Nonresponse in Business Tendency Surveys: Theoretical Discourse and Empirical Evidence

Christian Seiler

Dissertation zur Erlangung des Grades eines
Doktors der Naturwissenschaften (Dr. rer. nat.)
an der Fakultät für Mathematik, Informatik und Statistik
der Ludwig-Maximilians-Universität München

München, im Jahre 2013

Erstgutachter: PD Dr. Christian Heumann
Zweitgutachter: Prof. Dr. Kai Carstensen
Tag der mündlichen Prüfung: 12. Juni 2013

Danksagung

An dieser Stelle möchte ich bei den Personen bedanken, ohne deren Unterstützung diese Arbeit nicht hätte entstehen können. Dies gilt insbesondere meinen Betreuern Kai Carstensen und Christian Heumann. Beide hatten immer ein offenes Ohr und ich konnte jederzeit auf deren Meinung und Erfahrung vertrauen. Ein besonderer Dank gilt außerdem Meinhard Knoche, der mir die Möglichkeit gegeben hat als "Fachfremder" am ifo Institut zu promovieren.

Diese Arbeit hat sehr von der angenehmen Atmosphäre in meinem Arbeitsbereich 'Konjunktur und Befragungen' profitiert. Bedanken möchte ich mich bei meinen Kollegen Doris Hauke, Peter Jäckel, André Kunkel, Heike Mittelmeier, Johanna Plenk, Wolfgang Ruppert, Stefan Sauer, Sigrid Stallhofer und Annette Weichselberger für ihre Unterstützung. Desweiteren bedanken möchte ich mich bei Klaus Abberger, Teresa Buchen, Christian Grimme, Steffen Henzel, Nikolay Hristov, Michael Kleemann, Johannes Mayr, Steffen Elstner, Anna Wolf, Timo Wollmershäuser und Peter Zorn für die anregenden Diskussionen. Diese haben mich in meinem ökonomischen Denken vorangebracht und einen wichtigen Beitrag zum Gelingen dieser Doktorarbeit geliefert. Ein besonderer Dank gilt Klaus Wohlrabe für seine motivierende Unterstützung.

Dank gilt auch allen Teilnehmern des PR^2-Promotionsprogramm für den – insbesondere zu den beiden Sommerklausuren – intensiven Austausch über die Themenbereiche dieser Arbeit. Namentlich möchte ich an dieser Stelle besonders Felix Heinzl danken, der mich beim redaktionellen Teil dieser Arbeit unterstützt hat.

Ein besonderer Dank gilt meiner Familie und meiner Freundin Katharina, die mich zu jeder Zeit in meinen Entscheidungen bekräftigt haben und für mich immer ein großer Rückhalt und Ausgleich waren.

Zusammenfassung

Zur Beantwortung sozioökonomischer Fragestellungen nehmen Umfragen als Methode für den empirischen Erkenntnisgewinn eine zentrale Rolle ein. Ein weitverbreitetes Problem auf der Erhebungsseite ist jedoch das Auftreten fehlender Werte, welche zu verzerrten Ergebnissen führen können. Während es für Bevölkerungs- und Haushaltsbefragungen eine umfassende Literatur zu diesem Thema existiert, ist dieses Thema im Bereich von Unternehmensbefragungen in weit geringerem Maße erforscht worden. Diese Arbeit widmet sich den fehlenden Werten im *ifo Konjunkturtest*, welcher in ähnlicher Form in fast allen OECD-Ländern durchgeführt wird. Das prominenteste Ergebnis dieser monatlich durchgeführten Umfrage ist der *ifo Geschäftsklimaindikator*, ein Konjunkturindikator für die deutsche Wirtschaft, welcher große Beachtung bei Unternehmern, Analysten, Politikern, Journalisten, Wissenschaftlern und in der breiten Öffentlichkeit findet. Die Ergebnisse dieser Arbeit zeigen, dass Konjunkturindikatoren basierend auf dieser Form der Befragung sehr stabil bezüglich nicht-zufälligen Ausfallprozessen sind. Dies lässt sich sowohl mit Hilfe von Simulationsstudien als auch durch die Schätzung der fehlenden Werte zeigen. Insbesondere führen die fehlenden Werte nicht zu einer Verschlechterung der Prognoseleistung des ifo Geschäftsklimaindikators.

Abstract

Surveys are a widely used tool to answer socio-economic research question across disciplines. However, data collection can face certain problems such as nonresponding units. For household and population surveys, a large body of literature about the effects of nonresponse exist but only less is known in case of business surveys. This thesis deals with the missing values in the *Ifo Business Survey* which is conducted in similar form in nearly all OECD countries. The most prominent result of this survey is the *Ifo Business Climate Index*, a business cycle indicator for the German economy. This indicator is highly observed by entrepreneurs, analysts, politicians, journalists, academics and the general public. The results of this thesis show that business cycle indicators based on this type of questioning are very stable towards any kind of non-random missing data processes. This is shown by simulation studies as well as an estimation of the missing values. In particular, the missing values do not lead to a significant reduction in forecasting performance.

Nomenclature

Due to the cumulative structure of this thesis, a short description of the variables and abbreviations used is given in this section. Although all variables are explained in the appropriate chapter, this overview helps to avoid any misconceptions.

- In *Chapter 2* $\boldsymbol{Y}$ denotes the data set and $\boldsymbol{M}$ an indicator matrix for the observed and unobserved data points. i indices units, whereas j indices variables. Therefore, $\boldsymbol{y}_{i,j}$ and $\boldsymbol{m}_{i,j}$ are the appropriate cells in the matrices $\boldsymbol{Y}$ and $\boldsymbol{M}$. $f(\cdot)$ is a density function and ϕ unknown parameters.

- *Chapter 3* deals with a regression to explain effects on the response behaviour. Therefore, $y_{i,t}$ denotes a 1/0-dummy for (non-)response. Again, i indices units. t denotes the months of the used data set, i.e. it ranges from $1 \equiv$ January 1994 to $T = 192 \equiv$ December 2009. $\boldsymbol{x}_{i,t}$ are covariates. All other symbols represent model-specific parameters ($\beta_0, \boldsymbol{\beta}, \phi, \pi_{i,t}, \boldsymbol{\mu}_i, \boldsymbol{R}_i(\boldsymbol{\alpha}), \boldsymbol{V}_i, \hat{\boldsymbol{D}}_i, \hat{\boldsymbol{C}}_i, \boldsymbol{W}_i$) and functions ($g(\cdot), v(\cdot)$)

- In *Chapter 4* a theoretical discourse on the effects of selection biases on the indicators is given. $g(t)$ denotes an unobserved business cycle function (which will be later concretely defined as $g_j(t), j = 1, \ldots, 4$). s^* is the unobserved, s (and r) the observed 'state'. i indices units and t time points (which range from 1 to 500 in the simulation study). τ^s indicates unknown thresholds, $\Phi(\cdot)$ the cumulative density function of the standard normal distribution. As in Chapter 1, m is an indicator for the participation. The function $\pi^s(t)$ denotes the probability to participate given time t and being in state s. $\tilde{\rho}_C, \rho_C^{obs}$ and ρ_E^{obs} are different types of correlation in this chapter. For the simulation, z denotes the number of iterations and $\pi^C_{j,k_C}(t), \pi^S_{j,k_C}(t+k_T), \pi^L_{k_L}(t)$ and $\pi^M_{k_M}(t)$ different acceptance rates (with parameters k_C, k_L, k_M and k_T). v is defined as a dispersion measure. In Appendix B, b_t denotes the balance statistics, n_t the empirical number of observations at time t and $I(\cdot)$ an indicator function. a, a^+, a^-, b, b^+ and b^- are scalars for the proof in Appendix B.2.

- In *Chapter 5* the variables to be estimated are referred to $y_{i,t}$, where again i denotes the unit and t (and s) the time point as in Chapter 3. $d(\cdot)$ is a distance measure function. $\boldsymbol{P}$ with entries $p(\cdot)$ denotes a stochastic matrix with states $r, s \in S$ which is of order k. In contrast, $P(\cdot)$ is defined as a probability function. As in Chapter 1, $\eta_i, g(\cdot), h(\cdot), \mu_i, \tau_c, \boldsymbol{x}_i$ and $\boldsymbol{\beta}$ are parameters and function for the regression based imputation. The goodness-of-fit measure κ defined in Section 5.3.3 includes probabilities π_o and π_e. Finally, z_t denotes a time series and $\alpha, \phi_{i^*}, \theta_{j^*}$ and ϵ_t the parameters of the forecasting model. In Appendix C, $\widetilde{n}$ are the number of replies, b is the balance statistics, u and v are indices for the subsectors and ω sector weights.

In addition, the following table gives an overview on the abbreviations used in this thesis.

Abbreviation	
ADL	Autoregressive distributed lag
AR	Autoregressive
BE	Business Expectations
BS	Business Situation
CON	Construction
DGP	Data Generating Process
EU	European Union
GDP	Gross Domestic Product
GEE	Generalized Estimation Equation
GLM	Generalized Linear Model
IBS	Ifo Business Survey
IND	Industry
JD	Joint Distribution
kNN	k Nearest Neighbour
LOCF	Last observation carried forward
MAR	Missing at random
MC	Markov Chain
MCAR	Missing completely at random
MI	Multiple Imputation
NA	Not available (denotes a missing value)
NARHS	National HIV/AIDS and Reproductive Health Survey
NMAR	Not missing at random
NN	Nearest Neighbour
OECD	Organization for Economic Co-operation and Development
POM	Proportional odds model
RMSE	Root mean squared error
SDS	Synthetic data sets
SVD	Singular Value Decomposition
SVT	Singular Value Thresholding
TRA	Trade
U.S.	United States of America

Contents

List of Figures

List of Tables

Chapter 1

Introduction and motivation

"Data! Data! Data!" he cried impatiently.
"I can't make bricks without clay."

Sherlock Holmes in
The Adventure of the Copper Beeches

Today, data collection has gained substantive importance in research across academic disciplines. Theoretical considerations have to keep standing in front of 'hard' facts. But still it is problematic in many sciences to collect independent, unbiased information of high quality. In socio-economic sciences data collection commonly bases on surveys with respect to evaluate and analyse parameters of social and economic interaction. This practice has a long tradition in research but still open questions on reliability and quality of the data remain. One serious aspect is the question how the answering behaviour might influence survey results. In particular, nonresponses can cause significant biases when these are not independent from the research question. A wide-spread literature exists about effects of responding behaviour in population or household surveys but only less is known about the participation process in business surveys. As literature leaves a gap on this issue, this thesis deals with several aspects of nonresponse in business surveys, starting with effects which influence the response behaviour, a theoretical discourse on the effects of different types of nonresponse patterns on the survey outcomes and a comparison when these missing values are estimated. The data source of the empirical analysis in this thesis is a large, high-frequent business survey of German firms which exists since 1949. This data set allows to give answers on the questions stated above and can be used as a basis for future studies using this data sets if serious concerns arise that response behaviour might influence the results.

As writing a thesis is a dynamic process (like academics in general), parts of this work were already published and/or are currently under review. Therefore, this thesis bases on the following publications and working papers:

- C. Seiler (2010), 'Dynamic Modelling in Business Surveys', *Ifo Working Paper No. 93*. (submitted)
- S. A. Adebayo, L. Fahrmeir, C. Seiler and C. Heumann (2011), 'Geoadditive latent variable modelling of count data on multiple sexual partnering in Nigeria', *Biometrics* 67(2), 620-628.
- C. Seiler (2012), 'On the Robustness of the Balance Statistics with respect to Nonresponse', *Ifo Working Paper No. 126*. (submitted)[1]
- C. Seiler and C. Heumann (2012), 'Microdata Imputations and Microdata Implications: Evidence from the Ifo Business Survey', *Department of Statistics, Technical Report No. 119*. (submitted)
- C. Seiler (2012), 'Zur Robustheit des ifo Geschäftsklimaindikators in Bezug auf fehlende Werte', *ifo Schnelldienst* 65(17), 19-22.

Due to the cumulative nature of this thesis a short literature overview on the main topics will be provided in Chapter 2. We address for the research which was done before in this field with respect to sources of nonresponse in business surveys and give a short introduction into statistical theory. In addition, we briefly discuss possibilities to address for nonresponse in real-life data situations. Afterwards, we define the main scope of this thesis. Even if the whole thesis deals with missing data in business surveys, every chapter should be interpreted independently. For this reason,

[1]This paper was awarded with the *Isaac Kerstenetzky Young Economist Award 2012* at the 31st CIRET Conference.

a short description of the data set is included which introduces the topic to be focused. In addition, abstracts on the beginning of every chapter will give a short overview of the treated research question. Nevertheless, all chapters base on the results of the previous ones. In Chapter 6 we finally summarise the findings of all chapters and give a final conclusion of the results.

Chapter 2

Literature Review and Definitions

In this section, we want to give a short overview on the literature and the academic foundations of this thesis. Section 2.1 shows the historical development of business surveys in general and the Ifo Business Survey (IBS) which data sets are used in this thesis. After this, we give an introduction on survey nonresponse in Section 2.2. We briefly discuss the nonresponse problem with respect to business surveys and introduce some theoretical definitions on missing data patterns. Finally, a short outline is given on how to handle missing data. In Section 2.3 we display the main research questions of this thesis that will be discussed in Chapters 3, 4 and 5.

2.1 Business Surveys

2.1.1 A short historical overview

The collection of business-related data has a long tradition in research. Early surveys were conducted almost exclusively by official agencies mostly with the background to tax collection issues. The first census in North America, which also collected data about wealth of industry and agriculture, was the Census of New France performed in 1666. The first census in the United States was conducted in 1790 only some years after the foundation of the country. At this time, just with the beginning of the industrialisation era, the main focus was on the collection of agricultural statistics to improve farming practices (Allen et al., 1995). In Germany, first statistical offices were founded 1805 in Prussia, 1808 in Bavaria and 1820 in Württemberg. During the 19th century, official statistics, and therefore also the collection of business data, were established in many countries around the world (Worton and Platek, 1995).

2.1.2 The Ifo Business Survey

With the introduction of the Ifo Business Survey in 1949 the survey has gained a long tradition in the German economic research over the decades. To this day, the survey outcomes, in particular the *Ifo Business Climate Index*, are intensively observed and used by researchers, analysts, politicians and the general public. Interestingly, the survey was originally introduced for very pragmatic reasons: To close the gaps in official data which were relatively large in the early German postwar years (Goldrian, 2007). Another major advantage of this survey was to gain quick information on the current economic situation for those parameters which were evaluated by official statistics but only published with high delay and commonly were (and still are) affected by revisions. However, this advantage can only be achieved if the questionnaire can be filled fast which is obtained by asking tendency questions. Early analyses with the survey results can by found in Anderson (1951, 1952), Langelütke and Marquardt (1951) and Theil (1952, 1955). Becker and Wohlrabe (2008) give an overview on the research with the Ifo Business Survey results, Abberger and Wohlrabe (2006) on forecasting analyses with the Ifo Business Climate Index.

The success of the IBS led to a worldwide spread to this type of tendency surveys. Nardo (2003) states that this type of survey has been increased over the last decades. Nowadays, these surveys are conducted in Austria, Argentina, Australia, Belgium, Bolivia, Bulgaria, Brazil, Canada, China P.R., the Republic of China, Colombia, the Czech Republic, Denmark, the Dominican Republic, Ecuador, Estonia, Finland, France, Germany, Greece, Hong Kong, Hungary, India, Indonesia, Ireland, Italy, Japan, Kazakhstan, Republic of Korea, Latvia, Laos, Lebanon, Lithuania, Malaysia, Mexico, the Netherlands, New Zealand, Norway, Peru, Philippines, Poland, Portugal, Romania, Russia, Saudi Arabia, Singapore, Slovakia, Slovenia, South Africa, Spain, Sweden, Switzerland, Thailand, Turkey, Ukraine, the United Kingdom, the United States, Uzbekistan, Venezuela and Vietnam, see OECD (2003). For the European Union, the survey of their member states are harmonised in line with the Joint Harmonised EU Programme of Business and Consumer Surveys, see European Union (2006). This allows to calculate corporate results on EU level, such as the EU economic sentiment indicator, which is similar to Ifo's Business Climate Index. Within the EU, more than 50,000 firms are asked each month.

The main difference between these tendency surveys and other business surveys (including official statistics) is that are almost solely qualitative assessment questions on the questionnaire, usually on a 3-level Likert scale. Due to the general economic growth of wealth, these variables measure a change rather than the level (Theil, 1952), in particular with respect to questions regarded to expectations (Knöbl, 1974 and Carlson and Parkin, 1975). Thus, they measure the deviation of the growth, which commonly fluctuate over time. These fluctuations are understood as business cycles (Burns and Mitchell, 1946):

> *Business cycles are a type of fluctuation found in aggregate economic activity of nations that organize their work mainly in business enterprises: a cycle consists of expansions occurring at about the same time in many economic activities, followed by similarly general recessions, contractions, and revivals which merge into the expansion phase of the next cycle; this sequence is recurrent but not periodic; in duration business cycles vary from more than one year to ten to twelve years; they are not divisible into shorter cycles of similar character with amplitudes approximating their own.*

Besides the IBS, the Ifo Institute also performs other types of regular surveys which measure quantitative variables such as investments (Ifo Investment Survey) or innovations (Ifo Innovation Survey). Due to the fact that qualitative questions need more time to be answered, these surveys are conducted only (bi-)annually. However, in this thesis, we focus on the Ifo Business Survey and its most prominent result, the Ifo Business Climate Index.

2.2 Survey Nonresponse

Nonresponse is defined as a refusal to reply in surveys (Groves et al., 2002). This implies a made contact between the researcher and the asked participant. So, design-related selection processes, i.e. that someone has no positive probability to enter the study, are generally not included in this type of definition. Literature only differs between two types of nonresponse: *item* and *unit nonresponse*. In the first case, the respondent did generally agree to take part in the survey but refused to give answers to single questions (items) of the survey. This is often a concern in population surveys when asking cognitive complex or 'sensitive' questions (Tourangeau et al., 2000) like income, drug use or sexual behaviour. In contrast, unit nonresponse displays the fact when the asked person (or more general 'unit') even refused to take part in the survey. Regardless of the type of nonresponse it can cause certain problems such as biases in point estimators, inflation of variance of point estimators and biases in estimators of precision, see Dillman et al. (2002). While the problem of variance inflation can be weakened with a higher sample size or appropriate statistical models, biased point estimates are more serious due to fact that higher sample sizes did not solve the problem (Little and Rubin, 2002). In this case, unit weighting can help to reduce this effect (Dillman et al., 2002) but still biased estimates can occur when the missing-data mechanism is *not missing at random* (NMAR). Unbiasness can only be obtained if the data are *missing at random* (MAR) or, more restrictive, *missing completely at random* (MCAR). We now explain these mechanisms and introduce some formal defintions according to Little and Rubin (2002).

2.2.1 Definitions

Let $\boldsymbol{Y} = (y_{i,j}), i = 1, \ldots, n, j = 1, \ldots, K$ denote a $(n \times K)$ data set with n observations an K variables. We define $\boldsymbol{Y}$ as a data set without missing values. To indicate which data points are missing, we introduce a missing-data indicator matrix $\boldsymbol{M} = (m_{i,j})$ with $m_{i,j} = 1$ if $y_{i,j}$ is missing and $m_{i,j} = 0$ if not. Only values $y_{i,j}$ with $m_{i,j} = 0$ are actually observed.

The missing-data mechanism is defined by the distribution $f(\boldsymbol{M}|\boldsymbol{Y}, \phi)$ of $\boldsymbol{M}$ given the complete data set $\boldsymbol{Y}$ and unknown parameters ϕ. The missing-data mechanism does not depend on the values of $\boldsymbol{Y}$, i.e. is MCAR, if

$$f(\boldsymbol{M}|\boldsymbol{Y},\phi) = f(\boldsymbol{M}|\phi), \qquad \forall \boldsymbol{Y},\phi.$$

However, this is very restrictive and often not the case in real-life situations. A less strong assumption is MAR which allows the dependence of the missing-data mechanism from the *observed units* $\boldsymbol{Y}^{obs}$. That is,

$$f(\boldsymbol{M}|\boldsymbol{Y},\phi) = f(\boldsymbol{M}|\boldsymbol{Y}^{obs},\phi), \qquad \forall \boldsymbol{Y}^{mis},\phi.$$

If the mechanism also depends from the missing observations $\boldsymbol{Y}^{mis}$, $f(\boldsymbol{M}|\boldsymbol{Y},\phi)$ turns to

$$\begin{aligned} f(\boldsymbol{M}|\boldsymbol{Y},\phi) &= f(\boldsymbol{M}|\boldsymbol{Y}^{obs},\boldsymbol{Y}^{mis},\phi) \\ &= f(\boldsymbol{M}|\boldsymbol{Y},\phi), \qquad \forall \boldsymbol{Y},\phi, \end{aligned}$$

and is defined NMAR. Non-formal, this means:

- For **MCAR**, the missing-data mechanism is independent from the data (but might depend from other variables, i.e. survey design). For example, a general missing probability of 0.2 for all data points would be MCAR. In this case, the missing-data mechanism is called *ignorable*, i.e. standard statistical approaches can be used without obtaining biased results.

- **MAR** defines a missing-data mechanism which depends from observed values. This means that the missing values in one variable can be explained by an observed value in another variable. For example, a missing in an income question can be estimated if the school degree is known (and there exists a significant relationship between these two variables).

- In all other cases the missing-data mechanism is **NMAR**. For example, the probability to answer an income question depends on the income level and no other observations in the data set are observed to estimate the missing values.

It is important to note that these definitions are statements on the *whole data set*. As described above, single variables would be NMAR if they are analysed separately but the complete data set might be MAR. These definitions also explain the usage of imputation methods for item nonresponse in cross-sectional data sets since these models allow to estimate the missing values if the missing-data mechanism is at least MAR. Therefore, in case of unit nonresponse it is harder to find appropriate imputation models as only less (if any) information prevails.

2.2.2 Nonresponse in business surveys

The main difference between population and business surveys is the fact that the respondent in a business survey does not provide information on his or her personal circumstances but does act as an agent of the company. Depending on the position of the respondent and questions asked in the survey, the participation process may be more complex since collection of relevant information can be time-consuming, see Tomaskovic-Devey et al. (1994, 1995). A general participation framework was developed by Willimack et al. (2002) which extended the standard model for population surveys by Groves and Couper (1998). In contrast to population surveys, there exist only less papers, in particular empirical analyses, on nonresponse in business surveys, see also Janik and Kohaut (2012). In particular, there are only less studies concerning bias patterns in business cycle analyses. As defined in Section 2.2.1, biases can occur when the variable of interest, the business cycle in our case, leads to a different response behaviour. For population surveys, Harris-Kojetin and Tucker (1999) found such effects. We will analyse this question in Chapter 3. An overview on nonresponse studies in business surveys can be found in Willimack et al. (2002). For German-based nonresponse analyses, see Schnabel (1997), Hartmann and Kohaut (2000) and Janik and Kohaut (2012). A newer study with respect to the participation process in business surveys of large firms can be found in Willimack and Nichols (2010).

2.2.3 Adjustment for nonresponse

After defining missing patterns in Section 2.2.1, we now want to give an overview of statistical methods to account for possible nonresponse biases *after* data collection. An early method was proposed by Heckman (1979) which is known as Heckman correction and very popular in economic sciences. This type of bias correction requires a statistical model for the participation process. From this model, weights can be obtained by using the inverse probability to respond. So, the idea is to upweight units with high probability not to respond because it is assumed that their answers include more information (see Bethlehem, 1988 and Ekholm and Laaksonen, 1991). But this approach has the main disadvantage that it still only bases on the observed units and needs a correct specification of the missing data mechanism which is often hard to construct as this mechanism may be complex, see Subsection 2.2.2. Beyond, several papers found that the Heckman correction is inefficient when there exists a high correlation between the error term and the selection mechanism, see Puhani (2000).

