varepsilon}'\left(\mathbf{M}_{\boldsymbol{\alpha}}^{-1}\mathbf{B}^i\right)'\boldsymbol{\varepsilon}=\frac{1}{\sigma^2}\sum_{t=1+i}^{T}\sum_{s=1}^{t-i}\varepsilon_t\varepsilon_s\psi_{t-s-i},\\
s_{\beta,j} &= \frac{1}{\sigma^2}\left(\boldsymbol{\varepsilon}^+\right)'\left(\mathbf{M}_{\boldsymbol{\alpha}}^{-1}\mathbf{B}^j\right)'\boldsymbol{\varepsilon}=\frac{1}{\sigma^2}\sum_{t=1+j}^{T}\sum_{s=1}^{t-j}\varepsilon_t\varepsilon_s^+\psi_{t-s-j}.
\end{aligned}$$

Hence, the expectation of $s_{\alpha,i}s_{\beta,j}$ can be rewritten as

$$\mathbb{E}\left[s_{\alpha,i}s_{\beta,j}\right]=\frac{1}{\sigma^4}\sum_{t=1+i}^{T}\sum_{s=1}^{t-i}\sum_{l=1+j}^{T}\sum_{k=1}^{l-j}\psi_{t-s-i}\psi_{l-k-j}\mathbb{E}\left[\varepsilon_t\varepsilon_s\varepsilon_l\varepsilon_k^+\right].$$

Note that the above expectations are non-zero only if the four indices of ε_t are pairwise equal. More precisely, the only possible case is when $t=l$ and $s=k$. We thus obtain the following expression

$$\mathbb{E}\left[s_{\alpha,i}s_{\beta,j}\right]=\frac{\phi_2}{\sigma^2}\sum_{t=1+\max(i,j)}^{T}\sum_{s=1}^{t-\max(i,j)}\psi_{t-s-i}\psi_{t-s-j}=\frac{\phi_2}{\sigma^2}\operatorname{tr}\left[\left(\mathbf{M}_{\boldsymbol{\alpha}}^{-1}\mathbf{B}^i\right)\left(\mathbf{M}_{\boldsymbol{\alpha}}^{-1}\mathbf{B}^j\right)'\right].$$

(ii) Proof of the fact (1.13) goes along the same line. Rewrite expectation of $s_{\beta,i}s_{\beta,j}$ as

$$\mathbb{E}\left[s_{\beta,i}s_{\beta,j}\right]=\frac{1}{\sigma^4}\sum_{t=1+i}^{T}\sum_{s\le t-1}\sum_{l=1+j}^{T}\sum_{k\le l-1}\psi_{t-s-i}\psi_{t-s-j}\mathbb{E}\left[\varepsilon_t\varepsilon_s^+\varepsilon_l\varepsilon_k^+\right].$$

In this situations the expectations are non-zero only if the indices of ε satisfy conditions $t=$

$l \neq s = k$ and $s \neq k \neq t = l$. This in turn leads to (1.13) since

$$\begin{aligned}\mathbb{E}\left[s_{\beta,i}s_{\beta,j}\right] &= \frac{\phi_2}{\sigma^2}\sum_{t=1+\max(i,j)}^{T}\sum_{s=1}^{t-\max(i,j)}\psi_{t-s-i}\psi_{t-s-j} \\ &\quad +\frac{\phi_1^2}{\sigma^2}\sum_{t=1+\max(i,j)}^{T}\sum\sum_{1\leq s\neq k\leq t-\max(i,j)}\psi_{t-s-i}\psi_{t-k-j},\end{aligned}$$

where

$$\sum_{t=1+\max(i,j)}^{T}\sum_{s=1}^{t-\max(i,j)}\psi_{t-s-i}\psi_{t-s-j} = \operatorname{tr}\left[\left(\mathbf{M}_{\boldsymbol{\alpha}}^{-1}\mathbf{B}^{i}\right)\left(\mathbf{M}_{\boldsymbol{\alpha}}^{-1}\mathbf{B}^{j}\right)'\right],$$

$$\sum_{t=1+\max(i,j)}^{T}\sum_{s=1}^{t-\max(i,j)}\psi_{t-s-i}\sum_{k=1}^{t-\max(i,j)}\psi_{t-k-j} = \boldsymbol{l}'(\mathbf{M}_{\boldsymbol{\alpha}}^{-1}\mathbf{B}^{i})(\mathbf{M}_{\boldsymbol{\alpha}}^{-1}\mathbf{B}^{j})'\boldsymbol{l},$$

where $\boldsymbol{l}$ being a $T \times 1$ vector of ones.

Proof of Lemma 2

(i) Consider the following decomposition of $s_{\beta,i}$ elements into two terms

$$s_{\beta,i} = -\frac{1}{\sigma^2}\sum_{t=1+i}^{T}\sum_{s=1}^{t-i-1}\varepsilon_t\varepsilon_s 1\left(\varepsilon_{t-i}\geq 0\right)\psi_{t-s-i} - \frac{1}{\sigma^2}\sum_{t=1+i}^{T}\varepsilon_t\varepsilon_{t-i}^{+}, \tag{1.34}$$

for $i = 1, ..., p$. Hence, the expectation of $s_{\beta,i}s_{\beta,j}$ can be expressed as

$$\begin{aligned}\mathbb{E}\left[s_{\beta,i}s_{\beta,j}\right] &= \frac{1}{\sigma^4}\sum_{t=1+i}^{T}\sum_{s=1}^{t-i-1}\sum_{l=1+j}^{T}\sum_{k=1}^{l-j-1}\mathbb{E}\left[\varepsilon_t\varepsilon_s\varepsilon_l\varepsilon_k 1\left(\varepsilon_{t-i}\geq 0\right)1\left(\varepsilon_{l-j}\geq 0\right)\right]\psi_{t-s-i}\psi_{l-k-j} \\ &\quad +\frac{1}{\sigma^4}\sum_{t=1+i}^{T}\sum_{s=1}^{t-i-1}\sum_{l=1+j}^{T}\mathbb{E}\left[\varepsilon_t\varepsilon_s\varepsilon_l\varepsilon_{l-j}^{+}1\left(\varepsilon_{t-i}\geq 0\right)\right]\psi_{t-s-i} \\ &\quad +\frac{1}{\sigma^4}\sum_{t=1+j}^{T}\sum_{s=1}^{t-j-1}\sum_{l=1+i}^{T}\mathbb{E}\left[\varepsilon_t\varepsilon_s\varepsilon_l\varepsilon_{l-i}^{+}1\left(\varepsilon_{t-j}\geq 0\right)\right]\psi_{t-s-j} \\ &\quad +\frac{1}{\sigma^4}\sum_{t=1+i}^{T}\sum_{l=1+j}^{T}\mathbb{E}\left[\varepsilon_t\varepsilon_{t-i}^{+}\varepsilon_l\varepsilon_{l-j}^{+}\right].\end{aligned} \tag{1.35}$$

Consider first $i = j$. Then the second and the third term in (1.35) are both zero. The only relevant cases for the first term are when $t = l$, $s = k$ and for the fourth term when $t = l$.

These facts and the fact that

$$\mathcal{F}_0 := \mathbb{E}\left[1\left(\varepsilon_{t-i} \geq 0\right)\right] = \int_0^\infty dF_\varepsilon\left(x\right) = 1 - F_\varepsilon\left(0\right),$$

imply that

$$\begin{aligned}
\mathbb{E}\left[s_{\beta,i}s_{\beta,i}\right] &= \mathcal{F}_0 \sum_{t=1+i}^{T}\sum_{s=1}^{t-i-1}\psi_{t-s-i}^2 + \frac{\phi_2}{\sigma^2}(T-i) \\
&= \mathcal{F}_0 \sum_{t=1+i}^{T}\sum_{s=1}^{t-i}\psi_{t-s-i}^2 + \frac{\phi_2 - \sigma^2\mathcal{F}_0}{\sigma^2}(T-i) \qquad (1.36) \\
&= \mathcal{F}_0 \operatorname{tr}\left[(\mathbf{M}_\alpha^{-1}\mathbf{B}^i)(\mathbf{M}_\alpha^{-1}\mathbf{B}^i)'\right] + \frac{\phi_2 - \sigma^2\mathcal{F}_0}{\sigma^2}(T-i). \qquad (1.37)
\end{aligned}$$

When $i > j$, the second term in (1.35) is zero as well and the only relevant case for the first term is when $t = l$, $s = k$ and for the fourth term when $t = l$. However, the third term in (1.35) when $t = l$ and $s = t - i$ has non zero expectation and can be expressed as $\sigma^2\phi_2\left(1 - \mathcal{F}_0\right)\sum_{t=1+i}^{T}\psi_{i-j}$. This results in the following outcome

$$\begin{aligned}
\mathbb{E}\left[s_{\beta,i}, s_{\beta,j}\right] &= \mathcal{F}_0^2 \operatorname{tr}\left[(\mathbf{M}_\alpha^{-1}\mathbf{B}^i)(\mathbf{M}_\alpha^{-1}\mathbf{B}^j)'\right] \\
&\quad + \frac{(T-i)}{\sigma^2}\left((\phi_2 - \sigma^2\mathcal{F}_0)\mathcal{F}_0\psi_{i-j} + \phi_1^2\right). \qquad (1.38)
\end{aligned}$$

Finally, for $i < j$ the results are identical to those obtained for $i > j$ due to the symmetry of the variance covariance matrix.

(ii) Same techniques are used to find the covariance between $s_{\alpha,i}$ and $s_{\beta,j}$. For the case when $j < i$ we have that

$$\begin{aligned}
\mathbb{E}\left[s_{\alpha,i}s_{\beta,j}\right] &= \mathcal{F}_0 \sum_{t=1+i}^{T}\sum_{s=1}^{t-i}\psi_{t-s-i}^2 \qquad (1.39) \\
&= \mathcal{F}_0 \operatorname{tr}\left[(\mathbf{B}^i\mathbf{M}_\alpha^{-1})(\mathbf{B}^j\mathbf{M}_\alpha^{-1})'\right], \qquad (1.40)
\end{aligned}$$

and for $j \geq i$ additional terms enter the expression, i.e.,

$$\mathbb{E}\left[s_{\alpha,i}s_{\beta,j}\right] = \mathcal{F}_0 \operatorname{tr}\left[(\mathbf{B}^i\mathbf{M}_\alpha^{-1})(\mathbf{B}^j\mathbf{M}_\alpha^{-1})'\right] + \frac{(\phi_2 - \sigma^2\mathcal{F}_0)}{\sigma^2}\psi_{i-j}\left(T-j\right), \qquad (1.41)$$

which completes the proof of the Lemma.

Proof of Theorem 1

To ease the notation in what follows, we omit the argument $\boldsymbol{\theta}_0$. Rewrite the score vector as $\mathbf{s}_{\boldsymbol{\beta}} = \frac{1}{\sigma^2}\sum_t \boldsymbol{Z}_{t,T}$, where $\boldsymbol{Z}_{t,T} = \left(Z_{t,T}^{(1)}, ..., Z_{t,T}^{(p)}\right)'$ with $Z_{t,T}^{(i)}$ defined as

$$Z_{t,T}^{(i)} = \sum_{s=1}^{t-i} \varepsilon_t \varepsilon_s^+ \psi_{t-s-i} = \varepsilon_t \xi_{t-i},$$

and ξ_{t-i} denotes $\sum\limits_{s=1}^{t-1} \varepsilon_s^+ \psi_{t-s-i}$. To investigate the limiting behaviour the Cramer-Wold device is applied which tells that it is sufficient to study the limiting distribution of the sequence of scalars $\eta_{t,T} = \lambda' \boldsymbol{Z}_{t,T}$, where λ is a $p \times 1$ vector such that $\|\lambda\| = 1$ and $\|\cdot\|$ defines an L_2 vector norm.

The central limit theorem for martingale difference sequences (henceforth mds) applies to the $\{\eta_{t,T}\}$ if the following holds:[10]

(i) $\{\eta_{t,T}, \mathcal{F}_{t,T}\}$ is mds, where $\mathcal{F}_{t,T}$ is defined as an associated σ-algebra to the sequence $\eta_{t,T}$ such that $\eta_{t,T}$ is measurable with respect to $\mathcal{F}_{t,T}$;

(ii) $\mathbb{E}\left|\eta_{t,T}\right|^{2+r} < B < \infty$ for some $r > 0$ and all t;

(iii) Define $\overline{\sigma}_{\eta,T}^2 \equiv \frac{1}{T}\mathbb{E}\left[\left(\sum\limits_t \eta_{t,T}\right)^2\right]$, where $\overline{\sigma}_{\eta,T}^2 > r' > 0$ and

$$\frac{1}{T}\sum_t \eta_{t,T}^2 - \overline{\sigma}_{\eta,T}^2 \xrightarrow{p} 0.$$

It is straightforward to see that condition (i) is satisfied since $\mathbb{E}\left[\eta_{t,T}|\mathcal{F}_{t-1,T}\right] = \lambda'\mathbb{E}\left[\boldsymbol{Z}_{t,T}|\mathcal{F}_{t-1,T}\right] = 0$ and the given assumptions on ε_t assure that $\mathbb{E}\left|\eta_{t,T}\right| < \infty$. To verify condition (ii) notice first that by the Cauchy-Schwarz inequality and the Minkowski's inequality

$$\mathbb{E}\left|\eta_{t,T}\right|^{2+r} \leq \|\lambda\|^{2+r}\,\mathbb{E}\left\|\boldsymbol{Z}_{t,T}\right\|^{2+r} \leq \left(\sum_i \left(\mathbb{E}\left|Z_{t,T}^{(i)}\right|^{2+r}\right)^{\frac{1}{2+r}}\right)^{2+r}.$$

Hence, condition (ii) follows from uniform L_{4+r} boundedness of ε_t, uniform L_{4+r} boundedness

[10] see e.g. White (2001), Corollary 5.26

of ε_t^+ (implied by assumption 1) and the following arguments

$$\begin{aligned}\mathbb{E}\left|Z_{t,T}^{(i)}\right|^{2+r} &\leq \left(\mathbb{E}\left|\varepsilon_t\right|^{4+r}\mathbb{E}\left|\xi_{t-i}\right|^{4+r}\right)^{\frac{1}{2}} \\ &\leq C\left(\sum_{s=1}^{t-i}\left(\mathbb{E}\left|\varepsilon_s^+\psi_{t-s-i}\right|^{4+r}\right)^{\frac{1}{4+r}}\right)^{2+r} \\ &\leq C_1\left(\sum_{s=1}^{t-1}\left|\psi_{t-s-i}\right|\right)^{2+r} < \infty,\end{aligned}$$

where the second inequality follows from the Minkowski's inequality and the last one from invertibility and stability of the process.

Regarding the last condition (iii), it is clear that $\overline{\sigma}_{\eta,T}^2$ is bounded away from zero, i.e.,

$$\overline{\sigma}_{\eta,T}^2 = \frac{1}{T}\mathbb{E}\left[\left(\sum_t \lambda'\boldsymbol{Z}_{t,T}\right)^2\right] = \frac{1}{T}\lambda' V_{\boldsymbol{\beta}}\lambda > 0.$$

Finally, to show the convergence of $\frac{1}{T}\sum_t \eta_{t,T}^2 - \overline{\sigma}_{\eta,T}^2$ it is sufficient to show convergence of

$$\frac{1}{T}\sum_t Z_{t,T}^{(i)}Z_{t,T}^{(j)} - \frac{1}{T}V_{\boldsymbol{\beta}}(i,j) = \frac{1}{T}\sum_t\left(\varepsilon_t^2-\sigma^2\right)\xi_{t-i}\xi_{t-j} + \frac{1}{T}\sigma^2\sum_t X_{t-1}, \tag{1.42}$$

where $X_{t-1} \equiv \sum_t\left(\xi_{t-i}\xi_{t-j} - \gamma_2\left(\sum_{t=1+\max(i,j)}^T \overline{\psi}_t^{(i)}\overline{\psi}_t^{(j)} - J_{i,j}\right)\right)$. The first term on the r.h.s. of (1.42) satisfies the mds property and $\mathbb{E}\left|\left(\varepsilon_t^2-\sigma^2\right)\xi_{t-1}^2\right|^{2+r} < \Delta < \infty$. Therefore, the law of large numbers for mds gives that $\frac{1}{T}\sum_t\left(\varepsilon_t^2-\sigma^2\right)\xi_{t-i}\xi_{t-j} \xrightarrow{p} 0$. Moreover, the given assumptions with standard arguments (see e.g. Hamilton, 1994, Chapter 7, pp.192-193) imply that X_{t-1} is uniformly integrable L^1 mixingale which in turn gives that $\frac{1}{T}\sum_t X_{t-1} \xrightarrow{p} 0$.

Proofs of the limiting results for the AsAR model are similar to those given for the AsMA model and hence are omitted.

B Appendix: Tables

Table 1.1: Rejection frequencies (in %) under the null of no asymmetric effects for AsMA processes (left panel) and AsAR processes (right panel); the nominal size is 5%. $\varepsilon_t \sim N(0,1)$.

		MA(1)					AR(1)		
	α	LR	Wald	$\mathrm{LM}_T^{(IM)}$	$\mathrm{LM}_T^{(OPG)}$	$\mathrm{LM}_T^{(RB)}$	$\mathrm{LM}_T^{(IM)}$	$\mathrm{LM}_T^{(OPG)}$	$\mathrm{LM}_T^{(RB)}$
$T = 50$									
	0.0	6.0	6.7	7.9	6.3	6.5	5.8	6.0	4.9
	0.1	6.3	11.2	6.1	7.2	5.6	4.2	6.5	6.1
	0.2	7.2	16.2	7.1	5.8	4.7	4.4	5.5	5.1
	0.3	7.9	21.4	6.6	7.7	6.3	4.2	5.3	4.8
	0.4	7.5	29.0	6.7	8.0	6.9	3.8	5.5	4.3
	0.5	8.4	35.9	7.3	6.5	5.1	3.0	5.0	4.7
	0.6	10.3	45.7	7.4	7.8	5.8	2.6	6.0	5.2
	0.7	14.0	55.8	8.3	8.4	6.4	3.2	5.1	4.4
	0.8	21.7	67.0	8.7	10.9	7.4	1.9	5.3	4.7
	0.9	34.1	77.6	7.9	11.9	9.6	3.0	5.7	4.6
$T = 100$									
	0.0	5.5	5.4	5.9	5.2	4.7	5.1	5.8	5.3
	0.1	6.3	9.0	5.5	5.8	5.4	4.7	6.0	4.9
	0.2	5.5	13.4	5.0	5.5	4.5	4.1	5.7	4.6
	0.3	5.3	18.0	5.7	5.3	4.8	4.3	5.6	4.8
	0.4	5.2	25.2	5.0	5.3	4.7	3.8	5.9	4.9
	0.5	6.6	34.6	5.4	5.1	4.6	3.5	5.8	4.8
	0.6	6.2	40.7	4.5	6.4	5.7	3.6	5.3	5.3
	0.7	8.1	52.8	4.9	5.3	4.8	3.2	5.3	5.1
	0.8	9.0	64.7	5.2	6.2	6.4	3.3	5.7	4.8
	0.9	19.3	75.5	5.3	7.4	7.7	3.7	5.2	4.3
$T = 200$									
	0.0	4.7	5.2	4.8	6.1	5.9	5.0	5.5	5.1
	0.1	5.5	9.3	4.2	5.4	4.8	4.6	5.6	4.6
	0.2	5.8	13.9	5.0	5.1	5.1	4.8	5.4	5.0
	0.3	5.1	18.9	4.2	5.4	5.2	5.2	6.5	5.9
	0.4	5.9	27.7	4.3	6.7	6.5	4.1	4.7	4.4
	0.5	5.4	31.8	4.7	5.0	5.2	4.9	5.7	5.5
	0.6	5.5	41.0	4.0	5.1	5.0	4.1	4.8	4.5
	0.7	6.2	53.1	4.5	5.7	5.9	3.8	5.0	4.6
	0.8	6.3	62.0	4.4	4.9	4.6	4.1	5.7	5.3
	0.9	14.4	72.2	4.5	6.3	7.2	3.9	4.4	4.5

Table 1.2: Rejection frequencies (in %) under the null of no asymmetric effects for AsMA processes (left panel) and AsAR processes (right panel); the nominal size is 5%. $\varepsilon_t \sim \mathcal{B}(0, 1, 0.8, 3)$.

		MA(1)			AR(1)		
	α	$\mathrm{LM}_T^{(IM)}$	$\mathrm{LM}_T^{(OPG)}$	$\mathrm{LM}_T^{(RB)}$	$\mathrm{LM}_T^{(IM)}$	$\mathrm{LM}_T^{(OPG)}$	$\mathrm{LM}_T^{(RB)}$
$T = 50$							
	0.0	5.6	7.4	5.7	4.8	7.1	4.8
	0.1	6.6	7.7	5.9	4.6	7.3	5.9
	0.2	5.5	5.4	5.6	4.1	6.4	5.2
	0.3	6.4	8.0	6.1	3.4	7.3	5.2
	0.4	6.0	7.9	5.4	2.7	6.8	5.8
	0.5	6.8	7.1	4.9	2.5	6.1	4.8
	0.6	7.8	8.5	6.1	2.4	5.8	4.7
	0.7	8.1	9.9	6.7	2.1	7.0	5.1
	0.8	9.4	11.5	8.0	2.3	5.1	3.5
	0.9	8.9	13.9	9.0	2.6	5.6	4.5
$T = 100$							
	0.0	4.8	6.3	5.3	4.9	5.8	5.4
	0.1	4.5	6.2	4.9	4.7	6.2	5.3
	0.2	4.9	6.7	5.8	5.2	5.2	4.3
	0.3	4.0	5.8	4.5	4.5	6.6	5.5
	0.4	5.1	5.8	5.5	3.9	6.3	5.4
	0.5	4.4	5.0	4.6	3.5	5.6	4.6
	0.6	4.3	6.1	6.2	2.6	4.7	4.2
	0.7	4.4	6.1	5.7	3.0	5.7	4.7
	0.8	5.2	6.2	5.8	3.4	5.3	4.1
	0.9	5.9	8.5	6.3	3.2	5.2	4.1
$T = 200$							
	0.0	4.1	5.6	5.5	4.9	6.1	5.5
	0.1	4.7	6.3	5.8	5.5	5.5	5.3
	0.2	4.0	5.8	5.6	5.3	6.4	5.7
	0.3	3.8	4.6	4.1	4.9	6.0	5.9
	0.4	4.3	5.2	5.1	5.0	5.6	4.6
	0.5	4.4	5.7	5.2	3.8	5.2	4.2
	0.6	3.4	5.3	5.0	4.8	6.1	5.3
	0.7	3.8	6.4	6.8	3.5	5.3	4.6
	0.8	4.8	5.4	5.1	3.7	6.3	5.4
	0.9	5.4	5.6	5.5	3.5	4.4	4.1

Table 1.3: Rejection frequencies (in %) under the null of no asymmetric effects for AsMA processes (left panel) and AsAR processes (right panel); the nominal size is 5%. $\varepsilon_t \sim GARCH(1,1)$.