An alternative approach to account for missing data was developed by Donald Rubin (see Rubin (1987) for an overview). This method estimates the missing values and is called imputation. Although it seems attractive to generate observations 'out of nowhere', Dempster and Rubin (1983) argue that imputation has some pitfalls:

> *The idea of imputation is both seductive and dangerous. It is seductive because it can lull the user into the pleasureable state of believing that the data are complete after all, and it is dangerous because it lumps together situations where the problem is sufficiently minor that it can be legitimately handled in this way and situations where standard estimators applied to the real and imputed data have substantial biases.*

For these reasons, imputation methods have to be used with care. Every imputation model has to be evaluated carefully on the observed data to enable sufficient predictions of the missing values. In particular, the imputation model should not change the distributions of the variables if the data is MCAR. In addition, if the imputation method itself includes uncertainty it is recommended to reflect this with *multiple imputations* (MI), i.e.

more than one prediction for one missing value. With increase of computational performance, MI has been become more popular.

Nevertheless, imputation has substantial advantages as it allows a deeper insight into a possible selection mechanism and to test the data for robustness. Although a model for the estimation has to be specified, this is often more easy than to find a model for the missing-data mechanism. Little (1992) gives a review on regression with missing data and Horton and Lipsitz (2001) provide a comparison of software packages for imputation approaches. Again, Little and Rubin (2002) is the standard literature to enter imputation methods. In particular, imputation allows to control for high item nonresponse rates of single variables in regression analyses. We illustrate this with a small example from Adebayo et al. (2011):

Example for imputation of item nonresponse

The data used in this example are from the 2005 National HIV/AIDS and Reproductive Health Survey (NARHS) in Nigeria. The aim of this analysis was to find variables which have an effect on the number of sexual partners. Of the 4962 male respondents that participated in the survey, only 3174 had sex in the last 12 months prior to the survey. Information on *age at first sex* was only available for 2945 respondents, resulting in 7.2% missing observations. For the other covariates we had the following rather small missing proportions: *length of stay* 2.9%, *away from home* 1.8%, *marital status* 2.2%, *level of educational attainment* 0.9%. A list-wise deletion leads to only 2632 complete cases, so it seems plausible to address the missing problem by some imputation methods. The most critical or adversely affected covariate was *age at first sex*.

To impute the missing values, we used the package `Amelia II` of Honaker et al. (2012) in R to create 5 multiple imputed data sets. The main assumption is that the missing data are missing at random, as defined in Section 2.2.1. All covariates (including the completely observed ones) and the response variables were used in the imputation procedure. Since the variables *length of stay* and *age at first sex* are given only in full months and years we treated them as ordinal variables. All other (missing) covariates were treated as nominal variables. The analysis was then run for all five data sets. We did not combine the results to a final multiple imputation estimate but used the imputations only to see if there are any notable differences between the results.

To get a picture of the imputations we show in the following some plots of the relative frequencies of *age at first sex* (`agesex1`) and *length of stay* (`length.stay`) using (i) only the complete data and (ii) the imputed variables after imputation number 3 of the 5 imputations. There are no notable differences found in the distributions between complete and imputed data, see Figure 2.1. Therefore, it seems that the missing-data mechanism is rather small. To check whether the imputations have no significant impact on the estimation results, the non-linear effects for the complete-case and the imputed data set are displayed in Figures 2.2 and 2.3. It can be seen that the differences are very small.

Software overview

In this subsection, we want to give a short overview on software packages for imputation in practice. For R, several packages do exist. Besides the already mentioned `Amelia II` package by Honaker et al. (2012), package `imputation` (Wong, 2011) includes kNN (k Nearest Neighbour), Singular Value Decomposition (SVD) and Singular Value Thresholding (SVT) imputation. Multiple imputation can be done in R with package `mice` (Multivariate Imputation by Chained Equations) by van Buuren and Groothuis-Oudshoorn (2011) and `mi` (Multiple Iterative Regression Imputation) by Su et al. (2011). There also exist packages for special data types, e.g. `imputeMDR` (Multifactor Dimensionality Reduction) by Namkung et al. (2011) for missing data in gene-gene interactions and `imputeYn` (Khan and Shaw, 2012) for imputing the last largest censored datum in duration modelling. For STATA, add-on `mi` provides multiple imputation techniques (STATA-Corporation, 2009). In addition, there exists an enormous number of ado-files, for example `ice` (Royston, 2005), `stsurvimpute` (Royston, 2011) or `whotdeck` (Mander, 2003). However, most of these ado-files are limited to certain data structures and/or include only a small fraction of imputation methods. For SAS, the procedures `MI` (SAS-Institute, 2010) allows multiple imputation, `IVEware` (Raghunathan and van Hoewyk, 2002) fully conditional imputation approaches. For SPSS, the Missing Values add-on (MVA) module also in-

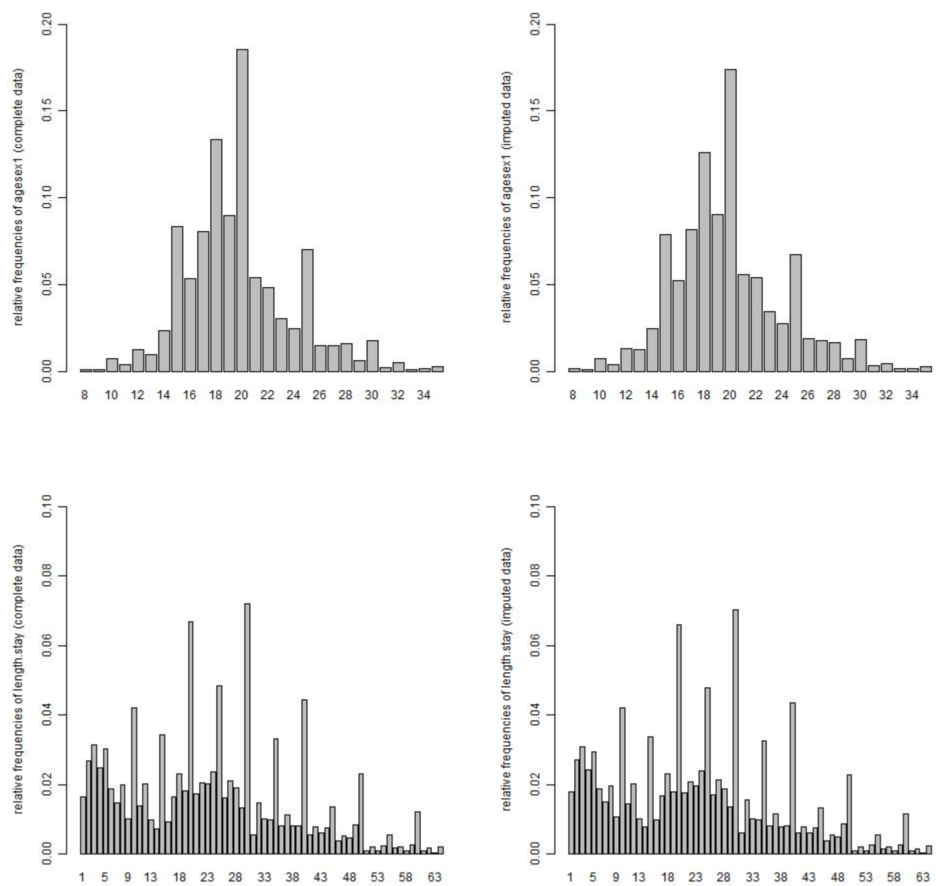

Figure 2.1: Relative frequency distribution of the variables *age at first sex* (`agesex1`, top panel) and *length of stay* (`length.stay`, bottom panel) for complete case data (left panel) and imputed data (right panel).

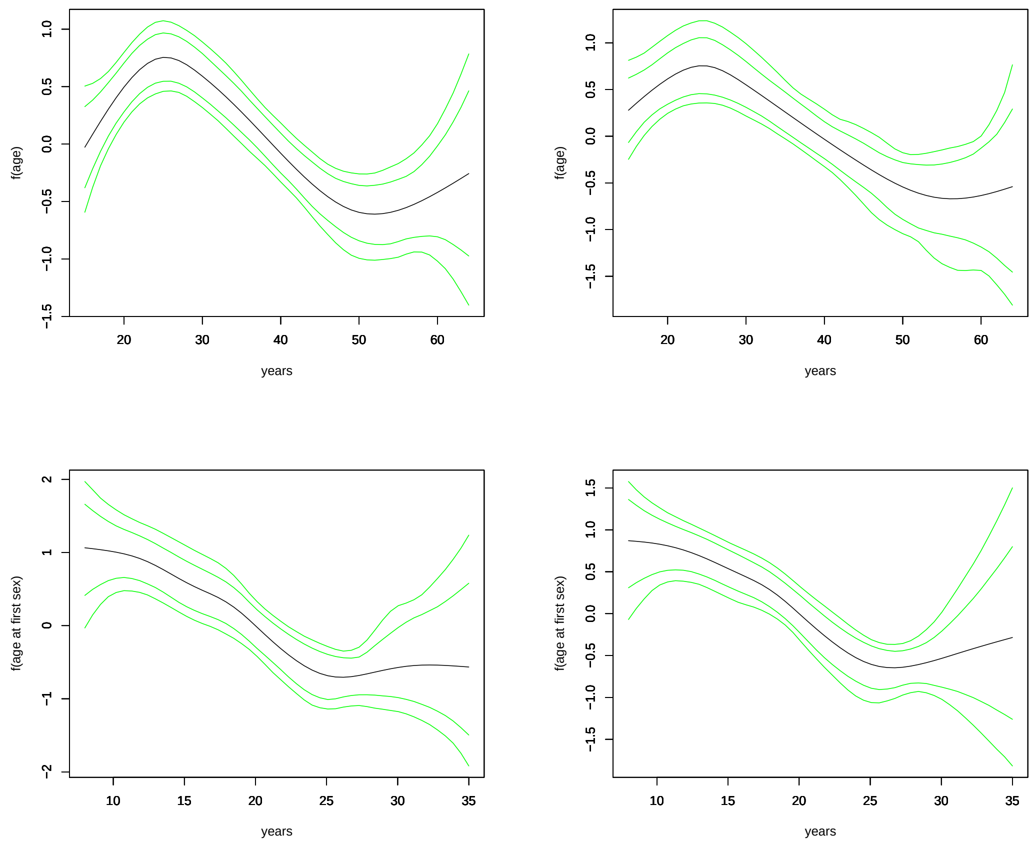

Figure 2.2: Non-linear effects of respondent's *age* (`age`, top panel) and *age at first sex* (`agesex1`, bottom panel) for complete case data (left panel) and imputed data (right panel)

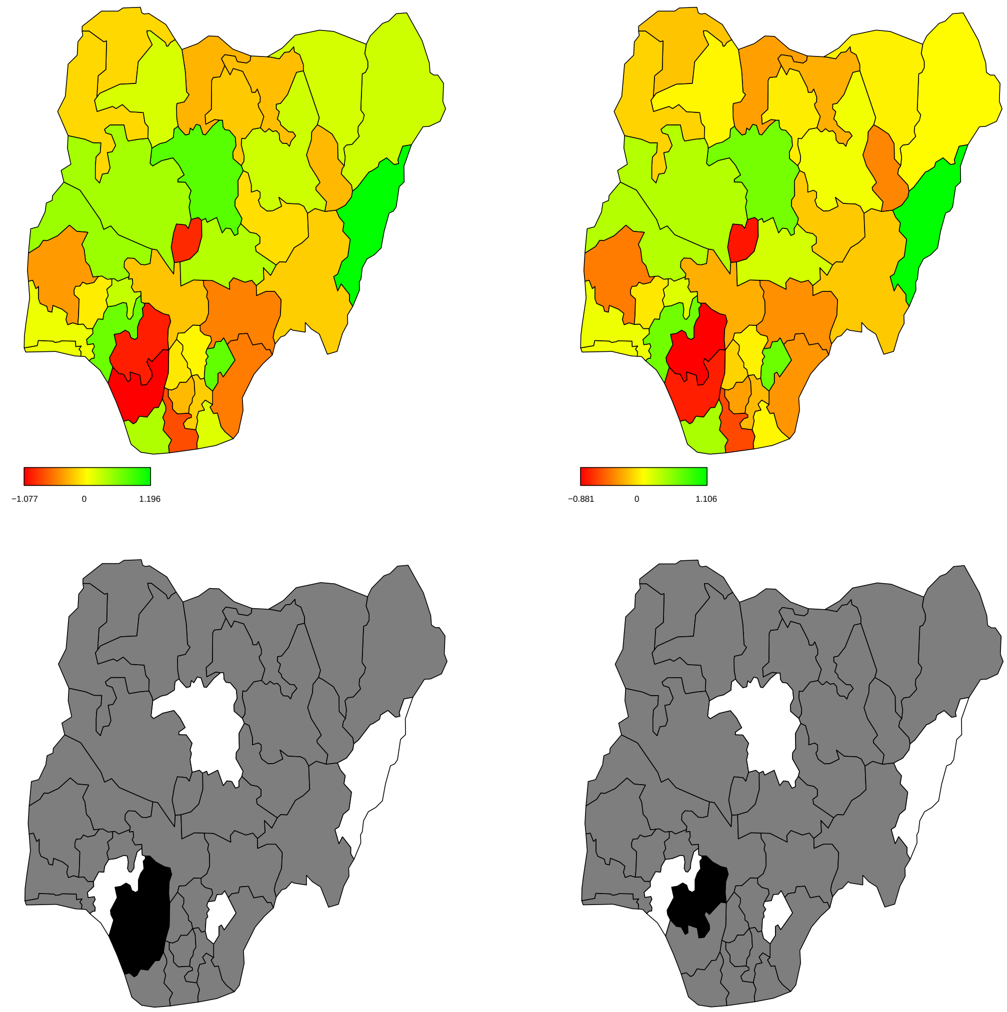

Figure 2.3: Non-linear spatial effects (top panel) for complete case data (left panel) and imputed data (right panel) and their maps of significance (bottom panel). States in dark colour have negative credible intervals, states in white colour have positive credible intervals while states in grey colour have credible intervals including zero.

cludes multiple imputation. As noted above, many software components are adjusted to different data situations and/or are limited with respect to the type of variables, e.g. continuous ones. Drechsler (2011b) noticed that several additional aspects have to be considered in practice when imputing real data.

2.3 Scope of the thesis

In this thesis, there are different aspects of nonresponse in business tendency surveys which will be analysed: In Chapter 3 we analyse the effects which change the response behaviour of the firms and might cause nonresponse. Besides the evaluation of differences in response behaviour according to business characteristics, we focus on the analysis of time-related effects as the results of the IBS are primarily interpreted in their time dimension. In particular, the dependence from the business cycle (for which the survey is conducted) is about to be examined. Chapter 4 deals with theoretical considerations on different types of missing patterns on the balance statistics indicators, such as the Ifo Business Climate Index. In this chapter, we analyse how different NMAR patterns affect these indicators. We focus our analysis on the correlation of the biased indicators with different cycle functions. Finally, the analysis shows that a decrease in correlation is seldom and the indicators are very robust towards different kinds of NMAR patterns. In Chapter 5 we finally analyse the effects of the missing values on the indicators when these are imputed. As the variables in the data set have a special structure, we developed different imputation strategies to enable predictions. We run several different imputation approaches for longitudinal categorical data and evaluate their predictive performance also for longer missing sequences. Afterwards, we recalculate the survey results and display the differences between the complete-case and the imputed indicators. Although the differences seem to be related to the business cycle, the shape of the indicators retain and no significant decrease in forecasting performance can be detected. Finally, Chapter 6 concludes with results of the preceding chapters and gives an outlook on future research questions.

Chapter 3

Sources of nonresponse in business surveys

Abstract It is well-known that nonresponse affects the results of surveys and can even cause biases due to selectivities if it cannot be regarded as missing at random. In contrast to household surveys, response behaviour in business surveys has been examined rarely in the literature. This chapter analyses a large business survey on microdata level for unit nonresponse. The data base is the Ifo Business Survey, which was established in 1949 and has about 7000 responding firms each month. The panel structure allows to use statistical modelling with the inclusion of different types of time dimensions as well as firm-specific effects. The results show that there are strong time-depending effects on the response rate and that nonresponse is more frequent in economic good times.

3.1 Introduction and motivation

The Ifo Business Survey is a monthly panel survey that has been conducted by the Ifo Institute first in 1949.[1] The IBS monitors German companies and collects data on different aspects of their business parameters, such as business situation, business expectations, demand situation or changes in staff (for an overview of the collected variables see Becker and Wohlrabe, 2008). The most well-known result of this survey is the *Ifo Business Climate Index*, one of the most prominent economic indicators for the German business cycle. Due to the fact that about 7000 respondents answer the questionnaire every month, the resulting indicators[2] have been proven to be very competitive for now- and forecasting the German economy, see for example Kholodilin and Siliverstovs (2006), Robinzonov and Wohlrabe (2010) or Drechsel and Scheufele (2010). In particular, the Ifo Business Climate Index (for the whole German economy) has a clear link with the growth rates of the German Gross Domestic Product (GDP), see Figure 3.1.

Due to the early availability of these macro level time series (in contrast to official data), these indicators were and still are strongly focused by econometricians, analysts, politicians and the general public. Abberger and Wohlrabe (2006) give an overview on the works with Ifo time series data. However, also the microdata sets have been analysed intensively, see Becker and Wohlrabe (2008) for an overview of the publications with Ifo microdata. In 2008, the Ifo Institute and the Ludwig-Maximilians-University of Munich founded the *LMU-ifo Economics & Business Data Center* to establish an easier access to the Ifo mirco data sets for reserachers. Since then, the microdata analyses have been increased, for example Bachmann et al. (2012), Pesaran and Timmermann (2009) and Rottmann and Wollmershäuser (2013).

Because of this increased usage of the Ifo microdata, it is important for all users to gain a precise insight into the survey process and possible selection problems due to unit nonresponse. In particular, the ques-

[1] Large parts of this paper base on C. Seiler (2010), 'Dynamic Modelling in Business Surveys', *Ifo Working Paper No. 93*.

[2] Because of the high number of observations, also indicators for sublevels, e.g. manufacturing of chemical products, are calculated.

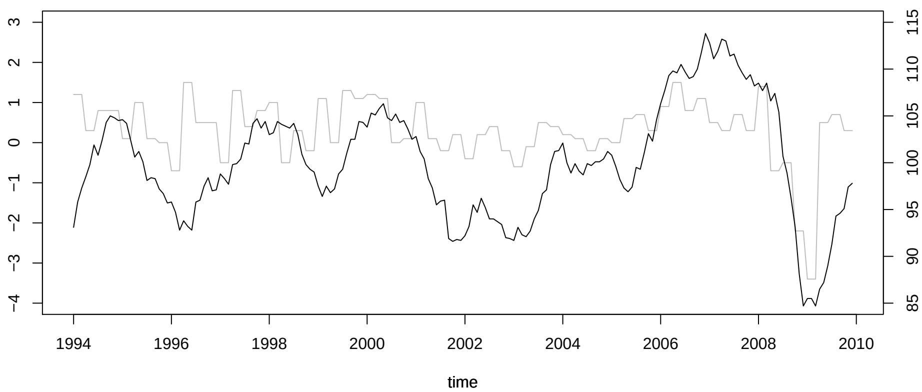

Figure 3.1: The Ifo Business Climate Index (black, left scale) and the growth rates of the German Gross Domestic Product (grey, right scale) 1994-2009.

tion arises if the nonresponding units in the IBS depend from the investigated variable, i.e. the business cycle, which can cause a nonresponse bias. To answer these questions, we give a detailed overview of the survey methodology and descriptive statistics in Section 3.2. In Section 3.3 we present a statistical model to explain effects on unit nonresponse in the IBS. Section 3.4 sums up the empirical findings and gives a short outlook.

3.2 The IBS data

3.2.1 Data collection

As noted in the previous section, the IBS was first conducted in 1949. Due to the transition of the German economy in the last 60 years, the structure of the participants in the IBS has changed. To preserve a stable number of participants, new companies were constantly asked to participate in the IBS. For this purpose, letters were sent with a request to participate and, if the company agreed, the firm was included into the monthly IBS. In most of the cases, a member of the head of the company (CEO, owner or board member) fills the questionnaire, see Abberger et al. (2009). To face the problem of self-selection, all macro level results in the IBS are weighted in two ways: First, by the companies' size (larger firms get a higher weight for the results) and second, by the added Gross value[3] of the appropriate subsector in accordance to the German Classification of Economic Activities (see Destatis (2008) for the most recent classification). In total, the share of German industrial production represented in the survey is about 40%. The construction sector is covered with 14%, the trade sector with 5% of total employment in Germany (Goldrian, 2007).

Although the survey was introduced in 1949, identification of single units is only possible since 1994. Therefore, we have to restrict our analysis to the period from January 1994 to December 2009. A specificity of the survey is that a single firm can answer more than one questionnaire if the company operates in various business areas which applies in particular to larger companies. A single unit in the IBS therefore is the enterprise as defined by the German official statistics. For each enterprise, the company is asked to fill a separate questionnaire which is done by different persons. We therefore assume that independence between two enterprises is given, even if they belong to the same company. For reasons of simplicity, in this paper each report is treated as coming from a different company, i.e. the number of 'respondents' ('companies', 'firms', 'enterprises', etc.) represents the number of sent questionnaires.

Because the firms respond on a voluntary basis, the data set is unbalanced, i.e. it is not recorded when a company was sent a mail but did not participate. To get the data set balanced, information about the first and the last month of intended survey participation is covered in an extra data base. With this information, months of non-participation can be reconstructed. Also complete drop-outs from the survey have been recorded: Companies either drop out because they do not exist any more, e.g. due to bankruptcy or takeover, or they are not interested any more in survey participation. For this reason, our analysis in Section 3.4 focuses on the effects on the response behaviour given that the firm does exist and in general is still interested in survey participation.

[3]Official statistics in Germany does weight their results in the same way, e.g. for the production statistics in manufacturing.

3.2.2 Descriptive analysis

In this section, we focus on the evaluation and analysis of unit and not item nonresponse, so we will provide descriptive statistics only for cases of unit nonresponse. Covering the period from January 1994 to December 2009, the total number of observations (including months of nonresponse) is 659,650 from 6822 enterprises in industry (with an average nonresponse rate of 14.5%), 204,318 from 3967 enterprises in construction (23.4%) and 277,256 from 4152 enterprises in trade (22.1%). The much lower nonresponse rate in industry maybe arises due to the focus of business cycle analysis on the manufacturing sector. For this reason, also the participants from this sector may be more motivated to participate.[4] Moreover, there is a significant spatial difference as can be seen in Figure 3.2. Firms from Eastern Germany have responded significantly less often than firms from Western Germany. However, this has been decreased throughout the years as shown in Figure 3.4. As the Eastern German firms have been included into the survey in 1991 after the German reunification, a habituation effect seems to prevail. In Table 3.1 we give an overview of the ratio of unit nonresponse for all variables which will later be used for the analysis and will be explained in detail in Section 3.3.1. Besides the differences in response behaviour mentioned above, it catches the eye that there exist huge differences according to the size of the company.[5] Figure 3.3 shows the nonresponse rates over time for the five different company sizes of the whole data set. It can clearly be seen that larger firms have a significant lower probability not to respond to the survey than smaller ones. Similar to the sector-specific effects, we presume that larger firms have greater interest to participate in the survey because they business is more affected by the general economic situation or they may have more staff and/or divisions to answer the survey regularly. We give a detailed explanation of all effects in Section 3.4.1.

[4] All participants receive results for their specific business area of the previous month. This information is much quicker available than official statistics.

[5] Notice that for the construction and manufacturing firms only the number of employees is available whereas for the trade firms only the yearly sales volume is recorded.

CATEGORY	VARIABLE	ABBREVIATION	VALUE	PERCENTAGE
Survey design	Number of questions	*questions*	-	
	Short time schedule in December	*short ts*	short time schedule	3.8
Business	Location	*east*	company located in Eastern Germany	24.7
	Size of the company	*size*		(see Table 3.2)
	Subsector of the company	*subsector*		(see Table 3.2)
External Environment	Business situation index in the (sub)sector	*business situation, bs*	-	
	Received an additional survey by Ifo	*add survey*	additional survey	8.0
Additional variables	Calendar time	*calendar time, ct*	-	
	Vacation days in the federal state	*vacation days*	-	
	Working days in the federal state	*working days*		-

Table 3.1: Description and distribution of non-sector specific variables

SECTOR	VARIABLE	ABBREVIATION	VALUE	PERCENTAGE
Industry	Size of the company (no. employees)	*size*	< 100 employees (smallest)	48.9
			100-199 employees (small)	17.7
			200-499 employees (medium)	18.1
			500-999 employees (large)	8.2
			>1,000 employees (largest)	7.2
	Subsector	*subsector*	Food & tobacco	6.4
			Textiles, textiles products & leather	7.1
			Wood	4.1
			Pulp, paper, publishing & printing	15.6
			Petroleum & chemical products	5.2
			Rubber & plastic products	6.8
			Other non-metallic mineral products	6.4
			Basic metals & fabricated metal products	12.3
			Machinery & equipment	16.0
			Electrical & optical equipment	12.2
			Transport equipment	2.9
			Furniture & manufacture n.e.c.	4.9
Construction	Size of the company (no. employees)	*size*	< 100 employees (smallest)	54.6
			100-199 employees (small)	24.0
			200-499 employees (medium)	12.0
			500-999 employees (large)	5.9
			>1,000 employees (largest)	3.6
Trade	Size of the company (annual sales volume)	*size*	< 1.0 million (smallest)	25.4
			1.0-5.0 million (small)	28.7
			5.0-12.5 million (medium)	18.6
			12.5-50.0 million (large)	20.4
			> 50.0 million (largest)	7.0
	Subsector	*subsector*	Sale, maintenance & repair of motor vehicles	10.4
			Wholesale trade	47.5
			Retail trade	42.1

Table 3.2: Description and distribution (by sector) of sector specific variables

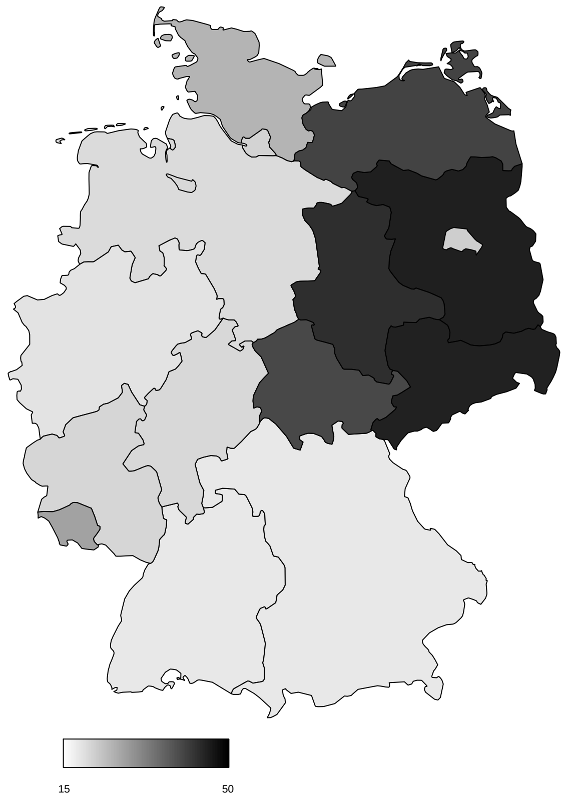

Figure 3.2: Nonresponse rates in percentage for the period 1994-2009 according to the German federal states in the IBS.

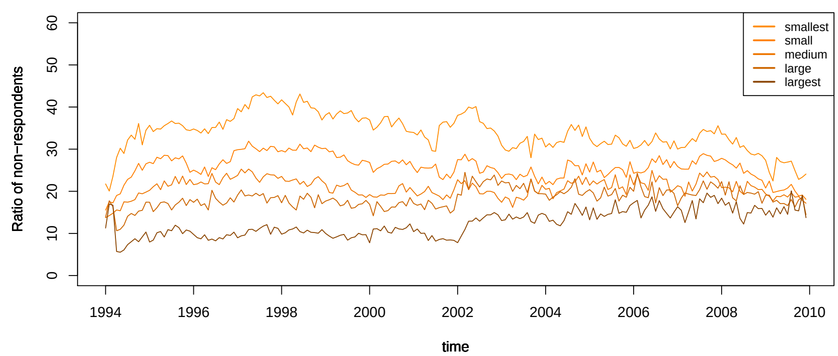

Figure 3.3: Nonresponse rates according to the companys' size in the IBS.

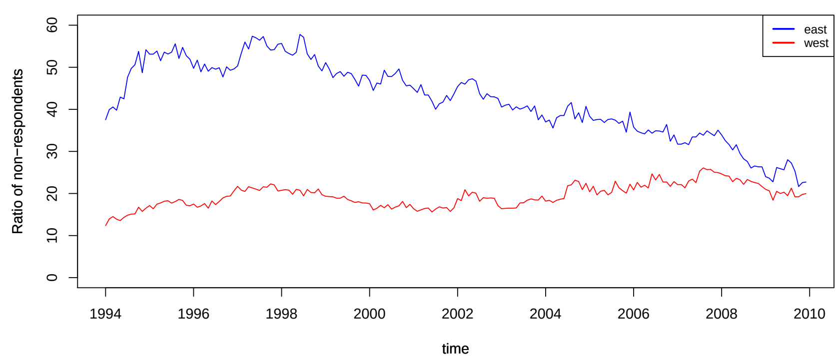

Figure 3.4: Nonresponse rates for Eastern and Western German firms in the IBS.

3.3 Explaining unit nonresponse

3.3.1 Variables

There are many 'risk factors' that may influence the response behaviour in business surveys. In this chapter, we categorise them in accordance to the conceptual framework of Willimack et al. (2002), see Figure A.1. This framework distinguishes two major categories of variables: First, factors which are under the control of the researcher, related to survey design (time schedule, instrument design, etc.) and second, factors out of researchers control. The latter can be divided into three groups: External environment (such as 'survey taking climate' and economic conditions), the business (characteristics, organisational structure) and finally the attributes of the respondent (authority, motivation). Based on this framework, it will be discussed which of these variables can be incorporated into the analysis and which additional variables will be included that cannot be classified into one of these categories. All variables which enter the final model in Section 3.4 are listed in Tables 3.1 and 3.2.

Survey design

Since the manufacturing sector can be regarded as the 'cycle maker' of the German economy, the IBS was first introduced in 1949 in industry. The extension of the survey to other sectors was carried out 1950 in trade and 1956 in construction. Due to the different structure of these sectors, the questionnaires are not identical, e.g. the question to the capacity utilization is not meaningful answerable for the trade companies. However, the questionnaire for each sector has undergone very minor changes. One of these small changes concerned the number of questions which consists of standard and special questions. The latter are asked each quarter, half year or once a year. A major change, which affected the level of content of the questionnaire, was in January 2002 when the survey was reorganised for the *Joint Harmonised EU Programme of Business and Consumer Surveys* (for more information see European Union, 2006). Before 2002, all questions asked in month t collected information on data from the prior reporting month $t-1$. This change has affected the content only marginally but clearly has implications for the time schedule. Since January 2002, potential respondents are asked to provide information from the current month t. This is a problem in December when the survey results have to be published five days before Christmas instead of five days before months' end. In our analysis, a dummy variable for short time schedule is introduced which indicates all Decembers since 2002. Recently, the number of days to answer the questionnaire would be interesting, but these data are only available since 2003. In order to avoid a strong reduction of the data set, this information will not be used in the analyses.

The Business

To control for effects of business characteristics, the size of the company and the subsector the company is working are included in the regression analysis. For the construction firms, controlling for different response behaviour across the subsectors is not possible because the companies report for all working areas in one questionnaire. In order to account for structural differences between the sectors, several weighting characteristics for the aggregation of the indicators are recorded in the survey: Firms from industry and construction are categorised by the number of employees whereas trade companies by their annual sales volume. Notice that this information is collected for the different subentities of the business and is updated once a year. However, it is likely that there are only minor changes within a year, so that this low frequency should be negligible. As we saw in Figure 3.4 that there exists a strong (and time-varying) difference between Eastern and Western German firms, we control for these time-varying effects in our regression model.

The Respondent

Tomaskovic-Devey et al. (1994) pointed out that the authority of the respondent is important for the answering behaviour. For the IBS, characteristics of the respondent, such as gender, age and position in the company are not available, even not on annual frequency. Abberger et al. (2009) undertook a meta survey directed to this question in spring 2009 with respect to trade firms. Since this was an one-time survey these data were not merged with the IBS panel; in particular, no information for older firms is available. Therefore, an authority variable cannot be included into the data analysis. The same applies to capacity and motivation of the respondent. However, in Section 3.3.3 we will show how we can reflect this firm-specific heterogeneity to a certain extent.

External Environment

An external aspect of response behaviour are economic conditions prevailing at the time of the survey. Harris-Kojetin and Tucker (1999) found lower cooperation in a population survey in periods of economic better times. As the IBS focuses on economic parameters of the companies, there is a variety of possible indicators for the current economic situation of a single firm. But obviously, there are no answers available in months of non-participation. Instead of this, economic indicators taken from the survey results can be used. The Ifo Institute computes business situation indicators for each (sub)subsector, so the indicators from the lowest available aggregation level (where each firm is classified to) are used as an approximation of the business situation of the single firm in the appropriate (sub)subsector. This approach can be problematic because these indicators are aggregated results from the participating subjects. Still it allows a deeper insight into possible selectivities related to the business cycle. If, in fact, the response behaviour depends on the business cycle, nonresponses depend from the investigated latent variable and thus, estimates can be biased. As mentioned above, there is no data for the subsectors of construction, so the indicator for the whole sector is integrated into the model. To validate our results, we also run a regression model containing the GDP growth rates of Germany, which are to be forecasted with the IBS results as described in Section 3.1.

Groves et al. (2004) mentioned that the intensity of survey research can be a reason for nonresponse. This 'survey taking climate' can be affected by the number of requests for survey participation the company receives each month. Lacking data about the total number of requests, there exists information about additional surveys conducted by the Ifo Institute, i.e. if the company received an extra questionnaire in a given month. Also the number of questions can be interpreted as an indicator for increasing intensity of survey research.

Additional variables

Several studies found evidence for declining interest in survey participation over the last decades (for an overview see de Leeuw and de Heer, 2002). Brehm (1994) pointed out that all institutions that organise surveys (academic, governmental, business and media) suffer from declining response rates. Therefore, the variable *calendar time* is included into the model, counting months since January 1994 (i.e. 1 for 01/1994, 2 for 01/1994, ..., 192 for 12/2009). This variable allows to control for general trends in response behaviour between 1994 and 2009. Besides calendar time, the *length of participation* in months t is available for all units, i.e. it represents the t-th month the company received a questionnaire. However, the reader has to keep in mind that the first month of participation is available (and makes it possible to calculate the exact participation month) even for all units which are left-censored due to the missing IDs before January 1994. As the IBS was established in 1949, there are still active companies which obtained more than 700 participation months. Another problem to face is the difference of vacation and working days, which speaks to the number of available days to respond. As the number of vacation days differ significantly between the German states, we include both variables into the analysis.

3.3.2 The statistical model

All variables described above have a panel structure, so the data set has the form $(y_{i,t}, \boldsymbol{x}_{i,t}), i = 1, \ldots, n$ and $t = 1, \ldots, T$, where $n = 14,941$ denotes the number of companies and $T = 192$ the waves of the survey since January 1994. Given that the dependent variable is an 1/0-dummy, $y_{i,t} = 1$ if company i did not answer the questionnaire in the t-th wave since January 1994 and $y_{i,t} = 0$ if it was observed in the data. The mean function $\pi_{i,t} = \mathrm{E}(y_{i,t})$ can be written as a *Generalized Linear Model* (GLM)

$$g(\pi_{i,t}) = \beta_0 + \boldsymbol{x}_{i,t}\boldsymbol{\beta} \tag{3.1}$$

with and an appropriate link function $g(\cdot)$, such as logit or probit, and a $(n \times K)$-matrix $\boldsymbol{x}_{i,t}$, with $K = 30$ as the total number of variables.

3.3.3 Unobserved correlation

Generalized Estimation Equation

As we analyse panel data, $y_{i,t}$ may be correlated over t so the i.i.d.-assumption could be violated. The probability for a missing given that a missing occurred in the previous month is 56%. This means that the missing structure in the data is relatively strongly correlated. To correct for such effects, we extend Equation (3.1) to a marginal model by using the *Generalized Estimation Equation* (GEE) approach developed by Liang and Zeger (1986). GEEs are part of the wide range of quasi-likelihood methods which were introduced first by Wedderburn (1974). Quasi-likelihood methods only require a given relationship between the dependent and the independent variable and the relation of the conditional mean and the variance of y. Therefore, the mean function in GEEs can be defined as in a GLM, i.e. of form (3.1). The variances $\mathrm{Var}(y_{i,t})$ have to be written as a function of the mean μ_i, i.e.

$$\mathrm{Var}(y_{i,t}) = \phi\, v(\boldsymbol{\mu}_i)$$

where ϕ is a common scale parameter and $v(\cdot)$ the known variance function. To obtain estimates for the slope parameters β, K 'quasi-score' functions

$$S_k(\boldsymbol{\beta}) = \sum_{i=1}^{n} \frac{\partial \boldsymbol{\mu}_i'}{\partial \boldsymbol{\beta}_k} \mathrm{Cov}(\boldsymbol{y}_i)^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_i) = 0, \qquad k = 1, \ldots, K,$$

have to be solved. If the model is correctly identified, $\mathrm{E}[S_k(\boldsymbol{\beta})] = 0$ and $\mathrm{Cov}[S_k(\boldsymbol{\beta})] = \frac{\partial \boldsymbol{\mu}_i'}{\partial \boldsymbol{\beta}_k} \mathrm{Cov}(\boldsymbol{y}_i)^{-1} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}_k'}, \forall k$. In cases of panel data, the form of the dependence across t has to be pretended. This is done by a specification of a $(T \times T)$ working correlation matrix $R_i(\boldsymbol{\alpha})$ which is completely described by $\boldsymbol{\alpha}$. Then,

$$\mathrm{Cov}(\boldsymbol{y}_i) = \phi\, \boldsymbol{V}_i^{1/2} \boldsymbol{R}_i(\boldsymbol{\alpha}) \boldsymbol{V}_i^{1/2}$$

is the corresponding working covariance matrix of $\boldsymbol{y}_i$ with $\boldsymbol{V}_i = \operatorname{diag}(v(\boldsymbol{\mu}_i))$ and $\dim(\boldsymbol{V}_i) = T \times T$, see Heagerty and Zeger (1996). Common working correlation matrices, in particular for small data sets, are $\boldsymbol{R}_i(\boldsymbol{\alpha}) = \boldsymbol{\alpha}$ ('exchangeable') or $\boldsymbol{R}_i(\boldsymbol{\alpha}) = \boldsymbol{\alpha}^{|t-s|}$ ('autoregressive', here AR(1)) $\forall t \neq s; s,t \in 1,\ldots,T$ since only one parameter $\boldsymbol{\alpha}$ has to be estimated.

Notice that the working correlation has to be specified properly to enable consistent estimates of $\operatorname{Var}(\hat{\boldsymbol{\beta}})$. A robust variant was proposed by Liang and Zeger (1986) by using a so-called Huber-White sandwich estimator (see Huber, 1967 and White, 1982):

$$\operatorname{Var}(\hat{\boldsymbol{\beta}}_k) = n\left(\sum_{i=1}^{n} \hat{\boldsymbol{D}}_i' \hat{\boldsymbol{C}}_i^{-1} \hat{\boldsymbol{D}}_i\right)^{-1} \left(\sum_{i=1}^{n} \hat{\boldsymbol{D}}_i' \hat{\boldsymbol{C}}_i^{-1} \boldsymbol{W}_i \hat{\boldsymbol{C}}_i^{-1} \hat{\boldsymbol{D}}_i\right) \left(\sum_{i=1}^{n} \hat{\boldsymbol{D}}_i' \hat{\boldsymbol{C}}_i^{-1} \hat{\boldsymbol{D}}_i\right)^{-1}$$

with $\hat{\boldsymbol{D}}_i' = \partial \boldsymbol{\mu}_i' / \partial \hat{\boldsymbol{\beta}}_k$, $\hat{\boldsymbol{C}}_i = \widehat{\operatorname{Cov}}(\boldsymbol{y}_i)$ and $\boldsymbol{W}_i = (\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_i)(\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_i)'$ as the empirical covariance estimator. This robust estimate is consistent even under misspecification of the correlation matrix and therefore widely used in literature, see Zorn (2001).

Fixed Effects model

Another possibility to account for unobserved heterogeneity are *fixed effects* models (FE models) on company level. Considering Equation (3.1), a FE model turns to

$$g(\pi_{i,t}) = \beta_0 + \boldsymbol{x}_{i,t}\boldsymbol{\beta} + \gamma_i \tag{3.2}$$

with individual-specific, time-invariant effects γ_i. In our case, a fixed effects model is preferred compared with a random effects model as the individual-specific effects γ_i might be correlated with the covariates $\boldsymbol{x}_{i,t}$, in particular the companies' size and the subsector. These do not change very often within each company.

3.3.4 Unit weighting

Our models specified above imply that all units in the data set have the same probability to enter the survey. A problem arises due to the fact that the probability weights for the inclusion in the survey are unknown. In addition, these weights may have changed over the time as the data set covers 16 years. A possible solution for this problem is to evaluate the number of enterprises in the appropriate year in Germany. The German Statistical Office provides annual numbers of enterprises in the appropriate subsector on 3-digit level which can be interpreted as strata weights $\omega_i^{subsector}$. To evaluate the stability of the results, we run our analyses also including these weights. As the estimation models force to define constant weights for each unit, the weights from the year in which the appropriate unit has entered the survey is used if known. In other cases, the weights of the year of the first appearance in the data are used.

3.4 Results and discussion

All variables described in Section 3.3.1 and listed in Tables 3.1 and 3.2 are potential factors that may influence the response behaviour. They enter the model as follows:

$$\begin{aligned}\eta_t \;=\;& \beta_0 + t\,\beta_t + \mathit{calendar\ time}\,\beta_{ct} + (\mathit{calendar\ time} \times \mathit{east})\,\beta_{ct\times east}\\ &+\mathit{east}\,\beta_{east} + \mathit{size}\,\boldsymbol{\beta}_{size} + \mathit{subsector}\,\boldsymbol{\beta}_{subsector}\\ &+\mathit{short\ ts}\,\beta_{short\ ts} + \mathit{vacation\ days}\,\beta_{vac\ days} + \mathit{working\ days}\,\beta_{work\ days}\\ &+\mathit{add\ survey}\,\beta_{add\ survey} + \mathit{questions}\,\beta_{questions} + \mathit{cycle\ indicator}\,\beta_{cycle}\end{aligned}$$

with a logit link function and potential additional individual-specific effects γ_i. Notice that $\boldsymbol{\beta}_{size}$ and $\boldsymbol{\beta}_{subsector}$ are vectors and the reference category for $\boldsymbol{\beta}_{subsector}$ is the construction sector. The interaction term $(\mathit{calendar\ time} \times \mathit{east})$ is included into the model as we saw in Section 3.2.2 that the response behaviour differs strongly between Eastern and Western firms over calendar time. As defined in Section 3.3.1 *cycle indicator* represents two different indicators: The lowest business situation indicator from the survey results and the GDP growth rates in Germany.

3.4.1 Interpretation of the results

The results for the GEE model are shown in Table 3.3 (model without weights) and 3.4 (model including weights). The appropriate results of the FE model are listed in Appendix A.2 in Tables A.1 (model without weights) and A.2 (model including weights). As the estimated effects do not differ to much between the models, we present the unweighted GEE in detail and discuss the differences afterwards. The results in Table 3.3 show that with rising participation time, the respondents more often answer the survey. Our results confirm those in Janik and Kohaut (2012), who also examine the response behaviour of German companies, but do not model time-dynamic effects since they only use the 2006 data from the IAB Establishment Panel. It can be supposed that in panel surveys, companies need some time until the collection of information (in which maybe various departments are involved) becomes regular. As different studies mentioned in Section 3.3.1, we also find evidence for a general declining trend in participation (see the coefficient for *calendar time*). However, our analysis shows that the willingness to participate has increased for the Eastern German firms. Still, this effect can be interpreted that there is a transition period when an existing panel is introduced into a new region and the survey has to become established with time. At this point, it should also be noted that the interaction term *calendar time* $\times$ *east* is necessary to include into the model as in these cases the main effect *calendar time* would change the sign.

With exception of the number of working days and the number of questions, all 'survey design related' variables show the supposed effects. However, the number of working days only have less variation and thus the 95% confidence interval includes the 0. Sending an additional survey to the respondents seems to increase the probability for nonresponse, but the effect is not significant at the 10% level. In contrast, an increasing number of vacation days reduces the willingness to participate with certain. It can be assumed that the respondent is more likely not in office in the holidays' season and therefore has less time to fill the questionnaire. Also, the short time schedule of the IBS in December since 2002 has a negative impact on the response rates.

The response behaviour also varies for different business' sizes: Basically, larger firms tend more likely to respond than smaller ones. Although organisational performance generally rises with the size of the company, we suppose that they may benefit more from the survey results than the smaller firms and therefore are more willing to respond regularly as presumed in Section 3.2.2. For the business areas we find different effects: The estimation results confirm that the trade companies have a significant higher probability not to respond than the construction firms, as in Figure 3.3. In addition, companies from the manufacturing sector mostly have a smaller probability for unit nonresponse. In the case of trade companies can be assumed that the topic of the survey (and their results) is not as interesting because the trade sector generally does not dependent on the business cycle so strongly than the other sectors.

After controlling for survey related and business specific effects, it can be seen that in economic good times the firms tend more to nonresponse which confirms the results of Harris-Kojetin and Tucker (1999) for household surveys. This effect holds when using the survey indicators as well as the GDP growth rates. We assume that this effect is driven by less times to answer the questionnaire in boom times because of many orders. Willimack and Nichols (2010) mention that for the respondent the 'priority is given to activities required to keep the business open and growing'.[6] So, filling the questionnaire might lose priority when the business situation becomes better. This can, but not has to, be a possible source of bias.

In general, our estimated effects do not differ substantially with respect to the assignment of weights as well as the usage of a fixed effects model. Table 3.4 shows the estimation results of the weighted GEE model. The effects mostly stay the same but some variables become insignificant. However, the (positive) effect of the business cycle on the nonresponse behaviour remains. In addition, the effects of the different time dimensions as well as the East/West-effect stays significant. The results of the fixed effects model in Tables A.1 and A.2 show a similar picture:[7] Again, the relationship between the business cycle and the response behaviour is positive. The effect of *size* has a similar pattern as in the GEE case. Larger firms tend more often not to respond to the survey than smaller ones (with

[6] Notice that these results are based on the evaluation of large firms.

[7] Notice that the intercept is excluded in this table as STATA automatically incorporates it into the individual-specific term.

exception of the largest firms, but this effect is also not significant in the GEE case). As in the GEE model, trade firms do have a significant higher probability not to respond to the survey.

3.5 Summary

In this chapter, (unit) nonresponse behaviour in the Ifo Business Survey was estimated with statistical models including the possibility to account for unobserved heterogeneity. The analysis shows that the risk for nonresponse decreases over participation time. Considering the framework of Willimack et al. (2002) and the magnitudes of the estimated effects, the main reasons for different response behaviour are among the business' characteristics since major differences were found across economic sectors and larger firms tend less to nonresponse than smaller ones. Survey characteristics, e.g. if an additional survey was sent to the firms or if the time schedule is short, seem to play a minor role in the participation process. After controlling for these 'survey design related' effects, the willingness to participate also depends to a small extent on the overall economic situation. In particular, in economic good times the companies respond less often. Since the IBS focuses on evaluating the state of the business cycle, this result can be critical in terms of biases. Although the results obtained here indicate a rather small bias as the effect is rather small, a simulation study can help to assess the magnitude of a possible nonresponse bias on the indicator which will be done in the next chapter.