	α	MA(1) $\mathrm{LM}_T^{(IM)}$	$\mathrm{LM}_T^{(OPG)}$	$\mathrm{LM}_T^{(RB)}$	AR(1) $\mathrm{LM}_T^{(IM)}$	$\mathrm{LM}_T^{(OPG)}$	$\mathrm{LM}_T^{(RB)}$
$T = 50$							
	0.0	9.1	6.6	8.8	7.8	6.3	8.9
	0.1	9.1	6.8	8.9	6.5	5.0	7.5
	0.2	8.9	6.4	9.1	6.9	6.2	9.0
	0.3	8.5	6.0	8.2	6.1	5.3	8.3
	0.4	8.4	5.0	8.8	5.3	5.2	7.5
	0.5	8.4	6.1	8.9	4.5	4.8	7.0
	0.6	8.1	6.0	9.1	3.5	5.0	7.2
	0.7	8.5	7.1	10.9	3.4	5.1	7.4
	0.8	8.6	8.8	12.2	4.1	5.5	7.7
	0.9	7.1	11.8	12.2	5.0	5.8	8.7
$T = 100$							
	0.0	8.0	6.1	8.9	7.9	5.2	8.2
	0.1	9.2	5.9	9.0	8.5	6.0	9.0
	0.2	8.7	6.1	8.7	7.3	5.8	8.1
	0.3	8.3	4.9	7.6	6.1	4.8	7.8
	0.4	8.5	4.8	8.7	6.4	5.5	7.8
	0.5	8.3	4.8	7.1	5.0	4.6	6.8
	0.6	7.6	5.4	8.5	5.4	5.7	7.4
	0.7	8.0	4.7	8.9	3.9	4.0	6.0
	0.8	7.2	5.5	8.9	5.4	5.0	7.2
	0.9	6.8	8.1	10.8	6.4	5.6	8.2
$T = 200$							
	0.0	8.3	5.1	8.4	9.4	6.2	9.5
	0.1	9.5	6.0	9.5	7.8	5.2	8.2
	0.2	8.2	4.6	7.8	9.3	6.3	9.4
	0.3	9.6	6.0	9.4	7.6	4.9	7.9
	0.4	7.8	4.7	7.7	7.1	5.3	8.0
	0.5	9.1	5.3	8.7	7.0	5.5	8.2
	0.6	8.0	5.1	8.1	6.1	4.8	7.9
	0.7	7.5	4.5	7.7	7.0	5.7	8.3
	0.8	7.3	5.2	8.5	7.2	6.5	9.2
	0.9	7.7	6.5	10.2	6.4	5.3	7.7

Table 1.4: Size of the $maxLM_T(IM)$ test without and with Bonferroni correction in case of an $MA(1)$; the nominal size is 5%

		$MA(1)$	
	α	$maxLM_T^{(IM)}$	$maxLM_T^{(IM)}$ (Bonferroni)
T=50			
	0.0	11.0	6.8
	0.1	10.1	6.1
	0.2	10.4	6.7
	0.3	10.8	6.5
	0.4	10.7	6.7
	0.5	10.3	6.2
	0.6	11.9	8.2
	0.7	13.7	8.9
	0.8	13.5	9.6
	0.9	11.4	7.3
T=100			
	0	9.1	5.7
	0.1	9.3	5.6
	0.2	10.4	6.3
	0.3	7.8	4.6
	0.4	7.3	4.2
	0.5	8.5	5.5
	0.6	8.4	5.1
	0.7	9.2	5.5
	0.8	9.3	5.4
	0.9	8.7	5.0
T=200			
	0	7.6	4.7
	0.1	8.7	4.9
	0.2	9.6	5.6
	0.3	7.6	4.4
	0.4	6.8	4.0
	0.5	7.0	4.1
	0.6	7.0	4.0
	0.7	7.2	4.2
	0.8	7.2	4.3
	0.9	8.4	4.9

Table 1.5: Model Selection based decision frequencies (in %) under different AsMA DGPs

$AsMA(2)$: $y_t = \varepsilon_t + 0.5\varepsilon^-_{t-1} + 0.4\varepsilon^-_{t-2} + 0.3\varepsilon^+_{t-1} + 0.2\varepsilon^+_{t-2}$						
	p=1	p=2	p=3	p=4	p=5	p=6
T=100						
AIC	3.90	67.00	12.20	6.30	5.60	5.00
BIC	20.05	75.95	3.20	0.50	0.15	0.15
HQ	9.85	77.50	7.60	2.75	1.25	1.05
T=200						
AIC	0.25	73.75	10.85	6.70	4.75	3.70
BIC	2.10	95.45	2.05	0.40	0.00	0.00
HQ	0.45	89.95	5.80	2.40	0.85	0.55
T=400						
AIC	0.00	74.40	11.75	6.90	3.95	3.00
BIC	0.00	97.85	1.90	0.25	0.00	0.00
HQ	0.00	90.90	6.40	2.00	0.40	0.30
T=600						
AIC	0.00	72.20	11.35	7.55	5.15	3.75
BIC	0.00	98.35	1.60	0.05	0.00	0.00
HQ	0.00	91.90	5.45	1.75	0.65	0.25
$AsMA(2)$: $y_t = \varepsilon_t + 0.5\varepsilon^-_{t-1} + 0.3\varepsilon^-_{t-2} + 0.1\varepsilon^+_{t-1} + 0.1\varepsilon^+_{t-2}$						
T=100						
AIC	19.45	53.20	10.40	6.95	5.65	4.35
BIC	49.95	46.70	2.35	0.80	0.15	0.05
HQ	32.65	55.80	6.55	2.70	1.65	0.65
T=200						
AIC	4.80	65.10	13.25	7.45	5.15	4.25
BIC	23.45	71.40	2.00	0.25	0.20	0.00
HQ	10.65	78.00	2.40	2.40	0.90	0.65
T=400						
AIC	0.15	67.95	14.05	7.75	5.30	4.80
BIC	2.65	94.30	2.70	0.35	0.00	0.00
HQ	0.50	88.10	7.50	2.40	0.95	0.55
T=600						
AIC	0.00	63.50	16.20	9.35	5.80	5.15
BIC	0.30	98.65	2.55	0.20	0.10	0.00
HQ	0.00	87.40	8.45	3.00	0.80	0.35

Table 1.6: Model Selection based decision frequencies (in %) under different AsAR DGPs

$AsAR(2)$: $y_t = \varepsilon_t + 0.5y^-_{t-1} + 0.4y^-_{t-2} + 0.3y^+_{t-1} + 0.2y^+_{t-2}$						
	q=1	q=2	q=3	q=4	q=5	q=6
T=100						
AIC	6.05	68.15	12.00	5.40	4.35	4.05
BIC	20.75	75.85	2.65	0.50	0.25	0.00
HQ	11.70	77.45	7.15	1.75	1.30	0.65
T=200						
AIC	0.15	73.55	12.65	6.00	4.50	3.15
BIC	2.35	94.95	2.60	0.05	0.05	0.00
HQ	0.75	88.50	7.85	1.40	1.25	0.25
T=400						
AIC	0.00	71.05	13.9	5.90	5.30	3.85
BIC	0.05	97.10	2.60	0.20	0.05	0.00
HQ	0.00	88.9	8.35	1.75	0.50	0.50
T=600						
AIC	0.00	67.55	15.65	5.95	6.20	4.65
BIC	0.00	97.30	2.35	0.35	0.00	0.00
HQ	0.00	88.15	8.90	1.50	1.05	0.40
$AsAR(2)$: $y_t = \varepsilon_t + 0.5y^-_{t-1} + 0.3y^-_{t-2} + 0.1y^+_{t-1} + 0.1y^+_{t-2}$						
T=200						
AIC	21.10	55.50	10.05	5.80	4.40	3.15
BIC	50.20	47.85	1.45	0.35	0.15	0.00
HQ	34.50	56.75	5.20	1.90	1.30	0.35
T=200						
AIC	5.05	69.35	11.30	5.90	5.05	3.35
BIC	22.05	76.15	1.30	0.45	0.05	0.00
HQ	10.90	81.20	4.70	2.05	0.85	0.30
T=400						
AIC	0.15	67.25	10.70	8.20	7.25	6.45
BIC	2.30	95.45	1.70	0.45	0.10	0.00
HQ	0.70	89.15	4.80	2.95	1.70	0.70
T=600						
AIC	0.00	62.75	8.80	8.80	10.30	9.35
BIC	0.10	98.90	0.50	0.35	0.15	0.00
HQ	0.05	90.55	4.35	2.35	1.90	0.80

C Appendix: Figures

Figure 1.1: Power of the LM_T Variants when $\varepsilon_t \sim N(0,1)$

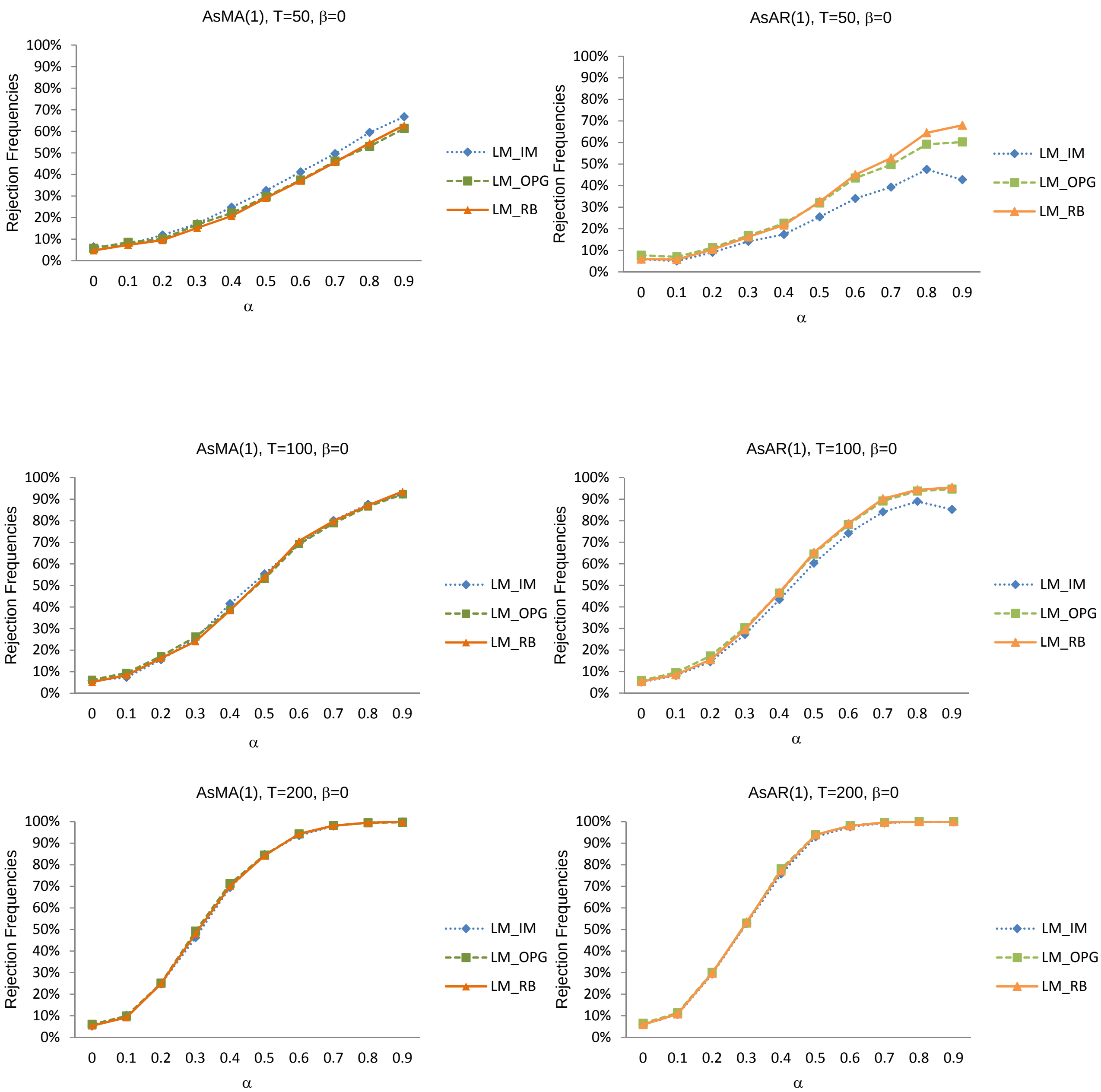


Figure 1.2: Kernel Density Estimate of $\varepsilon_t \sim \mathcal{B}(0, 1, 0.8, 3)$

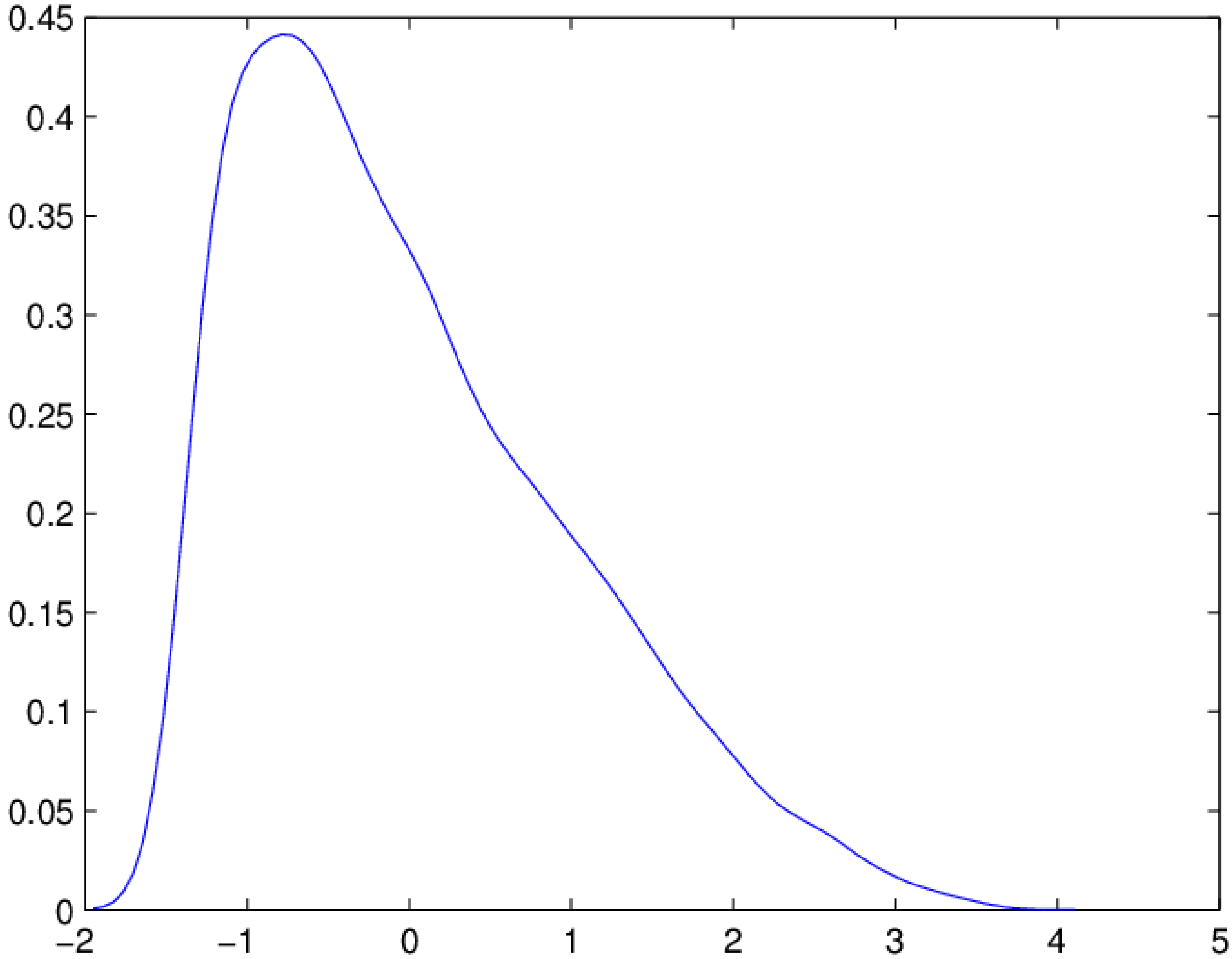


Figure 1.3: Power of the $LM_T^{(IM)}$ Test when $\varepsilon_t \sim \mathcal{B}(0, 1, 0.8, 3)$

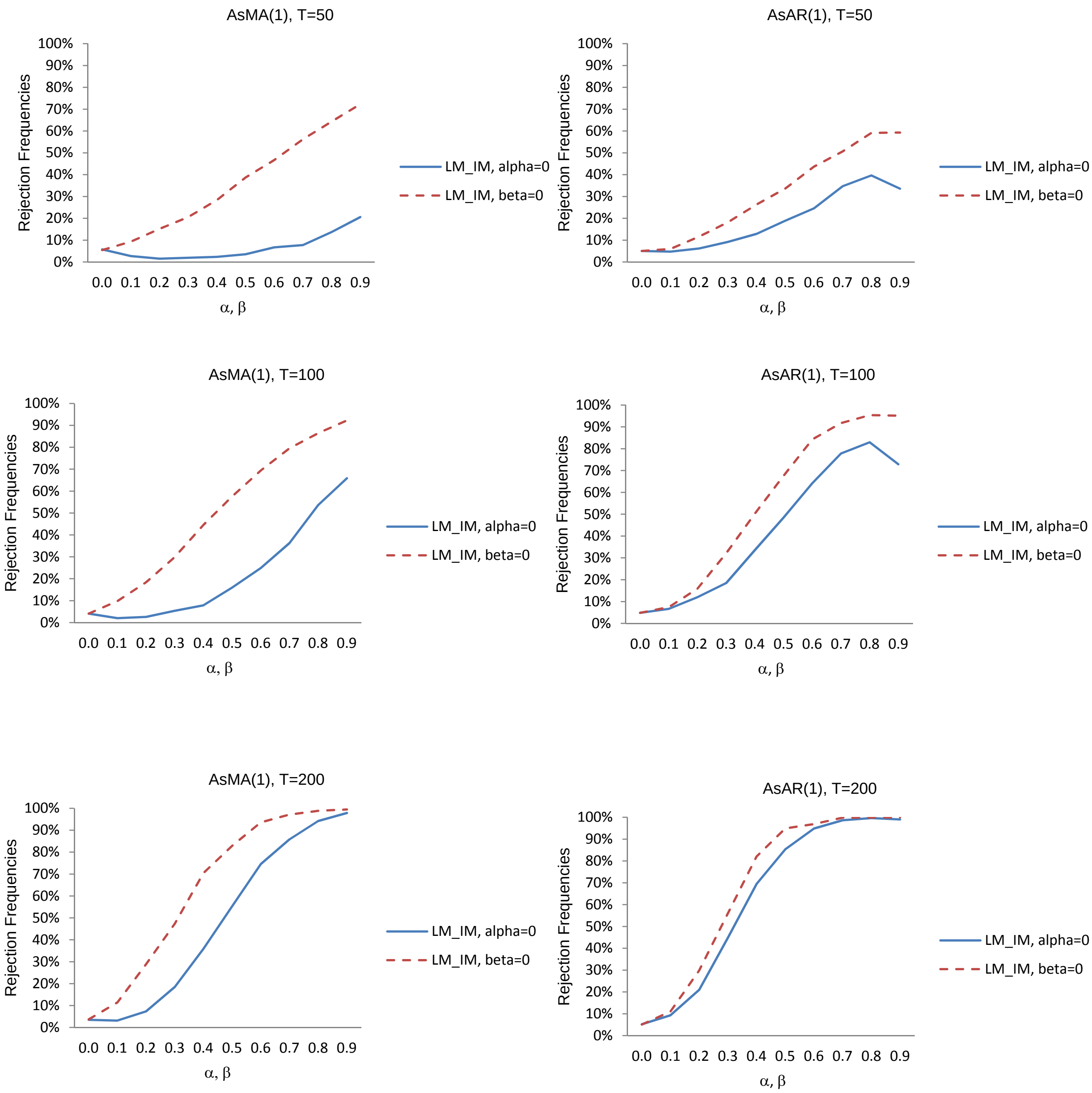


Figure 1.4: Power of the $LM_T^{(OPG)}$ Test when $\varepsilon_t \sim \mathcal{B}(0, 1, 0.8, 3)$

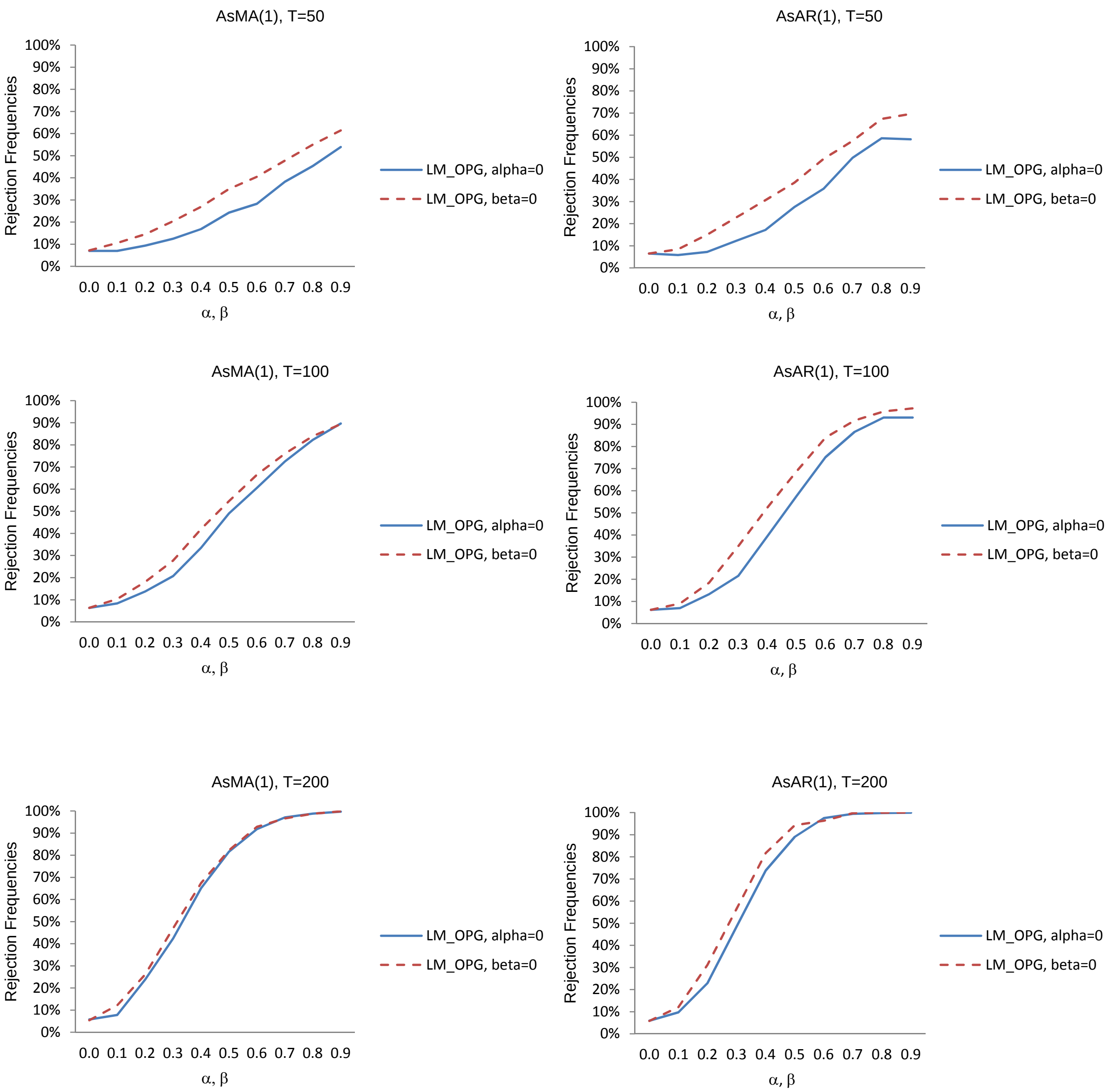


Figure 1.5: Power of the $LM_T^{(RB)}$ Test when $\varepsilon_t \sim \mathcal{B}(0, 1, 0.8, 3)$

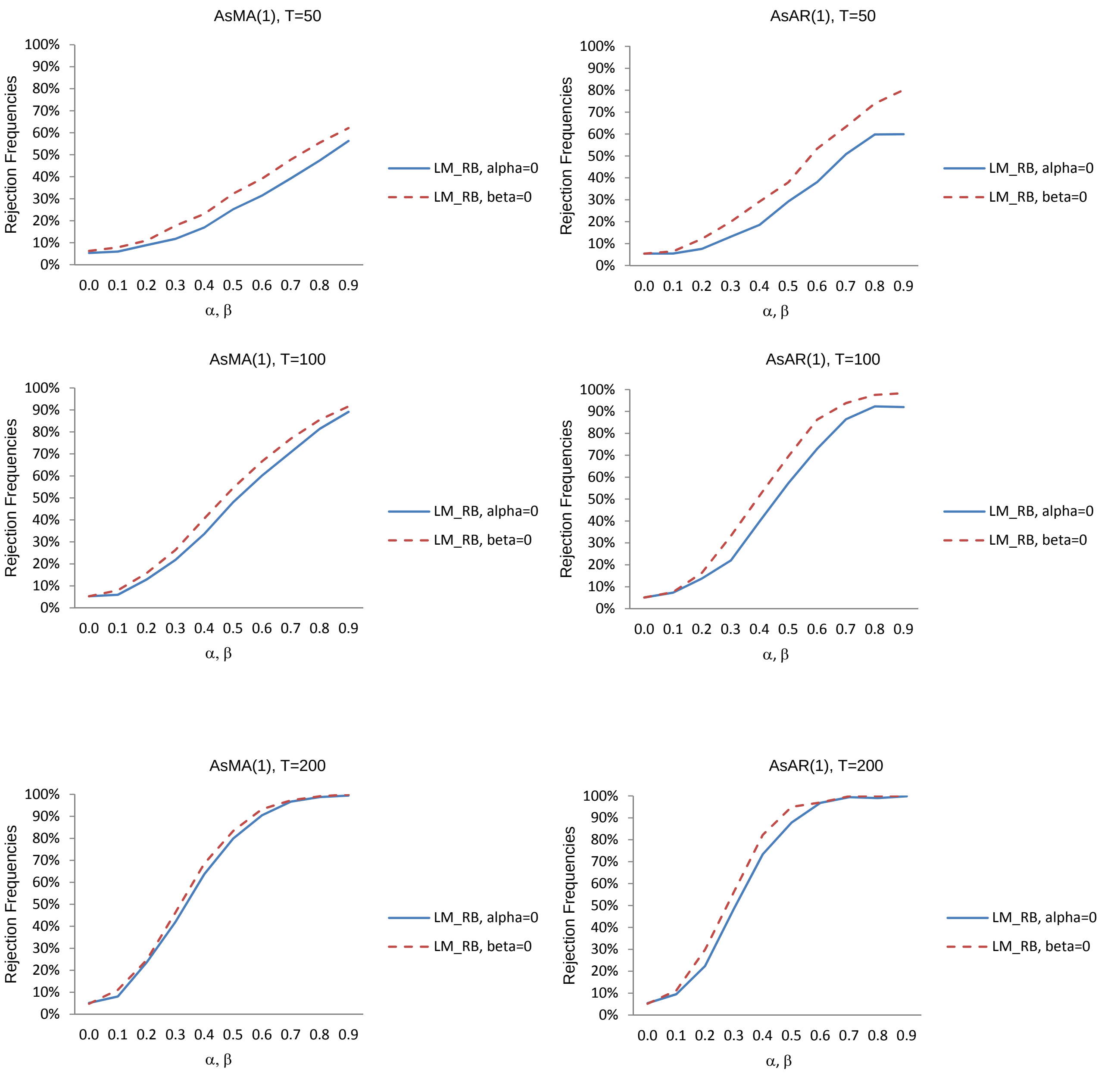


Figure 1.6: Power of the Robustified Version of the $LM_T^{(IM)}$ Test when $\varepsilon_t \sim \mathcal{B}(0, 1, 0.8, 3)$

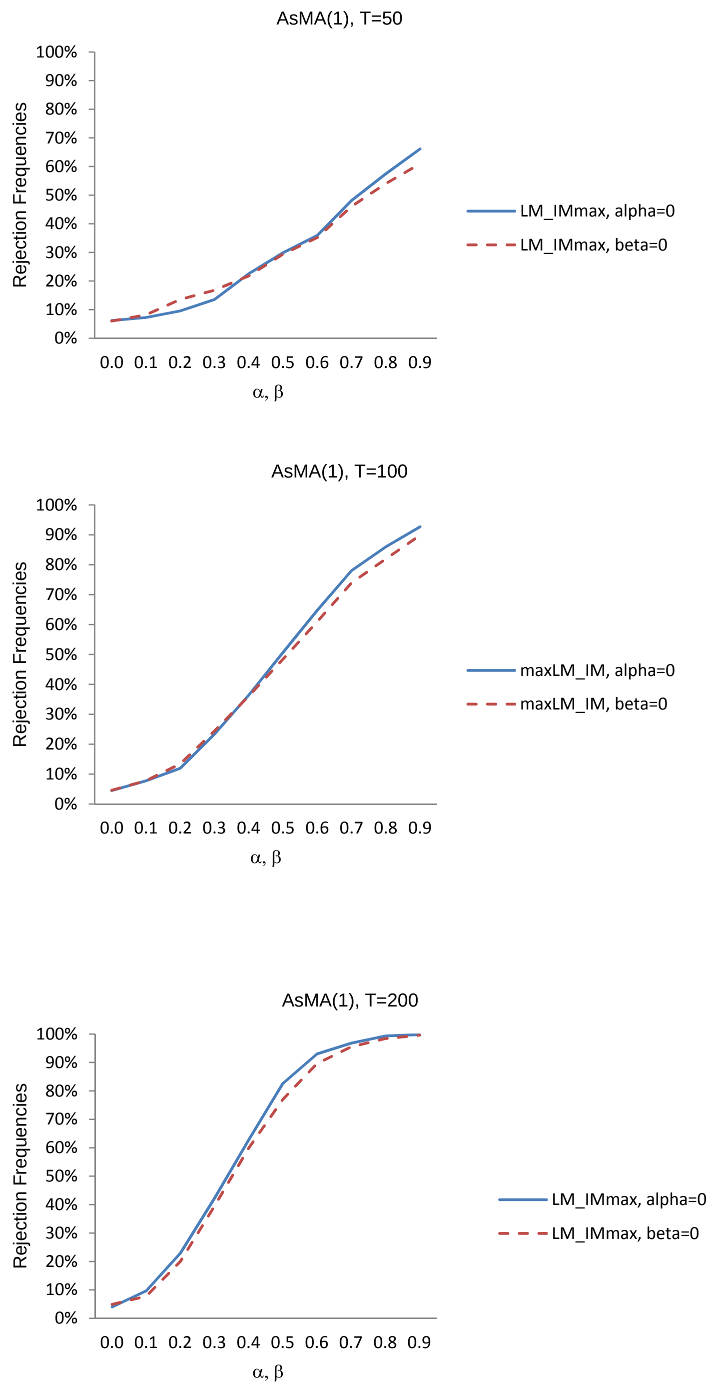


Figure 1.7: Power of the $LM_T^{(OPG)}$ Test when $\varepsilon_t \sim GARCH(1,1)$

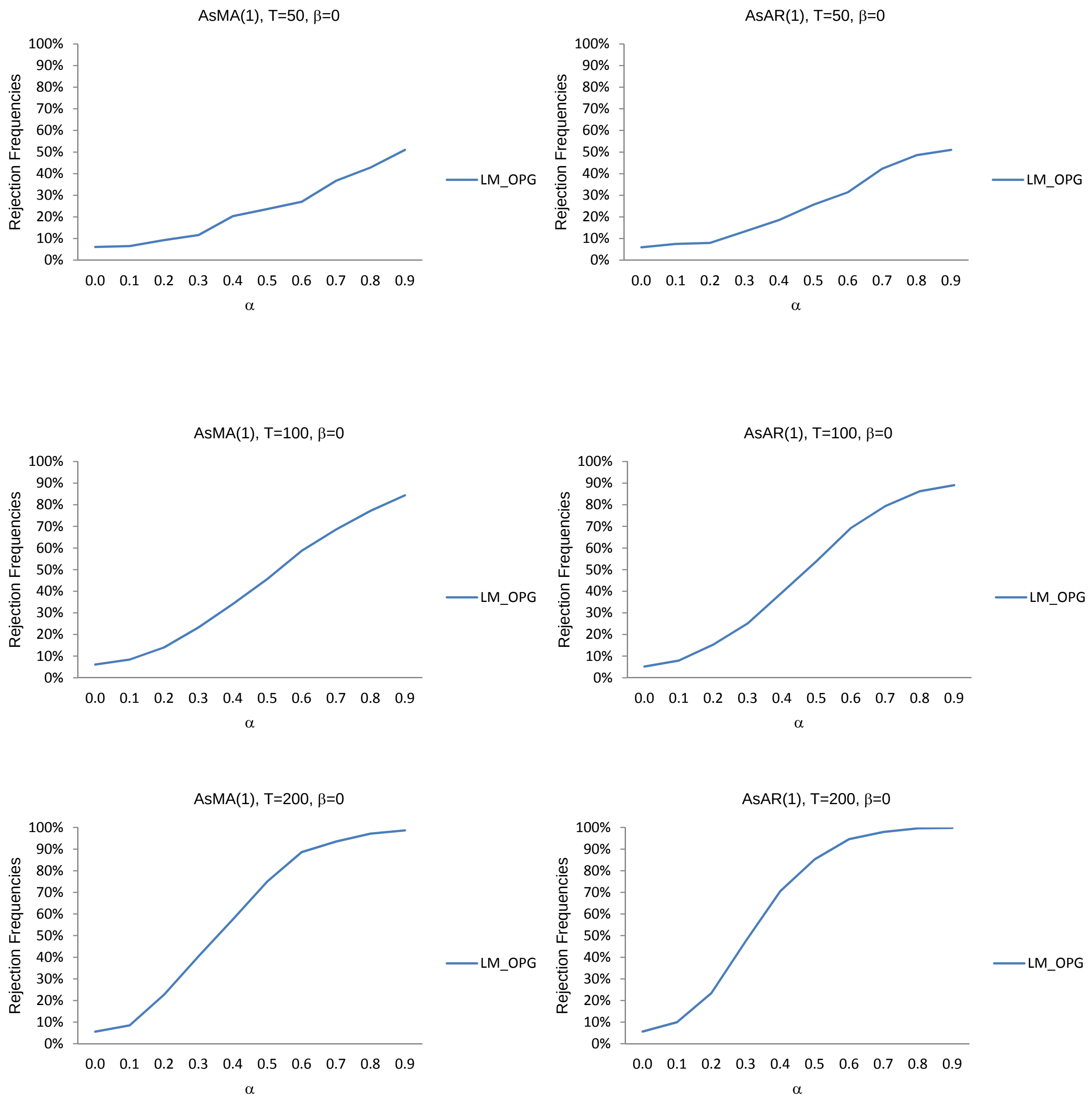


Figure 1.8: Power of the LM_T Variants when the Lag Length is Determined in a Preliminary Stage on the basis of the BIC

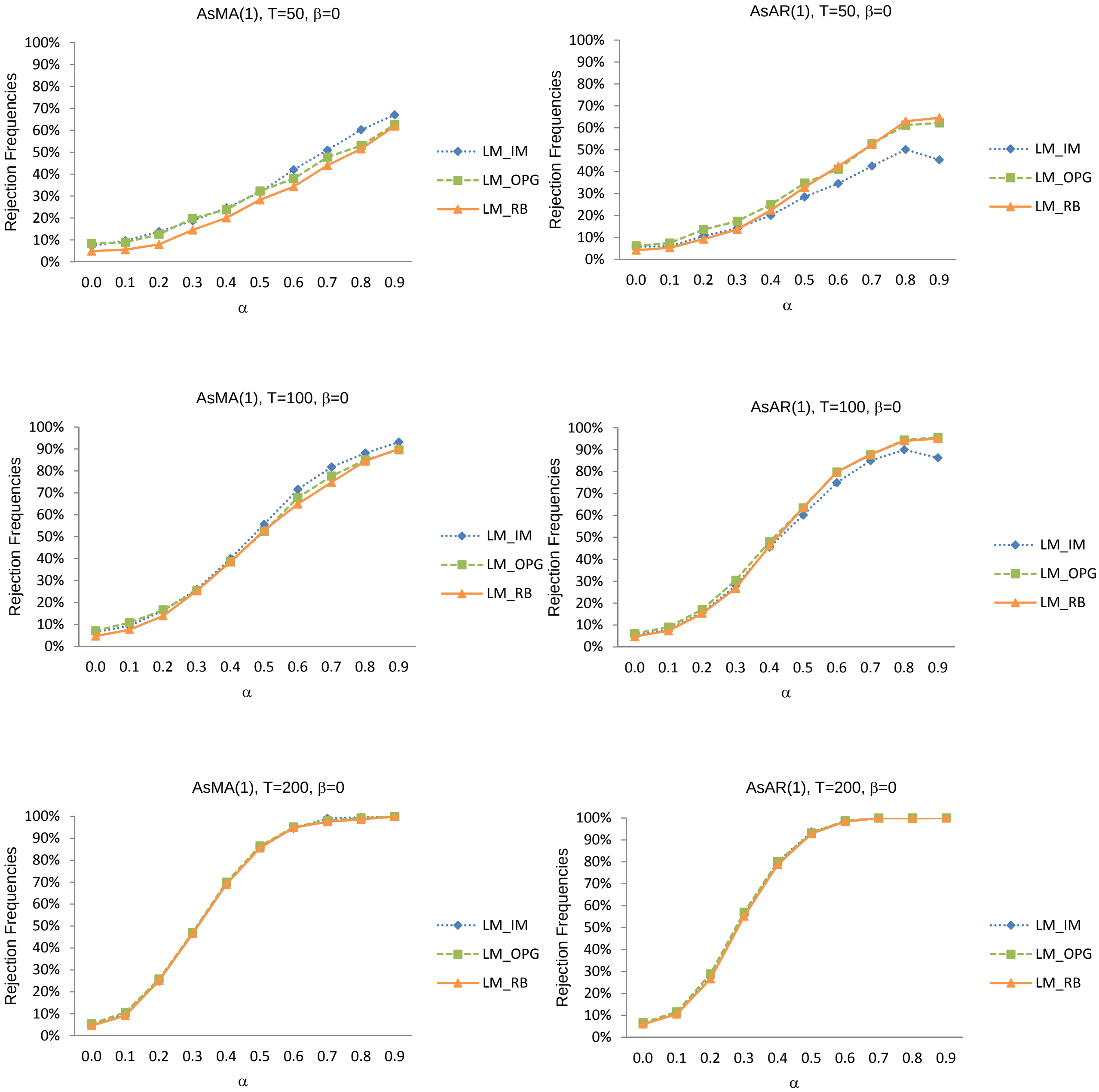
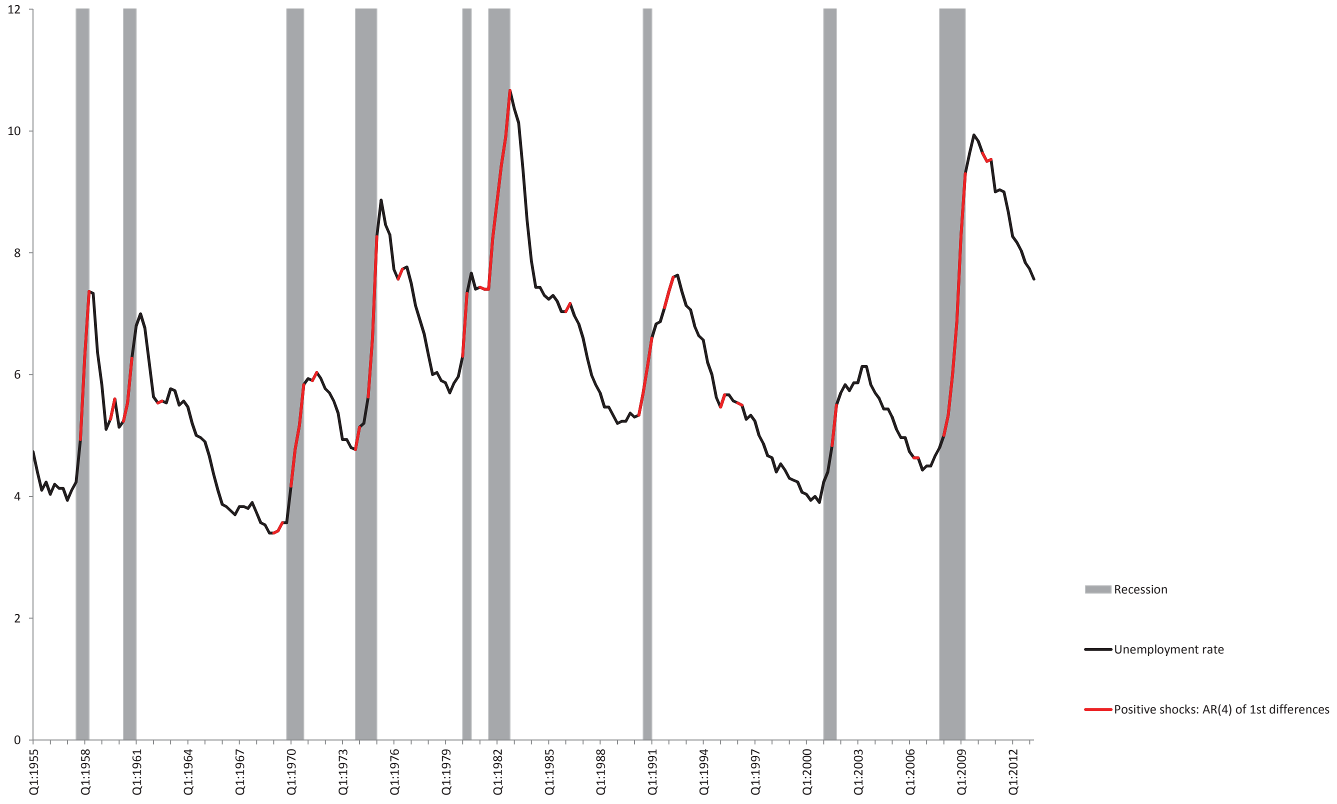
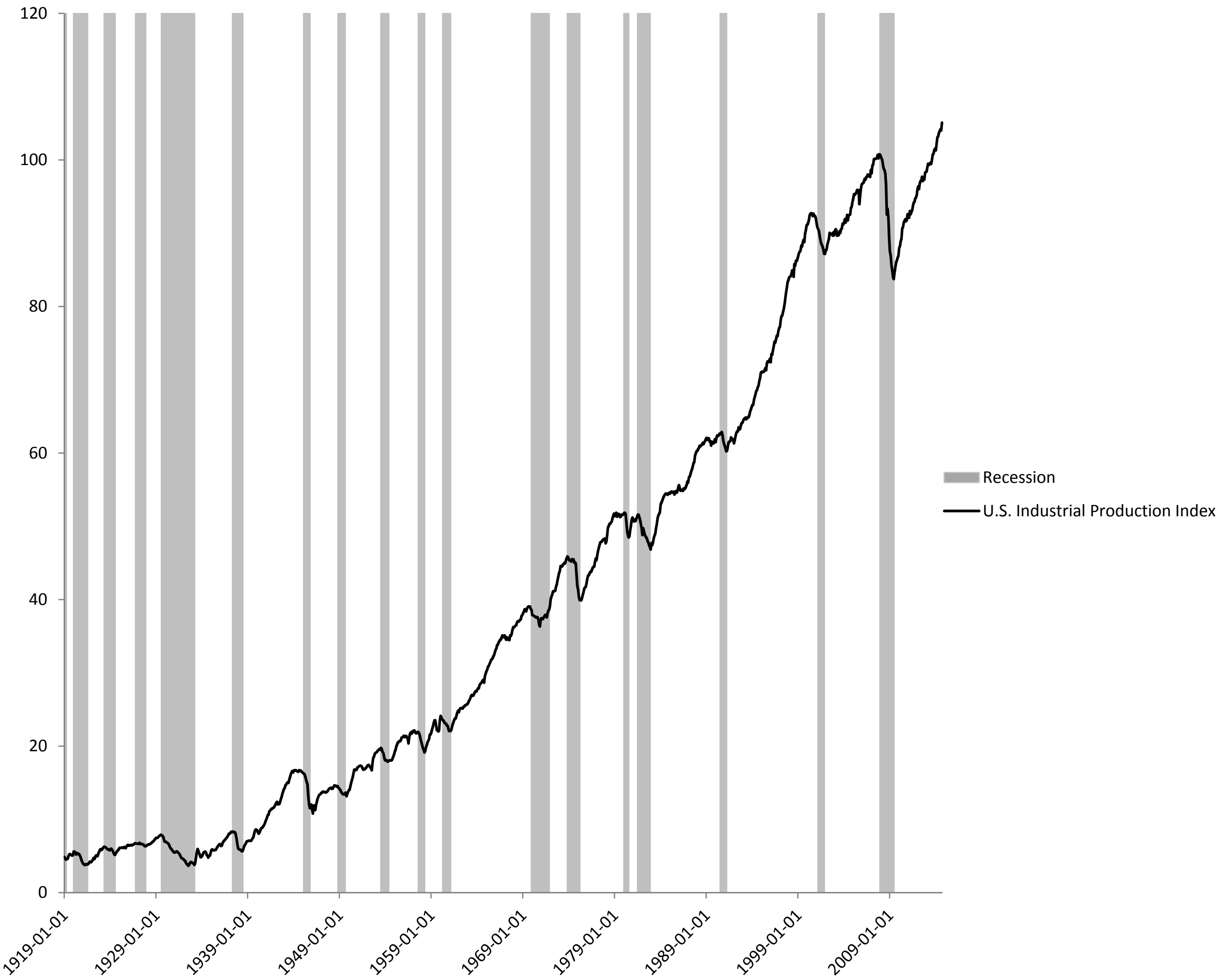


Figure 1.9: Quarterly U.S. Unemployment Rate in %



Notes: The observation period of the unemployment rate is 1955 Q1 to 2013 Q3. The data is seasonally adjusted. The grey shaded areas mark recessions.

Figure 1.10: Monthly U.S. Industrial Production Index



Notes: The observation period of the index is January 1919 to September 2014 and the basis year is 2007 (2007=100). The data is seasonally adjusted. The grey shaded areas mark recessions.

Chapter 2

Variance Bounds Tests and Forecast Evaluation

2.1 Introduction

More than three decades ago, Shiller (1981) and LeRoy and Porter (1981) initiated a discourse about whether stock prices are more volatile than traditional models would imply. Shiller derived a theoretical upper bound for the variance of the price of a stock in the present value model. He showed that the variance of the stock price has to be smaller than or equal to the variance of the ex post rational price. This descriptive measure is known as Shiller's variance bounds test. In an empirical application to the S&P 500 Composite Price Index, Shiller found that bound dramatically violated. As part of an explanation, the author pointed out that stock prices are five times too volatile to be accounted for by changes in fundamentals.

Although the evidence against the variance bound seemed to be striking at first glance, Shiller's findings were criticized in the subsequent discourse: The testing procedure was said to be prone to severe statistical problems under more realistic assumptions about the data generating process (Flavin, 1983; Kleidon, 1986a,b; Marsh and Merton, 1986). The critique led to the development of theoretical upper bounds under more general conditions (e.g. Mankiw et al., 1985; Engel, 2005; Lansing, 2015) as well as to a second generation of variance bounds tests in order to meet the aforementioned scepticism (e.g. West, 1988; Mankiw et al., 1991). Although empirical applications of these procedures still provided evidence against the view that stock prices are driven by fundamentals, the rejections of the hypothesis of market effiency are only marginal (Mankiw et al., 1991).

In this chapter, we show that the variance bounds test by Mankiw et al. (1991) is directly linked to the well known forecast encompassing test (e.g. Harvey et al., 1998). This observation gives rise to the idea of incorporating other evaluation techniques into the testing procedure. We argue that the Diebold-Mariano test (Diebold and Mariano, 1995) is a natural candidate: While the forecast encompassing test relies on a set of assumptions that are not necessarily satisfied in the financial market context, the Diebold-Mariano test (henceforth DM) is generally known to be a very robust evaluation tool. We apply the testing procedure to the monthly S&P 500 Composite Price Index. In contrast to Mankiw et al. (1991) we find that naive forecasts are not able to outperform the market price as a predictor for the ex post rational price.

Yet, using more sophisticated forecasting models allows us to find better predictors. Here, we exploit the fact that the price and the dividend series of the S&P 500 Composite Price Index are cointegrated. Based on this observation, Vector Autoregressive (henceforth VAR) and Vector Error Correction Models (henceforth VECM) are estimated in different specifications as a forecasting exercise. In addition, we employ two distinct forecasting approaches: First, we build a multi-step forecast by reiterating the one-step ahead forecast (iterated multi-step, henceforth IMS). Second, we construct forecasting models for specific time horizons (direct multi-step forecasts, henceforth DMS). Both procedures are commonly used in practice and the preference for one over the other of the two approaches usually is an empirical matter (see e.g. Stock and Watson, 2004; Marcellino et al., 2006). The results suggest that the VECM based on logarithmised price and dividend values and using IMS forecasts outperforms the actual stock price as a predictor for the ex post rational price at least for sufficiently long holding periods of the underlying stock.

The remainder of this chapter is structured as follows. In Section 2.2 and 2.3 we briefly discuss variance bounds tests of the first and second generation (for a more detailed overview see e.g. Gilles and LeRoy, 1991). In Section 2.4 we link these tests to the forecast evaluation literature. An application to the S&P 500 Composite Price Index is presented in Section 2.5.

In the last section we conclude.

2.2 First Generation Tests: Shiller's Test

In this section we describe the first generation test. In the following we first introduce the necessary notation before discussing Shiller's (1981) test procedure in more detail.

The standard present value model is given by:

$$\begin{aligned} P_t &= \gamma E_t(D_t + P_{t+1}) \\ &= \sum_{i=0}^{\infty} \gamma^{i+1} E_t D_{t+i} \\ &= E_t(\sum_{i=0}^{\infty} \gamma^{i+1} D_{t+i}) \\ &= E_t(P_t^*), \end{aligned} \tag{2.1}$$

where

$$\begin{aligned} P_t &= \text{the price of the stock at time } t; \\ D_{t+i} &= \text{the dividend paid at the end of period } t+i; \\ E_t &= \text{the expectation operator conditional on information available at time } t; \\ \gamma &= \text{the discount factor } \frac{1}{1+r}, \text{ with } r \text{ being the constant rate of return;} \\ P_t^* &= \text{the ex post rational price.} \end{aligned}$$

In that standard present value model the price of a stock is equal to the expected present value of its future dividends. This implies that the price P_t is an optimal forecast for the ex post rational price P_t^*. The forecast error ε_t is then defined as:

$$\varepsilon_t = P_t^* - P_t.$$

Under rational expectations, the forecast error is uncorrelated with information up to time t, i.e. $E_t(\varepsilon_t|\mathcal{I}_t) = 0$ where $\mathcal{I}_t$ corresponds to an information set including all information available at time t. It follows that

$$Var(P_t^*) = Var(P_t) + Var(\varepsilon_t).$$

Based on the non-negativity property of variances, that decomposition leads to the following inequality (Shiller, 1981):

$$Var(P_t^*) \geq Var(P_t).$$

In other words, the variance of the forecast P_t is smaller than the variance of the forecasted variable P_t^*.

Shiller assumes the dividend D_t to be stationary or at least stationary after detrending the dividend series. If that assumption holds, we know that D_t has a moving average representation:[1]

$$D_t = \phi(L)\varepsilon_t,$$

where $\phi(L)$ is the moving average operator with $\phi_0 = 1$ and $Var(D_t) = \sum_{i=0}^{\infty} \phi_i^2 \sigma^2$.