Variable	Business Sit. coef.	Business Sit. p-value	GDP growth coef.	GDP growth p-value
Intercept	-3.071	0.000	-3.105	0.000
Participation time	-0.002	0.000	-0.002	0.000
Calendar time	0.003	0.000	0.003	0.000
Calendar time $\times$ East	-0.007	0.000	-0.007	0.000
East	1.373	0.000	1.382	0.000
Cycle indicator	0.001	0.000	0.028	0.000
Additional survey	0.010	0.150	0.009	0.222
Number of questions	0.000	0.966	0.000	0.648
Short time schedule	0.058	0.000	0.060	0.000
Working days	0.000	0.912	0.000	0.850
Vacation days	0.002	0.000	0.002	0.000
Size:				
Smallest	0.238	0.000	0.237	0.000
Small	0.088	0.016	0.087	0.018
Large	-0.109	0.013	-0.113	0.010
Largest	-0.045	0.458	-0.044	0.470
Subsector:				
Food & tobacco	-0.225	0.042	-0.231	0.037
Textiles & textiles products	-0.328	0.009	-0.335	0.007
Wood	-0.312	0.079	-0.365	0.038
Pulp, paper, publishing & printing	-0.238	0.065	-0.254	0.049
Petroleum & chemical products	-0.152	0.172	-0.135	0.227
Rubber & plastic products	0.092	0.394	0.085	0.433
Other non-metallic mineral products	-0.129	0.236	-0.144	0.184
Basic metals & fabricated metal products	0.136	0.322	0.129	0.346
Machinery & equipment	-0.242	0.071	-0.237	0.077
Electrical & optical equipment	-0.200	0.035	-0.203	0.034
Transport equipment	-0.222	0.337	-0.225	0.333
Furniture & manufacture n.e.c.	0.079	0.486	0.061	0.594
Sale, maintenance & repair of motor vehicles	0.324	0.002	0.291	0.003
Wholesale trade	0.239	0.001	0.218	0.002
Retail trade	0.279	0.000	0.258	0.000

Table 3.3: Estimation results of the unweighted GEE model

VARIABLE	BUSINESS SIT. COEF.	BUSINESS SIT. P-VALUE	GDP GROWTH COEF.	GDP GROWTH P-VALUE
Intercept	-3.979	0.000	-3.997	0.000
Participation time	-0.003	0.000	-0.003	0.000
Calendar time	0.003	0.000	0.004	0.000
Calendar time $\times$ East	-0.007	0.000	-0.007	0.000
East	1.755	0.000	1.759	0.000
Cycle indicator	0.001	0.147	0.020	0.045
Additional survey	0.012	0.483	0.011	0.491
Number of questions	0.004	0.000	0.004	0.000
Short time schedule	0.083	0.003	0.085	0.002
Working days	0.007	0.064	0.007	0.065
Vacation days	0.004	0.000	0.004	0.000
Size:				
Smallest	0.033	0.756	0.033	0.757
Small	0.047	0.539	0.047	0.542
Large	-0.182	0.112	-0.181	0.112
Largest	-0.061	0.610	-0.060	0.618
Subsector				
Food & tobacco	0.056	0.814	0.055	0.820
Textiles & textiles products	-0.195	0.481	-0.205	0.461
Wood	0.244	0.332	0.219	0.386
Pulp, paper, publishing & printing	-0.361	0.380	-0.358	0.385
Petroleum & chemical products	0.046	0.846	0.056	0.811
Rubber & plastic products	0.318	0.202	0.316	0.206
Other non-metallic mineral products	0.194	0.428	0.187	0.448
Basic metals & fabricated metal products	0.499	0.061	0.497	0.063
Machinery & equipment	0.092	0.726	0.097	0.712
Electrical & optical equipment	0.163	0.487	0.162	0.489
Transport equipment	0.559	0.444	0.560	0.446
Furniture & manufacture n.e.c.	1.836	0.000	1.818	0.000
Sale, maintenance and repair of motor vehicles	0.440	0.100	0.439	0.101
Wholesale trade	0.480	0.065	0.468	0.074
Retail trade	0.510	0.049	0.502	0.054

Table 3.4: Estimation results of the weighted GEE model

Chapter 4

Theoretical considerations on missing data patterns in business tendency surveys

Abstract Business cycle indicators based on the balance statistics are a widely used method for monitoring the recent economic situation. In contrast to official data, indicators from business surveys are available early and typically not revised after their initial publication. But as surveys can be in general affected by distortions arising from the response behaviour, these indicators can also be biased. In addition, time-dependent nonresponse patterns can produce even more complex forms of biased results. This chapter examines a framework which kind of nonresponse patterns lead to biases and decreases in performance. An extensive simulation study is performed to analyse their effects on the indicators. Our analyses show that these indicators are extremely stable towards nonresponse biases.

4.1 Introduction

Monitoring and forecasting of economic activity is nowadays high on the agenda of both public and private institutions.[1] As official data are commonly released with a long delay, timely business cycle indicators are needed. Balance statistics indicators based on survey data play a central role in this context. Their major advantages are timely availability and the fact that they are subject to almost no revisions. Indeed, aggregated business survey data has proved to be one of the most competitive indicators for analysing macroeconomic variables, e.g. the Consumer Confidence Index in the United States (Ang et al., 2007), the Economic sentiment indicator for the European Union (Gayer, 2005) or the Ifo Business Climate Index for Germany (Drechsel and Scheufele, 2010, Kholodilin and Siliverstovs, 2006 and Robinzonov and Wohlrabe, 2010)

Most papers using survey based business cycle indicators (usually not explicitly) assume that the indicator is measured without any nonresponse bias, i.e. the survey results are not affected by any nonresponse bias. While a large body of literature exists concerning nonresponse in household or individual surveys, less is known about the processes and reasons for participation and responding behaviour in business surveys (see Janik and Kohaut, 2012), particularly in the case of surveys aimed at evaluating the business cycle. Therefore, the indicators cannot generally be said to be unbiased. Chapter 3 gives evidence for a dependence of the responding behaviour in the Ifo Business Survey from the business cycle which confirmed the results of Harris-Kojetin and Tucker (1999) for population surveys. So, the question arises how a potential nonresponse bias affects the results of these type of business surveys. How do indicators look when they are biased by systematic nonresponse patterns? How strongly is the performance reduced?

To answer these questions, this chapter is organised as follows: In Section 4.2 we define the methodologic framework for the calculation of the balance statistics and show how nonresponse biases affect the indicators. We subsequently introduce measurements to explore the magnitude of the biases. In Section 4.3, we perform an extensive simulation study for a wide

[1] Large parts of this paper base on C. Seiler (2012), 'On the Robustness of the Balance Statistics with respect to Nonresponse', *Ifo Working Paper No. 126*.

variety of nonresponse patterns and different types of business cycles. We show that the bias is minimal even for very strong bias structures in the data. Finally, Section 4.4 sums up the results.

4.2 Methodological framework

In market economies, one can typically observe that general long-term growth is accompanied by temporary fluctuations. In general, the classical business cycle is divided into four phases: upswing, boom, downswing and recession. In its ideal form, the cycle would have a sinus-like shape. The business cycle itself is a theoretical construct which can not be measured directly, but can be visualised by observing economic indicators, such as GDP growth rates or production in the manufacturing sector. We can think of a business cycle $g(t)$ as a stationary function in time, e.g. $g(t) = \sin(t)$. Due to the fact that official data are published with a long delay (and often revised after being first released too), business cycle tendency surveys can monitor the recent economic situation considerably more quickly. The Ifo Institute for Economic Research was one of the first institutes to conduct such surveys and this method has been widely accepted in the OECD countries, for an overview see OECD (2003). In line with the Joint Harmonised EU Programme of Business and Consumer Surveys (see European Union, 2006), the indicators base on two variables (business situation and business expectations) which are measured on a 3-level Likert scale representing a good, equal or bad state, i.e. $s \in S = \{+, =, -\}$. Due to the construction of the questions in the questionnaire, the indicators obtained actually measure the business cycle without trend, see OECD (2003). For more theoretical aspects of the balance statistics see Anderson (1951, 1952) and Theil (1952).

For the current study, we abstract from any formal definition of the business cycle. We define a macroeconomic time series for which a qualitative survey with a 3-level Likert scale is constructed or referenced to. Usually, a variable on a 3-level Likert scale is influenced by an unobserved process (e.g. the business cycle in our case). However, continuous variables may also be surveyed on a 3-level Likert scale, realisations of which are only available on a very delayed basis, e.g. the World Economic Survey of the Ifo Institute asks for inflation in the appropriate country (see Stangl, 2007). In this Section, we leave the exact shape of $g(t)$ undefined, but show the effects for different types of $g(t)$ in Section 4.3.

4.2.1 Construction of the balance statistics

All surveys asking on a 3-level Likert scale mentioned above calculate a so-called *balance statistics* after data collection. The balance statistics is defined as the fraction of positive answers subtracted by the fraction of negative answers for a certain variable. We assume that respondent i is affected in his opinion formation by the cycle function $g(t)$, for which the survey is conducted, and an individual error term ϵ_i:

$$s^*_{i,t} = g(t) + \epsilon_i, \qquad \epsilon_i \sim N(0, \sigma^2), \tag{4.1}$$

$i = 1, \ldots n, t = 1, \ldots, T$ with

$$\mathrm{E}_t(s^*_{i,t}) = g(t), \qquad \forall i, \tag{4.2}$$

where $\mathrm{E}_t(\cdot)$ denotes the expected value at time t. As each survey participant i is restricted to give answers on a 3-level Likert scale, we observe

$$s_{i,t} = \begin{cases} + & \text{if} \quad s^*_{i,t} > \tau^+ \\ = & \text{if} \quad \tau^+ \geq s^*_{i,t} > \tau^- \\ - & \text{if} \quad \tau^- \geq s^*_{i,t} \end{cases}$$

where τ^s define time-independent thresholds.[2] This is the standard model for latent variable models (Bartholomew and Knott, 1999). The assumption of time- and individual-independent thresholds as well as time- and individual-independent normality of ϵ_i is required for the calculation of the balance statistics. In addition, τ^s are assumed to be symmetric around zero, i.e. $-\tau^- = \tau^+$ (see Wollmershäuser and Henzel, 2005). Therefore, the balance statistics is a special case according to the Carlson-Parkin framework (Carlson and Parkin, 1975) which allows for differences for thresholds and error terms in time and across individuals. These very

[2] $\tau^=$ leaves undefined.

strong assumptions have been criticised by several papers, see Nardo (2003) for an overview. However, as these thresholds are 'contained' in the cumulative distribution function of ϵ_i, we do not go more into detail as this is less of an issue in terms of nonresponse. For a discussion of the shifts of thresholds τ^s see Stangl (2009). The same applies to the distribution of ϵ_i which can be generalised to a wider class of distributions. Due to the fact that these extensions would affect both the biased indicator and the unbiased indicator in the same way, and that this analysis focuses on a relative comparison between unbiased and biased indicators, we consider τ^s and $\epsilon_i := \epsilon$ as independent from time and individuals.

To enable the calculation of means, we define $'+' \equiv 1$, $'=' \equiv 0$ and $'-' \equiv -1$. Then, the expected mean of $s_{i,t}$ for given t is defined as $\mathrm{E}_t(s_{i,t}) = \sum_{s=-1}^{1} s \cdot P(s_{i,t} = s)$. With $\mathrm{E}_t(s) = \mathrm{E}_t(s_{i,t})$ we follow

$$
\begin{aligned}
\mathrm{E}_t(s) = \mathrm{E}_t(s_{i,t}) &= \sum_{s=-1}^{1} s \cdot P(s_{i,t} = s) \\
&= 1 \cdot P(s^*_{i,t} > \tau^+) + (-1) \cdot P(s^*_{i,t} \leq \tau^-) \\
&= P(g(t) + \epsilon_i > \tau^+) - P(g(t) + \epsilon_i \leq \tau^-) \\
&= P(\epsilon_i > \tau^+ - g(t)) - P(\epsilon_i \leq \tau^- - g(t)) \\
&= [1 - P(\epsilon_i < \tau^+ - g(t))] - P(\epsilon_i \leq \tau^- - g(t)) \\
&= \left[1 - \Phi\left(\frac{\tau^+ - g(t)}{\sigma}\right)\right] - \Phi\left(\frac{\tau^- - g(t)}{\sigma}\right) \\
&= \Phi^+(t) - \Phi^-(t)
\end{aligned}
\tag{4.3}
$$

with $\Phi(\cdot)$ as the cumulative distribution function of the standard normal distribution.[3] $\Phi^+(t) := 1 - \Phi((\tau^+ - g(t))/\sigma)$ denotes the probability for a positive reply whereas $\Phi^-(t) := \Phi((\tau^- - g(t))/\sigma)$ the probability for a negative reply. $\mathrm{E}_t(s) = \mathrm{E}_t(s_{i,t})$ in Equation (4.3) is given since $\left[1 - \Phi\left(\frac{\tau^+ - g(t)}{\sigma}\right)\right] - \Phi\left(\frac{\tau^- - g(t)}{\sigma}\right)$ includes no individual information i. In the following, we call $\mathrm{E}_t(s)$ the unbiased mean function or the unbiased indicator. In Appendix B.1 we show that (4.3) is the same as the balance statistics for a 3-level Likert scale.

[3] We abstract from different types of weighting that may occur in real data sets.

4.2.2 Inclusion of response decision

The mean function $\mathrm{E}_t(s)$ in Equation (4.3) is the expected function in time for $n \to \infty$ when no nonresponse bias is present. To include the decision of the respondent to participate in the survey, we define a binary variable m which indicates this decision, so that $m_{i,t} = 1$ if the observation i is missing at time t and $m_{i,t} = 0$ if not. According to Little and Rubin (2002), not missing at random, i.e. a nonresponse bias, occurs if the probability for missing $P(m_{i,t} = 1)$ depends on the outcome of the variable of interest s, i.e. $P(m_{i,t} = 1|s_{i,t} = s, \xi)$, with ξ as additional parameters which are unrelated to the variable of interest (including t in our case). Therefore, the missing process would depend on the state s. To evaluate the effects of the bias patterns on the balance statistics, we assume that each respondent has the same probability not to respond to the survey for given s (and t), i.e. $P(m_{i,t} = 1|s, \xi) = P(m = 1|s), \forall i$. Then, $\pi^s(t) = P(m_{i,t} = 0|s_{i,t} = s)$ is the (not necessarily time-dependent) probability to be observed at time t being in state s. Therefore, the mean function of the *observed units* $s_{i,t}^{obs}$ is given by

$$\begin{aligned}
\mathrm{E}_t(s^{obs}) = \mathrm{E}_t(s_{i,t}^{obs}) &= \sum_{s=-1}^{1} s \cdot P(s_{i,t} = s|m_{i,t} = 0) \qquad (4.4)\\
&= \sum_{s=-1}^{1} s \cdot \frac{P(m_{i,t} = 0|s_{i,t} = s) \cdot P(s_{i,t} = s)}{P(m_{i,t} = 0)}\\
&= \sum_{s=-1}^{1} s \cdot \frac{P(m_{i,t} = 0|s_{i,t} = s) \cdot P(s_{i,t} = s)}{\sum_{s=-1}^{1} P(m_{i,t} = 0|s_{i,t} = s) \cdot P(s_{i,t} = s)}\\
&= \frac{\pi^+(t) \cdot [1 - \Phi(\frac{\tau^+ - g(t)}{\sigma})]}{\sum_{s=-1}^{1} \pi^s(t) \cdot \Phi^s(t)} - \frac{\pi^-(t) \cdot \Phi(\frac{\tau^- - g(t)}{\sigma})}{\sum_{s=-1}^{1} \pi^s(t) \cdot \Phi^s(t)}\\
&= \frac{\pi^+(t)}{\bar{\pi}(t)} \cdot \Phi^+(t) - \frac{\pi^-(t)}{\bar{\pi}(t)} \cdot \Phi^-(t)
\end{aligned}$$

with $\bar{\pi}(t) := \sum_{s=-1}^{1} \pi^s(t) \cdot \Phi^s(t)$, $\Phi^=(t) := 1 - [\Phi^+(t) + \Phi^-(t)]$ and $0 \leq \pi^s(t) \leq 1$. Note that $\pi^s(t)$ is the probability to answer ('acceptance rate') being in state s at time t and $\Phi^+(t)$ and $\Phi^-(t)$ directly depend on $g(t)$ and beyond only on time-invariant variables. A comparison of Equation (4.3) and (4.4) clearly shows that $E_t(s^{obs})$ is the *weighted balance statistics* with weights $\frac{\pi^s(t)}{\bar{\pi}(t)}$, i.e. the relation between the acceptance rate $\pi^s(t)$ of state s and the sum $\bar{\pi}(t)$ of all (weighted) acceptance rates at time t. Therefore, NMAR occurs in cases of $\exists t \in \{1, \ldots, T\} \cap \exists (r, s) : \pi^r(t) \neq \pi^s(t)$ with $r, s \in S$. Note that the probability $\pi^=(t)$ for the center category enters the mean function $E_t(s^{obs})$ in $\bar{\pi}(t) = \sum_s \pi^s(t) \cdot \Phi^s(t)$. Thus, $E_t(s^{obs})$ is the mean function in t of our observed (and potentially biased) indicator for $n \to \infty$. Therefore, $E_t(s^{obs})$ is designated as the observed or biased mean function or the observed or biased indicator. In addition, $E_t(s)$ denotes the unbiased case of $E_t(s^{obs})$.

4.2.3 Correlation with the cycle function

When inspecting nonresponse biases and their effects on the observed indicator $E_t(s^{obs})$, we have to define which type of bias we want to analyse. The first 'natural' comparison would be to evaluate the difference between the observed indicator $E_t(s^{obs})$ and the unbiased mean function $E_t(s)$, e.g. by (absolute) differences. However, we have two notes on this issue: First, survey indicators that are constructed by balance statistics are, at least, artificial. Unlike official data, for example, their level does not reflect a certain quantity. This particularly holds for latent variables such as the 'business situation'. At best, a positive/negative value indicates an increase/decrease. Second, these indicators are constructed to display the cyclical development $g(t)$ for different economic parameters over time. A shift of the whole time series $E_t(s)$ or a stretch/compression by a constant factor would leave the relationship between $E_t(s)$ and the cycle function $g(t) = E_t(s^*)$ unaffected. Therefore, the main subject of investigation is (in contrast to most bias analyses) not the level but the correlation over t.[4]

Before we start to evaluate the effects on the correlation with the cycle function $g(t)$, we have to think about the correlation that can be obtained in the unbiased case for $n, T \to \infty$. We define

$$\lim_{n,T \to \infty} \rho(E_t(s), g(t)) = \rho_C^{unbiased}, \tag{4.5}$$

[4] In Appendix B.5, a small simulation study on the effects of the nonresponse patterns on the forecasting performance of the indicators is given.

where $\rho_C^{unbiased}$ denotes the maximum correlation in cases of an unbiased mean function $\mathrm{E}_t(s)$ with the cycle function $g(t)$. As Pearsons correlation coefficient measures the linear relationship between two variables, it is only invariant towards linear transformations $h(\cdot)$. Since $\mathrm{E}_t(s)$ only depends on the cycle function $g(t)$ and $\Phi(\cdot)$ is continuous, but not linear in Equation (4.3), $\rho_C^{unbiased} < 1$. This result is not surprising since a 3-level trait always includes less information. For this reason, we have to rescale $\rho_C^{obs} = \rho(\mathrm{E}_t(s^{obs}), g(t))$ when inspecting possible bias effects to

$$\tilde{\rho}_C := \tilde{\rho}_C^{obs} = \frac{\rho_C^{obs}}{\rho_C^{unbiased}}. \tag{4.6}$$

We notice that $\tilde{\rho}_C$ is not a real correlation, but the adjustment in Equation (4.6) is necessary to display the differences in correlation. However, in nearly all of the cases $\rho_C^{unbiased}$ is close to 1. The non-linearity arises from the cumulative density function of ϵ. Linearity can only be received in cases of a rectangular distribution $\epsilon \sim U(c_l, c_u)$. However, this might be unrealistic in most settings. Stangl (2009) evaluated both variables of the Ifo index on a visual analog scale and found evidence that the distribution of s^* is similar to a bell curve. A consequence of $\rho_C^{unbiased} < 1$ is that an observed biased indicator $\mathrm{E}_t(s^{obs})$ may have a higher correlation with the cycle function $g(t)$ than the unbiased indicator $\mathrm{E}_t(s)$. We can not analytically derive in which cases this effect would appear, but we show in Appendix B.2 that a perfect correlation can never be obtained.[5]

4.3 Simulation study

4.3.1 Definition of cycle functions

Before we run our simulation study, we have to specify our cycle functions $g(t)$. As noted in Section 4.2, the ideal case of a business cycle is a sinus function in time, i.e. $\sin(t)$. In addition, historical business cycle theory identifies four types of overlapping cycles: Kitchin (3-5 years), Juglar (7-

[5] In Appendix B.3 we also show that the correlation of $\mathrm{E}_t(s^{obs})$ with the unbiased indicator $\mathrm{E}_t(s)$ is affected by a decrease in any case of NMAR.

11), Kuznets (15-25) and Kondratiev (45-60).[6] We first want to inspect the effects of our acceptance functions $\pi^s(t)$ on a simple sinus-function for the cycle $g(t)$, so $g(t) \approx \sin(t)$. We refer t to represent months and define our simple sinus-function as a Juglar cycle with 10 years, i.e. 120 months. So, we have to re-scale $\sin(t)$ by a constant k to $\sin(t/k)$. The first full cycle is reached at 2π, so $\sin(120/k) \approx \sin(2\pi)$.[7] With this, $k \approx 120/2\pi \Rightarrow k = k_{Jug} \approx 20$. Then, our first cycle $g(t)$ to inspect is defined as

$$g_1(t) := \sin(t/k_{Jug}) = \sin(t/20).$$

Secondly, we define a cycle function $g(t)$ including all four cycles mentioned above. One full Kitchin 4-years cycle covers 48 months, one Kuznets 20-years cycle 240 months and one Kondratiev 50-years cycle 600 months. As above, the scaling parameters k turn to: $k_{Kit} \approx 8$, $k_{Kuz} \approx 38$ and $k_{Kon} \approx 95$. Therefore, the cycle function is defined as

$$\begin{aligned} g_2(t) &:= \sin(t/k_{Kit}) + \sin(t/k_{Jug}) + \sin(t/k_{Kuz}) + \sin(t/k_{Kon}) \\ &= \sin(t/8) + \sin(t/20) + \sin(t/38) + \sin(t/95) \end{aligned}$$

Of course, $g_1(t)$ and $g_2(t)$ display a very ideal case of a business cycle, which is usually not observed in reality. To come closer to a more realistic process, we define a third cycle function $g_3(t)$ by a stationary AR(1)-process

$$g_3(t) := 0.9 \cdot g_3(t-1) + u_t$$

with $g_3(0) = 0$ and $u_t \sim N(0,1)$. The series $g_3(t)$ can be regarded as a monthly growth rate. One example for such a latent process is the inflation question of Ifo's World Economic Survey, which is also asked on a 3-level Likert scale. Our fourth cycle function is based on a real data example, the U.S. industrial production. We define $g_4(t)$ by extracting the output gap with the Hodrick-Prescott band pass filter for time series (see Hodrick and

[6]For a recent study to the 2008-2009 economic crisis with respect to the different cycles see Korotayev and Tsirel (2010).

[7]We notice that this procedure is inaccurate but the cycles, as defined in the literature, do also not have an exact length. The cycles in this section are only selected for representation of the nonresponse bias effects.

Prescott, 1997).

All cycle functions $g_1(t), g_2(t), g_3(t)$ and $g_4(t)$ are displayed in Figure 4.1 for $T = 500$. For an easier comparison, we standardise all cycle functions to $g_j(t) \in [-1, 1]$. The thresholds τ^s are defined as $\tau^+ = 1/3$ and $\tau^- = -1/3$ so that the range of every function is divided into three parts of equal size which fit the assumptions for the calculation of the balance statistics in Section 4.2.1.

4.3.2 Definition of response functions

After defining cycle functions $g_j(t)$ we introduce different nonresponse patterns determined by the acceptance functions $\pi^s(t)$. Of course, the number of different functions $\pi^s(t)$ is infinite (and thus the number of combinations), as well as the number of cycle functions $g_j(t)$. Nevertheless, we examine the effects of some main types of $\pi^s(t)$ on the correlations and perform a simulation study on this.