Consider now the innovation operator $\delta_t = E_t - E_{t-1}$, where E_t again denotes the conditional expectation operator. Applying this operator to the present value model one can show that the innovation in price is related to the innovation in dividends as follows:

$$\begin{aligned}
\delta_t P_t &= \sum_{k=0}^{\infty} \gamma^{i+1} \delta_t D_{t+i} \\
&= \gamma \varepsilon_t + \gamma^2 \phi_1 \varepsilon_t + \gamma^3 \phi_2 \varepsilon_t + ... \\
&= \gamma \phi(\gamma) \varepsilon_t.
\end{aligned}$$

Further, one can use the Cauchy-Schwarz inequality in order to derive the following inequality:

$$\left(\sum \gamma^{i+1} \phi_i \right)^2 \leq \left(\sum \gamma^{2(i+1)} \right) \left(\sum \phi_i^2 \right).$$

[1] We set the mean of the dividends to zero since a nonzero mean would drop out of all variance expressions.

The above leads to an upper bound for the variance of $\delta_t P_t$ given the variance of D_t:

$$\begin{aligned} Var(\delta_t P_t) &\leq \frac{\gamma^2}{1-\gamma^2} Var(D_t) \\ &\leq \frac{Var(D_t)}{(1+r)^2 - 1}. \end{aligned}$$

These upper bounds are known as 'Shiller's variance bounds test', although the author does not provide a significance test in a statistical sense. In particular, Shiller's violation of the variance bounds result is purely based on descriptive statistics.

Moreover, Shillers' procedure is prone to some econometric pitfalls.[2] Flavin (1983) outlines that the variances of the actual price P_t and the ex post rational price P_t^* are estimated with downward bias. When estimating the variance of a population with unknown mean the estimator is given by

$$\widehat{\sigma}^2 = \sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{n-1}.$$

Yet, this estimator is only unbiased if the x_i are uncorrelated. However, both P_t and P_t^* are positively autocorrelated and, as Flavin (1983) points out, the persistence of P_t^* is stronger than that of P_t. As a result, the variance of P_t^* is estimated with a larger downward bias. Based on that argument, violations of the variance bounds might be due to an increased probability of a type I error.

Moreover, Flavin finds that Shiller's method to calculate an approximation for the unobservable variable P_t^* induces an additional bias towards rejection.[3] In order to see why, notice that the ex post rational price is the solution to the following recursive problem:

$$P_t^* = \beta(P_{t+1}^* + D_t)$$

satisfying

$$lim_{t\to\infty} \beta^t P_t^* = 0.$$

[2] For a more detailed overview see e.g. Gilles and LeRoy (1991).
[3] See also Shea (1989) for a more general discussion.

On the contrary, Shiller calculates P_t^* by solving the recursion satisfying the terminal condition $P_T^* = \frac{1}{T}\sum_{t=1}^{T} P_t$. In general, the resulting estimates for P_t^* are biased. However, this issue can be overcome when generating $P_{t|T}^*$ by the recursion and the terminal condition $P_{T|T}^* = P_T$. Yet it is important to notice that even then the sample variance is still biased. However, this effect is negligible in sufficiently large samples.

Kleidon (1986a) shows that the variance bounds test is problematic for small sample sizes. The author finds that the test tends to be biased towards rejection of the null hypothesis of market efficiency.

In another article, Kleidon (1986b) points out that Shiller's inequalities are based on a cross sectional and not on a time series perspective: If one was able to replicate the economy a large number of times, one would observe the following:

$$\widehat{Var}(P_t) = \frac{\sum_i (P_{it} - \overline{P_t})^2}{n-1} \leq \widehat{Var}(P_t^*) = \frac{\sum_i (P_{it}^* - \overline{P_t}^*)^2}{n-1},$$

where the summation refers to replications $i = 1, ..., N$. Kleidon finds that if N=1, as in the present case, one can not be sure that the variance bound is satisfied.

Marsh and Merton (1986), Kleidon (1986b) and Durlauf and Phillips (1988) argue that Shiller's variance bounds rely on the stationarity of the underlying series. In the context of prices and dividends that assumption is highly questionable even after deflating and detrending the series (see e.g. Nelson and Plosser, 1982). In that line of argumentation, Engel (2005) derives a reverse result to Shiller's inequality for nonstationary time series expressing prices in first-differences. However, Lansing (2015) more recently shows that this reversal does only hold for a very specific setting.

2.3 Second Generation Tests: Mankiw, Shapiro and Romer's Test

The critique of Shiller's test opened ground for the so called second generation tests which are described in this section. The procedure by Mankiw et al. (1985, 1991) is the most prominent among the tests of the second generation. The idea is the following: Suppose P_t is an optimal predictor for P_t^* with $E_t(P_t^* - P_t) = 0$. Let P_t° be any other (naive) predictor that can be derived from information which is available to agents at time t. Consider the following identity:

$$P_t^* - P_t^\circ \equiv \underbrace{(P_t^* - P_t)}_{\varepsilon_t} + (P_t - P_t^\circ).$$

Alternatively, accounting for potential heteroscedasticity in the variables, that identity can be written as:

$$E_t\left(\frac{P_t^* - P_t^\circ}{P_t}\right)^2 = E_t\left(\frac{P_t^* - P_t}{P_t}\right)^2 + E_t\left(\frac{P_t - P_t^\circ}{P_t}\right)^2. \tag{2.2}$$

The test statistic can now be derived in two steps. First, define:

$$\begin{aligned} d_t &= \left(\frac{P_t^* - P_t^\circ}{P_t}\right)^2 - \left(\frac{P_t^* - P_t}{P_t}\right)^2 - \left(\frac{P_t - P_t^\circ}{P_t}\right)^2 \\ &= \frac{P_t^* P_t^\circ + P_t^2 - P_t^* P_t - P_t P_t^\circ}{P_t^2}. \end{aligned} \tag{2.3}$$

Second, notice that equation (2.2) implies that $E(d_t) = 0$. That leads to a linear regression model of d_t:

$$d_t = \alpha + \eta_t,$$

where α is a constant and η_t is a zero mean error term. That model allows to test if $\alpha = 0$ and thus if the theoretical results hold up.

Mankiw et al. (1985) suggest that their test is unbiased[4] and does not rely on the stationarity of the dividend series. Moreover, they incooperated Flavin's (1983) ideas by using noncentral instead of central variances.[5] Yet, Shea (1989) critizises that the test procedure is sensitive to the choice of the terminal date in the approximation of the ex post rational price P_t^*. Therefore,

[4]As noted by Gilles and LeRoy (1991), the terms 'test' and 'unbiased test' in the variance bounds literature are not equivalent with their common meaning in the econometric literature (this is also in line with the observation that Shiller's procedure is only based on descriptives). These variance bounds tests reject the null hypothesis of market efficiency, if the sample estimates of the variances don't reflect the theoretical variance bounds inequalities. A confidence region is not specified. Furthermore, 'unbiased test' means that the expectation of the test statistic is unbiased. This does not ensure that the test is unbiased in the usual econometric sense.

[5]Mankiw et al. (1985) do not center the variances around the sample mean but around a naive forecast.

Shea suggests to use a fixed holding period h instead of $h = T - t$ to calculate P_t^* in

$$P_t^{*h} = \sum_{i=0}^{h-1} \gamma^{i+1} D_{t+i} + \gamma^h P_{t+h}. \tag{2.4}$$

There are two obvious disadvantages of that approach. On the one hand, the resulting $P_{t|t+h}^*$ does not equal the expectation of P_t^* conditional on the whole sample. On the other hand the loss of observations leads to less precise estimates. Taking these notions into account, Mankiw et al. (1991) calculate their test statistic on the basis of different holding periods $h = 1, 2, 5, 10, T - t$ using annual S&P 500 data.

2.4 Variance Bounds Tests and Forecast Evaluation

In this section, we link the variance bounds tests to forecast evaluation results. In particular, it is interesting to note that the variance bounds test by Mankiw et al. (1991) is directly linked to the well-known forecast encompassing test (e.g. Harvey et al., 1998). In the current context, the test regression for the classical forecast encompassing test can be written as follows:

$$\begin{aligned} \underbrace{\left(P_t^{*h} - P_t\right)}_{e_{1t}} &= \lambda \left(P_t^{\circ} - P_t\right) + \varepsilon_t \\ &= \lambda \underbrace{\left(\left(P_t^{\circ} - P_t^{*h}\right) + \left(P_t^{*h} - P_t\right)\right)}_{e_{1t} - e_{2t}} + \varepsilon_t, \end{aligned}$$

where e_{it} denotes the respective forecast error from the actual price P_t or the naive forecast P_t° and ε_t is a zero mean error term. We are interested in testing the null hypothesis whether λ equals zero: Under the null hypothesis, the naive forecast does not add any valuable information. In that case, Granger and Newbold (1973) describe the actual stock price as 'conditionally efficient' with respect to the naive forecast. The resulting OLS estimator for λ is given by:

$$\widehat{\lambda} = \frac{\frac{1}{T} \sum_{t=1}^{T} \left(P_t^{\circ} - P_t\right) \left(P_t^{*h} - P_t\right)}{\frac{1}{T} \sum_{t=1}^{T} \left(P_t^{\circ} - P_t\right)^2}.$$

Expanding the numerator yields:

$$\left(P_t^{\circ} - P_t\right)\left(P_t^{*h} - P_t\right) = P_t^{*h} P_t^{\circ} + P_t^2 - P_t^{*h} P_t - P_t P_t^{\circ}.$$

Comparing this expression with the numerator of equation (2.3) demonstrates the direct link between the two statistics. However, as Harvey et al. (1998) point out, the classical forecast encompassing test relies on the restrictive assumption of normality distributed forecast errors. This assumption is highly optimistic in a financial market context, where one would intuitively expect fat-tailed error distributions. A violation of the normality assumption results in size distortions of substantial magnitude.

Therefore, in order to receive reliable results, one needs to apply more robust techniques in the assessment of variance bounds. One obvious candidate is the Diebold-Mariano test (Diebold and Mariano, 1995). As before, let P_t and P_t° denote two competing forecasts for the ex post rational price P_t^*. The forecast erros from the two models can be written as:

$$\begin{aligned} \varepsilon_t^1 &= P_t^* - P_t^{\circ} \\ \varepsilon_t^2 &= P_t^* - P_t. \end{aligned}$$

The accuracy of a forecast can be measured with a chosen loss function, e.g. squared error loss. Let q_t denote the loss differential:

$$\begin{aligned} q_t &= (P_t^* - P_t^{\circ})^2 - (P_t^* - P_t)^2 \\ &= (\varepsilon_t^1)^2 - (\varepsilon_t^2)^2. \end{aligned}$$

Based on that loss differential we are able to evaluate whether the current stock price is a better predictor for the ex post rational price than any other forecast. In statistical terms that means we can formulate the null and alternative hypothesis in terms of the expected loss differential:

$$H_0 : E(q_t) \geq 0 \qquad vs. \qquad H_1 : E(q_t) < 0.$$

An appropriate test statistic is developed in Diebold and Mariano (1995):

$$DM = \sqrt{T}\frac{\overline{q}}{\sigma_q} \sim \mathcal{N}(0,1),$$

where $\overline{q}$ corresponds to the mean of q_t, σ_q to a consistent estimate of the standard deviation of q and T to the sample size. Hence, we reject the null hypothesis of market efficiency at the 5% significance level if DM< -1.64.

Note that the Diebold-Mariano as well as the forecast encompassing test are companion tests: While the latter tests an orthogonality condition of the forecast errors that allows for assessing whether a combined forecast has a lower squared error than one of the single forecasts, the DM test compares the accuracy of the competing forecasts in a more direct approach. Moreover, the DM test has the advantage over the forecast encompassing test of good size and power properties under minimal assumptions. The only assumption that needs to be satisfied refers to the covariance stationarity of the loss differential.

2.5 Application

In the following, we apply the DM test to a monthly series of the S&P 500 Composite Price Index.[6] The data set contains monthly dividend and price series from January 1871 to June 2013. Here, we use real prices and dividends to account for inflation specific effects. In order to deflate the series, Shiller used the Producer Price Index (PPI) published by the U.S. Bureau of Labor Statistics from 1913 onwards and the Warren and Pearson price index for the years before.

The ex post rational price P_t^* is a latent variable. Hence, one needs to approximate the true value of this variable. In order to calculate the ex post rational price we apply the following trading strategies: The S&P 500 is bought at time t and held for h periods. Up to period h, a stream of dividends is paid. We consider holding periods of $h = 1, 6, 12$ and $T - t$ months.

[6] The data can be downloaded from the website `http://www.irrationalexuberance.com/` by Robert Shiller.

Thus, an approximation for the ex post rational price can be written as

$$P_t^{*h} = \sum_{i=0}^{h-1} \gamma^{i+1} D_{t+i} + \gamma^h P_{t+h}.$$

The discount factor γ corresponds to the inverse of $1 + r_{12}$ where r_{12} is the monthly analog of the constant annual interest rate r. The rate of return is a latent variable and hence we follow Mankiw et al. (1991) in calibrating the interest rate to the values $r = 0.05, 0.06, 0.07$. Alternatively, one can estimate the constant interest rate r on the basis of equation (2.1) using observable variables as described in West (1988) or Gürkaynak (2005). In any case, the corresponding estimate of the interest rate $\widehat{r}$ is afflicted with high uncertainty and varies between 2% and 6%.

2.5.1 Simple Forecasts

In order to construct our naive forecast P_t°, we assume that the dividend series follows a random walk or a random walk with drift, respectively. The respective forecasts are hence given by (e.g. Evans, 1991):

$$\begin{aligned} P_t^{RW} &= \frac{D_t}{r} \\ P_t^{RWD} &= \frac{1+r}{r^2}\mu + \frac{D_t}{r}. \end{aligned}$$

These naive types of forecasts are often used as benchmarks for competing forecasts in the literature. In the following, we define μ as the average month-to-month difference of the dividend series over the past six and twelve months, respectively:[7]

$$\begin{aligned} \mu_t^6 &= \frac{1}{6} \sum_{i=t-6+1}^{t} \Delta D_i \\ \mu_t^{12} &= \frac{1}{12} \sum_{i=t-12+1}^{t} \Delta D_i. \end{aligned}$$

[7] We do not consider actual month-to-month differences directly in order to reduce the influence of possible outliers.

We apply the Diebold-Mariano test for the forecasts as described above. We correct for the autocorrelation of the multi-period forecast errors: An efficient h-period forecast has forecast errors that follow an MA(h-1) process. As recommended by Diebold and Mariano, we hence use a Newey-West type estimator for the sample variance of the loss differential.

The results are shown in the left panel of Table 2.1 for the baseline version of the test and for the normalized variant in the right panel of Table 2.1, repectively. The normalization is based on the non-covariance stationarity of the loss differential in the standard test with $h = T - t$. The forecast errors are normalized by dividing them by P_t as in Mankiw et al. (1991). As can be seen in Table 2.1, values for the test statistic exceed the critical value of -1.64. That means that we don't find evidence for simple forecast models, as the random walk or the random walk with drift, being able to outperform the actual price in terms of the mean squared error. This is indeed an important observation, since it contradicts Mankiw et al. (1991) who found that naive forecasts often outperform the market price at least over long holding periods.

2.5.2 Cointegration

As outlined in Section 2.2, the stationarity assumption in Shiller (1981) with respect to the dividend and the price series has been challenged in the literature (e.g. Kleidon, 1986b). In line with that, the plots of both series in Figure 2.1 seem to confirm the scepticism about the stationarity. In order to evaluate statistically whether we have unit roots in our data, we apply the Augmented Dickey-Fuller (ADF) test to the different components of the net present value model.[8]

The lag length selection for this testing procedure was conducted using an automatic search based on the Schwarz Information Criterion (SIC). The results are illustrated in Table 2.2. The null hypothesis of a unit root is accepted for all series. Furthermore, taking a closer look at the p-values acceptance of nonstationarity is not a marginal issue. Therefore, we have to bear in mind that we have to forecast in a nonstationary environment.

The notions above lead directly to the concept of cointegration: Any reasonable forecast P_t°

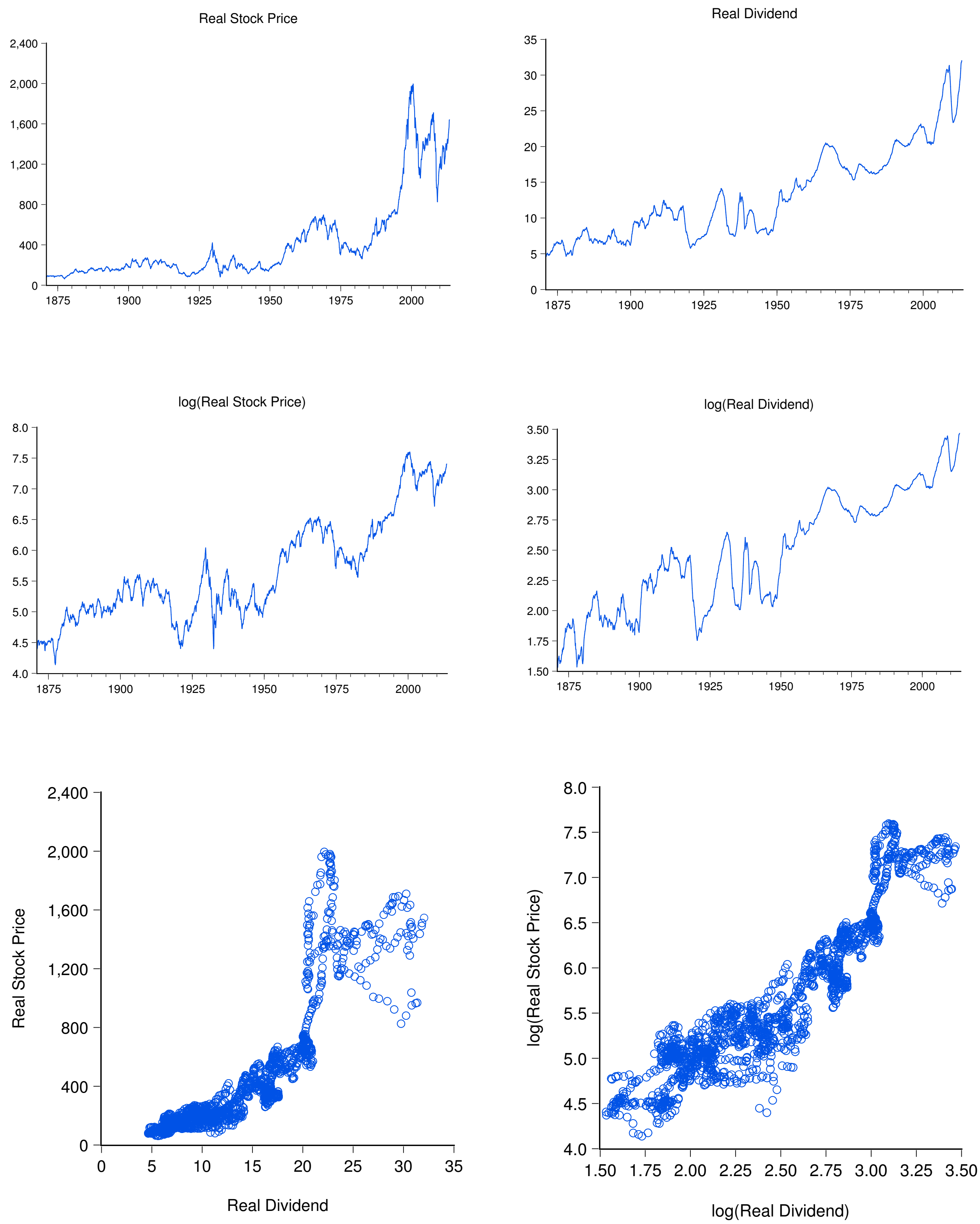
[8]For the calculations of the test statistics for the ex post rational price P_t^* we assume an interest rate r of 5%. The results remain qualitatively the same for $r = 6\%$ and $r = 7\%$.

Table 2.1: Diebold-Mariano Test: Simple Forecasts

	Standard			Normalized		
	Random walk	*Random walk with drift* (μ_t^6)	*Random walk with drift* (μ_t^{12})	*Random walk*	*Random walk with drift* (μ_t^6)	*Random walk with drift* (μ_t^{12})
			r=5%			
h						
1	14.464	13.638	13.918	26.661	13.679	27.301
6	0.707	4.055	3.836	4.517	3.835	3.577
12	1.210	3.702	3.323	3.806	3.724	3.582
T-t	-	-	-	-0.972	3.661	3.403
			r=6%			
h						
1	15.319	13.851	14.041	34.121	13.602	34.301
6	0.757	3.716	3.344	4.526	3.774	2.985
12	1.193	3.372	2.965	3.608	3.698	3.581
T-t	-	-	-	2.665	3.533	3.264
			r=7%			
h						
1	16.004	14.592	14.749	42.450	13.724	14.623
6	0.744	3.305	2.846	5.319	3.761	3.018
12	1.187	3.002	2.566	4.394	3.728	3.666
T-t	-	-	-	6.361	3.495	3.251

Notes: The table contains DM statistics for the simple forecasting models described in section 2.5.1. The models are evaluated for different holding periods h and different interest rates r. The null hypothesis is rejected at the 5% significance level if DM< -1.64.

Figure 2.1: Real Stock Price and Real Dividend Series



Notes: Whilst the upper panel shows the price and the dividend series in levels, the middle panel shows both series in log-levels. The two panels at the bottom contain scatterplots of both series in levels and in log-levels respectively.

Table 2.2: Augmented Dickey-Fuller Test: Price, Dividend, Ex Post Rational Price

	t-statistic	p-value
P_t	-0.112	0.946
D_t	-0.182	0.938
$log(P_t)$	-1.002	0.754
$log(D_t)$	-1.331	0.617
P_t^{*1}	-0.187	0.938
P_t^{*6}	-0.400	0.907
P_t^{*12}	0.147	0.969
P_t^{*T-t}	4.035	1.000

Notes: The table contains test statistics and p-values for the augmented Dickey-Fuller Test without a trend in the test regression.

should be cointegrated with the variable being predicted. In technical terms this means that the resulting forecast error series $P_t^{*h} - P_t^{\circ}$ is $I(0)$ and hence the variance exists. In order to test whether there exists a cointegration relationship between the variables can be evaluated by means of cointegration tests. Here we use the Johansen as well as the Engle-Granger cointegration tests applied to the entire sample period.

The order of the VAR for the Johansen test is determined as follows: First of all, a search method based on the SIC is used up to a lag of twelve periods. Afterwards, we check for remaining autocorrelation using the LM test and increase the number of lags until no further autocorrelation can be detected. Yet, in the case of the level price and the dividend series it is not possible to entirely eliminate the autocorrelation[9], whereas in the case of log-levels this is not a concern. For the Engle-Granger cointegration test we use an automated search based on the SIC as well. The results for both testing procedures are shown in Table 2.3. As recommended by Gonzalo and Lee (1998), we have used both tests in order to avoid pitfalls of the respective testing procedure. As such, the Johansen test collapses for $h = T - t$ due to a (near) singular error covariance matrix in the VAR representation.

All combinations of the different series are cointegrated of order one at least at the 10% significance level. That might not be a suprising but still an interesting result: It does not only tell us that all series stick in some sense to the predicted variable, but also that alternative

[9] One higher order lag was still significant.

forecasts based on the respective cointegration relation can be constructed.

Table 2.3: Johansen and Engle-Granger Cointegration Test

Johansen Cointegration Test					
	r=	L_{trace}	p-value	$L_{eigenvalue}$	p-value
D_t, P_t^{*1}	0	14.679	0.066	14.630	0.044
	1	0.049	0.824	0.049	0.824
D_t, P_t^{*6}	0	20.431	0.008	20.277	0.005
	1	0.154	0.695	0.154	0.695
D_t, P_t^{*12}	0	19.650	0.011	19.595	0.007
	1	0.055	0.814	0.055	0.814
D_t, P_t^{*T-t}	0	-	-	-	-
	1	-	-	-	-
D_t, P_t	0	13.939	0.085	13.916	0.057
	1	0.023	0.879	0.023	0.879
$log(D_t), log(P_t)$	0	34.324	0.000	33.503	0.000
	1	0.821	0.365	0.000	0.365

Notes: The table contains empirical values and p-values of the Johansen trace and the Johansen maximum-eigenvalue statistic for different combinations of prices and dividends; r denotes the cointegration rank. Cases with singularity issues are marked by '-'.

Engle-Granger Cointegration Test				
	τ-statistic	p-value	z-statistic	p-value
D_t, P_t^{*1}	-3.866	0.011	-30.860	0.006
D_t, P_t^{*6}	-4.061	0.006	-35.706	0.002
D_t, P_t^{*12}	-3.905	0.010	-32.827	0.004
D_t, P_t^{*T-t}	-4.176	0.004	-35.567	0.002
D_t, P_t	-3.514	0.032	-25.473	0.018
$log(D_t), log(P_t)$	-4.474	0.001	-39.822	0.001

Notes: The table contains empirical values and p-values of the Engle-Granger cointegration τ- and z-statistic for different combinations of prices and dividends.

2.5.3 Multivariate Forecasts

In this section we describe and estimate a panoply of multivariate forecasting models. A common feature of the models described in the following is that they all exploit the cointegration relation between the actual stock price, the ex post rational price and the dividend series, even if to a different extend.

For now, consider the following regression equation:

$$P_t^{*h} = \mu + \alpha D_t + u_t, \tag{2.5}$$

where μ is an intercept and u_t is a zero mean error term. This equation reflects the long run relation between the ex post rational price and the dividend series. It is equivalent to the first step in the Engle-Granger procedure for error correction models. We estimate the parameters of this model by using a rolling window of 300 observation. The rolling window allows for changing parameters over the sample.

The dividend payment for the next period is known in advance. Hence, we can create a forecast for the ex post rational price using the estimates from the preceding observation period in the following manner

$$P_t^{\circ} = \widehat{\mu} + \widehat{\alpha} D_t.$$

The first column of Table 2.4 (REG) shows the results of the DM test comparing the performance of the forecasts from this model with the actual stock market prices as predictors for the ex post rational price. All results are statistically significant at the 5% level: Our forecast is a better predictor for the ex post rational price than the actual stock price. Compared to earlier studies which found significance only for long holding horizons of the underlying asset, this is a very interesting result. However, the results presented here are only valid for holding periods of one month. In terms of longer holding periods, one has to keep in mind that P_t^* is based on information from future values of P_t and D_t. Therefore, the results for longer horizons are inflicted by the assumption of known future stock payments. Consequently, we turn to procedures that avoid these complications.

In the following, we consider forecasts based on bivariate vector autoregressions (VARs). In the case stated here the model can be written as

$$y_t = c + A_1 y_{t-1} + ... + A_p y_{t-p} + \varepsilon_t, \tag{2.6}$$

where $y_t = (P_t, D_t)'$ denotes the vector of observed variables including the price and the dividend series, c is a vector of equation specific constants, A_i (i=1,...,p) are (2×2) parameter matrices and ε_t is a zero mean error term. Notice that in contrast to model (2.5) this model does not explicitly include the ex post rational price P_t^* and as a result avoids the aforementioned difficulties.

For the forecasting exercise we choose again a rolling regression approach with a moving window of 600 observations. The lag order of the VAR is determined by an automated search based on the SIC up to a maximum of twelve lags. Based on that choice of lags we reestimate the parameters of the model for every period and construct out of sample forecasts for the dividend and the price series up to a horizon of twelve months:

$$\widehat{y}_{t+h|t} = \widehat{c} + \widehat{A}_1 \widehat{y}_{t+h-1|t} + ... + \widehat{A}_p \widehat{y}_{t+h-p|t}, \tag{2.7}$$

where $\widehat{y}_{t+h|t}$ denotes the h-step ahead forecast of y_t at time t, $\widehat{A}_i$ with i=1,...,p represents the estimated coefficient matrices of the system, and $\widehat{y}_{t+h-i|t} = y_{t+h-i}$ if $h - i \leq 0$. Hence, in the case of a one-step ahead forecast (h=1), the prediction can be derived by just inserting the observed values of the time series in equation (2.7). For longer horizons, predicted values of the dividend and the price series are recursively inserted into the formula, resulting in iterated multi-step forecats (IMS).

We exploit levels and log-levels of both series respectively for our forecasting exercise. As noted by Lütkepohl and Xu (2012) the log-transformation, although often applied without fundamental reasons, can help to stabilize the volatility and therefore improves the forecasting ability of our model. That stabilizing effect is illustrated in Figure 2.2 at least for the price which is the main building block in the construction of a forecast for the ex post rational price (at least for the holding periods being studied here).

The forecasted values, and in the case of the log-levels the forecasted values transformed back to levels respectively, are inserted into equation (2.4). This allows us to construct a forecast for the ex post rational price P_t^* out of these values.

Another, and potentially more natural way to proceed is to create a vector error correction model (VECM) for the price and the dividend series. Phillips (1998) shows that forecasts based on a VECM that explicitly estimates cointegrating relationships -as far as cointegration is present- are consistent and asymptotically optimal. Here, the VECM can be represented as follows:

$$\Delta y_t = \Pi y_{t-1} + \Sigma_1 \Delta y_{t-1} + ... + \Sigma_k \Delta y_{t-k} + \eta_t, \tag{2.8}$$

where $\Delta y_t = y_t - y_{t-1}$ denotes the first difference of y_t, Π represents the long-run parameters, Σ_j the short-run parameters for $j = 1, ..., k$ and η_t a zero mean error term of the VECM. To specify the model, we use the same procedure as in the case of the VAR. Hence, the h-step ahead forecast is given by:

$$\Delta \widehat{y}_{t+h|t} = \widehat{\Pi} \widehat{y}_{t+h-1|t} + \widehat{\Sigma}_1 \Delta \widehat{y}_{t+h-1|t} + ... + \widehat{\Sigma}_k \Delta \widehat{y}_{t+h-k|t}, \tag{2.9}$$

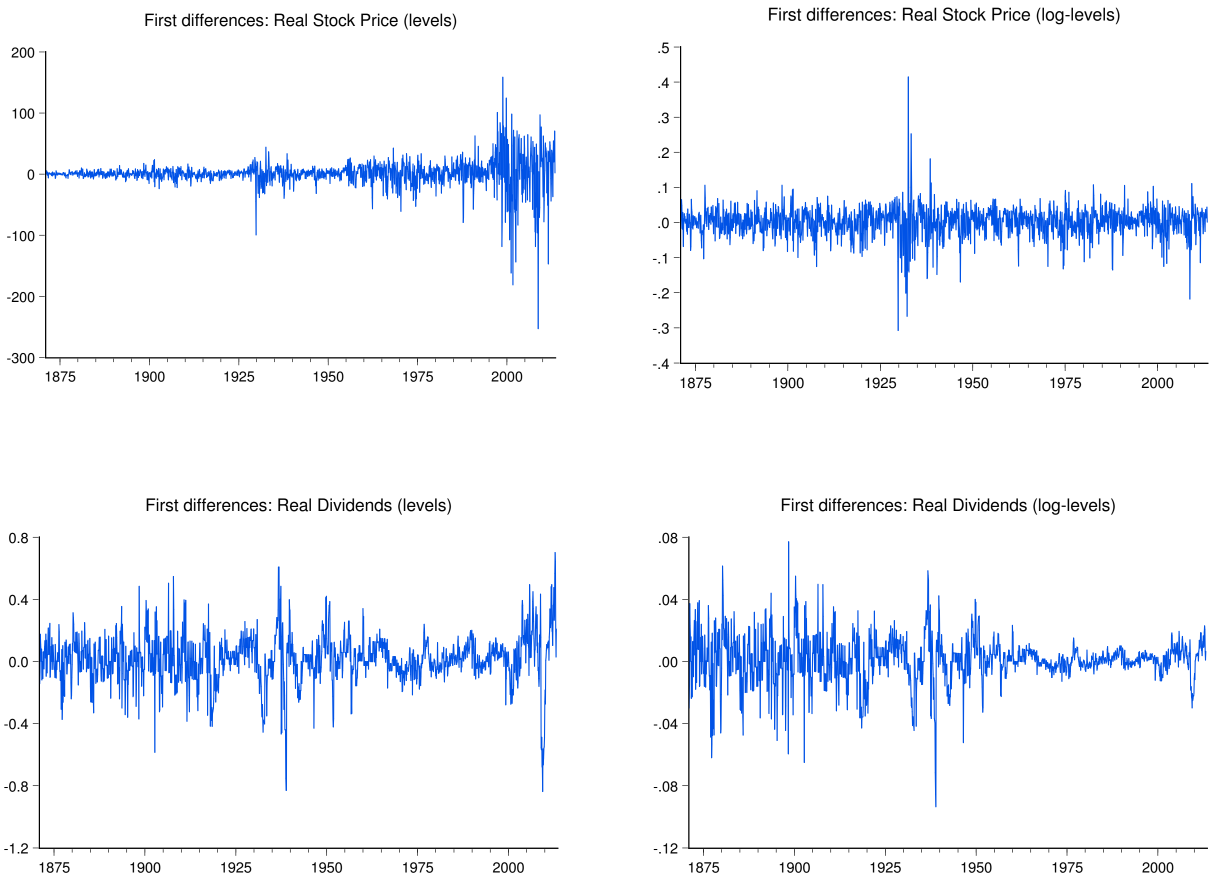
where $\widehat{y}_{t+h-j|t} = y_{t+h-j}$ if $h - j \leq 0$. As before, multi-step forecasts are constructed by reiterating the one-step ahead forecast.

Columns 3 to 6 of Table 2.4 report results of the Diebold-Mariano test for the different VAR and VECM specifications. The results show that the VECM approach leads to forecasts that outperform the actual price as a predictor when using log-levels for holding periods of 6 and 12 months. All of the resulting values are smaller than the critical value of -1.64 and hence we have strong rejections at the 5% significance level.

These findings are in line with past empirical evidence that shows that in the presence of cointegration VECMs are superior to VARs for long horizon forecasts, while the evidence is somewhat mixed for the very short run. For all other specifications the resulting forecasts for the ex post rational price are less successful compared to the actual market price. However, the perfomance of the VAR and VECM forecasts seems to improve with increasing forecast horizon. We discuss that finding in more detail later on.

We test the robustness of our results by choosing different numbers of observations for the rolling window of the VAR and the VECM. The results remain qualitatively the same.

Figure 2.2: Differenced Real Stock Price and Real Dividend Series



Notes: The left panel contains plots of the differenced price and dividend series in levels. The right panel shows the same for log-levels.

Table 2.4: Diebold-Mariano Test: Multivariate Forecasts

		Iterated multi-step forecasts				Direct multi-step forecasts			
Functional form	REG(normalized) levels	VECM levels	VECM logs	VAR levels	VAR logs	VECM levels	VECM logs	VAR levels	VAR logs
				r=5%					
h									
1	-11.301	8.339	6.617	8.590	8.502	8.339	6.617	8.590	8.502
6	-1.758	2.877	-2.588	2.635	3.304	3.510	3.513	2.797	3.067
12	-2.145	1.590	-3.979	1.896	1.877	1.963	1.974	2.830	2.734
T-t	-2.911	-	-	-		-	-	-	-
				r=6%					
h									
1	-11.264	8.400	6.675	8.652	8.556	8.400	6.675	8.652	8.556
6	-1.731	2.634	-2.162	2.730	4.661	3.605	3.594	2.852	3.104
12	-2.079	2.161	-3.892	2.352	1.882	1.508	1.518	2.856	2.699
T-t	-2.902		-	-	-	-	-	-	-
				r=7%					
h									
1	-11.228	8.457	6.730	8.711	8.606	8.457	6.730	8.711	8.606
6	-1.703	2.963	-2.030	3.007	4.697	3.623	3.598	2.901	3.134
12	-2.008	2.600	-3.802	2.732	1.884	1.121	1.128	2.874	2.655
T-t	-2.902	-	-	-	-	-	-	-	-

Notes: The table contains DM statistics for the multivariate forecasting models described in section 2.5.3. The models are evaluated for different holding periods h and different interest rates r. Moreover we distiguish between iterated and direct multi-step forecasts. The null hypothesis is rejected at the 5% significance level if DM< -1.64.

Direct Forecasts

Up to this point we have built our multi-step forecasts on iterating the one-step ahead forecasts. The one-step ahead estimation minimizes the square of the one-step ahead residuals. Further, one might be interested in whether iterating the one-step ahead forecast yields the best predictors for multi-step forecasts.

An alternative to the above is the direct multi-step forecast approach (DMS). The intuition behind that method is to estimate a model for a specific forecast horizon h and thereby minimizing the square of the h-step ahead residuals. From a theoretical perspective DMS can be superior to IMS in specific scenarios (see e.g. Clements and Hendry, 1996; Schorfheide, 2005; Chevillon and Hendry, 2005). However, there is a trade-off when choosing the best suited approach: On the one hand, the IMS approach leads to more efficient parameter estimates. On the other hand it can be severely biased if the model for the one-step-ahead forecast is misspecified. In contrast, the DMS approach is more robust to model misspecification. Therefore, the question of superiority of one of the two approaches remains an empirical one.[10]

The *direct* h-step multi period forecasting model is obtained from regressing y_{t+h} on p lags of y_t. As before, p is determined by the SIC with a maximum number of twelve lags. Based on the resulting models, we forecast the stock price and the dividend series for all horizons. We then substitute these values into equation (2.4) in order to construct a forecast for the ex post rational price.

The right panel of Table 2.4 shows the results from comparing the forecasts obtained by using the DMS approach and the actual stock price. We find no evidence that the DMS forecasts -neither in VAR settings nor in VECM settings- outperform the actual stock price as a predictor for the ex post rational price. However, as before, we observe that the performance of the actual stock price deteriorates with the length of the holding period of the underlying asset.

[10] This is reflected by several empirical studies that cannot determine whether IMS or DMS is preferable (see e.g. Stock and Watson, 2004; Marcellino et al., 2006).

Horizons

The above results raise the question about the mechanisms between the forecast horizon and the declining empirical Diebold-Mariano test values. A plausible explanation can be inferred by a close inspection of Figure 2.1 as well as the formula of the ex post rational price approximation, restated here for convenience:

$$P_t^{*h} = \sum_{i=0}^{h-1} \gamma^{i+1} D_{t+i} + \gamma^h P_{t+h}$$

Comparing the graphs of the price and the dividend series it becomes apparent that the latter follows a smoother pattern. Thus, we argue that our forecasting models account for the dynamics of the dividend series better than those of the price series resulting in lower forecast errors.

The equation above describes the ex post rational price (approximation) P_t^{*h} as a weighted average of future dividends and the terminal stock payment. Increasing the holding period of the underlying stock implies giving more weight to the dividend payments compared to the market price in the final period. Again, we argue that the dividend series is easier to forecast than the price series. Hence, giving more weight to this series in the construction of the forecast of the ex post rational price lowers the overall forecast error, which is one potential explanation for declining DM statistics with increasing holding periods.

2.6 Conclusion

In this chapter we review the two most prominent approaches to assessing variance bounds as implied by the present value model, namely the procedures by Shiller (1981) and Mankiw et al. (1985). A direct link between Mankiw et al. (1991) test and the well known forecast encompassing test is established. Moreover, we suggest the Diebold-Mariano test, which works under a minimal set of assumptions, as an alternative and more robust tool in this context.

In an application to the monthly S&P 500 Composite Price Index, we find no evidence that naive forecasts, such as the random walk or the random walk with drift, can outperform the market price as a predictor for the ex-post rational price. This result is in conflict with findings in earlier studies (e.g. Shiller, 1981; Mankiw et al., 1991).

In a next step, we show that the price and the dividend series are cointegrated. Based on the cointegration relation between these two series, we estimate a panoply of multivariate forecasting models for different transformations of the data. In the case of multi-step ahead forecasts we consider iterated as well as direct forecasting approaches. The resulting forecast for the price and the dividend series serve as components in the construction of forecasts for the ex post rational price.

For the case of a VECM in logarithms, these forecasts perform better than the actual price in terms of mean squared error loss at least for longer holding periods. We find clear rejections, and hence provide evidence against the view that stock prices are driven by fundamentals only. Moreover, we provide an explanation for the horizon specific improvements in the forecasting performance.

Chapter 3

Social Networks and the Distribution of Wealth

3.1 Introduction

Do social networks play a role in determining the distribution of wealth in an economy? In this paper, we study the role of one particular social network, namely church, on wealth. The literature so far has provided some evidence for the importance of religious attitudes for economic choices (see e.g. Iannaccone, 1998; Noussair et al., 2013). According to Keister (2003), two mechanisms directly link religion to wealth: First, religion and in particular church affiliation provide individuals with a network of information and (potentially) risk sharing. In that sense, one expects religiously active individuals to be advantaged with respect to financial decisions and hence to accumulate more wealth over time than the non-religious counterfactual. Imagine for example a person who attends church services on a regular basis and socialises with the church community. That person has access to a wider social network and may receive information for example about local real estate investment opportunities that he or she would otherwise have missed out on. Second, religious attitudes are to some extend related to risk attitudes in general and risk aversion in particular. Now, more risk averse individuals may on the one hand be more likely to save and on the other hand be more likely to invest in less risky assets than risk neutral or risk loving individuals. If less risky assets lead to lower returns, then the more risk averse person accumulates on average less wealth over time than the less risk averse counterfactual.

In this paper, we use data from a novel dataset, the German Panel on Household Finances (PHF), to estimate the effect of church as a social network on wealth and its distribution in Germany. In particular, we are able to differentiate between church affiliation and actual church attendance. In addition, the dataset provides measures of risk aversion and further personality traits which we exploit in order to disentangle the two mechanisms described above. Indeed we find that the effect of church affiliation on wealth is driven by what we call network effects. Moreover, we are able to disaggregate our results considering different types of assets and we find that the effect of religion on wealth is heterogeneous across asset types.

In estimating the impact of church affiliation on wealth we have to overcome the problem of self-selection. Religious individuals may differ in attitudes towards financial risks compared to less religious individuals. This in turn may affect savings and investment decisions, and hence wealth accumulation. In order to identify the effect of interest and to quantify the impact of religion on wealth we exploit the natural experiment of the German Reunification: Between 1949 and 1990, Germany was split into two parts, the Federal Republic of Germany (FRG) and the German Democratic Republic (GDR). The separation of the country was an exogenous shock to the population, as was the reunification of the country in 1990. Up to that point, political and cultural systems in both parts of the country were substantially different from each other, a fact that has been used in the literature to explain long-lasting differences in preferences and economic behaviour. In this paper, we argue that growing up in different political and cultural systems has contributed to differences in religious attitudes. We exploit that natural experiment (see Section 3.4 for more details) in order to quantify the impact of church as a social network on wealth in Germany.

We show that even 20 years after Germany's reunification there are differences with respect to wealth and religiosity between individuals who used to live in the FGR and the GDR in 1989, one year before the reunification. We discuss different measures of religion and church affiliation and show first stage results for each specification. Our first stage estimates indeed reveal differences in church affiliation between both sub-populations. Exploiting the natural experiment with a quasi-random allocation of the population to different political and cultural systems allows us to identify the impact of religion on wealth in Germany in an instrumental variable (IV) setting. Our results show that religion has a positive effect on wealth. We find evidence that this effect is associated with the social aspects of church attendance rather than risk aversion or other personal characteristics.

The IV model identifies the *average* effect of religion on wealth. Yet, we are particularly interested in the heterogenous impact of social networks over the distribution of wealth. Therefore, we estimate conditional and unconditional quantile regression models which allow us to quantify the impact of church affiliation on wealth at different *quantiles* of the wealth distribution, as opposed to the *mean* wealth. We find that individuals in the higher quantiles of the wealth distribution have higher benefits from the social networks described above than those individuals in the lower quantiles of the distribution. Again, we also demonstrate heterogeneous effects across asset types.

The remainder of this paper is structured as follows. We give an overview of the related literature in the next section. Section 3.3 and Section 3.4 describe the data and the institutional setting, respectively. The empirical approach is presented in Section 3.5, the respective results are discussed in Section 3.6. The last section concludes.

3.2 Previous Literature

Our paper relates to different strands of the literature. First, the paper contributes to the literature that analyses the impact of religion on wealth and the mechanisms as described above. Second, the paper contributes to the literature that studies the natural experiment of Germany's division and reunification. In the following, we discuss research papers in both strands of the literature.

The literature which links religion to economic outcomes dates back to Weber (*[1905]*, 1958) in Sociology and to Smith (*[1776]*, 1965) in Economics. Since that time, the relationship between religion and economic outcomes has been documented in a growing number of articles, as for example reviewed by Iannaccone (1998). Our study relates to Keister (2003) who in her analysis links religion to wealth. The author finds two channels through which religion directly affects wealth ownership: First, religion shapes values and priorities which in turn affects wealth. Second, religion potentially provides important social contacts. In this study, we base our argumentation on these findings and break the impact of religion on wealth down into a risk aversion and an information and social network effect.

Noussair et al. (2013) present evidence of a relationship between religion and risk aversion from incentivised experimental measures. Based on the LISS panel, they find that risk aversion is positively correlated with religiosity, as measured by church membership. The authors find that the correlation is driven by social aspects rather than by religious beliefs. Bartke and Schwarze (2008) analyse the relationship between risk aversion on the one hand and religion and nationality on the other hand using data from the German Socio Economic Panel (GSOEP). The authors find that individuals with a religious affiliation appear to be significantly less risk tolerant than atheists. Using the same dataset, Köbrich Leon and Pfeifer (2013) provide evidence that religious individuals are significantly more risk averse than non-religious individuals. Moreover, their findings document an increasing relationship between risk aversion and religious activities: Risk aversion increases with religious activities, and the higher the risk aversion the less individuals are willing to invest in risky assets.

Previous research has been done in order to identify the relationship between risk aversion and economic outcomes exploiting the experiment variation from the German Reunification. The majority of these studies uses data from the GSOEP. Fuchs-Schündeln and Schündeln (2005) study household savings in response to the German Reunification. They base their analysis on the assumption that income risk was not correlated with risk aversion in the communist system of the former GDR. In that context, the reunification induced an exogenous reassignment of income risk to different occupational groups for individuals in East Germany. Exploiting that institutional setting, the authors show that risk aversion influences occupational choice and that individuals act in line with the predictions from the precautionary savings theory. Based on the same dataset and using data from 1992 to 2000, Fuchs-Schündeln (2008) finds that individuals who lived in the former GDR still show higher saving rates than those who lived in the FRG before 1990. The author demonstrates that the difference is more pronounced for older birth cohorts and decreases over time for every cohort. She shows furthermore that a standard life-cycle model is able to predict these stylised facts. Again based on the GSOEP, Alesina and Fuchs-Schündeln (2007) present empirical evidence of long-lasting differences in preferences resulting from the division of Germany. In particular, the authors show that individuals from East Germany are more likely to favour state interventions than those from West Germany. They find that the difference in preferences is largely due to the communist political system in the former GDR. Hence, one of their main results is that the politico-economic system which individuals experience in their direct environment profoundly shapes their respective preferences. In our analysis we use a similar argument, however studying religious attitudes rather than political preferences.

For the following analysis we follow the arguments by Keister (2003). We proxy the two channels through which religion directly affects wealth by a measure of risk aversion on the one hand[1] and the frequency of attending religious events on the other hand.

3.3 Data

We use data from the German Panel on Household Finances (PHF) for our empirical analysis. The dataset is provided and managed by the Deutsche Bundesbank. The PHF is a novel dataset and an integral part of the Household Finance and Consumption Survey (HFCS) which consists of surveys on households' finances and consumption in every country of the Euro area. It is designed as a panel dataset of which the first wave was conducted in 2010 and 2011. Up to date, only one wave of the survey is available and hence only cross-sectional variation can be exploited at this point.[2]

The PHF comprises 3,565 randomly selected households. Household heads as well as other household members are surveyed. In this paper, we focus our analysis on data related to household heads. Individuals respond to survey questionnaires which consist of a number of

[1] Our measure of risk aversion may be a somewhat narrower concept than the idea introduced in Keister (2003). In our approach we follow standard measures of risk aversion in the Economics literature.

[2] The second wave started in spring 2014 and was not yet available at the time of this analysis.

modules. Modules are linked to topics such as demographics, consumption, real assets and their financing, other liabilities and credit constraints, private businesses and financial assets, intergenerational transfers and gifts, employment, pension and insurance policies as well as income.

The PHF is designed to oversample wealthy households by enhancing the selection probability of these units. That is done by using a number of different stratifications on the basis of micro-geographic information. The procedure allows for a meaningful econometric analysis in the wealthy subpopulation, as it overcomes the problem of limited representation of wealthy households in a fully random sampling design. A complex weighting scheme is then used in order to correct for the bias that is introduced by the oversampling of wealthy households (design weights), for the bias that is introduced by non-responding units (non-response weights) and for a bias due to non-representativeness (calibration weights). Moreover, the PHF provides a solution to the non-response problem by a multiple imputations approach: Missing values due to non-response are simulated by repeatedly drawing from a sample of estimates from the conditional distribution of the data. That procedure is carried out five times to improve the efficiency of the estimates.

In this study, we are specifically interested in wealth and wealth components. The PHF provides two alternatives to measure overall wealth: On the one hand, the PHF includes a variable on self-reported overall wealth. On the other hand, we are able to construct a wealth variable by adding up the self-reported components of wealth. We prefer the latter for two reasons: First, we believe that adding up the components leads to a more accurate measure of wealth. Individuals are likely to be informed about the precise components of wealth which they hold, e.g. bank accounts and the value of real estate or stocks, but may introduce substantial measurement error when providing an ad hoc overall estimation of their wealth. Second, one contribution of this paper is to estimate heterogeneous effects of social networks and risk aversion with respect to disaggregated wealth. Hence, it is consistent with the empirical design to define aggregate wealth as the sum of the disaggregated wealth components. The PHF is very rich in information on wealth and its component, and hence we construct the aggregate wealth variable as follows:

$$\textit{net wealth} = \textit{safe assets} + \textit{risky assets} + \textit{non-financial assets} + \textit{real estate} - \textit{debt} \tag{3.1}$$

The categorisation of the different assets is documented in Table 3.1 of the Appendix. In this study we measure all quantities in levels. Taking logarithms would erase information about households with negative wealth.[3]

In this study, we are interested in the effect of social networks on wealth. As detailed before, we study religion or church affiliation as one example for a strong social network. In order to proxy religion for our empirical analysis, we construct a dummy variable indicating church membership based on church tax. All members of the (Christian) church in Germany pay a monthly church tax which is linked to the income tax. That means that we can identify church membership by church tax payments and construct a dummy variable as follows: If an individual pays church tax the dummy takes the value 1, and 0 otherwise. One might be worried that some religious groups do not have a separate tax system and hence our variable does not capture religious group membership for the whole population. However, the majority of religious groups in Germany are Christian and hence are subject to tax payments. In addition, and perhaps more convincing, we can show that our results hold when we use a different measure of religion which is independent of church taxes (see below for details).