We focus on fixed (time-independent), cycle-dependent, cycle-shifted, monotone and random types of nonresponse patterns in t. All time-independent structures describe general differences in the responding behaviour when being in the appropriate state. For example, one can imagine that the respond rate is higher when the firms' situation is bad and the company wants to complain about that. The other effect can occur, when the firms' state is good and they want to let the others know it. Trends, nonlinear as well as linear, may be possible as the nonresponse rates in general raised throughout the last decades (see Brehm, 1994). So, it is possible that these trends can also be different according to the state of the company. The probability to respond may have decreased for those firms which are in a good situation. A dependence of the overall response rate from the business cycle in general was found in Chapter 3 for the Ifo Business Survey and Harris-Kojetin and Tucker (1999) for the U.S. Current Population Survey. Although nothing is known directly about the state of the company, it would be as conceivable that the company has a higher probability for nonresponse when the firm remarks that their situation is bad in relation to the market. A dependence from the cycle, as well as cycle shifts, may occur as the decision to respond may be affected by the latent variable. For robustness checks, we also include random probabilities for

$\pi^s(t)$. Regardless of the type of $\pi^s(t)$, we have to define $0 \leq \pi^s(t) \leq 1, \forall t$, to obtain a correct specified probability function.

4.3.3 Simulation study

As we are able to evaluate the mean function for the observed units $\mathrm{E}_t(y^{obs})$ directly with Equation (4.4), we focus our simulation study on analysing different situations of nonresponse patterns. In Section 4.3.2, we defined five general types of acceptance rates: random, fixed, cycle-dependent, cycle-shifted and monotone. For every cycle function $g_j(t)$ and each of these bias patterns, we draw $Z = 1000$ different situations for the triple $\pi^+(t), \pi^=(t)$ and $\pi^-(t)$. For example, for a time-independent bias pattern we draw a probability for each $\pi^s(t)$ and fix these over t. In contrast, a random bias pattern would lead to a draw of each $\pi^s(t)$ for every t. For cycle-dependent and cycle-shifted bias patterns, $\pi^s(t)$ are functions of $g_j(t)$ whereas monotone and linear bias patterns are functions in t. In every iteration z, different scaling parameters (which are defined in the subsequent paragraphs) are drawn. Therefore, we are able to reflect very different situations of nonresponse patterns. For each of these patterns, we calculate the correlations defined in Section 4.2.3.

Random and fixed

We start our analysis with random, unstructured response functions $\pi^s(t)$. For every iteration z, response function $\pi^s(t)$ and t, we draw from the Uniform distribution $U(0,1)$ to obtain randomness, i.e. $\pi^s_z(t) \sim U(0,1), \forall s, t, z$. For the time-independent case, we draw only once for every iteration z and state s and fix this value over t, i.e. these functions may be different across s but do not fluctuate over t which leads to $\pi^s_z(t) = \pi^s_z \sim U(0,1)$.

Cycle-dependent

In contrast to random and fixed bias patterns, all cycle-related as well as monotone patterns are functions in $g_j(t)$ or t. As we have to ensure that $0 \leq \pi^s(t) \leq 1, \forall t$, we can not use $g_j(t)$ directly and have to standardise it. In Section 4.3.1 we defined $g_j(t) \in [-1,1]$. So, our cycle-dependent

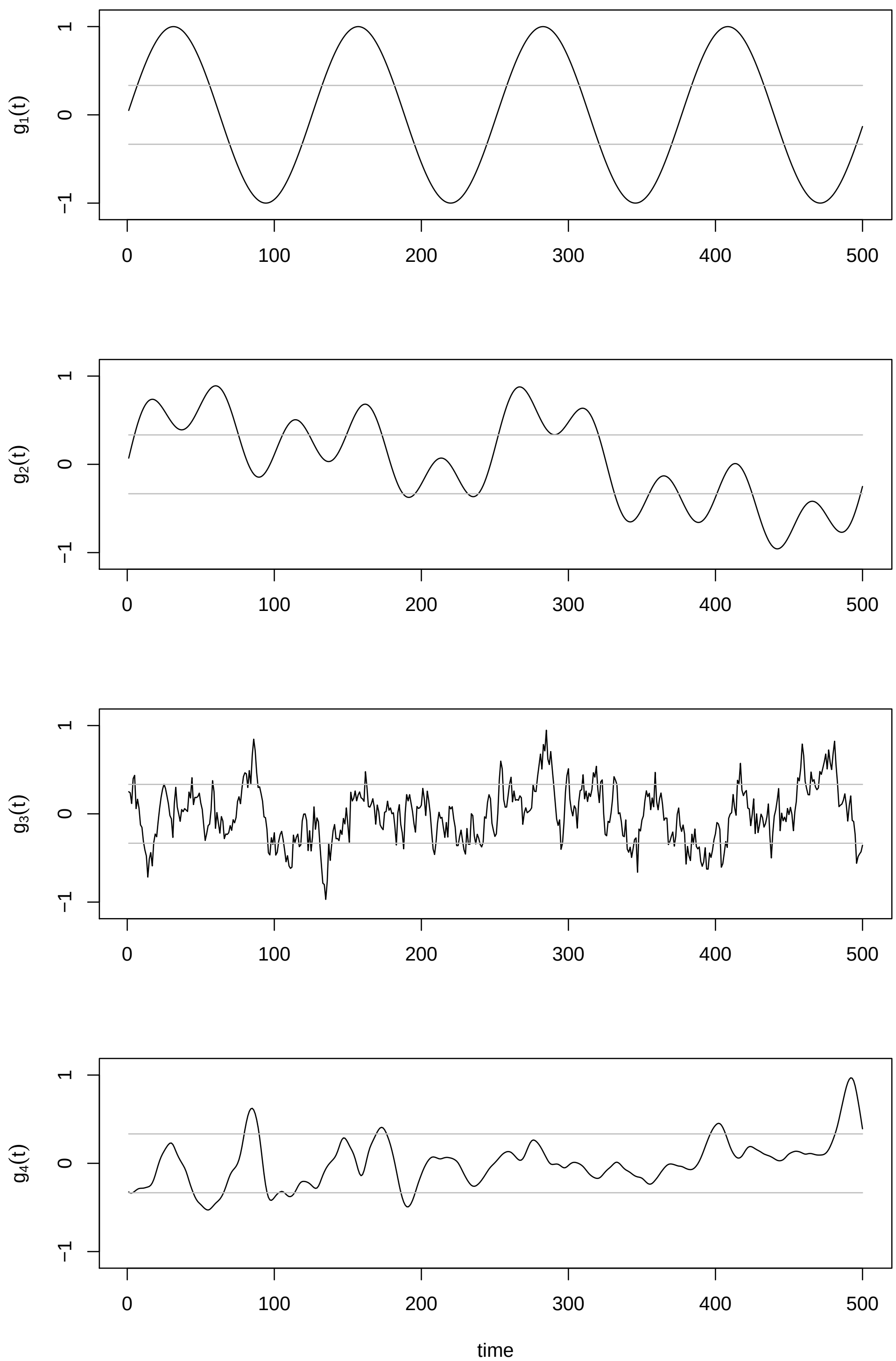

Figure 4.1: Different cycle functions with the appropriate thresholds τ^s (grey lines).

acceptance rate function has the form

$$\pi^C_{j,k_C}(t) := \frac{g_j(t)+1}{2k_C}, \tag{4.7}$$

for $j = 1,\ldots,4$. For our simulation study, we draw scaling parameters $k_C \in \{1,\ldots,10\}$. As Equation (4.7) defines a positive relationship between $\pi^s(t)$ and $g_j(t)$, we also allow negative relationships $1-\pi^C_{j,k_C}(t)$. This is done by drawing a 0/1-dummy d_I if to use $\pi^C_{j,k_C}(t)$ or $1-\pi^C_{j,k_C}(t)$. d_I will also be used for the cycle-shifted and monotone response functions.

Cycle-shifted

As survey business cycle indicators including expectation questions are commonly used for now- and forecasting, i.e. should contain leading information, we now assume that the selection process also includes leading or lagging information. It can be assumed that the respondents might anticipate coming development[8] or might react on the most recent development. To reflect this, we define the acceptance rate functions $\pi^s(t)$

$$\pi^S_{j,k_C}(t+k_T) := \pi^C_{j,k_C}(t+k_T) \tag{4.8}$$

for $j = 1,\ldots,4$. Equation (4.8) has the same form as the cycle-dependent acceptance rates in (4.7), but is shifted in time by k_T units. Therefore, we first draw $k_T \in \{-12,-11,\ldots,12\}$, i.e. we allow a maximum of 12 months for leads or lags. In addition, we draw k_C and d_I as in the cycle-dependent case.

Monotone

Cycle-dependent and cycle-shifted response rates $\pi^s(t)$ are direct transformations of $g(t)$, i.e. the underlying cycle function. To allow $\pi^s(t)$ to be dependent from t but not from $g(t)$, we specify functions $\pi^s(t)$ which display general trends of monotone or linear form. Therefore, we define

[8]The participants of the Ifo Business Survey are asked every month to give an assessment on the future business situation. This might also affect the decision to respond.

two different types of acceptance rates $\pi^s(t)$:

$$\begin{aligned}\pi^L_{k_L}(t) &:= \frac{t+1000}{k_L} \quad \text{and} \\ \pi^M_{k_M}(t) &:= \Phi\left(\frac{t-250}{k_M}\right),\end{aligned}$$

where $\Phi(\cdot)$ is the cumulative density function of the standard normal distribution. $\pi^L_{k_L}(t)$ was chosen to have values between 0 and 1 for $t \in \{1, \dots, T\}$. k_L and k_M are scaling parameters which control for the steepness of $\pi^M_{k_M}(t)$ and $\pi^L_{k_L}(t)$. In a first step, we draw d_L which decides to use $\pi^L(t)$ or $\pi^M(t)$. Conditional on d_L we draw $k_L \in [2000, 4000]$ or $k_M \in [100, 500]$. Finally, we draw d_I as in the cycle-dependent case to allow for decreasing trends e.g. by $1 - \pi^L(t)$.

Mixed
Of course, the five main bias patterns described above can also appear simultaneously. To reflect this situation, we mix these patterns by introducing a variable d_P which specifies the bias pattern to use. After this classification, the simulations are done as described above.

4.3.4 Dispersion measure

From Section 4.2.2 we know that MCAR occurs in cases of $\pi^r(t) = \pi^s(t), \forall r, s \in S$, and $\forall t \in \{1, \dots, T\}$. We want to analyse if a higher average variation in the acceptance rates $\pi^s(t)$ between the different states for given t leads to a lower correlations. To evaluate this variation for the whole time series, we introduce a variation measure

$$v = \frac{1}{T}\sum_{t=1}^{T} \mathrm{Var}_s(\pi^s(t)|t). \tag{4.9}$$

The inner part $\mathrm{Var}_s(\pi^s(t)|t)$ is the variance across $\pi^s(t)$ for given t which is then averaged over t. As $\pi^s(t) < 1, \forall t$, $\mathrm{Var}_s(\pi^s(t)|t) < 1$ and therefore also $v < 1$.

4.3.5 Results

Figure 4.2 shows the scatterplots for the pairs $(\tilde{\rho}_C, v)$ according to the five different main bias patterns plus the mixed case as defined in Section 4.3.3.[9] The vertical axis are scaled to $[-1, 1]$ whereas the horizontal axis are not scaled due to the fact that the range of v differs strongly between the various patterns. In addition, boxplots for the correlations and the dispersion parameters are drawn and the underlying cycle functions $g_j(t)$ are displayed with different colors. The boxplots on the y-axis clearly show that the majority of biased indicators still receives high correlations close to 1. This holds in particular for the fixed, cycle-dependent, cycle-shifted and monotone types of $\pi^s(t)$. On average, the lowest correlations appear in the random and mixed case. Furthermore, clusters according to the different cycle functions can be seen in the random case. For this bias pattern type, the ideal cycle functions $g_1(t)$ and $g_2(t)$ seem to be less affected by the nonresponse bias. As the random bias pattern includes more unstructured uncertainty in the data, the correlations become smaller on average. With exception of the monotone bias types, there seems to be no clear connection between the dispersion v and the correlations. Correlations remain high, even if the the dispersion increases.

To get an idea of how the acceptance rates $\pi^s(t)$ transform the biased indicator $\mathrm{E}_t(s^{obs})$, we show some extreme cases for every main bias type in Figures 4.3 and 4.4. We select the cycle function $g_2(t)$ to display the effects (the effects for the other cycle functions $g_1(t), g_2(t)$ and $g_3(t)$ are drawn in Figures B.1 to B.6 in Appendix B.4). On the left side, the cycle function $g_2(t)$ (in grey), the unbiased mean function $\mathrm{E}_t(y)$ (in green) and the biased mean function $\mathrm{E}_t(s_{obs})$ (in red) are drawn. The right side shows the appropriate acceptance rates $\pi^+(t)$ (- - -), $\pi^=(t)$ (—) and $\pi^-(t)$ $(\cdots)$. In Figure 4.3, a random, fixed and cycle-dependent case, including correlations $\tilde{\rho}_C$ as well as the dispersion parameter v, is displayed. The random bias pattern clearly shows that the correlations decrease, but the general underlying structure from the cycle function is still present. Smoothing approaches, such as the Hodrick-Prescott band pass filter, could still be used in this case to reduce this effect. However, we have to notice here that sampling

[9] The appropriate scatterplots for (ρ_E^{obs}, v) as defined in Appendix B.3 are drawn in Figure B.7

size effects will include even more uncertainty. The second row displays the effect of time-independent acceptance rates $\pi^s(t)$. With $\pi^-(t) = 0.1$ we chose a low acceptance for the negative replies. This led to a shift of $E_t(s^{obs})$ upwards.[10] However, the correlation still remained high, as we already saw in the scatterplots in Figure 4.2. In the last row, the cycle-dependent case is shown. $\pi^s(t)$ were chosen to display a very extreme case with high negative correlations. This effect appears when the acceptance rates are anti-cyclical, i.e. negative replies are seldom in recessions and positive ones in boom times. In our case, we have $\pi^+(t) = 1 - \pi^C_{2,1}(t)$ and $\pi^-(t) = \pi^C_{2,1}(t)$.[11] Figure 4.4 shows the other three bias types: cycle-shifted, monotone and mixed cases. Although the correlations remained high, it can clearly be seen for the cycle-shifted case that the lag/lead structure of the observed indicator has changed. The bias pattern led to a shift into the future which caused a decrease of leading information of the indicator. The second row shows the effect of monotone and linear acceptance rates $\pi^s(t)$ on the indicator. This type introduced a clear trend movement of the indicator which is diametrical to the underlying cycle. This was caused by an increase of acceptance of the positive replies and a simultaneous decrease of the negative replies. In this case, the correlations are also negative. The last type of bias pattern is the mixed case where we combined a cycle-shifted (for $\pi^+(t)$), a monotone (for $\pi^=(t)$) and a fixed (for $\pi^=(t)$) bias pattern. Although the acceptance rates are of very different kind, the correlations still remain high.

4.4 Conclusion

In this chapter, a methodological framework for the widely used balance statistics indicators for economic time series was built. As these indicators are based on surveys, we included nonresponse into this framework to evaluate their effects. Due to the fact that the indicators are artificial, we focused our analyses on the effects of nonresponse on correlation. The anal-

[10] A stretch of the indicator can appear in cases where $\pi^=(t) << \pi^+(t), \pi^-(t)$ whereas a compression of the indicators appears in cases of $\pi^=(t) >> \pi^+(t), \pi^-(t)$.

[11] As the negative replies have higher probability in recession times,$\pi^C_{2,1}(t)$ is anti-cyclical for the '$-$'-category.

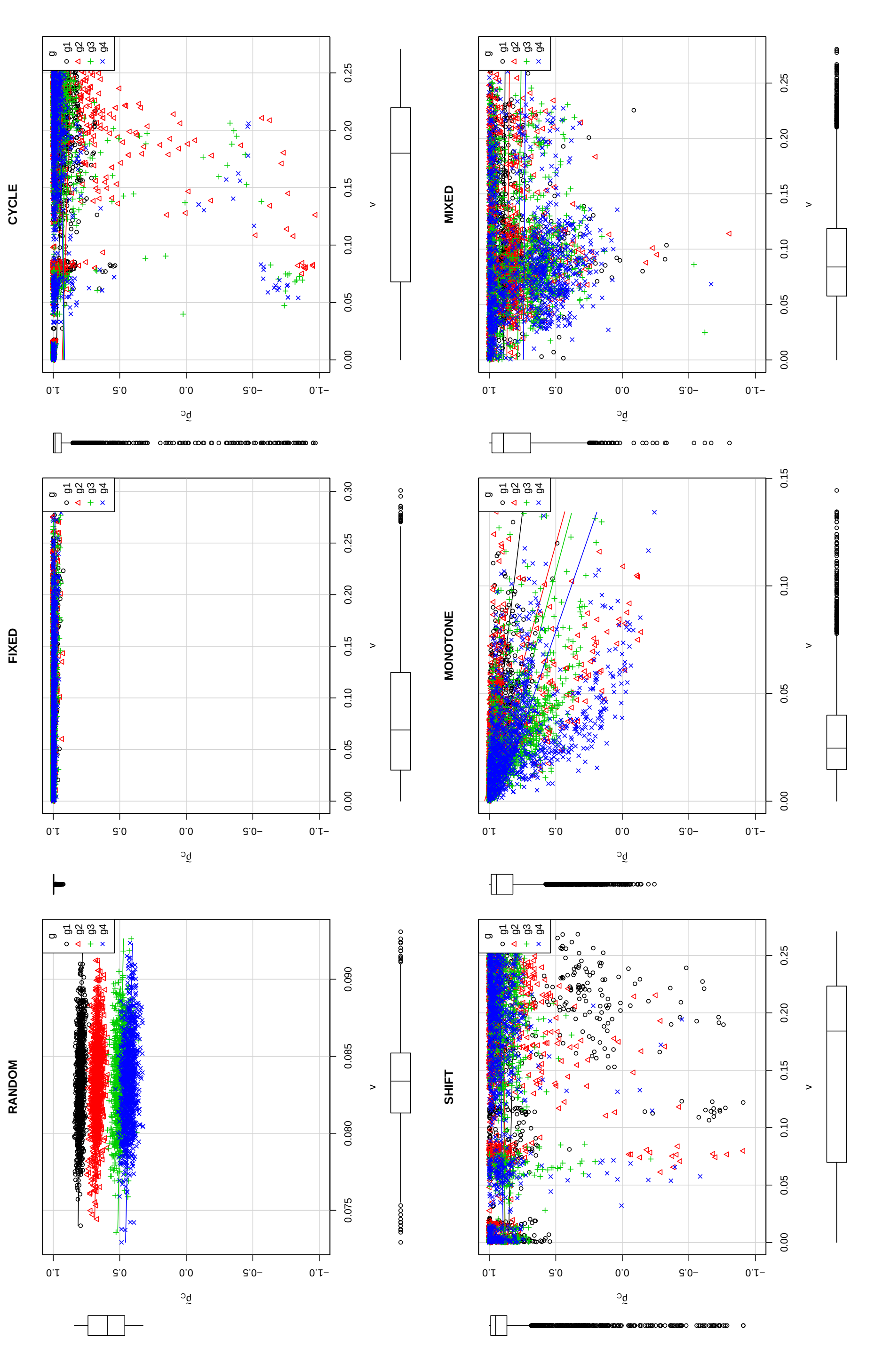

Figure 4.2: Scatterplots for correlations $\tilde{\rho}_c$ and dispersions v for 6 different types of acceptance rates $\pi^s(t)$ as defined in Section 4.3.3.

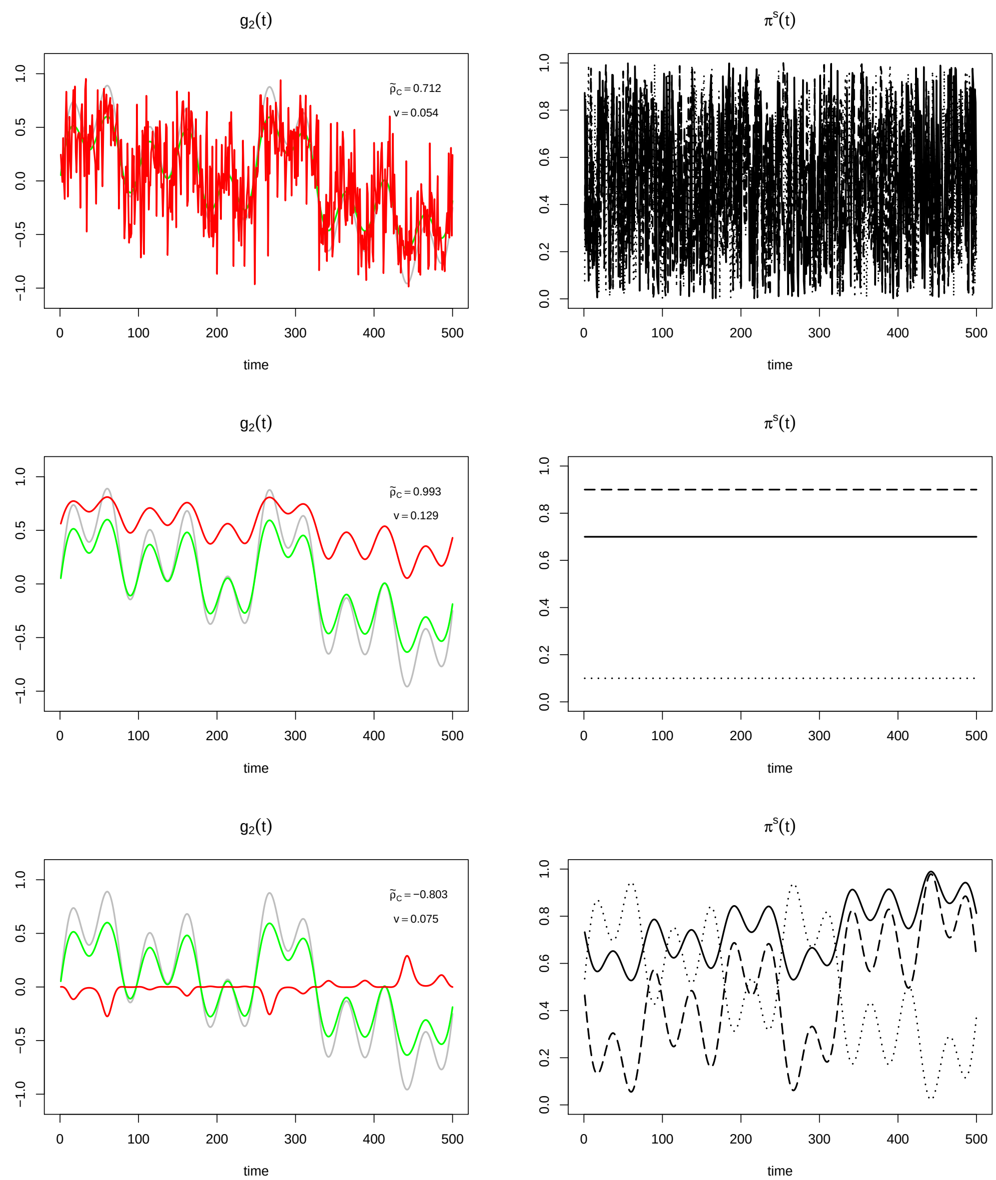

Figure 4.3: Effects of bias patterns on cycle function $g_2(t)$ - part I. Left: cycle function $g_2(t)$ (—), unbiased mean function $\mathrm{E}_t(s)$ (—) and observed mean function $\mathrm{E}_t(s^{obs})$ (—). Right: acceptance rates $\pi^+(t)$ (- - -), $\pi^=(t)$ (—) and $\pi^-(t)$ ($\cdots$). From top to bottom: Random, fixed (with $\pi^+(t) = 0.9$, $\pi^=(t) = 0.7$ and $\pi^-(t) = 0.1$) and cycle-dependent (with $\pi^+(t) = 1 - \pi^C_{2,1}(t)$, $\pi^=(t) = 1 - \pi^C_{2,2}(t)$ and $\pi^-(t) = \pi^C_{2,1}(t)$) patterns.