As pointed out in the previous section, we follow the argumentation of Keister (2003) and proxy the channels through which church affiliation affects wealth by the intensity of church attendance on the one hand and risk aversion on the other hand. Church offers a platform for social contacts through a number of social and church related activities. We assume that the development and intensity of these networks depend on the frequency of attending these events. Therefore, church attendance serves as a proxy for networking effects. We construct a dummy variable that takes the value 1 if the individual attends church on a regular basis, and 0 if seldom or never. This measure is independent of being a church tax payer or not.

An essential ingredient of our analysis is the measurement not only of religion but also of risk aversion. The PHF contains a question asking the respondent to rate her risk aversion on a scale from 1 (risk averse) to 10 (risk loving). As the survey response may be reference point dependent and hence to some extent subject to measurement error, we construct two measures

[3]For the interested reader, we have also run regressions with logarithmised wealth and income; the results are qualitatively robust.

from that self-assessment: *'Risk aversion a'* takes the value 1 if the risk assessment is smaller or equal to 3, and 0 otherwise. *'Risk aversion b'* has a lower threshold; it takes the value 1 if risk aversion ranges from 0 to 5, and 0 otherwise. In other words, *'risk aversion a'* selects those individuals who are severely risk averse, while *'risk aversion b'* is a more moderate measure. We similarly construct variables to measure individuals' trust, patience, and life satisfaction. Using these measures as well as our measure of church as a social network, we are thus able to disentangle the two mechanisms described above.

3.4 The Institutional Background

In order to identify the impact of church affiliation on individuals' wealth we have to overcome the problem of self-selection: Unobserved factors that are correlated with religious attitudes *and* financial success (wealth) may confound the effect of interest. Such factors could be general attitudes towards risks or types of financial investments, for example. In this paper, we exploit the natural experiment of the German reunification in order to identify and quantify the impact of church affiliation on wealth. In the following, we describe the institutional environment and discuss the identifying assumptions of our empirical analysis.

Between 1949 and 1990 Germany was split into two parts, the Federal Republic of Germany (FRG) and the German Democratic Republic (GDR) as shown in figure 3.1. The separation of the country was an exogenous shock to the population following the events from WWII. The politico-economic systems in both parts of the country subsequently differed substantially one from each other, a fact that has been used in the literature to explain long-lasting differences in preferences and economic behaviours: While the FRG was shaped by the *German economic miracle*, the politico-economic system in the GDR has typically been described as a communist regime. Importantly, there was no preference-based selection into one part of the country or the other. The border between the two parts of Germany was strictly enforced and migration from East to West Germany practically impossible.[4] The reunification of the country in October 1990 again could not have been anticipated by the population before the fall of the Berlin Wall

[4]For more details and the exceptional case of refugees see e.g. Sheffer (2007) and Alesina and Fuchs-Schündeln (2007).

Figure 3.1: West and East German Territories

Notes: The figure shows a map of the West and East German territories. The light shaded area represents the West German territories before 1989/1990, the darker shaded area the East German territories. The darkest shaded area represents Berlin which was divided into two parts (West and East Berlin).

on November 9, 1989.

We argue that the different, exogenously assigned politico-economic environments shaped individuals religious attitudes which in turn has contributed to long-lasting differences in economic behaviour. Whilst East Germany had traditionally been a mostly Protestant area, the political environment of the GDR is typically described as hostile towards religion and church leading to a repression of both in the East German territories before the reunification.[5]

In our sample, we observe whether individuals lived in West or East Germany in 1989. Based on the observations above, that offers implicit information about the location of the individual in the years before 1989. We exploit that information by instrumenting church affiliation by the individual's location in 1989: At the first stage, we estimate the effect of living in West Germany before the reunification on individual's church affiliation today. We then rely on the exogenous variation introduced by the instrument to estimate the impact of religion on wealth

[5]For an overview see e.g. Sammet (2012).

at the second stage.

The instrumental variable approach heavily relies on two assumptions. First, we assume that individuals did not select into living in the FGR or GDR based on their current wealth. Second, we assume that living in the FGR or GDR in 1989 is exogenous with respect to the individual's current church affiliation. To reiterate, we exploit a natural experiment in which selection into one part of Germany or the other was not possible and the separation of the country was an exogenous shock to the population. We thus argue that the assumptions for a suitable instrument hold and that our empirical strategy identifies the causal impact of church affiliation on wealth.

3.5 Empirical Approach

In the following, we describe our empirical approach for this study. For ease of exposition, we start with the standard treatment effect notation. Let $Y_i(1)$ and $Y_i(0)$ denote the potential outcomes for the head of household i, where $i = 1, ..., N$. As explained before, the outcome variable Y_i stands for household wealth and its respective components. Consequently, $Y_i(1)$ is the outcome for the head of the household i if assigned to the treatment group, and $Y_i(0)$ if assigned to the control group. We denote the realisation of the treatment assignment by a binary variable $R_i \in \{0, 1\}$. To reiterate, treatment here is church affiliation. We additionally test for potentially confounding treatments such as risk aversion and other personal characteristics. The treatment effect on household i is then given by $\delta_i = Y_i(1) - Y_i(0)$. Only one of the two potential outcomes is observed in practice, leading to the well known problem of unobserved counterfactuals. In our case, for example, one individual can only be religious or not at one fixed point in time.

As explained in the above, there is a potential risk that the estimated effect is contaminated by a self-selection bias. We tackle this issue using an instrumental variable approach. Let $Z_i \in \{0, 1\}$ denote our binary instrument where the dummy takes the value 1 for individuals living in West Germany in 1989, and 0 otherwise. As is well known from the literature, this model identifies the local average treatment effect (LATE) which can be formulated as $\beta^{IV} =$

$E[Y_i(1) - Y_i(0)|complier]$. The validity of that approach crucially depends on the exclusion restriction and the random assignment to treatment assumption as discussed in section 3.4. If the assumptions hold, then our baseline specification, written as follows, identifies the LATE of church affiliation on wealth:

$$Y_i \quad = \quad \beta_0 + \beta_1 R_i(Z_i) + \beta_2' X_i + \varepsilon_i \tag{3.2}$$

where Y_i denotes wealth (or its respective component), R_i is a dummy for church affiliation (or one of the alternative treatments), Z_i is the binary instrument indicating location in 1989, and X_i is a vector of individual and household characteristics.

For the purpose of this study, we use the LATE model as a starting point only. We are particularly interested in whether the treatment effect is heterogenous along the distribution of the outcome variable. In other words, is the impact of church as a social network on wealth different at different points of the wealth distribution? Quantile treatment effect models (QTEs) allow us to answer that question as discussed in the following. In particular, the QTE estimators (other than the LATE estimator) are robust to outliers in the data - a feature that is highly desirable when working with wealth data as in this study.

A suitable method to estimate conditional quantile treatment effects was introduced by Koenker and Bassett (1978). Subsequently, the econometric literature has suggested estimators which are valid in the case of endogenous regressors (e.g. Abadie et al., 2002; Chernozhukov and Hansen, 2005). These conditional QTEs are very useful, yet one may furthermore be interested in unconditional QTEs. Firpo (2007) introduced estimators applicable in the case of an exogenous treatment variable, whereas Frölich and Melly (2010) and Frölich and Melly (2013) extend that to the case of endogenous treatment variables. It is important to stress at that point that conditional or unconditional QTEs lead to different interpretations. The conditional QTE identifies the treatment effect on the distribution of the dependent variable *conditional* on the covariates. Hence, adding further covariates may shift the distributional location of an individual and thus may change the limit of the estimated conditional QTEs. In contrast to the standard average treatment effect models this may even be true when the covariates are independent of the treatment. Compared to that, the unconditional QTE describes the difference in the quantiles for different values of the variable of interest, irrespective of the covariates. However, the addition of covariates improves the efficiency of the estimate and helps the plausibility of the identifying assumptions. Following Abadie et al. (2002) and Frölich and Melly (2010, 2013), we estimate both conditional as well as unconditional QTEs. As especially the latter method is relatively new in the literature and as applications up to date are scarce, we briefly summarise the key assumptions and features in what follows.

Let us denote the conditional QTEs by $\delta^\tau = Q^\tau_{Y(1)|X} - Q^\tau_{Y(0)|X}$ and the unconditional QTEs by $\Delta^\tau = Q^\tau_{Y(1)} - Q^\tau_{Y(0)}$ where Q^τ_Y refers to the τ-Quantile of the distribution of Y. Same as in the case of the LATE model, the QTE model includes an endogenous regressor which we instrument for, using the binary instrument from above. That implies that the model still identifies the treatment effect for the compliers only.

In terms of the conditional QTE model, Abadie et al. (2002) suggest a linear potential outcomes model which can be written as:

$$Y_i \;=\; \delta^\tau R_i + (\beta^\tau)' X_i + \varepsilon_i \qquad \text{and} \qquad Q^\tau_{\varepsilon_i} = 0.$$

In order to identify the quantile treatment effect in our model, the standard IV assumptions must hold. In particular, for almost all values of X it must hold that:

Ass.1 (independence):	$(Y(0), Y(1), R(0), R(1)) \perp\!\!\!\perp Z|X$
Ass.2 (non-trivial assignment):	$0 < Pr(Z = 1|X) < 1$
Ass.3 (relevance):	$E(R(1)|X) \neq E(R(0)|X)$
Ass.4 (monotonicity):	$Pr(R(1) \geq R(0)|X) = 1$

The conditional QTE estimator can be derived as the solution of the following weighted minimization problem with weights $W_i^{AAI} = 1 - \frac{R_i(1-Z_i)}{1-Pr(Z=1|X_i)} - \frac{(1-R_i)Z_i}{Pr(Z=1|X_i)}$ and check function $\rho_\tau(u) = \tau u^+ + (1-\tau)u^-$ where $u^+ = I(u \geq 0) \cdot |u|$ and $u^- = I(u < 0) \cdot |u|$:

$$(\widehat{\beta}^\tau, \widehat{\delta}^\tau) \;=\; \underset{\beta^\tau, \delta^\tau}{\arg min} \sum_i W_i^{AAI} \times \rho_\tau(Y_i - X_i\beta^\tau - R_i\delta^\tau) \qquad (3.3)$$

Frölich and Melly (2013) provide an unconditional QTE estimator which - contrary to the conditional case - can be estimated at the $\sqrt{n}$ rate without any further parametric assumptions. For the unconditional case, the relevance assumption can be slightly relaxed and has to hold unconditionally only. The new minimization problem with weights $W_i^{FM} = \frac{Z_i - Pr(Z=1|X_i)}{Pr(Z=1|X_i)[1-Pr(Z=1|X_i)]}(2R_i - 1)$ is given by the following:

$$(\widehat{\alpha}^{\tau}, \widehat{\Delta}^{\tau}) = \underset{\alpha^{\tau},\Delta^{\tau}}{\arg min} \sum_i W_i^{FM} \times \rho_{\tau}(Y_i - \alpha^{\tau} - R_i\Delta^{\tau}) \tag{3.4}$$

We estimate the conditional and unconditional QTE models as outlined above for overall wealth as well as for the wealth components, i.e. different asset types. Yet, one has to be cautious with comparing the results with respect to the different outcome variables: One problem that arises with the QTE approach is that the distribution of households may change with respect to the different asset types. For example, a household that lies in the top percentile of the overall wealth distribution may hold a larger share in financial than in non-financial assets. Hence, the same household finds itself in different percentiles of the distribution of wealth in financial and non-financial assets, respectively. Whilst it is still reasonable to assume that positions in the respective distribution do not differ substantially from the position in the overall wealth distribution, we nevertheless construct a robustness check for that scenario. In particular, we estimate a standard IV model interacting the treatment variable with a dummy indicating the household's quantile in the overall wealth distribution:

$$Y_i = \gamma_0 + \gamma_1^{\tau}(R_i \times q_i^{\tau}) + \gamma_2' X_i + \varepsilon_i \tag{3.5}$$

where q_i^{τ} corresponds to a dummy variable that takes the value 1 if household i belongs to the τ-th quantile in the overall wealth distribution. The advantage of that approach is that it allows us to hold constant the position of the household with respect to the overall wealth distribution and hence improves the comparability across asset types. Yet, the disadvantage of that specification is that multiplying the two dummy variables (treatment and quantile dummy) may lead to a very small number of treated observations per subgroup. That may contribute to high standard errors and potentially to weak instrument issues. For that reason we do not rely on the results of that approach too heavily but suggest to interpret them as an additional robustness check.

3.6 Empirical Evidence

In the following, we present the empirical evidence. We first provide descriptive evidence before discussing the results from estimating the econometric models outlined above.

3.6.1 Descriptive Evidence

Figures 3.2 to 3.7 show the empirical distributions of total wealth as well as its components, each separately for the West and the East German sample. The West German sample comprises those individuals who were located in the FRG in 1989, the East German sample those who were located in the GDR, respectively. The assignment to West and East Germany depends on the location in 1989 only, and not on the current location.

Figure 3.2 shows the empirical distribution of total wealth in our sample, measured according to equation (3.1). Compared to the East German wealth distribution, the West German wealth distribution is shifted to the right. Most importantly, the figure illustrates that 20 years after the Reunification there are still differences in wealth between individuals who had been located in the FRG and the GDR, respectively. Figures 3.3 to 3.7 show the empirical distributions of different wealth components, i.e. asset types. For all the positive wealth components considered here (safe assets, risky assets, non-financial assets, real estate) the distribution for the West German sample is shifted to the right compared to the East German distribution. The most striking differences can be seen for the distribution of real estate wealth, with a substantially higher level being held by West Germans compared to East Germans. To reiterate at this point, West and East German here refers to individuals who lived in West and East Germany, respectively, in 1989 independent of their current location. In terms of debt, the peak of the distribution for the West German sample is again shifted to the right compared to the East German sample. In this paper, we argue that these observations can partially be explained by

Figure 3.2: Total Wealth

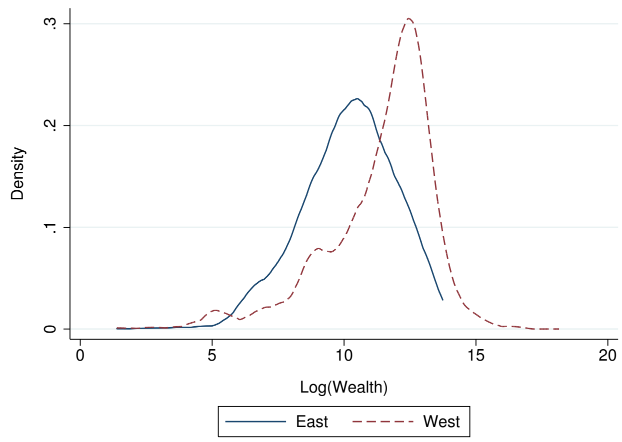


Figure 3.3: Safe Assets

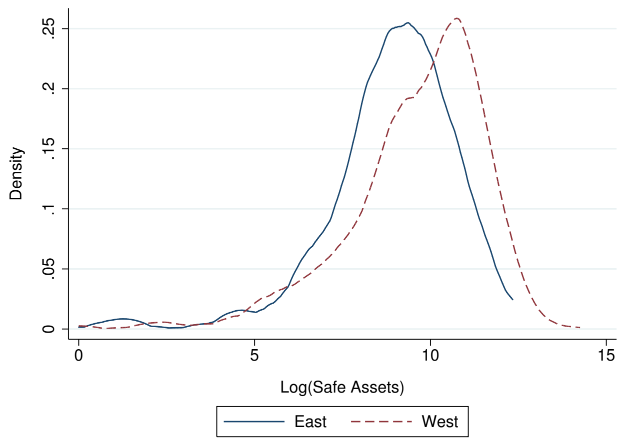


Figure 3.4: Risky Assets

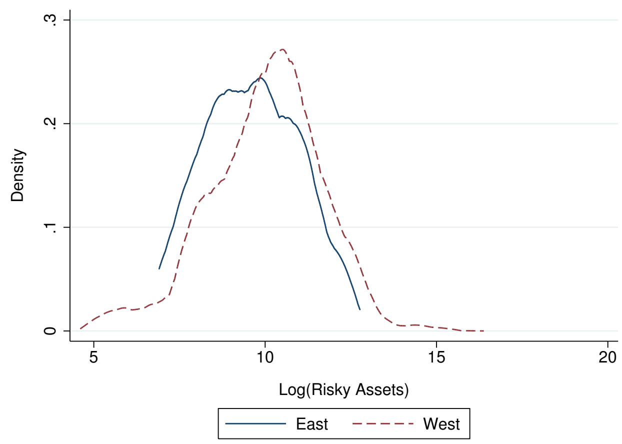


Figure 3.5: Real Estate

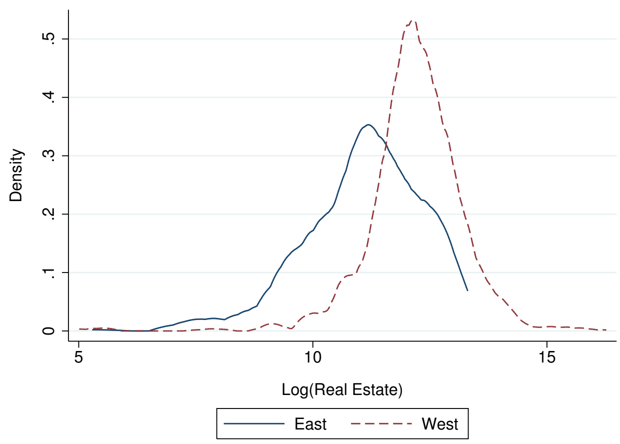


Figure 3.6: Non-Financial Assets

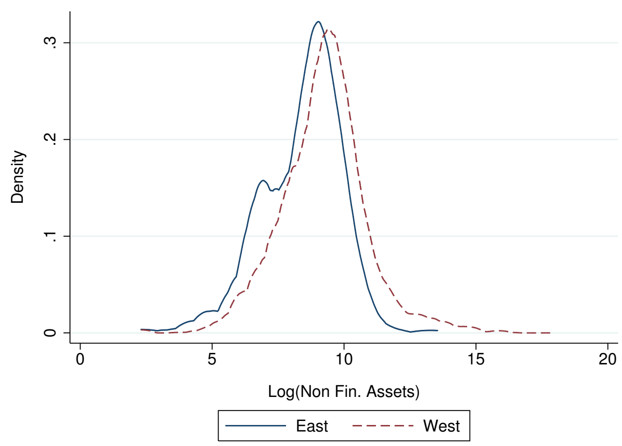
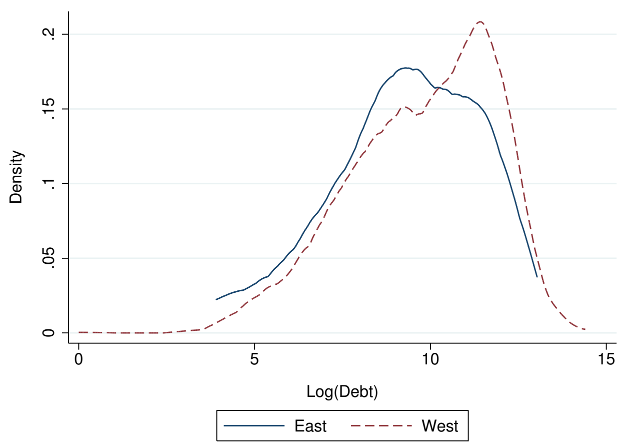


Figure 3.7: Debt



Notes: The solid line represents the empirical distribution of the overall wealth and the respective components for the East German sample, the dashed line for the West German sample. *Source:* 2010 PHF and own calculations.

underlying differences in social networking and the mechhanisms described above.[6]

Table 3.2 (Appendix) shows the (weighted) descriptive statistics for our sample. In particular, the table displays sample means as well as the standard deviations in parentheses for each variable. Column (1) refers to the entire sample, columns (2) and (3) to the West and East German sample, respectively. The remaining columns show descriptive statistics separately for the treatment and control groups with respect to the two treatments church affiliation (our proxy for religion) and church attendance (our proxy for church related network effects).[7]

In line with our argumentation above, we indeed find that the share of church affiliated individuals is higher in the West than in the East German sample (59% versus 25%). The same is true for church attendance (54% versus 23%). In terms of risk aversion, we find that for both alternative measures the share of risk averse households in the West and East German sample is almost identical (60% versus 56% for *risk aversion a* and 83% versus 82% for *risk aversion b*). In terms of risk aversion, the same pattern emerges when we compare the share of risk averse individuals across those who are church affiliated and those who are not (56% versus 57%) as well as those who regularly attend church and those who do not (84% versus 80%).

Our instrument is a binary variable that indicates whether the individual was located in the FRG or in the GDR in 1989. For the entire sample, we find that 20% of the individuals are assigned to the East German sample. The share of individuals who are assigned to the East German sample is higher among those who are classified as not affiliated to church compared to those affiliated to church. That is perfectly in line with the observations above, and a similar conclusion emerges from comparing individuals who regularly attend church to those who do not.

All wealth and income variables are measured in 10,000 Euros. Mean overall wealth in the entire sample amounts to 244,900 Euros. Confirming the graphical evidence, overall wealth is on average substantially higher in the West compared to the East sample. Again, differences in

[6]In this study, we explicitly study the effect of social networks approximated by church affiliation on wealth. We investigate alternative mechanisms that could potentially confound our results, but we do not investigate all potential mechanisms that contribute to differences in wealth distributions. Indeed, that will be an interesting strand for future research.

[7]For reasons that become obvious in what follows, we do not show any additional columns for the treatment risk aversion here.

wealth held in non-financial assets, safe assets and real estate seem to be particularly striking. Both samples, West and East, are balanced in terms of observable characteristics. Differences in the covariates' sample means are within one standard deviation, and not significant.

3.6.2 First Stage, Reduced Form and IV Estimation

In the following, we present the results from the instrumental variable estimation according to equation (3.2). Column 1 of Table 3.3 shows the first stage estimation results with respect to our main variable of interest, church affiliation. Living in the FRG compared to the GDR before the Reunification increases the probability of being affiliated to church by about 34%. The effect is statistically significant at the 1% level. This is in line with what we argued before: The different socio-economic conditions in the FRG and the GDR shaped cultural landscapes and had a significant impact on attitudes towards religion and church affiliation.

As discussed in the above, we study the mechanisms through which religion - or church affiliation in particular - affect wealth: Risk aversion on the one hand and a network and information mechanism on the other hand. Columns 2 to 4 of Table 3.3 refer to the first stage estimations of these specifications based on the variables *risk aversion a*, *risk aversion b* and *church attendance*.[8] With respect to risk aversion, neither of the specifications shown in columns 2 and 3 yields results that are statistically different from zero.[9] In contrast to that, column 4 shows a strong first stage effect for church attendance - our proxy for the network and information mechanism. Living in the FGR compared to the GDR before the Reunification increases the probability of regular church attendance by about 32% - which is in terms of magnitude almost identical to the effect for our main variable church affiliation as shown in column 1.

Up to this point, our sample includes individuals who at the time of the Reunification were at least 25 years old. This is based on the notion that individuals should have developed

[8]For more details on the variables, see Section 3.3.

[9]Again, we would like to point out that risk aversion might be a somewhat narrower concept than what Keister (2003) has in mind: She claims that religion shapes values and priorities which in turn affect wealth. Therefore, we also checked other personality related variables like patience, trust and life satisfaction as a robustness check. However, none of the variables resulted in a strong first stage. The results can be found in Table 3.4.

their general religious attitudes at that age and are not affected by the Reunification in that sense. Yet, one may be concerned that individuals may develop religious attitudes up to a younger or older age than the age of 25. In that case our age cutoff would be too high or too low, respectively. We can rule that concern out by re-estimating the first stage equation using different age cutoffs. Columns 5 to 8 of Table 3.3 show the first stage results when we restrict the sample to individuals who at the time of the Reunification are at least 20 or 30 years old, respectively. The results prove to be very robust to these specifications. Hence, we adhere to our original restriction; all following results refer to the sample with an age cutoff of 25 years.

Column 1 of Table 3.5 shows the estimation results from the reduced form, i.e. the long-term effect of being located in the FGR compared to the GDR in 1989 on wealth in 2010. Confirming the graphical evidence, we find that even 20 years after the Reunification there are substantial wealth differences: Living in the FGR in 1989 on average increases wealth by 149,720 Euros compared to living in the GDR. The effect is statistically highly significant.[10]

Columns 2 to 9 of Table 3.5 show the results of the OLS and IV (2SLS), i.e. the impact of church affiliation on wealth as well as the mechanisms risk aversion and church attendance. The OLS estimation with respect to our main variable of interest - church affiliation - suggests a positive correlation between church affiliation and overall wealth. When we instrument church affiliation by the location in 1989, we find that being affiliated to church is associated with a higher overall wealth of about 444,000 Euros. The estimated effect is statistically significant at the 1% level. The OLS estimates are downwards biased compared to the IV estimates. As outlined in the previous sections, the direction of the bias can be rationalised by an underlying self-selection process: Individuals may self-select in and out of church based for example on latent moral attitudes that correlate with their investment behaviour. When we do not take that notion into account, we underestimate the impact of religion on wealth and hence we find a

[10]One may be concerned that the results are partly driven by the current location, based on the observation that former West Germany performs economically better than former East Germany. Unfortunately, the data does not allow us to control for the exact location of the individuals, but only for the region (North, West, South and East). Yet, the current location as proxied by these regional indicators is very highly correlated with the location in 1989 ($\sim 80\%$). In order to avoid inaccurate estimates due to multicollinearity issues, we thus do not control for the current location but include an extensive set of individual control variables in the regressions. We argue that these variables reflect the individual's economic environment, but - given the data used for this study - we are not able to evaluate the impact of the current location beyond that point.

downward bias of the OLS compared to the IV estimates. Further, in terms of the magnitude of the bias, one should keep in mind that we identify a local average treatment effect - which alters the interpretation of the IV coefficient compared to the interpretation of the OLS coefficient.[11]

We are further interested in understanding the mechanisms which drive the effect of church affiliation on wealth. Again, we separately estimate the OLS and the IV (2SLS) models for risk aversion and church attendance. We do not find a significant impact of risk aversion on overall wealth, yet that might be due to a weak instrument problem for that specification. However, studying the impact of regular church attendance on wealth, we find very similar results - both in terms of magnitude and statistical significance - to the baseline specification. The results suggest that the effect of church affiliation on overall wealth is indeed driven by the network and information mechanism, i.e. the social network, and in our case cannot be attributed to the correlation between church affiliation and risk aversion.

In the specifications which we have described in the above, we have estimated the effect of church affiliation on *level* wealth in a linear model. One may be interested in whether the results are sensitive to the specific functional form of equation (3.2). Table 3.6 shows the results of our estimations when we change the functional form of the estimation equation to a log-level relationship: We estimate the impact of church affiliation on *log* wealth and hence obtain semi-elasticities of overall wealth with respect to church affiliation and church attendance, respectively.[12] Again, we find a statistically significant negative effect at the first stage. Estimating the semi-elasticities at the second stage of our IV (2SLS) model, we find that being affiliated to church is associated with 280% higher overall wealth, and attending church with 250% higher overall wealth.

3.6.3 Conditional and Unconditional QTE Estimation

The instrumental variable model identifies the local average treatment effect: The average effect of church affiliation on overall wealth. However, the impact of church affiliation could

[11] See Section 3.5 for more details.

[12] By computing the logarithms of our wealth variable we exclude negative or zero wealth observations from our sample.

reasonably be very different for individuals at distinct points of the wealth distribution. In this study, we argue that church activities provide an individual with social contacts and offer a network for information exchange and cooperations. Yet, that information might not be of exactly the same value for everyone: For example, wealthier individuals might be more able to exploit these networks for investments than less wealthy individuals. In that sense, we expect the relatively high magnitudes of the IV model to be driven by the higher quantiles of the wealth distribution.

These effects might again differ by type of investment. Someone who regularly attends church services and socialises within the church community for example potentially receives information about local real estate investment opportunities that he would have missed out on otherwise, but not on riskier investment opportunities. Again, this could be a potential driver for the higher magnitudes of the results from the IV model. In order to quantify the heterogeneous effects with respect to the wealth distribution and the different wealth components, we estimate quantile treatment effect models for the different asset types and report the results in what follows.

The results in this section correspond to equations (3.3) and (3.4). Having established in the above that the wealth enhancing impact of church affiliation is almost entirely driven by the church related network effects and not by risk aversion, we restrict our analysis to our main variable of interest (church affiliation). We estimate the conditional and unconditional quantile treatment effects on overall wealth and on particular asset types. In particular, we test the following one-sided hypotheses, each evaluated at the respective point in the wealth distribution:

(H1): Church affiliation is associated to higher levels in wealth.

(H2): Church affiliated individuals hold more safe assets.

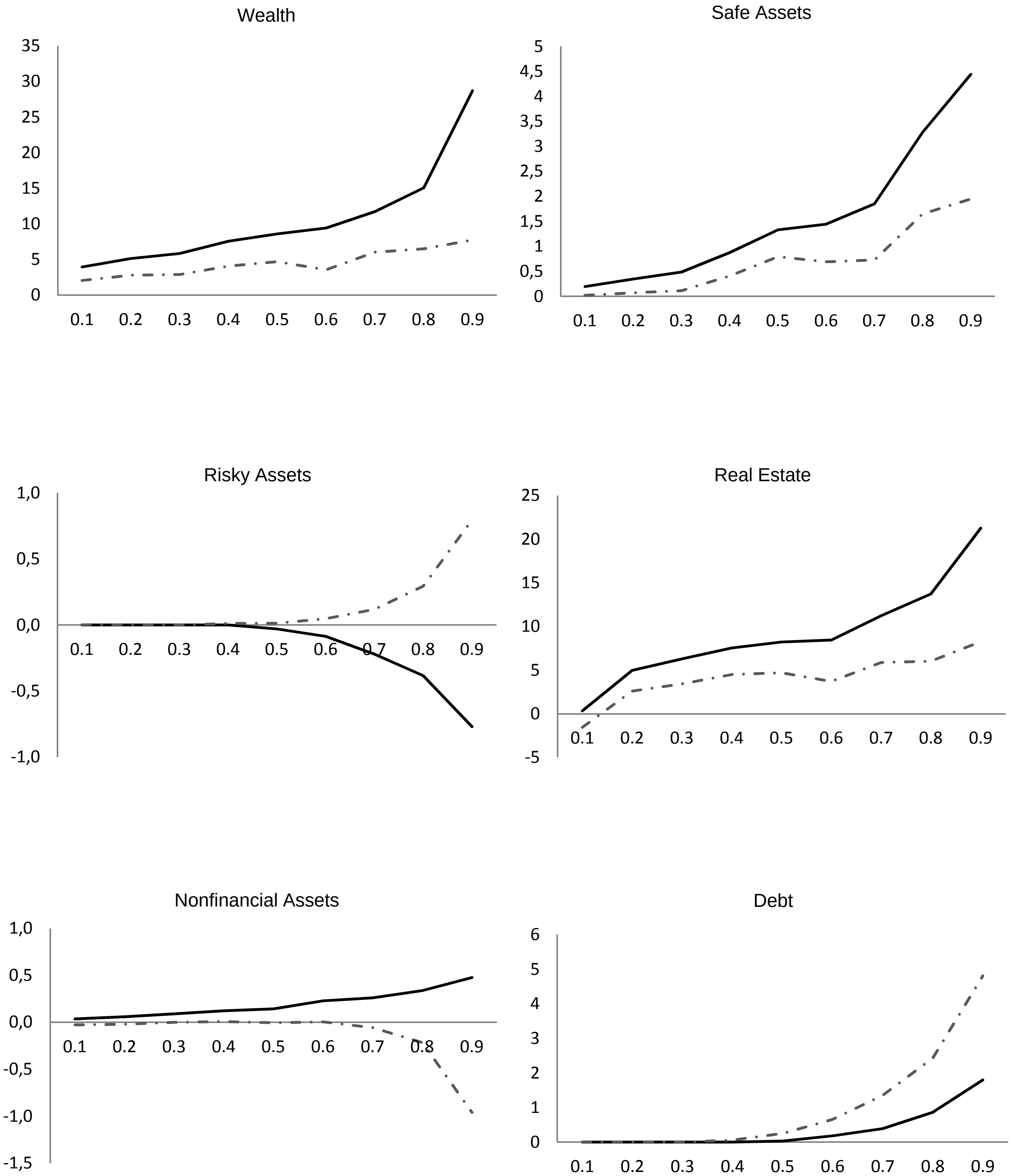
(H3): Church affiliated individuals hold less risky assets.

(H4): Church affiliated individuals own more real estates.

(H5): Church affiliation is associated to higher levels in non-financial assets.

(H6): Church affiliated individuals hold less debt.

Figure 3.8: Conditional Quantile Treatment Effects



Notes: The solid line represents the estimated coefficient for each quantile (specified on the x-axis), the dashed line the 95%-confidence interval for the respective (one-sided) hypothesis based on 1000 bootstrap replications. *Source:* 2010 PHF and own calculations.

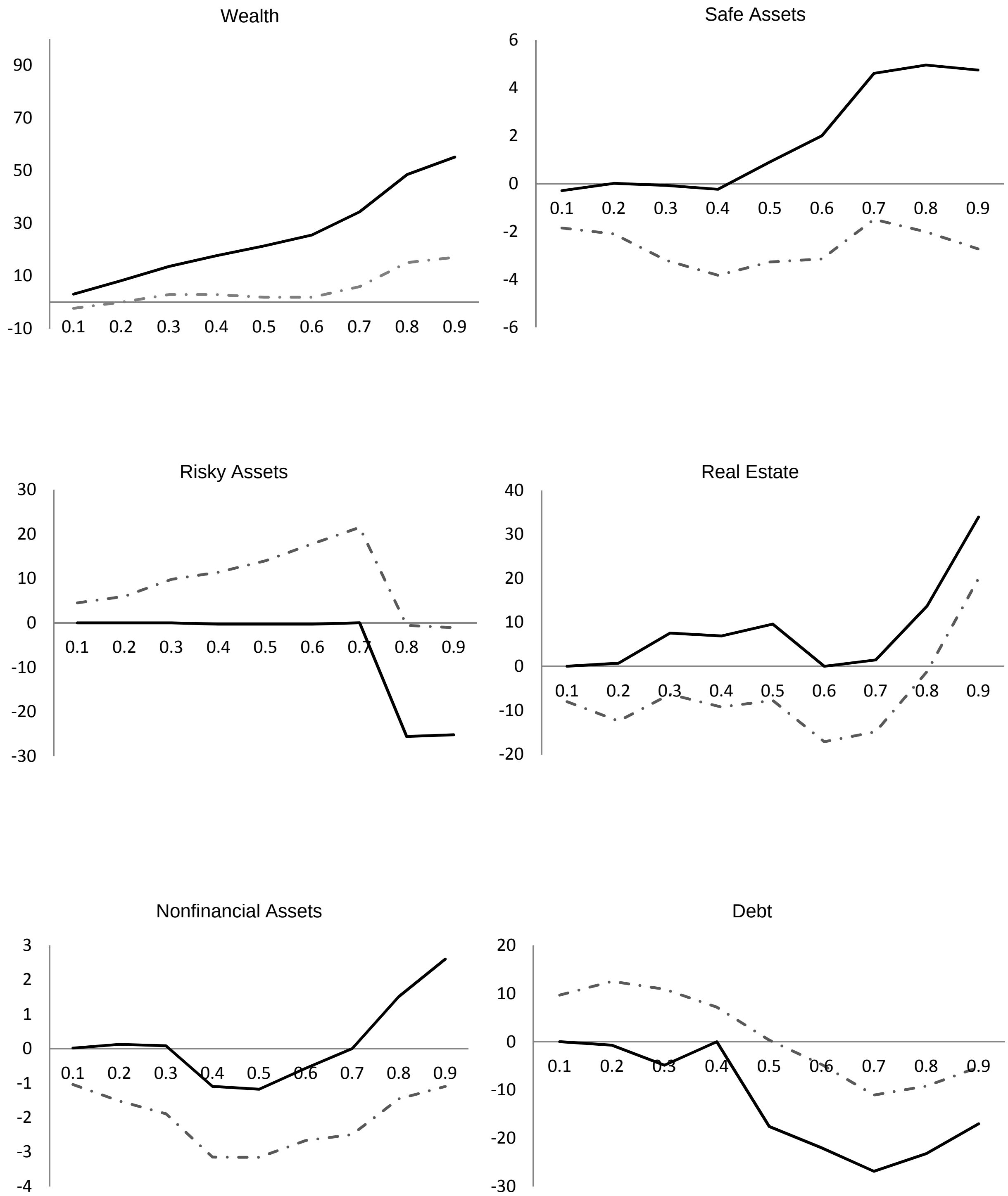
Figure 3.8 presents the results of the conditional quantile treatment effect estimation. For overall wealth, the impact of church affiliation is indeed significantly larger than zero at every point of the wealth distribution. In particular, the impact of church affiliation is substantially larger for the higher quantiles of the wealth distribution: The information network that is provided through church activities seems to be particularly useful for those individuals who are initially wealthy enough to use the information efficiently.

When we disaggregate our wealth measurement into its components and estimate the quantile treatment effect of church affiliation on wealth with respect to the particular asset type, we also find heterogeneous results. For safe assets, the impact of church affiliation is significantly larger than zero everywhere along the asset specific wealth distribution. Again, the effect is substantially - more than 5 times - larger in the higher quantiles of the distribution. For risky assets, the picture is quasi-mirrored: Church affiliated individuals seem to possess less risky assets than their counterfactual, holding constant their position in the risky asset wealth distribution. However, the coefficient is not significantly negative. That provides (albeit weak) support for the mechanisms outlined above, and in particular for the information network channel: If individuals are better informed about risks, they may restrain from it.[13] In terms of real estate assets, we find that church affiliated individuals hold more real estate assets than their counterfactuals. The effect again increases in magnitude for the higher quantiles compared to the lower quantiles of the real estate wealth distribution. The findings are in line with the explanations and examples given above: Wealthier individuals benefit more from the information sharing that the church related social network offers which seems to be particularly plausible in the case of real estate. No significant impact of church affiliation is found on wealth held in non-financial assets. In contrast, we find that we reject the hypothesis that church affiliated individuals hold less debt.

So far, we have presented the results of the conditional QTE estimation. As discussed in Section 3.5 we are further interested in the unconditional QTE of church affiliation on wealth and the different asset types, keeping in mind the differences in interpretation between both

[13] At this point, the correlation between risk aversion and church affiliation also plays a role of course. If church affiliated individuals are more risk averse on average, that could be mirrored in the fact that they invest less in risky assets.

Figure 3.9: Unconditional Quantile Treatment Effects



Notes: The solid line represents the estimated coefficient for each quantile (specified on the x-axis), the dashed line the 95%-confidence interval for the respective (one-sided) hypothesis based on 1000 bootstrap replications. *Source:* 2010 PHF and own calculations.

models.[14] Figure 3.9 shows the corresponding estimation results. Similar to the conditional QTE results, we find that church affiliation significantly increases overall wealth and that the coefficients are larger for the upper tail of the wealth distribution. When we estimate the unconditional QTE of church affiliation on wealth being held in the different types of assets, the results differ to a certain extent from the results found in the conditional QTE estimation. Contrary to the conditional QTE estimation, we do not find any significantly larger than zero effects of church affiliation on wealth held in safe assets. For risky assets, the impact of church affiliation is significantly negative for the higher deciles of the distribution. The pattern is similar to the conditional QTE estimation, yet the precision is higher for the coefficients corresponding to the higher quantiles. Similarly, we find that for real estate wealth the impact of church affiliation increases with the position in the distribution. However, in that case the unconditional QTE yields less precise estimates with coefficients significantly larger than zero for the highest quantiles only. We do not find significantly positive effects for non-financial assets, confirming the results from the conditional QTE estimation. In terms of debt, the unconditional QTE estimation results suggest that indeed church affiliation decreases debt holdings. Again, the coefficients are statistically significant for the higher quantiles of the debt distribution only.

3.6.4 IV Estimation by Wealth and Asset Type

As mentioned before, the quantile treatment effects are estimated for the asset specific wealth distribution. One problem that arises with that approach is that the position in the distribution may not be the same for each individual and each asset type. For example, a person who is in the 5th decile of the overall wealth distribution may be in another decile when looking at the distribution of safe assets only. In order to further test the heterogenous effects of church affiliation for the different asset types, we therefore estimate IV models holding constant the position in the overall wealth distribution. That approach in particular allows us to estimate

[14] The unconditional QTE integrates out the covariates upon which the interpretation of the conditional QTE relies. Our preferred specification and interpretation is the unconditional model. For further details, see Section 3.5.

the impact of church affiliation on asset wealth relative to the individual's overall wealth. The results correspond to equation (3.5) are shown in Table 3.7.

Column 1 of Table 3.7 shows the results for overall wealth. We find a significantly positive impact of church affiliation on wealth for the lowest wealth decile as well as for the five highest deciles. A similar pattern emerges for wealth in safe assets, non-financial assets and real estate, though magnitudes are much smaller in that case. We only find significantly positive coefficients for the highest two deciles in the risky asset wealth distribution, and we do not find any significantly non-zero effects for debt. The results presented in this section must be interpreted with caution, for the reasons explained in Section 3.5. Whilst we can not directly compare the results from this IV estimation to the results from the QTE estimation for methodological reasons, we can nevertheless observe some parallels. Most strikingly, the results presented here suggest that church affiliation indeed has an impact on wealth and that the impact is substantially stronger for wealthier individuals.

3.7 Conclusion

In this study, we analyse the impact of social networks on wealth using church as an example for a strong social network. In order to quantify the relationship of interest, we exploit the unique event of the German Reunification which offers a natural experiment with exogenous variation in religiosity and church affiliation. Using a novel data set, the first wave of the German Panel of Household Finances, we are not only able to study the impact of church affiliation on wealth overall, but furthermore we can disaggregate our measurement of wealth into its components (different types of assets) and gain more detailed insights.

In a first step, we estimate IV models and show that church affiliation indeed has a wealth enhancing impact. The mechanisms through which religion affects wealth are non-singular. We follow Keister (2003) who argues that religion impacts on wealth ownership by shaping values and priorities as well as by providing important social contacts. In order to disentangle the two mechanisms, we approximate the first by different measurements of risk aversion and the second by the frequency of church attendance. In our empirical analysis we find evidence for the importance of the social network mechanisms, yet we do not find any evidence in support of the risk aversion hypothesis.

In line with that observation, we argue that church activities widen individuals' social contacts and hence give rise to an exchange of information about financial and non-financial products and investment opportunities. Whilst that information is particularly interesting for those individuals who are easily able to invest, i.e. wealthier individuals, others might not be affected to the same extend. In order to estimate these heterogeneous effects of religion on wealth, we hence use an IV-QTE framework (conditional as well as unconditional QTE). Our results suggest that church affiliation has a positive impact on wealth with respect to the entire wealth distribution. Yet, the effects seem to be much stronger for the wealthier quantiles, confirming the notion above. Our results furthermore suggest that the differences - at least partially - stem from higher wealth in safe assets and real estate as well as from less debt.

Appendix: Tables

Table 3.1: Wealth and Wealth Components

Risky financial assets:	Shares (hd1010) Securities (dhd0750) Investment funds (dhd2420h) Certificates (dhd910) Bonds (dhd2520) Stocks (dhd2610) Other securities (dhd2310)
Safe financial assets:	Current account (dhd3200) Saving account (hd1210) Occupational pension plans-except Riester or Rürup plans and direct insurance policies (dpf1300f) Non-state-subsidised whole life insurance policies (dpf1300h) Other non-state-subsidised private pension plans (dpf1300i) Riester or Rürup bank saving plans (dpf1300r) Riester or Rürup savings and loan contracts (dpf1300s) Riester or Rürup mutual fund saving plans (dpf1300t) Classic Riester or Rürup pension plans (dpf1300u) Riester or Rürup occupational pension plans - excluding direct insurance policies (dpf1300w) Other Riester or Rürup plans (dpf1300z) Riester or Rürup plans (dpf1300y) Building savings agreement (dhd0610)
Non-financial assets:	Car (dhb0810) Other vehicle (hb4600) Other valuables likeWorks of art, Antiques, valuable jewellery etc. (hb4710) Business equity (hd701, hd702,hd703, hd801, hd802, hd803, dhd851, dhd852, dhd853) Value of additional assets in managed accounts (hd1620) Other assets not mentioned in other categories (hd1920) Debts owed to the household (hd1710)
Total real estate:	Main residence (dhb200a, dhb200b, hb0900, hb0500) Other properties (hb2701, hb2702, hb2703, hb2801, hb2801, hb2803, hb2900)
Debt:	Debt on credit line (hc0220) Debt on credit card (hc0320) Student loan, Bafög (dhc0720) Debt on mortgage (hb1701, hb1702, hb1703, hb3701, hb3702, hb3703) Mortgage: Money still owed on all other loans (hb2100, hb4100) Outstanding balance of loan (dhc6301, dhc6302, dhc6303) Outstanding bills (dhc1150) Total principal amount of all other loans (dhc0900)

Notes: Code identifiers correspond to `http://www.bundesbank.de/Redaktion/EN/Downloads/Bundesbank/Research_Centre/phf_codebook_en.pdf?__blob=publicationFile`

Table 3.2: Weighted Descriptive Statistics

		By West 89		By Church Affiliation		Church Attendance	
	Entire Sample	West	East	Affiliated	Not Affiliated	Regular	Seldom
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment:							
Church Affiliation	0.52	0.59	0.25			0.67	0.38
	(0.53)	(0.55)	(0.38)			(0.51)	(0.51)
Church Attendance	0.48	0.54	0.23	0.61	0.32		
	(0.53)	(0.55)	(0.37)	(0.55)	(0.47)		
Risk Aversion A	0.57	0.56	0.60	0.56	0.57	0.57	0.57
N=2437	(0.50)	(0.52)	(0.40)	(0.56)	(0.50)	(0.54)	(0.52)
Risk Aversion B	0.82	0.82	0.83	0.84	0.80	0.85	0.80
N=2437	(0.41)	(0.40)	(0.31)	(0.41)	(0.40)	(0.39)	(0.43)
Instrument:							
West 1989	0.20			0.09	0.31	0.09	0.29
	(0.42)			(0.33)	(0.46)	(0.32)	(0.48)
Outcome Variables:							
Wealth	24.49	28.66	7.46	34.20	13.91	29.30	20.12
	(76.24)	(87.40)	(11.57)	(103.49)	(34.77)	(87.04)	(65.11)
Safe Assets	4.03	4.50	2.08	5.20	2.75	4.34	3.74
	(7.75)	(8.66)	(3.23)	(8.35)	(6.85)	(7.18)	(8.18)
Risky Assets	1.61	1.86	0.61	1.93	1.26	1.76	1.48
	(11.73)	(13.49)	(2.44)	(11.40)	(11.86)	(10.34)	(12.80)
Non-financial Assets	4.29	5.16	0.77	6.53	1.86	5.24	3.43
	(38.55)	(44.56)	(2.98)	(54.88)	(10.73)	(46.90)	(29.43)
Real Estate	17.17	19.92	5.94	23.71	10.05	20.58	14.08
	(47.44)	(54.21)	(9.37)	(63.21)	(24.19)	(45.32)	(48.94)
Debt	2.62	2.78	1.95	3.17	2.01	2.62	2.61
	(8.20)	(9.09)	(4.39)	(9.72)	(6.42)	(8.71)	(7.72)
Covariates:							
Income	0.24	0.25	0.17	0.27	0.20	0.24	0.23
	(0.21)	(0.24)	(0.09)	(0.25)	(0.17)	(0.23)	(0.20)
Age	62.88	62.70	63.63	62.21	63.61	64.68	61.25
	(13.03)	(13.49)	(10.86)	(13.87)	(12.05)	(13.07)	(12.75)
Age squared	4102.80	4079.07	4199.69	4022.19	4190.54	4328.70	3897.64
	(1695.01)	(1747.14)	(1436.96)	(1806.67)	(1565.43)	(1707.15)	(1653.91)
Number of Kids	0.11	0.12	0.05	0.13	0.08	0.12	0.09
	(0.47)	(0.52)	(0.26)	(0.52)	(0.42)	(0.53)	(0.42)
Number of Adults	1.75	1.77	1.67	1.87	1.62	1.77	1.74
	(0.85)	(0.91)	(0.61)	(0.97)	(0.70)	(0.89)	(0.81)
Married	0.50	0.51	0.47	0.55	0.46	0.53	0.48
	(0.53)	(0.55)	(0.44)	(0.56)	(0.50)	(0.54)	(0.53)
Sex	0.52	0.53	0.48	0.54	0.50	0.47	0.57
	(0.53)	(0.55)	(0.44)	(0.56)	(0.50)	(0.54)	(0.52)
Urban	0.58	0.61	0.47	0.53	0.64	0.54	0.62
	(0.53)	(0.54)	(0.44)	(0.56)	(0.48)	(0.54)	(0.51)
Retired	0.51	0.50	0.53	0.45	0.57	0.58	0.44
	(0.53)	(0.55)	(0.44)	(0.56)	(0.50)	(0.53)	(0.52)
A-level	0.17	0.16	0.22	0.18	0.16	0.15	0.19
	(0.40)	(0.40)	(0.37)	(0.43)	(0.37)	(0.39)	(0.41)

Notes: The sample is drawn from the 2010 PHF. The observational units are household reference persons aged 45 and older. Standard deviations are in parentheses.