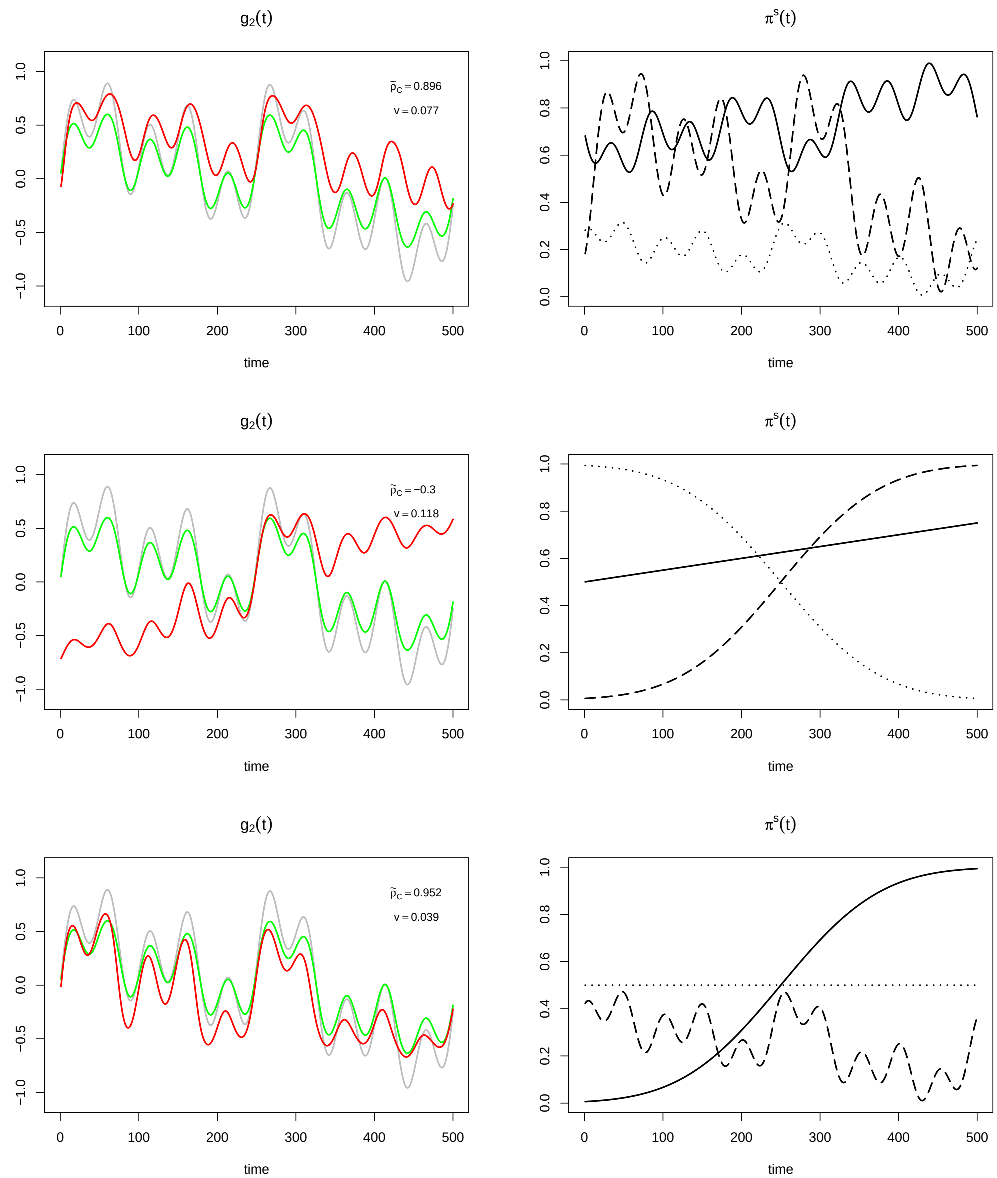

Figure 4.4: Effects of bias patterns cycle function $g_2(t)$ - part II. Left: cycle function $g_2(t)$ (—), unbiased mean function $\mathrm{E}_t(s)$ (—) and observed mean function $\mathrm{E}_t(s^{obs})$ (—). Right: acceptance rates $\pi^+(t)$ (- - -), $\pi^=(t)$ (—) and $\pi^-(t)$ ($\cdots$). From top to bottom: Cycle-shifted (with $\pi^+(t) = \pi^S_{2,1}(t-12)$, $\pi^=(t) = 1 - \pi^S_{2,2}(t+3)$ and $\pi^-(t) = \pi^S_{2,3}(t+12)$), monotone (with $\pi^+(t) = \pi^M_{100}(t)$, $\pi^=(t) = \pi^L_{2000}(t)$ and $\pi^-(t) = 1 - \pi^M_{100}(t)$) and mixed (with $\pi^+(t) = \pi^S_{2,2}(t+12)$, $\pi^=(t) = \pi^M_{100}(t)$ and $\pi^-(t) = 0.5$) patterns.

yses showed that the correlations between the observed indicators and the underlying cycle still remain high in nearly all of the cases. Of course, some bias patterns exist where the shape of the indicator is transformed strongly, but we notice that these patterns are seldom and unrealistic to appear in real-life situations. However, some patterns might cause certain problems such as a complete switch of the indicator or a shift which might affects the leading performance. The analyses also showed that there exists only a small connection between the dispersion in the acceptance rates and the correlations. If the acceptance rates differ strongly for random or monotone transformations of t, the correlation with the underlying cycle decreases on average. The reasons for this robustness are probably due to the fact that all states contain the oscillation of the underlying cycle. All of the three states are affected by $g(t)$ and therefore include its information. When the probability for one state to be observed decreases, the other states still include enough information. Only for very extreme, in particular time- or cycle-dependent cases, significant decreases in correlation might appear. Although the results in this chapter support the usage of these indicators, the impact of the missing data in a real-life situation will be evaluated in the next chapter. To this end, several imputation methods are developed to enable a consistent estimation for the missing data.

Chapter 5

Effects of imputed observations on business survey results

Abstract A widespread method for forecasting economic macro level parameters such as GDP growth rates are survey-based indicators which contain early information in contrast to official data. But surveys are commonly affected by nonresponding units which can cause biased results. Many papers have examined the effect of nonresponse in individual or household surveys, but less is known in the case of business surveys. For this reason, we analyse and impute the missing observations in the Ifo Business Survey, a large business survey in Germany. The most prominent result of this survey is the Ifo Business Climate Index, a leading indicator for the German business cycle. To reflect the underlying latent data generating process, we compare different imputation approaches for longitudinal data. After this, the microdata are aggregated and the results are compared with the original indicators to evaluate their implications at the macro level. Finally, it is shown that the differences between the original and imputed indicators are small.

5.1 Introduction

The usage of survey-based indicators for monitoring and forecasting economic parameters has a long tradition in research.[1] For more than 60 years these tendency surveys exist and their number has increased throughout the last decades (Nardo, 2003). However, as surveys are commonly affected by nonresponding units, a serious problem may occur when there are major differences between respondents and nonrespondents in their answering behaviour. The evaluation and correction for possible nonresponse biases is especially a concern in household or population surveys and therefore has been discussed extensively in the literature. In contrast, Janik and Kohaut (2012) mention that less papers exist with respect to missing data and their effects in business surveys. An exception is the paper by Drechsler (2011a) who imputes the business data of the 2007 IAB Establishment Panel. In general, biases may lead to a lower accuracy in forecasting performance of business survey indicators. To fill this gap, we analyse the missing observations in the Ifo Business Survey, a large monthly business survey in Germany with about 7000 responding firms each month. Although the IBS has high return rates around 80%, nonresponses can cause problems. For example, Schafer (1997) suggests that missing observations should be analysed if their fraction is higher than 5%. In general, the missing data mechanism is only ignorable if the data are missing at random, i.e. the parameters for the missing data generating process are unrelated to the values the units would have answered, see Schunk (2008).

To evaluate the effects of nonresponding units, imputation methods can be used. We develop different imputation strategies for the missing microdata sets in the IBS and analyse their effects on the aggregated macro indicators, i.e. the Ifo Business Climate Index, a leading indicator for the German business cycle. After imputation, we are able to investigate and compare the difference of the original and the imputed indicators, also with respect to forecasting issues. Although a general problem in studies

[1] Large parts of this paper base on C. Seiler and C. Heumann (2012), 'Microdata Imputations and Microdata Implications: Evidence from the Ifo Business Survey', *Department of Statistics, Technical Report No. 119*. We thank Lisa Möst and Gunther Schauberger for their help.

regarding imputation analysis prevails that the MAR assumption can not be tested when there exists no additional information about the data (see Manski (2003) and Cameron and Trivedi, 2005), we evaluate the predictive accuracy of several different imputation approaches to find an appropriate model which reflects the inherent dynamics and leads to good estimates.

Therefore, this chapter is organised as follows: The data set and its specifics are described in Section 5.2. We provide some descriptive statistics and show how the survey is performed and structured according to EU regulations. In Section 5.3 we develop different imputation strategies and compare the effectiveness of these imputation approaches. Section 5.4 shows and compares the aggregated results after imputation of the missing values. We analyse these macrodata time series also for the subareas and finally compare the forecasting performance of the original and imputed Ifo Business Climate Index. Section 5.5 sums up our empirical findings.

5.2 Data

5.2.1 The survey

The development of survey-based business cycle indicators has its seeds in the need of early information about the economic development. As official data are commonly published with delay and also may be revised after the first publication, business cycle tendency surveys can monitor the recent economic situation more quickly. The Ifo Institute was one of the first conducting its *Ifo Business Survey* in 1949 and within the last 60 years this method has been widely accepted see OECD (2003) and Nardo (2003). In line with the Joint Harmonised EU Programme of Business and Consumer Surveys (European Union, 2006), these indicators are based on two variables which are measured on a 3-level Likert scale representing a good, equal or bad state. Due to the construction of the questions in the questionnaire, the resulting indicators in fact measure the business cycle without trend (OECD, 2003).

The data used in this chapter are from the Ifo Business Survey, the German part of the Joint Harmonised EU Programme. The most well-known result of this survey is the *Ifo Business Climate Index*, a monthly indicator for the German business development which is widely used for forecasting analyses. Every month about 7000 companies respond. For further methodological information on this survey see Goldrian (2007) and the early works of Anderson (1951, 1952) and Theil (1952). Becker and Wohlrabe (2008) give an overview on the collected variables and Abberger and Wohlrabe (2006) on the literature with respect to forecasting analyses with the Ifo index.

As stated above, the Ifo index is constructed using only two variables of the survey: The current business situation (BS) in the appropriate month and the business expectations (BE). Both are measured on a 3-level Likert scale with values 'good'/'better' (indicated by $+$), 'satisfactory'/'about the same' ($=$) or 'bad'/'worse' ($-$). To calculate the index, the answers are weighted by the companies' size (which is updated once a year) and the business area the firm is classified to according to the official classification from the German Statistical Office. To achieve the final value of the index, the fraction of negative replies is substracted from the positive ones in a first step and then the harmonic mean is calculated to construct the 'business climate' from the 'business situation' and 'business expectation' balance. The aggregation scheme is presented in Appendix C.1. In principle, the index can be interpreted as a weighted mean. Due to about 7000 respondents every month, also indicators for lower aggregation levels are calculated.

Since more than 99.9% of missing values for BS and BE are due to unit nonresponse, we do not perform an analysis by imputing only item but not unit nonresponse because we do not expect any major differences to the original results with missing data. Therefore, no other variables from the survey (such as demand or production) could be used as covariates, as we have no answers from the corresponding company at the time of unit nonresponse. However, in Section 5.3.2 we show how these strong assumptions can be relaxed so that regression approaches are possible. In addition, we make use of the dynamics of the business cycle and additionally are able to examine relations based on the characteristics of the firms.

5.2.2 Some descriptive statistics

To get an idea of the extent of the variables and the missing values in the IBS, we provide some descriptive statistics in this subsection. Our variables of interest are both measured on a 3-level Likert scale, so every company changes over time between these three states and nonresponse can be treated as a fourth state. Table 5.1 shows the transition matrices for BS and BE for six transition periods $t-1 \rightarrow t, \ldots, t-6 \rightarrow t$ with t representing months. The probability for staying in the same state is relatively high, so the state change is slow in relation to the survey frequency. It can also be seen that the business expectations change more often than the business situation which is rather unsurprising. It seems that the probability for changing from response to nonresponse from period $t-1$ to t is different depending on the state of the company at time $t-1$. The probability not to respond in period t is almost twice as high after responding 'bad' in contrast to 'good' in $t-1$. This is contrarily to macro level results, where nonresponses are more frequent in economic better times, see Harris-Kojetin and Tucker (1999) and Chapter 3. In addition, switching from nonresponse to response also seems to be selective since the probabilities of the categories are not equal. If the firms leave the nonresponse-'state', e.g. $P(y_t \equiv +|y_{t-1} \equiv NA)$, only 4% of the firms replied a positive business situation.

Since we have repeated measures from the same units, we have to get an idea of the presence of missing sequences. Figure 5.1 shows the length of successive unit nonresponse. 56% of the missing data is due to nonresponse only for a single month. More than 80% of nonresponse appears within 3 months and 90% within 6 months. Thus, for most values prevailing information is still available. Depending on the imputation method (in particular conditional models), probably not all missing values can or should be estimated as one can expect that the predictive accuracy of the model decreases when many successive missings occur. Therefore, we validate our results in Section 5.4 according to different horizons h of successive nonresponse.

$t-1/t$	+	=	-	NA
+	0.68	0.24	0.03	0.05
=	0.06	0.75	0.12	0.07
-	0.01	0.15	0.75	0.09
NA	0.04	0.20	0.20	0.56

$t-1/t$	+	=	-	NA
+	0.58	0.32	0.04	0.06
=	0.07	0.76	0.11	0.07
-	0.02	0.27	0.62	0.09
NA	0.05	0.26	0.12	0.56

$t-2/t$	+	=	-	NA
+	0.62	0.29	0.04	0.05
=	0.07	0.71	0.15	0.07
-	0.01	0.19	0.71	0.09
NA	0.04	0.22	0.21	0.52

$t-2/t$	+	=	-	NA
+	0.50	0.38	0.05	0.07
=	0.08	0.72	0.13	0.08
-	0.04	0.31	0.56	0.09
NA	0.06	0.29	0.13	0.52

$t-3/t$	+	=	-	NA
+	0.58	0.32	0.05	0.05
=	0.07	0.68	0.17	0.07
-	0.02	0.21	0.68	0.09
NA	0.05	0.24	0.22	0.49

$t-3/t$	+	=	-	NA
+	0.45	0.41	0.07	0.07
=	0.08	0.70	0.14	0.08
-	0.05	0.34	0.52	0.09
NA	0.07	0.31	0.14	0.49

$t-4/t$	+	=	-	NA
+	0.54	0.34	0.06	0.05
=	0.08	0.66	0.18	0.08
-	0.02	0.22	0.66	0.10
NA	0.05	0.26	0.23	0.46

$t-4/t$	+	=	-	NA
+	0.41	0.43	0.09	0.07
=	0.08	0.69	0.15	0.08
-	0.06	0.36	0.49	0.10
NA	0.07	0.32	0.14	0.46

$t-5/t$	+	=	-	NA
+	0.52	0.35	0.07	0.06
=	0.08	0.65	0.19	0.08
-	0.02	0.23	0.65	0.10
NA	0.06	0.27	0.23	0.44

$t-5/t$	+	=	-	NA
+	0.38	0.44	0.10	0.07
=	0.09	0.68	0.15	0.08
-	0.07	0.37	0.47	0.10
NA	0.07	0.34	0.14	0.44

$t-6/t$	+	=	-	NA
+	0.49	0.36	0.09	0.05
=	0.08	0.64	0.20	0.08
-	0.02	0.24	0.63	0.10
NA	0.06	0.28	0.24	0.42

$t-6/t$	+	=	-	NA
+	0.36	0.45	0.12	0.07
=	0.09	0.67	0.16	0.08
-	0.07	0.38	0.45	0.10
NA	0.08	0.35	0.15	0.42

Table 5.1: Transition matrices for business situation (left) and business expectation (right) for $t-1 \rightarrow t, t-2 \rightarrow t, \ldots, t-6 \rightarrow t$ (top to bottom) evaluated from the observed period (1994-2009).

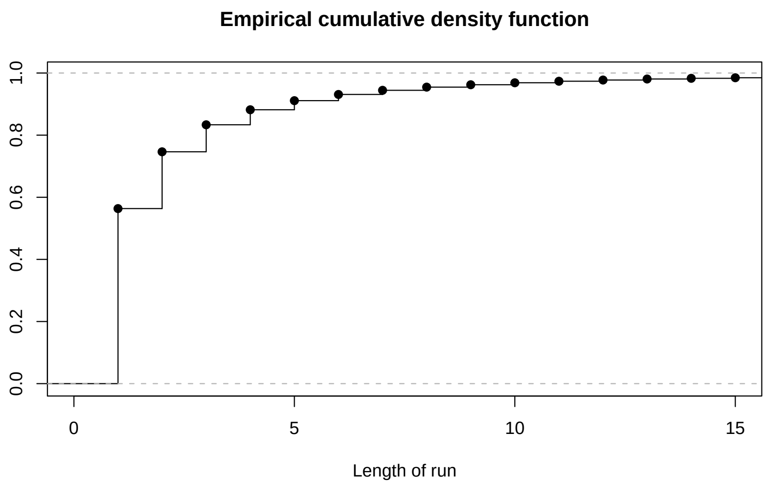

Figure 5.1: Empirical cumulative density function for the length of successive unit nonresponse

5.3 Methodology

5.3.1 Requirements on the imputation methods

The methodological beginnings of imputation models for missing observations are mostly associated with Donald Rubin, see in particular Rubin (1987) and Little and Rubin (2002) for an overview and definition of missing data patterns. In literature, a wide variety of different imputation methods exist but the data in this thesis has particular structure so a viable imputation approach has to account for this. Our variables of interest are measured on a 3-level Likert scale but both can be regarded to be influenced by latent variables which change over time depending on the business cycle. This fact implies two requirements on the imputation methods: First, as we analyse panel data, we have to use imputation methods which can reflect the inherent dynamics of the underlying latent process. This means that t, the calendar time, should be included in some form into the imputation model. Engels and Diehr (2003) and Kleinke et al. (2011) give an overview on the imputation of panel data but also mention that most standard approaches implemented in statistical software packages are limited to handle incomplete panel data as they mostly do not include the individual past and/or need covariates which are not available in our case to due unit nonresponse. In particular, the method proposed by Little and Su (1989) has been proven to be useful in panel surveys (Watson and Starick, 2011) but is limited to item nonresponse and also not appropriate in our case since the variables to be imputed should include a long-term trend to estimate non-zero values. Second, as our variables of interest have only three different states, we have to choose methods which impute plausible values. For this reason, many approaches such as simple mean imputation can not be used in our case as they require a continuous variable to impute, see Finch (2010) for an overview. Based on this structure we consider our data as Markov chains for every unit with time-inhomogenous transition matrices. Other imputation models for longitudinal data, for example autoregressive models for univariate time series as in Shumway and Stoffer (1982), are limited to continuous data or need covariates (Jennrich and Schluchter, 1986) and are also not an appropriate solution in our case.

In addition to the requirements stated above, we have to face some other issues in our analysis. Because missing observations appear almost solely due to unit nonresponse there are hardly any cases where covariates are available at the same point in time, so that a regression analysis (or more general: external information) can, in principle, not be used at the company level. This would reduce the number of eligible imputation methods enormously, but we will later show how these strong assumptions can be relaxed. Basically, two general approaches to include explanatory information remain: First, using the individual past and their inherent dynamics. Second, using attributes from similar companies at the same time after defining a similarity structure. Another problem is the extent of the data set when running multiple imputations. Covering the microdata sets from 1994-2009, we receive more than 1.6 million observations which cause enormous computing time to run multiple imputations (MI).[2] Graham et al. (2007) notice that the number of imputations

[2]The different number of observations compared to Chapter 3 results from the questionnaire. For the construction firms, all business areas are asked on one questionnaire. The answers of all business areas enter the calculation of the Ifo index and therefore the

depends on the researchers tolerance of precision and the computing time to run these multiple imputations. The most common choice in literature is to set the number of multiple imputations to 5 which is also done here. In general, MI is only appropriate for probabilistic approaches, because deterministic methods only lead to the same value. So the strategy is as follows: We try to find the imputation method which reflects the data generating process best and use this method to impute the missing values. To decide between the approaches, we introduce a measure to evaluate the predictive accuracy in Section 5.3.3.

5.3.2 Imputation methods for ordinal panel data

Last observation carried forward

One of the easiest ways to impute longitudinal data is the *last observation carried forward* (LOCF) method, in which the last recorded observation of the nonresponding unit is used. Let $y_{i,t}^{mis} := y_{i,t} | m_{i,t} = 1$ be the missing observation of unit i in the t-th wave with a missing indicator $m_{i,t}$ as defined in Section 2.2.1. Then, the imputed value by LOCF $y_{i,t}^{imp,LOCF}$ is

$$y_{i,t}^{imp,LOCF} = y_{i,t'}^{obs}$$

with $t' < t$ and $y_{i,t'+1}^{mis}, \dots, y_{i,t}^{mis}$, i.e. all data points for unit i from wave $t' + 1$ to t are missing. This method makes the strong assumption that the value remains unchanged in case of nonresponse, i.e. $y_{i,t'} = y_{i,t}$. Little and Rubin (2002) argue that this assumption is unrealistic in many settings. In recent years, this approach came more and more under criticism, see for example Cook et al. (2004) and Saha and Jones (2009). Nevertheless, LOCF is widely used, particularly in clinical studies (see Woolley et al., 2009). We assume that this method leads to relatively good results in cases of ordinal data with less states and high persistence. Also, this method is more appropriate if the number of successive missing values is not too long. For our data, both arguments seem to be the case. From Section 5.2.2 we know that long runs of missings are seldom but can occur. Therefore, we have

missings are to be imputed. For the nonresponse analysis in Chapter 3, we treat the whole questionnaire as a single unit.

to evaluate the power of this method according to the length of successive missing values, because it is plausible that predictive power decreases if the last recorded observation dates back several months. Due to its intensive use of LOCF, it is also a good proxy for other imputation methods. We expect that a structured approach (including additional information from covariates) should be able to produce better estimates than LOCF.

Nearest neighbour

The *nearest neighbour* method (NN) is a very wide class of imputation approaches and is also one of the most commonly used. Chen and Shao (2000) give an overview on the consistency of NN imputation. The basic idea behind is to find a 'donor' for the 'recipient', i.e. an observation with same or similar properties and a recorded value which is then transferred to the nonresponding unit. Let $\boldsymbol{x}_{i,t} = (x_{i,t,1}, \ldots, x_{i,t,K})'$ be the K-dimensional vector of covariates of unit i at time t with a missing value $y_{i,t}^{mis}$ for the variable of interest y. To find one or more possible nearest neighbours, a metric $d(i,j)$ for the covariates $\boldsymbol{X}_t = (\boldsymbol{x}_{t,1}, \ldots, \boldsymbol{x}_{t,K})$ of the t-th wave has to be defined. Typical functions for $d(i,j)$ are $d(i,j) = \max_k |x_{i,k} - x_{j,k}|$ (maximum deviation) or $d(i,j) = (\boldsymbol{x}_i - \boldsymbol{x}_j)'\hat{\Sigma}_{xx}^{-1}(\boldsymbol{x}_i - \boldsymbol{x}_j)$ (Mahalanobis distance) where $\hat{\Sigma}_{xx}$ is the estimated covariance matrix of $\boldsymbol{x}_i$. Then, the imputed value by NN $y_{i,t}^{imp,NN}$ at time t is

$$y_{i,t}^{imp,NN} = y_{j^*,t}^{obs} \tag{5.1}$$

for

$$d(i,j^*) = \min_{j \in J_t^{obs}} d(i,j) \tag{5.2}$$

where J_t^{obs} is the set of units with observed values for variable y at time t. If more than one possible donor j fit Equation (5.2) or more donors should be included into the imputation model, this approach is known as kNN imputation where k is the number of possible donors. In this case, several possibilities remain: Imputing with a centrality measure, i.e. mean, mode or median or drawing from their distribution. In the latter case, it is highly recommended to perform multiple imputation as drawing from a distribution includes uncertainty.