Table 3.3: First Stage Estimation Results

	i) First stage estimation				ii) First stage estimation: Robustness check			
Treatment (R_i):	Church affiliation	Risk aversion a	Risk aversion b	Church attendance	Church Affiliation age ≥ 40	Church Affiliation age ≥ 50	Church Attendance age ≥ 40	Church Attendance age ≥ 50
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
West89	0.337***	0.001	-0.046	0.325 ***	0.344***	0.298***	0.329***	0.313***
	(0.036)	(0.030)	(0.040)	(0.036)	(0.034)	(0.040)	(0.033)	(0.040)
Income	0.231**	-0.091	-0.108	0.004	0.240***	0.326***	-0.028	0.095
	(0.090)	(0.067)	(0.095)	(0.068)	(0.087)	(0.091)	(0.072)	(0.063)
Age	-0.009	0.000	0.010	0.029 *	-0.016	-0.025	0.015	0.040*
	(0.015)	(0.011)	(0.017)	(0.015)	(0.011)	(0.023)	(0.011)	(0.024)
Age squared	0.000	0.000	-0.000	-0.000	0.000*	0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Number of Kids	-0.007	-0.008	-0.001	0.076***	0.008	-0.025	0.115***	0.010*
	(0.031)	(0.028)	(0.032)	(0.029)	(0.026)	(0.058)	(0.023)	(0.053)
Number of Adults	0.076***	0.003	0.037*	0.011	0.062***	0.050**	-0.007	-0.018
	(0.021)	(0.018)	(0.022)	(0.021)	(0.020)	(0.025)	(0.020)	(0.026)
Married	-0.043	0.011	-0.057	0.054	-0.022	-0.039	0 .078**	0.046
	(0.036)	(0.030)	(0.038)	(0.037)	(0.034)	(0.040)	(0 .035)	(0.041)
Sex	0.032	-0.043*	-0.050*	-0.105 ***	0.027	0.024	-0.105***	-0.098***
	(0.029)	(0.023)	(0.030)	(0.029)	(0.027)	(0.033)	(0.028)	(0.033)
Urban	-0.139***	-0.008	0.034	-0.097 ***	-0.149***	-0.145***	-0.097***	-0.106***
	(0.029)	(0.024)	(0.031)	(0.029)	(0.027)	(0.032)	(0.027)	(0.032)
Retired	-0.154***	0.033	0.079*	0.049	-0.148***	-0.126***	0.065	0.037
	(0.043)	(0.039)	(0.046)	(0.045)	(0.042)	(0.047)	(0.043)	(0.049)
A-level	0.042	-0.064*	-0.141***	-0.007	0.041	0.028	-0.000	-0.025
	(0.037)	(0.035)	(0.040)	(0.040)	(0.034)	(0.039)	(0.038)	(0.044)
Constant	0.662	0.782**	0.187	-0.476	0.920***	1.269	-0.010	-0.851
	(0.478)	(0.387)	(0.541)	(0.482)	(0.353)	(0.779)	(0.356)	(0.840)
N	2448	2448	2448	2437	2715	2084	2704	2073
F-Stat	21.05	2.73	4.61	11.51	24.59	14.92	15.93	8.76
R-Squared	0.134	0.025	0.039	0.113	0.142	0.119	0.122	0.110

Notes: The left hand side variable is the respective treatment. The sample is drawn from the 2010 PHF. The observational units are household reference persons aged 45 and older if not stated otherwise. Robust standard errors are in parentheses.

Table 3.4: Additional First Stage Estimation Results

	First stage estimation					
Treatment (R_i):	Trust a	Trust b	Patience a	Patience b	Life Satisfaction a	Life Satisfaction b
	(1)	(2)	(3)	(4)	(5)	(6)
West89	0.050	0.066*	0.010	-0.029	0.017	0.182***
	(0.033)	(0.039)	(0.041)	(0.036)	(0.025)	(0.037)
Income	0.159***	0.060	0.157***	0.276***	0.184***	0.364***
	(0.044)	(0.097)	(0.048)	(0.053)	(0.054)	(0.074)
Age	-0.002	-0.010	0.020	0.026**	-0.006	-0.018
	(0.012)	(0.016)	(0.016)	(0.013)	(0.008)	(0.014)
Age squared	0.000	0.000	-0.000	-0.000*	0.000	0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Number of Kids	-0.018	-0.007	-0.014	-0.002	-0.013	-0.009
	(0.025)	(0.031)	(0.033)	(0.030)	(0.016)	(0.044)
Number of Adults	-0.017	-0.020	-0.028	-0.020	0.014	0.008
	(0.017)	(0.022)	(0.022)	(0.020)	(0.016)	(0.019)
Married	0.040	-0.033	0.103***	0.084**	0.015	0.087**
	(0.030)	(0.038)	(0.038)	(0.036)	(0.021)	(0.034)
Sex	-0.010	-0.002	-0.002	0.015	-0.023	-0.030
	(0.025)	(0.030)	(0.031)	(0.028)	(0.018)	(0.027)
Urban	0.020	-0.018	0.047	0.035	0.002	0.052**
	(0.025)	(0.031)	(0.031)	(0.027)	(0.018)	(0.027)
Retired	-0.011	-0.096**	-0.051	-0.026	-0.041	0.008
	(0.039)	(0.046)	(0.046)	(0.044)	(0.037)	(0.045)
A-level	0.041	0.124***	-0.066*	-0.037	0.006	0.119***
	(0.029)	(0.040)	(0.039)	(0.036)	(0.023)	(0.031)
Constant	0.741**	0.630	-0.120	-0.591	0.933***	0.817*
	(0.375)	(0.514)	(0.508)	(0.420)	(0.268)	(0.440)
N	2445	2445	2446	2446	2445	2445
F-Stat	2.80	2.32	2.48	5.04	2.64	12.22
R-Squared	0.018	0.021	0.019	0.029	0.038	0.105

Notes: The left hand side variable is the respective treatment. The sample is drawn from the 2010 PHF. The observational units are household reference persons aged 45 and older if not stated otherwise. Robust standard errors are in parentheses.

Table 3.5: Reduced Form, OLS and IV (2SLS) Results

Treatment (R_i):	(i) West89	(ii) Church affiliation		(iii) Risk aversion a		(iv) Risk aversion b		(v) Church attendance	
Model:	*Reduced form*	*OLS*	*2SLS*	*OLS*	*2SLS*	*OLS*	*2SLS*	*OLS*	*2SLS*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Treatment	14.972 ***	11.911***	44.427***	-10.143***	-322.835	-6.069	0.000	7.915**	45.611***
	(2.568)	(3.389)	(8.053)	(3.780)	(281.865)	(4.117)	(25.208)	(3.469)	(8.803)
Income	85.200***	88.883 ***	74.922***	91.483***	50.317	92.271***	90.709	92.050***	84.932***
	(30.249)	(27.368)	(29.275)	(27.293)	(41.436)	(27.612)	(91.018)	(28.036)	(31.879)
Age	2.849*	3.104**	3.236***	3.098**	6.029	3.065**	2.914	2.935**	1.583
	(1.553)	(1.488)	(1.556)	(1.483)	(5.836)	(1.518)	(20.341)	(1.475)	(1.646)
Age squared	-0.018	-0.020*	-0.022***	-0.020*	-0.039	-0.019*	-0.018	-0.019*	-0.010
	(0.011)	(0.011)	(0.011)	(0.011)	(0.043)	(0.011)	(0.126)	(0.011)	(0.012)
Number of Kids	-4.657 *	-4.668**	-4.361	-4.506**	-5.069	-4.418**	-4.025	-5.032**	-7.997***
	(2.518)	(2.192)	(2.814)	(2.159)	(10.486)	(2.140)	(8.249)	(2.297)	(3.043)
Number of Adults	0.805	-0.426	-2.576	.822	12.657	.456	1.046	.016	.437
	(3.151)	(3.098)	(3.269)	(2.950)	(12.216)	(2.996)	(11.846)	(3.032)	(3.268)
Married	9.574	10.140*	11.503	8.957	-8.885	9.599*	9.129	9.290	7.318
	(6.007)	(5.922)	(6.263)	(5.668)	(20.055)	(5.798)	(8.299)	(5.685)	(5.852)
Sex	1.071	1.082	-0.335	.885	-15.180	1.116	1.485	2.319	6.104
	(3.925)	(3.801)	(4.047)	(3.874)	(18.045)	(3.720)	(36.570)	(3.585)	(4.034)
Urban	-1.473	1.609	4.691	.586	9.436	.302	.156	.835	3.041
	(3.470)	(3.648)	(4.100)	(3.451)	(12.891)	(3.450)	(7.215)	(3.518)	(4.057)
Retired	-13.084**	-9.817	-6.234	-10.723*	12.532	-11.479*	-12.786	-12.200*	-15.398**
	(6.521)	(5.856)	(6.370)	(6.113)	(27.066)	(6.220)	(10.364)	(6.340)	(6.779)
A-level	8.603	5.437	6.726	3.968	-36.782	4.955	6.672	5.788	8.954
	(5.736)	(5.256)	(5.561)	(5.352)	(42.212)	(5.304)	(56.944)	(5.251)	(5.960)
Constant	-100.021**	-116.380**	-129.438	-107.374**	-39.538	-106.830**	-107.578	-109.422 **	-79.594
	(49.701)	(47.943)	(49.836)	(47.509)	(179.881)	(49.374)	(51.075)	(47.702)	(52.668)
N	2448	2527	2437	2527	2448	2527	2448	2515	2437
F-Stat	17.84	15.64	14.99	14.89	2.36	14.88	20.94	15.28	14.34
R-Squared	0.098	0.105	0.053	0.103	0.007	0.099	0.091	0.101	0.029

Notes: The left hand side variable is overall wealth. The sample is drawn from the 2010 PHF. The observational units are household reference persons aged 45 and older if not stated otherwise. Robust standard errors are in parentheses.

Table 3.6: Robustness Check - Functional Form (Log-Level)

Treatment (R_i):	Church affiliation	Church attendance	Church affiliation	Church attendance	Church affiliation	Church attendance
Model:	*i) First stage (2SLS)*		*ii) Second stage (2SLS)*		*iii) OLS*	
	(1)	(2)	(3)	(4)	(5)	(6)
West89	0.302***	0.326***				
	(0.039)	(0.039)				
Church affiliation			2.753***		0.595***	
			(0.534)		(0.114)	
Church attendance				2.522***		0.577***
				(0.483)		(0.112)
log(Income)	0.145***	0.004	1.200***	1.606***	1.709***	1.816***
	(0.030)	(0.029)	(0.225)	(0.215)	(0.186)	(0.189)
Age	-0.002	0.025	0.086	0.020	0.103	0.098
	(0.015)	(0.015)	(0.067)	(0.074)	(0.066)	(0.068)
Age squared	0.000	-0.000	-0.001	-0.000	-0.001	-0.001
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)
Number of Kids	-0.006	0.069**	-0.040	-0.226*	-0.132	-0.163*
	(0.032)	(0.029)	(0.135)	(0.116)	(0.099)	(0.095)
Number of Adults	0.059***	-0.006	-0.155	-0.002	-0.025	-0.017
	(0.021)	(0.022)	(0.107)	(0.097)	(0.091)	(0.089)
Married	-0.095***	0.055	0.594***	0.201	0.300**	0.216
	(0.037)	(0.039)	(0.181)	(0.156)	(0.137)	(0.137)
Sex	0.032	-0.106***	0.143	0.506***	0.236**	0.309***
	(0.030)	(0.031)	(0.124)	(0.133)	(0.109)	(0.110)
Urban	-0.148***	-0.098***	-0.171	-0.334**	-0.465**	-0.506***
	(0.030)	(0.031)	(0.148)	(0.139)	(0.116)	(0.115)
Retired	-0.134***	0 .021	0.074	-0.338*	-0.198***	-0.285
	(0.045)	(0.047)	(0.211)	(0.203)	(0.176)	(0.183)
A-level	-0.007	-0.012	0.445***	0.451***	0.298	0.293*
	(0.038)	(0.042)	(0.164)	(0.168)	(0.155)	(0.156)
Constant	0.833*	-0.339	-0.972	2.122	0.291	0.765
	(0.504)	(0.502)	(2.057)	(2.193)	(2.085)	(2.129)
N	2330	2320	2330	2320	2401	2390
F-Stat	20.29	9.27	31.10	31.34	40.88	38.21
R-Squared	0.141	0.105	0.119	0.159	0.369	0.371

Notes: The sample is drawn from the 2010 PHF. The observational units are household reference persons aged 45 and older if not stated otherwise. Robust standard errors are in parentheses.

Table 3.7: IV (2SLS) Estimation - Wealth Decile Interactions

Treatment (R_i):	Church affiliation					
Outcome:	Wealth	Safe assets	Risky assets	Non-financial assets	Real estate	Debt
Decile: 0.1	855.03**	105.26**	32.76	132.04*	588.58**	3.62
	(345.12)	(43.39)	(22.82)	(69.02)	(236.59)	(23.70)
Decile: 0.2	-7233.52	-890.53	-277.16	-1117.10	-4979.40	-30.66
	(44172.29)	(5435.54)	(1697.43)	(6828.79)	(30404.80)	(266.90)
Decile: 0.3	-1843.36	-226.94	-70.63	-284.68	-1268.93	-7.81
	(3839.57)	(473.51)	(149.06)	(598.90)	(2643.47)	(52.35)
Decile: 0.4	-2805.59	-345.40	-107.50	-433.28	-1931.31	-11.89
	(7702.37)	(948.38)	(301.00)	(1202.77)	(5298.53)	(84.51)
Decile: 0.5	601.00*	73.99*	23.03	92.82	413.72**	2.55
	(313.75)	(38.95)	(17.57)	(58.04)	(216.55)	(16.79)
Decile: 0.6	704.42	86.72	26.99	108.79	484.91	2.99
	(510.06)	(62.91)	(24.91)	(86.81)	(352.24)	(19.77)
Decile: 0.7	318.05**	39.15**	12.19	49.12*	218.94**	1.35
	(133.38)	(16.04)	(8.43)	(26.98)	(92.61)	(8.89)
Decile: 0.8	202.20***	24.89***	7.75	31.23**	139.19***	0.86
	(49.45)	(6.80)	(4.67)	(31.23)	(33.86)	(5.63)
Decile: 0.9	174.63***	21.50***	6.69	26.97**	120.21***	0.74
	(34.67)	(4.59)	(3.81)	(10.76)	(24.39)	(4.86)

Notes: The sample is drawn from the 2010 PHF. The observational units are household reference persons aged 45 and older if not stated otherwise. Robust standard errors are in parentheses.

Bibliography

Abadie, A., J. Angrist, and G. Imbens (2002). Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings. *Econometrica 70*(1), pp. 91–117.

Alesina, A. and N. Fuchs-Schündeln (2007). Goodbye Lenin (or Not?): The Effect of Communism on People's Preferences. *American Economic Review 97*(4), 1507–1528.

Bartke, S. and R. Schwarze (2008). Risk-Averse by Nation or by Religion?: Some Insights on the Determinants of Individual Risk Attitudes. SOEPpapers 131, DIW.

Beaudry, P. and G. Koop (1993, April). Do Recessions Permanently Change Output? *Journal of Monetary Economics 31*(2), 149–163.

Blanchard, O. J. and L. H. Summers (1987, April). Fiscal Increasing Returns, Hysteresis, Real Wages and Unemployment. *European Economic Review 31*(3), 543–560.

Brännäs, K. and J. G. De Gooijer (1994). Autoregressive-Asymmetric Moving Average Models for Business Cycle Data. *Journal of Forecasting 13*(6), 529–544.

Brännäs, K., J. G. De Gooijer, C. Lönnbark, and A. Soultanaeva (2012, January). Simultaneity and Asymmetry of Returns and Volatilities: The Emerging Baltic States' Stock Exchanges. *Studies in Nonlinear Dynamics & Econometrics 16*(1), 1–24.

Brännäs, K., J. G. De Gooijer, and T. Teräsvirta (1998). Testing Linearity against Nonlinear Moving Average Models. *Communications in Statistics, Theory and Methods 27*, 2025–2035.

Chernozhukov, V. and C. Hansen (2005, January). An IV Model of Quantile Treatment Effects. *Econometrica 73*(1), 245–261.

Chevillon, G. and D. F. Hendry (2005). Non-Parametric Direct Multi-Step Estimation for Forecasting Economic Processes . *International Journal of Forecasting 21*(2), 201 – 218.

Clements, M. P. and D. F. Hendry (1996). Multi-Step Estimation for Forecasting. *Oxford Bulletin of Economics and Statistics 58*(4), 657–684.

Davidson, R. and J. MacKinnon (1983). Small Sample Properties of Alternative Forms of the Lagrange Multiplier Test. *Economics Letters 12*, 269–275.

Davidson, R. and J. G. MacKinnon (2001, January). Artificial Regressions. Working Papers 1038, Queen's University, Department of Economics.

Diebold, F. and R. Mariano (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics 13*, 253–263.

Durlauf, S. N. and P. C. B. Phillips (1988, November). Trends versus Random Walks in Time Series Analysis. *Econometrica 56*(6), 1333–54.

Elwood, S. K. (1998). Is the Persistence of Shocks to Output Asymmetric? . *Journal of Monetary Economics 41*(2), 411 – 426.

Engel, C. (2005, October). Some New Variance Bounds for Asset Prices. *Journal of Money, Credit and Banking 37*(5), 949–55.

Evans, G. W. (1991, September). Pitfalls in Testing for Explosive Bubbles in Asset Prices. *American Economic Review 81*(4), 922–30.

Firpo, S. (2007). Efficient Semiparametric Estimation of Quantile Treatment Effects. *Econometrica 75*(1), 259–276.

Flavin, M. A. (1983). Excess Volatility in the Financial Markets: A Reassessment of the Empirical Evidence. *Journal of Political Economy 91*, 929–956.

Frölich, M. and B. Melly (2010, September). Estimation of Quantile Treatment Effects with Stata. *Stata Journal 10*(3), 423–457.

Frölich, M. and B. Melly (2013, July). Unconditional Quantile Treatment Effects Under Endogeneity. *Journal of Business & Economic Statistics 31*(3), 346–357.

Fuchs-Schündeln, N. (2008). The Response of Household Saving to the Large Shock of German Reunification. *American Economic Review 98*(5), 1798–1828.

Fuchs-Schündeln, N. and M. Schündeln (2005, August). Precautionary Savings and Self-Selection: Evidence from the German Reunification Experiment. *The Quarterly Journal of Economics 120*(3), 1085–1120.

Gelfand, I. M. and G. Shilov (1964). *Generalized Functions, Properties and Operations*, Volume 1. New York and London: Academic Press.

Gilles, C. and S. F. LeRoy (1991). Econometric Aspects of the Variance-Bounds Tests: A Survey. *The Review of Financial Studies 4*(4), pp. 753–791.

Gonzalo, J. and T. H. Lee (1998). Pitfalls in Testing for Long Run Relationships. *Journal of Econometrics 86*(1), 129–154.

Granger, C. W. J. and P. Newbold (1973). Some Comments on the Evaluation of Economic Forecasts. *Applied Economics 5*(1), 35–47.

Gürkaynak, R. S. (2005). Econometric Tests of Asset Price Bubbles: Taking Stock. Finance and Economics Discussion Series 2005-04, Board of Governors of the Federal Reserve System (U.S.).

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.

Hansen, G. D. and E. C. Prescott (2005). Capacity Constraints, Asymmetries, and the Business Cycle . *Review of Economic Dynamics 8*(4), 850 – 865.

Harvey, D. I., S. J. Leybourne, and P. Newbold (1998, April). Tests for Forecast Encompassing. *Journal of Business & Economic Statistics 16*(2), 254–259.

Iannaccone, L. R. (1998, September). Introduction to the Economics of Religion. *Journal of Economic Literature 36*(3), 1465–1495.

Karras, G. and H. H. Stokes (1999). On the Asymmetric Effects of Money-Supply Shocks: International Evidence from a Panel of OECD Countries. *Applied Economics 31*(2), 227–235.

Köbrich Leon, A. and C. Pfeifer (2013). Religious Activity, Risk-Taking Preferences and Financial Behaviour: Empirical Evidence from German Survey Data. Working Paper 269, University of Lüneburg.

Keister, L. A. (2003). Religion and Wealth: The Role of Religious Affiliation and Participation in Early Adult Asset Accumulation. *Social Forces 82*(1), 175–207.

Keynes, J. M. (1936). *The General Theory of Employment Interest and Money.* Harcourt, Brace and Company.

Kilian, L. and R. J. Vigfusson (2011, November). Are the Responses of the U.S. Economy Asymmetric in Energy Price Increases and Decreases? *Quantitative Economics 2*, 419–453.

Kleidon, A. W. (1986a). Bias in Small Sample Tests of Stock Price Rationality. *Journal of Business 59*, 237–261.

Kleidon, A. W. (1986b). Variance Bounds Tests and Stock Price Valuation Models. *Journal of Political Economy 94*, 953–1001.

Koenker, R. W. and G. W. Bassett (1978). Regression Quantiles. *Econometrica 46*(1), 33–50.

Koutmos, G. (1999). Asymmetric Index Stock Returns: Evidence from the G7. *Applied Economics Letters 6*(12), 817–820.

Lansing, K. J. (2015). On Variance Bounds for Asset Price Changes . *Journal of Financial Markets*, forthcoming. http://dx.doi.org/10.1016/j.finmar.2015.06.002.

LeRoy, S. F. and R. D. Porter (1981). The Present Value Relation: Tests Based on Implied Variance Bounds. *Econometrica 49*, 555–574.

Lütkepohl, H. (1996). *Handbook of Matrices.* John Wiley & Sons, Ltd.

Lütkepohl, H. and F. Xu (2012, June). The Role of the Log Transformation in Forecasting Economic Variables. *Empirical Economics 42*(3), 619–638.

Mankiw, N. G., D. Romer, and M. D. Shapiro (1985). An Unbiased Reeximination of Stock Market Volatility. *Journal of Finance 40*, 677–678.

Mankiw, N. G., D. Romer, and M. D. Shapiro (1991). Stock Market Forecastability and Volatility: A Statistical Appraisal. *The Review of Economic Studies 58*(3), 455–477.

Marcellino, M., J. H. Stock, and M. W. Watson (2006). A Comparison of Direct and Iterated Multistep AR Methods for Forecasting Macroeconomic Time Series . *Journal of Econometrics 135*(12), 499 – 526.

Marsh, T. A. and R. C. Merton (1986). Dividend Variability and Variance Bounds Tests for the Rationality of Stock Market Prices. *American Economic Review 76*, 483–503.

Neftci, S. N. (1984, April). Are Economic Time Series Asymmetric over the Business Cycle? *Journal of Political Economy 92*(2), 307–28.

Nelson, C. R. and C. I. Plosser (1982). Trends and Random Walks in Macroeconmic Time Series : Some Evidence and Implications. *Journal of Monetary Economics 10*(2), 139–162.

Noussair, C., S. Trautmann, G. van de Kuilen, and N. Vellekoop (2013). Risk Aversion and Religion. *Journal of Risk and Uncertainty 47*(2), 165–183.

Pelagatti, M. M. and F. Lisis (2009). *"Variance Initialisation in GARCH Estimation" Complex Data Modeling and Computationally Intensive Statistical Methods for Estimation and Prediction.*

Phillips, P. (1991, December). A Shortcut to LAD Estimator Asymptotics. *Econometric Theory 7*(04), 450–463.

Phillips, P. (1997). Towards a Unified Asymptotic Theory for Autoregression. *Biometrica 74*(3), 535–547.

Phillips, P. C. B. (1998). Impulse Response and Forecast Error Variance Asymptotics in Nonstationary VARs. *Journal of Econometrics 83*(1), 21 – 56.

Sammet, K. (2012). *Religious Identity and National Heritage Empirical-Theological Perspectives*, Chapter Atheism and Secularism: Cultural Heritage in East Germany, pp. 269 –288. Brill Academic Pub.

Schorfheide, F. (2005). VAR Forecasting under Misspecification . *Journal of Econometrics 128*(1), 99 – 136.

Shea, G. S. (1989, April-Jun). Ex-Post Rational Price Approximations and the Empirical Reliability of the Present-Value Relation. *Journal of Applied Econometrics 4*(2), 139–59.

Sheffer, E. (2007, 6). On Edge: Building the Border in East and West Germany. *Central European History 40*, 307–339.

Shiller, R. J. (1981). Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends? *American Economic Review 71*(3), 421–436.

Smith, A. (1965). *An Inquiry into the Nature and Causes of the Wealth of Nations*. New York: Modern Library.

Stock, J. H. and M. W. Watson (2004, September). Combination Forecasts of Output Growth in a Seven-Country Data Set. *Journal of Forecasting 23*(6), 405–430.

Ullah, A. (2004). *Finite Sample Econometrics*. Oxford University Press.

Weber, M. (1958). *The Protestant Ethic and the Spirit of Capitalism*. Translated by Talcott Parsons, New York: Free Press.

Wecker, W. E. (1981). Asymmetric Time Series. *Journal of the American Statistical Association 76*(373), 16–21.

West, K. D. (1988, January). Dividend Innovations and Stock Price Volatility. *Econometrica 56*(1), 37–61.

White, H. (2001). *Asymptotic Theory for Econometricians: Revised Edition.* Academic Press.