In our setting, we define similarities according to the same business area in an appropriate month. We assume that business' from the same area may be affected by similar effects (demand, orders, etc.). Therefore, this variable is used to form so-called adjustment cells (Little and Rubin, 2002). All possible donors are collected in a specific cell according to the defined similarity structure. Then, the simple metric

$$d(i,j) = \begin{cases} 0 & \text{if} \quad i,j \text{ in same cell} \\ 1 & \text{if} \quad i,j \text{ in different cells} \end{cases}$$

is used. Usually, the number of possible donors in kNN imputation is set to a fixed value prior to the analysis. In our data set, the number of possible donors differ between business areas and survey waves. The main reason why the number of donors has to be flexible in our case is that there are no other sensible variables which can be used to define a similarity structure. As only one variable (the business area) remains, we use all possible units for this approach. So, our imputation strategy is as follows: For every month, we calculate the distribution for the three states $(+, =, -)$ according to a specific business area. For the nonresponding firms, we draw from this distribution.[3] This means, we assume that missing units behave as the observed companies from the same business area.

Markov Chains

As can be seen in Table 5.1, the probability for staying in the same state is relatively high even after six months. In order to use this fact, we consider individual state changes as a *Markov Chain* (MC). Therefore, let $y_i = \{y_{i,t}, t \in T\}, i = 1, \ldots, n$, be a stochastic process for every unit i representing *BS* or *BE* by a given probability space $(\Omega, \mathcal{F}, \mathcal{P})$. In our case, we are interested in calculating the stochastic matrices

$$P = (p_{r,s}), \quad r, s \in S,$$

[3] The mode is not used in this case, because at almost every time the proportion of '='-responses is largest and hence we would impute only '='-values.

$p_{r,s} = P(y_t = s|y_{t-1} = r)$ and $S = \{+, =, -\}$. Because of the presumption made in Section 5.3.1 that there exists an underlying dynamic process, we assume that the stochastic matrix P is time-dependent, so

$$P = P(t) = (p_{r,s}(t))$$

which means that y_i is an inhomogeneous Markov Chain. In Table 5.1 we showed that the probabilities for staying in the same state are quite high, so that this method would not make any difference to LOCF if we take the mode and assume that the highest probabilities are on the main diagonal for every t. For this reason, we take a step beyond and extend the Markov Chains to order k. So, the stochastic matrix is

$$P^k(t) = (p_{(r_{t-1},\ldots,r_{t-k},s)}(t)),$$

$p_{(r_{t-1},\ldots,r_{t-k},s)}(t) = P(y_t = s|y_{t-1} = r_{t-1}, \ldots, y_{t-k} = r_{t-k}, t)$. Notice that $\dim(P^k(t)) = |S|^k \times |S|$, so when k increases by 1 the number of rows of $P^k(t)$ increase by the factor $|S| = 3$. This means that we evaluate the runs of answers of the last k months and calculate the probabilities for different states in t. This procedure is done for every t, so we produce 'rolling' stochastic matrices. We assume that with this method a good classification of the companies can be obtained and we receive high probabilities for at least one of the three states in every row of $P^k(t)$. After evaluating the stochastic matrices $P^k(t)$, every company with missing data in t is classified by their past k values and finally we draw from this distribution or take the mode, i.e. the state with the highest probability in t, to impute the missing value. The higher k is set, the higher the specialisation is but the more transitions have to be evaluated. For $k = 5$ there are $3^5 = 243$ transitions. In spite of the large data set, many transitions do not occur in the data and therefore we set the maximum for k to 4, i.e. 81 possible transitions. We notice that this approach is uncommon in imputation analysis but results from data structure and effectively is the equivalent to an AR process on macro level which plays a major role in forecast analyses of time series. In fact, this approach is the same as a nearest neighbour imputation with a similarity structure defined on the past k months, i.e. $\boldsymbol{x}_{i,t} = (y_{i,t-k}, \ldots, y_{i,t-1})'$ at time t and $\boldsymbol{X}_t = (\boldsymbol{y}_{t-k}, \ldots, \boldsymbol{y}_{t-1})$. As in Section 5.3.2, we build adjustment cells based on the rows of the transition matrices $P^k(t)$.

Joint distribution

The assumptions for the MC approach are relatively restrictive as the exact transition of the answers is needed to be a possible candidate for imputation. A more flexible method would be to focus on the *joint distribution* (JD) of BS and BE and the individual past $t-1,\dots,t-k$ of both variables, i.e.

$$f_{JD,k}(t) := f(BS_t, BS_{t-1}, \dots, BS_{t-k}, BE_t, BE_{t-1}, \dots, BE_{t-k}, t). \tag{5.3}$$

Therefore, the imputed value by JD $y_{i,t}^{imp,JD}$ at time t is

$$y_{i,t}^{imp,JD} \sim f_{JD,k}(t). \tag{5.4}$$

The most frequent approach to obtain such joint probability functions $f_{JD,k}(t)$ in order to impute missing data is done with the `Amelia II` package, version 1.5-4 developed by Honaker et al. (2012) and originally proposed in King et al. (2001). `Amelia II` requires that the joint distribution in Equation (5.3) is multivariate normal, which is obviously violated in our case. Fortunately, the `Amelia II` package also provides imputation of ordinal variables. As we analyse panel data, `Amelia II` enables to specify time and cross sectional variables. In addition, time-varying effects as well as lags of the interesting variables can be included into the imputation model, see Honaker and King (2010). Therefore, this approach is very flexible and can reflect both, the individual state change as well as the overall underlying latent process.

Regression approaches

All approaches mentioned above are relatively easy to implement but do not exploit economic relationships at the company level for the current time period as this information is not available for nonrespondents. Due to the ordinal structure of our variables of interest, a regression-based imputation approach would have the form of a *proportional odds model* (McCullagh, 1980)

$$\eta_{i,t} = g(\mu_{i,t}) = \log \frac{P(y_{i,t} \leq c|\boldsymbol{x}_{i,t})}{P(y_{i,t} > c|\boldsymbol{x}_{i,t})} = \tau_{c,t} - \boldsymbol{x}_{i,t}\boldsymbol{\beta}_{c,t}, \qquad c = 2, \dots, C, \quad (5.5)$$

where

$$\mu_i = \mathrm{E}(y_{i,t}|\boldsymbol{x}_{i,t}) = h(\eta_{i,t}), \qquad h(\cdot) = g^{-1}(\cdot)$$

with $y_{i,t}$ as the variable of interest and $\boldsymbol{x}_{i,t}$ as the covariates of unit i. $C = 3$ as both variables of interest are measured on a 3-level Likert scale. Then, the imputed value by POM $y_{i,t}^{imp,POM}$ at time t is

$$y_{i,t}^{imp,POM} = \begin{cases} + & \text{if} \quad \mathrm{E}(y_{i,t}|\boldsymbol{x}_{i,t}) \geq \tau_+ \\ = & \text{if} \quad \tau_+ > \mathrm{E}(y_{i,t}|\boldsymbol{x}_{i,t}) \geq \tau_- \\ - & \text{if} \quad \tau_- > \mathrm{E}(y_{i,t}|\boldsymbol{x}_{i,t}) \end{cases} \quad (5.6)$$

Model (5.5) has the advantage that it models a latent variable and calculates thresholds τ_c which fits our concept in Chapter 4, Equation (4.1). As noted in Section 5.2.1, due to unit nonresponse no covariates are available. But from Section 5.2.2 we know that the companies remain relatively long in the same state and they change their answers slowly in relation to the survey frequency. It is assumed that this applies also for the other variables of the survey, which are also mainly be measured on a 3-level Likert scale. So, besides the individual past, $\boldsymbol{x}_{i,t} = \boldsymbol{x}_{i,t-1}$ contains additional variables asked in the survey from the preceding month.[4] As the individual past of the dependent variable BS or BE is included into $\boldsymbol{x}_{i,t-1}$, model (5.5) enables to check whether the inclusion of additional explanatory variables improves the estimation of BS and BE.

The major disadvantage of this approach is the fact that model (5.5) can only be estimated when all variables are observed. In cases of two or more successive unit nonresponse $\boldsymbol{x}_{i,t-1}$ is missing, i.e. $\boldsymbol{x}_{i,t-1}$ itself has to be imputed. This exacerbates the problem since we have to find an appropriate model for every variable in $\boldsymbol{x}_{i,t-1}$ which is not done in this chapter. The

[4] Considering Section 1, we interpret wave $t-1$ as a 'representative' for wave t due to the high frequency of the survey.

analyses are therefore restricted to impute only the first month after the firm responded, so we are only able to impute at least 56% of our missing data. Another issue is that the questions[5] on the questionnaire differ between the business areas. For example, the degree of capacity utilisation is asked in construction but obviously not in trade. For this reason, we calculate a different model of form (5.5) for each of the three sectors with sector-specific covariates $\boldsymbol{x}_{i,t-1}^{sec}$ which are listed in Table C.1. In addition, we need to evaluate different models depending on t to reflect the inherent dynamics. Thus, a separate model

$$\eta_{i,t}^{sec} = \log \frac{P(y_{i,t} \leq c|\boldsymbol{x}_{i,t-1}^{sec})}{P(y_{i,t} > c|\boldsymbol{x}_{i,t-1}^{sec})} = \tau_{c,t}^{sec} - \boldsymbol{x}_{i,t-1}^{sec}\boldsymbol{\beta}_{c,t}^{sec}$$

with $t = 2, \ldots, T, c = 2, 3$, for each t is calculated. $T = 192$ as the data ranges from January 1994 to December 2009.

5.3.3 Goodness of fit

When comparing different imputation methods we would ideally estimate a value and compare it with the value the respondent would have reported. Obviously, this is not possible. To decide which imputation method explains the data best, some papers set a fraction of the observed values to missing and test their approaches on these (see for example Watson and Starick, 2011). In this analysis, we go a step further and make use of the leaving-one-out principle, i.e. we treat every data point (i, t) as missing and construct the imputation method based on this reduced data set. Afterwards, we estimate this value with the imputation method. This approach de facto leads to an overimputation of the whole data set.[6] To decide between the approaches, we introduce a statistical measure to

[5]Moreover, all variables included in $\boldsymbol{x}_{i,t-1}$ are restricted to those who are measured monthly.

[6]However, for JD, we have to adapt this principle. As JD calculates time and lag effects in closed form for the whole data set, leaving-one-out would increase computing time enormously as 1.6 million observations enter the imputation model and beyond this has to be done for all of these 1.6 million observations. To calculate Cohens kappas for this approach, we randomly drop about 20% of our observed data and test the power of this imputation method with these values.

evaluate the goodness of fit for the estimators of the different imputation methods. As our variables of interest are discrete, we can count the number of correct and incorrect predicted values in a 3×3-matrix. Therefore, we introduce *Cohen's kappa* (Cohen, 1960) which is defined as

$$\kappa = \frac{\pi_o - \pi_e}{1 - \pi_e},$$

where $\pi_o = \sum_{c=1}^{C} \pi_{cc}, C = 3$, is the relative observed correspondence of the estimators, and $\pi_e = \sum_{c=1}^{C} \pi_{c\cdot}\pi_{\cdot c}$ the hypothetical probability of correspondence when there is no relationship between the original and imputed values. There also exists a weighted version of Cohen's kappa with weights w_{cd}, leading to $\pi_o = \sum_{c=1}^{C}\sum_{d=1}^{C} w_{cd}\pi_{cd}$ and $\pi_e = \sum_{c=1}^{C}\sum_{d=1}^{C} w_{cd}\pi_{c\cdot}\pi_{\cdot d}$. If you use, for example, quadratic weights

$$w_{cd} = 1 - \frac{(c-d)^2}{(c-1)^2} = 1 - \frac{(c-d)^2}{4},$$

these would give a weight of 1 to the diagonal elements, i.e. the correct imputed values, 0.75 to the adjacent categories (to 2 if 1 or 3 is correct and to 1 and to 3 if 2 is correct) and 0 in the other cases. In this thesis, the unweighted version of κ is calculated. This is more restrictive than the weighted version but since there are only 3 different possible states the imputation method should be good enough to estimate the observed value, in particular as only 2 out of 9 combinations would have a weight of 0.

For $\kappa > 0$, the estimator provides an improvement over a pure random estimate. Note that the theoretical maximum of 1 is only reached when row and column sums are identical. In all other cases, $\max(\kappa)$ is smaller than 1. Due to large number of observations in our data the maximum is actually close to 1.

In Section 5.2.2, Figure 5.1, we showed that 56% of our missing values are missings after at least one missing occurred in the previous wave. If we use an imputation approach depending on the individual past (i.e. $f(y_{i,t}^{mis}|y_{i,t-1}^{mis}, \ldots, y_{i,t-k})$) and an unit which has more than one missing in a row, we are able to impute the first missing as usual. But if the second missing in a row is about to be imputed, we would depend this method on an imputed value, i.e. $f(y_{i,t+1}^{mis}|y_{i,t}^{imp}, y_{i,t-1}, \ldots, y_{i,t-k+1})$, which might cause more uncertainty. Therefore, we have to check how each imputation method works on longer runs of missings. This is done by calculating κ's for a 'forecasting horizon' h of up to 6 months. In the next section, we also calculate the indicators based on an imputation for different lengths of successive missings, i.e. for $h = 1, 3, 6$ and $\max(h)$, to display the differences.

5.3.4 Comparison

Tables 5.2 and 5.3 show Cohen's kappas for BS and BE and the different imputation approaches. To assess the strength of imputation we make use of Landis and Koch's (1977) rule of thumb: $\kappa < 0$ indicates no agreement, for $0 \leq \kappa \leq 0.2$ agreement is slight, $0.2 < \kappa \leq 0.4$ is fair, $0.4 < \kappa \leq 0.6$ is moderate, $0.6 < \kappa \leq 0.8$ is substantial, and $0.8 < \kappa \leq 1$ is almost perfect. First, notice that κ_{BE} is always smaller than κ_{BS}. This result is not surprising since expectations are directed to the future and therefore more difficult to assess than the present situation. It is striking that LOCF, which is a rather simple method, produces relatively good estimates. Up to six months the agreement is still moderate for BS and fair for BE. It is not surprising that the quality of imputation becomes worse as more consecutive missing values have to be imputed. Note that for the Markov Chain approaches, $\kappa_h = \kappa_{k+1}$ if $h \geq k+1$ because after k months these approaches depend only on estimated values.

In contrast to LOCF, the nearest neighbour approach clearly performs worse. Since imputations are drawn from the populations' distribution, we only replicate these probabilities. As no preceding information enters this method, all κ's are equal regardless of horizon h. Therefore, this method is only slightly better than randomisation (drawing from $\{+, =, -\}$ with equal probabilities), since specialisation by business area seems to contain only minor information. A higher specialisation is possible, but this would increase computing time enormously and it is unlikely that higher values of κ compared to LOCF can be obtained. All proportional odds models perform relatively well but on average they are not better than LOCF. In addition, all of these models are restricted to an imputation of the first month a missing value occurs. A good performance (for BS) is obtained when using the Markov Chain approach and taking the mode.[7] In general, drawing from the distribution of the calculated stochastic matrices is always worse than taking the mode. For BS, the best result can be achieved by using the modes of the Markov Chains of order 2. Although LOCF performs slightly better compared to MC2 (M) for $h = 2, 3$ and 4, MC2 (M) has the main advantage that for $h \geq 3$ it imputes 67% of our data correct although it only depends on estimated values. However, a higher differentiation by including more months into the stochastic matrices does not lead to a better imputation performance. Therefore, only MC2 (M) provides an improvement over LOCF as it seems to reflect the current dynamics better. The joint distribution evaluated by the `Amelia II` package seems to include even more uncertainty as the MCk (D) approaches and also performs worse. For BE, none of the other approaches outperforms LOCF. For this variable, it seems to be hard to find a real structured model. However, the reader should keep in mind that imputation models are predictive and not causal (Honaker et al., 2012), so every imputation approach is only measured by its estimation performance.

5.4 Results

5.4.1 Visual inspection

After evaluating the performance of the different imputation approaches, it is now assumed that *the data set to be MAR given that preceding observations contain enough information to explain the missing ones.* We impute the missing values according to MC2 (M) (for BS) and LOCF (for BE) and run the aggregation scheme displayed in Appendix C.1. Now, we are able to compare the indices with imputed missing values with the original ones. We also run imputations for four different horizons h, as mentioned in Section 5.3.3. For level 0, the indices for business situation, business expectations and the composed business climate are displayed in Figure 5.2. For industry, construction, retail and whole sale trade the indicators are

[7] As the fraction of 'equal' answers is the highest in nearly all of the months, we did not calculate MC1 (M) because this is equal to LOCF. If more than one state is the mode, then it is randomly drawn from these states.

Abbr.	Method	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$
JD	Joint Distribution	0.222 (48%)	0.008 (34%)	0.007 (34%)	0.011 (34%)	0.001 (33%)	0.010 (34%)
LOCF	Last observation carried forward	0.668 (78%)	0.590 (73%)	0.542 (69%)	0.504 (67%)	0.476 (65%)	0.452 (63%)
MC1 (D)	Markov Chains (last month, distribution)	0.492 (66%)	0.238 (49%)	0.238 (49%)	0.238 (49%)	0.238 (49%)	0.238 (49%)
MC2 (D)	Markov Chains (last 2 months, distribution)	0.470 (65%)	0.344 (56%)	0.179 (45%)	0.179 (45%)	0.179 (45%)	0.179 (45%)
MC3 (D)	Markov Chains (last 3 months, distribution)	0.489 (66%)	0.389 (59%)	0.008 (34%)	0.004 (34%)	0.004 (34%)	0.004 (34%)
MC4 (D)	Markov Chains (last 4 months, distribution)	0.044 (36%)	0.035 (36%)	0.036 (36%)	0.035 (36%)	0.033 (36%)	0.033 (36%)
MC2 (M)	Markov Chains (last 2 months, mode)	0.674 (78%)	0.570 (71%)	0.501 (67%)	0.501 (67%)	0.501 (67%)	0.501 (67%)
MC3 (M)	Markov Chains (last 3 months, mode)	0.655 (77%)	0.581 (72%)	0.240 (49%)	0.212 (47%)	0.212 (47%)	0.212 (47%)
MC4 (M)	Markov Chains (last 4 months, mode)	0.391 (59%)	0.348 (57%)	0.307 (54%)	0.256 (50%)	0.219 (48%)	0.219 (48%)
NN-II	Nearest Neighbour (level II)	0.066 (38%)	0.066 (38%)	0.066 (38%)	0.066 (38%)	0.066 (38%)	0.066 (38%)
NN-III	Nearest Neighbour (level III)	0.093 (40%)	0.093 (40%)	0.093 (40%)	0.093 (40%)	0.093 (40%)	0.093 (40%)
POM-IND	Proportional odds model (industry)	0.640 (76%)	-	-	-	-	-
POM-CON	Proportional odds model (construction)	0.700 (84%)	-	-	-	-	-
POM-TRA	Proportional odds model (trade)	0.552 (72%)	-	-	-	-	-

Table 5.2: Overview of imputation methods for variable *business situation* (BS): Cohens kappas and the fraction of correct imputed values (in brackets under the values of κ) by horizon h for the whole time period (1994-2009). For Markov Chains approaches, (M) denotes usage of the mode whereas (D) denotes drawing from the calculated distribution. κ's for all probabilistic methods (JD, MC (D) and NN) are calculated by the average of 5 replications.

Abbr.	Method	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$
JD	Joint Distribution	0.122 (41%)	0.003 (34%)	0.003 (34%)	0.007 (34%)	0.001 (33%)	0.013 (34%)
LOCF	Last observation carried forward	0.544 (70%)	0.435 (62%)	0.398 (60%)	0.358 (57%)	0.329 (55%)	0.308 (54%)
MC1 (D)	Markov Chains (last month, distribution)	0.295 (53%)	0.150 (43%)	0.150 (43%)	0.150 (43%)	0.150 (43%)	0.150 (43%)
MC2 (D)	Markov Chains (last 2 months, distribution)	0.118 (41%)	0.087 (39%)	0.033 (36%)	0.033 (36%)	0.033 (36%)	0.033 (36%)
MC3 (D)	Markov Chains (last 3 months, distribution)	0.124 (42%)	0.102 (40%)	0.006 (34%)	0.000 (33%)	0.000 (33%)	0.000 (33%)
MC4 (D)	Markov Chains (last 4 months, distribution)	0.013 (34%)	0.008 (34%)	0.010 (34%)	0.011 (34%)	0.010 (34%)	0.010 (34%)
MC2 (M)	Markov Chains (last 2 months, mode)	0.173 (45%)	0.155 (44%)	0.038 (36%)	0.038 (36%)	0.038 (36%)	0.038 (36%)
MC3 (M)	Markov Chains (last 3 months, mode)	0.172 (45%)	0.158 (44%)	0.062 (37%)	0.038 (36%)	0.038 (36%)	0.038 (36%)
MC4 (M)	Markov Chains (last 4 months, mode)	0.098 (40%)	0.089 (39%)	0.077 (38%)	0.051 (37%)	0.021 (35%)	0.021 (35%)
NN-II	Nearest Neighbour (level II)	0.035 (36%)	0.035 (36%)	0.035 (36%)	0.035 (36%)	0.035 (36%)	0.035 (36%)
NN-III	Nearest Neighbour (level III)	0.056 (37%)	0.056 (37%)	0.056 (37%)	0.056 (37%)	0.056 (37%)	0.056 (37%)
POM-IND	Proportional odds model (industry)	0.464 (66%)	-	-	-	-	-
POM-CON	Proportional odds model (construction)	0.320 (59%)	-	-	-	-	-
POM-TRA	Proportional odds model (trade)	0.310 (58%)	-	-	-	-	-

Table 5.3: Overview of imputation methods for variable *business expectations* (BE): Cohens kappas and the fraction of correct imputed values (in brackets under the values of κ) by horizon h for the whole time period (1994-2009). For Markov Chain approaches, (M) denotes usage of the mode whereas (D) denotes drawing from the calculated distribution. κ's for all probabilistic methods (JD, MC (D) and NN) are calculated by the average of 5 replications.

shown in Figures C.1-C.4.

It can easily be seen that the difference between both indices is small. The maximum difference is about 0.02. As we can not display all time series for all of the sublevels, we draw boxplots for the absolute differences according to level and horizon in Figure 5.3. In general, the absolute differences increase with the level. This is not surprising as the number of observations get lower and the imputed firms obtain more weight in the subgroups' indicators. Nevertheless, the maximum difference found in our data is around 0.15. Also, the difference rises with a higher horizon h. As more missing values are imputed, the average difference between two indicators increases. However, as we also imputed up to $h = \max(h)$, one can see that the average difference does not rise too strongly compared to h.

Even if the absolute difference is small, Figure 5.2 shows that the differences seem to depend from the underlying variable. To check this assumption and to evaluate the magnitude of the dependence from the underlying variable, we calculate Spearmans correlation coefficient ρ_{GDP} between the difference of both indicators and the growth rates of the German Gross Domestic Product, which are the most common 'expression' of the business cycle. For level 0 and $h = 1$, $\rho^0_{GDP} = 0.452$, which means that the difference is relatively strongly correlated with the business cycle. Figure 5.4 shows the boxplots according to level and horizon h. With increasing h, the correlations ρ_{GDP} do not seem to become higher. However, the correlations decrease with the levels, but this may due to a lower dependence from the business cycle in the sublevels. Since it is hard to find a time series for every business area which reflects the business cycle in this area at best, we also calculate the correlations between the differences and the imputed indicators ρ_{IND}. Figure 5.5 shows their distributions. The correlations are higher than for the correlations with the GDP and rise, on average, with horizon h. In general, the visual inspection shows that the difference is minimal but seems to be related to the underlying variable.

However, Figure 5.2 also suggests that the indicators are stretched. These effects may for example occur when the 'equal'-category is underrepresented. As these indicators are artificial by definition, such a stretch would not lead to a substantial change in interpretation as the absolute value of the indicator does not reflect a certain quantity (e.g. in contrast

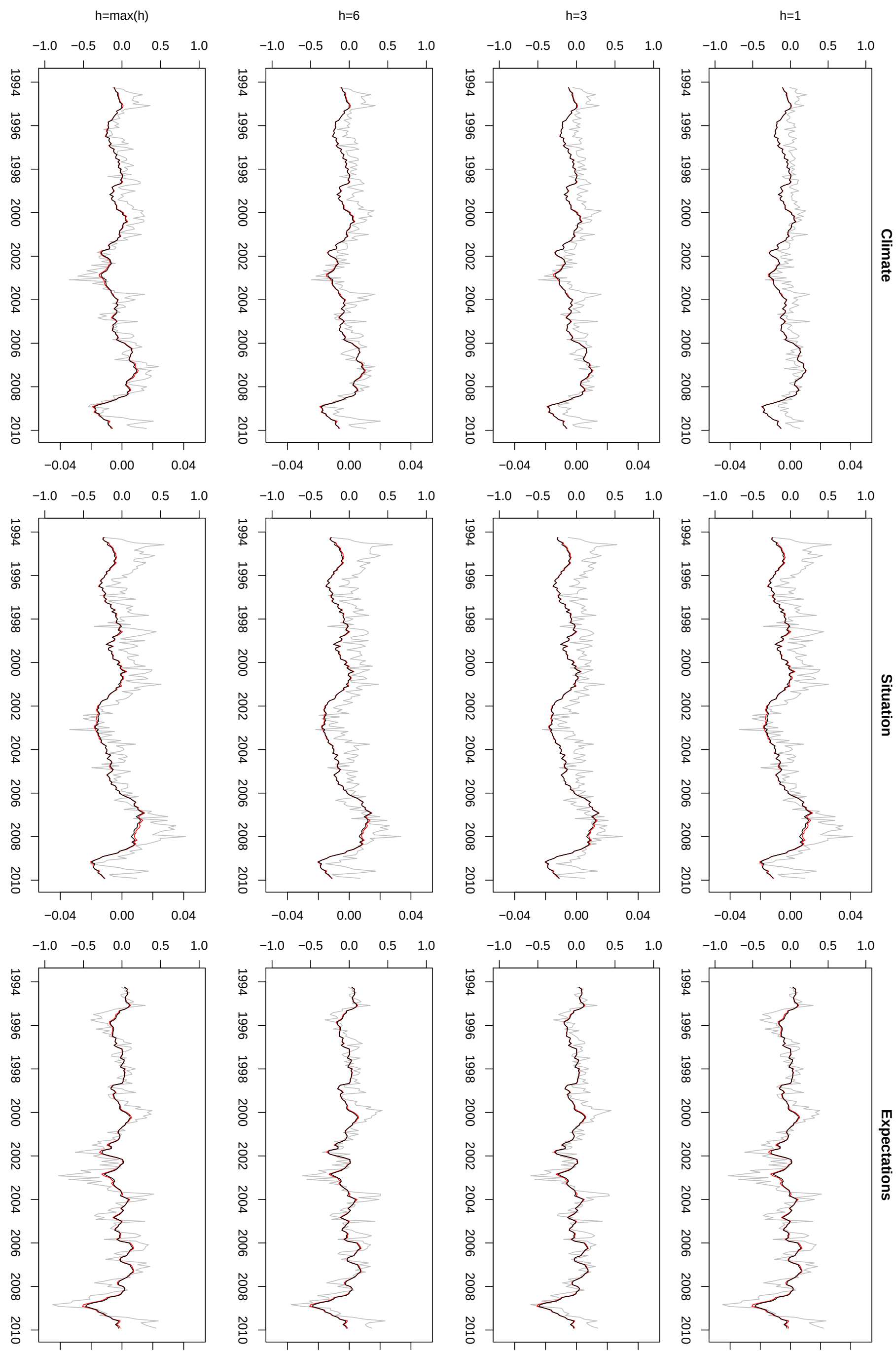

Figure 5.2: Original (black) and imputed (red) Ifo indicators and their difference (grey, right scale)

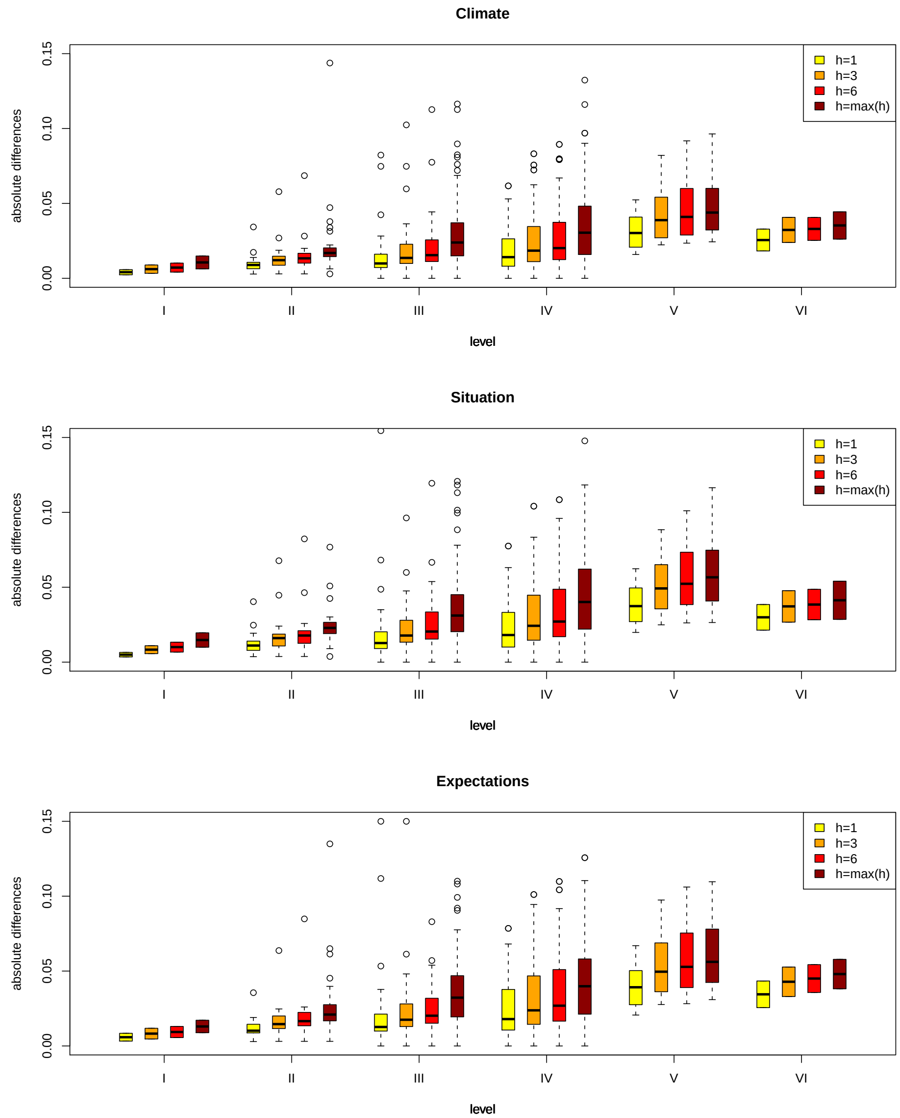

Figure 5.3: Boxplots for the distribution of the absolute differences between the original and the imputed indicators for different aggregation levels and horizons h.

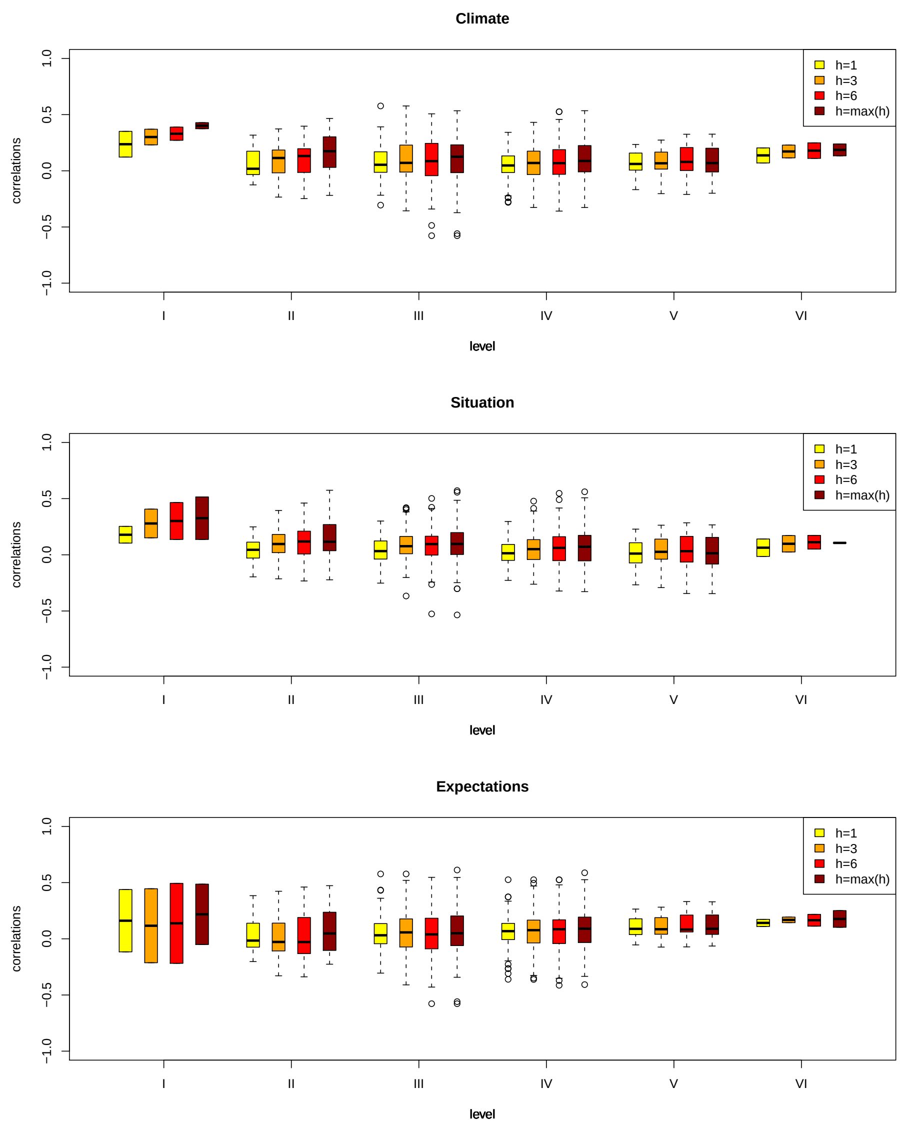

Figure 5.4: Boxplots for the distribution of ρ_{GDP} for different aggregation levels and horizons h.

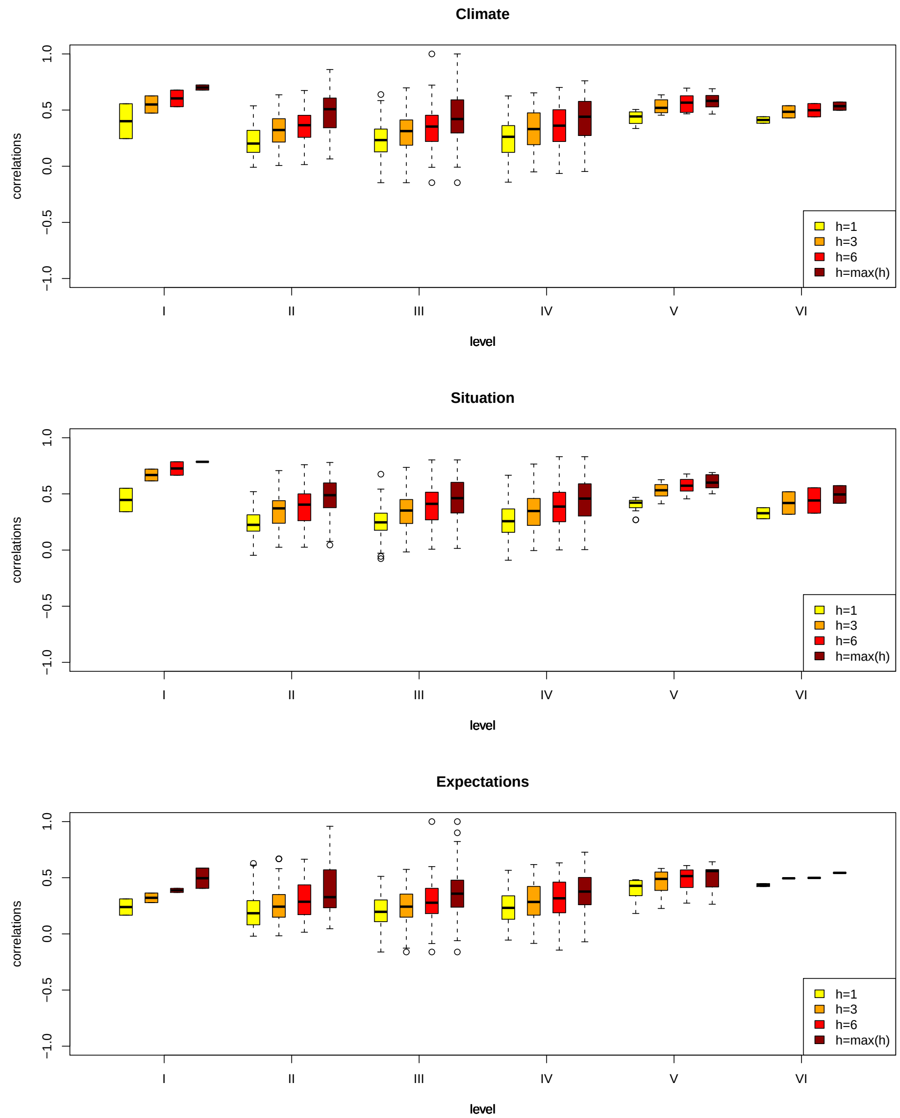

Figure 5.5: Boxplots for the distribution of ρ_{IND} for different aggregation levels and horizons h.

to the GDP). Therefore, we standardise all indicators (original as well as imputed) by

$$\tilde{z}_t = \frac{z_t - \bar{z}_t}{\sqrt{\mathrm{Var}(z_t)}}$$

for any indicator z_t. Figure C.5 shows the standardised indicators and their differences.[8] It can easily be seen that the cycle dependence of the differences has vanished after standardisation. For example, for the standardised Ifo index and $h = 1$, $\rho^0_{GDP} = -0.114$ and $\rho^0_{IND} = -0.029$. Also for the subsectors we receive similar results on average, see Figures C.10 and C.11. Of course, it is highly discussible if a standardisation leads to a 'fairer' comparison between the original and the imputed indicators. However, a more distinctive comparison may be achieved by evaluating the difference in forecasting power which is done in the following section.

5.4.2 Forecast comparison

Although it was shown in Section 5.3 that previous wave(s) contain enough information to produce relatively good estimates and the difference seems to be very small, we have to compare our indicators for their out-of-sample performance. Usually, the Ifo index is used as a leading indicator for the German GDP growth rates. Therefore, we will perform a comparison between the original and the imputed indicators to test if forecasting performance of the imputed indicator is better than the original indicator. We consider a standard autoregressive distributed lag (ADL) model

$$z^{GDP}_{t+h^*} = \alpha + \sum_{i^*=1}^{p} \phi_{i^*} z^{GDP}_{t+1-i^*} + \sum_{j^*=1}^{q} \theta_{j^*} z^{IFO}_{t+1-j^*} + \epsilon_t$$

with horizon $h^* = 1$. z^{GDP}_t denotes the quarterly growth rate of the German GDP and z^{IFO}_t the averaged quarterly Ifo index. For both variables, we allow the maximum of $p = q = 4$ lags and select the best model by

[8] The standardised indicators for industry, construction, retail and whole sale trade are shown in Figures C.6-C.9.

AIC. We receive a RMSE ratio of 0.942, i.e. the imputed indicator leads, on average, to slightly better forecasts than the original indicator. To test whether this difference is statistical significant, we perform a Giacomini-White test (Giacomini and White, 2006) which gives a p-value of 0.348. Therefore, we can conclude that the imputed indicator does not lead to significant better forecasts.

5.5 Summary and discussion

In this chapter, we developed different imputation strategies for a large business survey with time-dependent latent process (the business cycle) and ordinal outcomes. Although the missing observations in our data set were caused in nearly all of the cases due to unit and not item nonresponse, we received good estimates by using the individual past as covariates for every unit. Also the predictive power of the imputations for runs of successive missings for a single firm was evaluated. But the analysis also showed that the strength of the imputation method is not always the same for every question in the survey. Questions relevant to the recent situation seem to be imputed with more certainty than questions with respect to future developments. This is an intuitive result as the latter inherent more uncertainty. After imputing missing observations with respect to different horizons of successive months of nonresponse, we recalculated the survey outcomes. The comparison with the original indicators showed that the difference is minimal but in general increases with rising horizon and for indicators in sublevels. For the correlations with the business cycle, we also found a similar effect as for the differences, so that the correlation rises when more values are imputed. To check our results with respect to forecasting power, we also performed a comparison between the original and the imputed indicator. These results showed that the imputed indicator has a slightly better forecasting accuracy, but this effect is not significant according to a Giacomini-White test.

However, our results do not hold if the indicators are standardised, in particular cycle dependence of the differences vanishes. This may not lead to the conclusion that a nonresponse bias is not present but confirms the results in Chapter 4 that the bias of a business cycle indicator is small even for very different patterns of NMAR. So, usage of the Ifo Business Climate index for monitoring and forecasting the German economy is secured under measurement error aspects due to nonresponse. Of course, the question remains how such patterns may arise. Figure 5.6 shows the fractions of imputed values over time. For the business situation, a small cycle dependence can be seen. Another very interesting issue is that LOCF seems to introduce a strong seasonal pattern in the business expectations variable but we notice that LOCF is no explicit model and therefore the results for this variable have to be interpreted with care. However, bad and equal states are imputed considerable more often than good states by LOCF. The same, with exception of the seasonal pattern, regards to the business situation: For BS, we can see that the fractions of imputed values are very different according to the three states and are slightly cycle dependent across t which confirms the results of Chapter 3 and Harris-Kojetin and Tucker (1999). This concludes that in general the nonresponse rate increases with the business cycle but still firms more likely respond if their situation and expectations are positive. But how does this fit to our results in Section 5.4? The pattern found here reduces the amplitudes of the indicators in boom times because more equal and negative values are imputed. But it also reduces the amplitudes in bad times as the imputed 'equal' values shift the indicator upwards. Figure 5.2 shows that the difference is, on average, positive, especially for the business situation indicators. Therefore, it can be concluded that a possible nonresponse bias, i.e. the differences between the three states, is more or less stable across time but the general decision to respond seems to be slightly correlated with the cycle. This bias leads to an overestimation of the indicators' amplitudes in extreme economic times (boom or recession). Since the level of these indicators is artifical and the forecasting performance is not reduced significantly, we conclude that the possible bias pattern in our data is ignorable for this type of surveys and macro level results. However, micro level analyses as well as other surveys including quantitative information may be affected stronger by such biases.

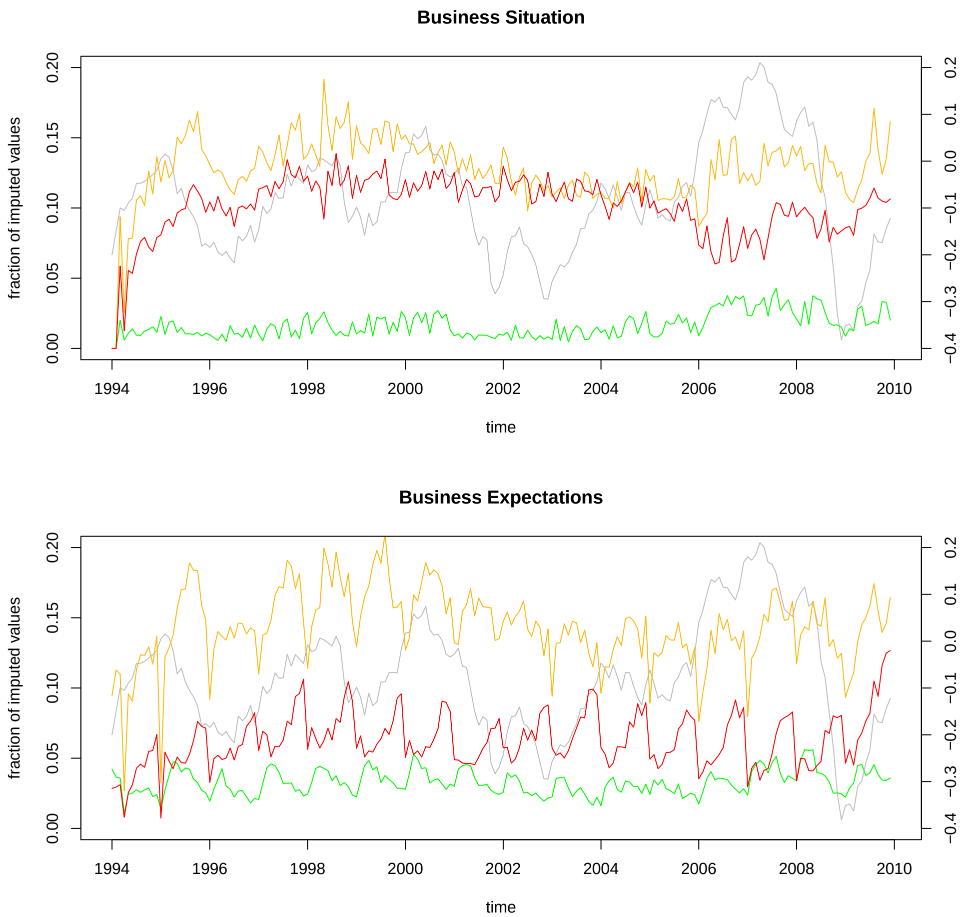

Figure 5.6: Fraction of imputed values (green line for ’+’, yellow line for ’=’ and red line for ’-’, left scale) in relation to the total number of contacted companies over time and the Ifo index (grey, right scale).

Chapter 6

Final conclusions and outlook

In this thesis, different aspects of missing data in business surveys were analysed. Chapter 3 focused on the reasons for nonresponse in the Ifo Business Survey. We showed that the decision to respond to the survey is not completely random. Differences across business parameters such as business area and size exist. In particular, larger firms have a lower probability not to respond to the survey. Also, survey-related variables, e.g. the number of questions on the questionnaire or a shorter time schedule, have an impact on the responding behaviour. A surprising fact is that strong time-dependent effects prevail. The probability not to respond has raised within the past 20 years. However, the opposite effect occurred for the firms in Eastern Germany. We found a high probability for nonresponse when the survey was introduced and a declining trend over time. Nevertheless, research has to be done to explore more information on the participation process. For example, the position of the respondent as well as a change of the responsible person in the company could also have a high impact on the decision to respond. In addition, a precise evaluation of the reasons for those participants who did not answered the survey should be done.

A more serious result of Chapter 3 was the fact that there exists a dependence of the respond behaviour from the business cycle, i.e. the interesting underlying variable. As the IBS and other tendency surveys which are constructed in a similar way calculate a so-called balance statistics to visualise the business cycle over time, this result might cause certain problems. To display considerations for the effects of nonresponse patterns on the resulting indicators, we built a theoretical framework in Chapter 4. In this chapter, we performed an extensive simulation study to simulate a wide range of nonresponse scenarios. The results of this study showed that these indicators are extremely stable towards nearly all kind of nonresponse patterns. Even for high differences in the nonresponse rates and very unrealistic scenarios, the balance statistics is robust and still holds a strong correlation with the underlying latent variable, in particular for cycle-dependent nonresponse probabilities. The source of this robustness seems to be driven by the fact that the oscillation structure is included in all three states. For example, focussing only on the change over time of one state, i.e. to look only on the fraction of the positive replies, displays the business cycle to some extent. However, some (very rare) nonresponse patterns exist which transform the final indicator in a extreme way, e.g. lead to a negative correlation or a shift in the lag/lead structure. Therefore, an extension to more than 3 levels and continuous variables would be an interesting topic.

As shown in Chapter 4, the balance statistics is generally robust towards nonresponse biases but problematic transformations by the missing data process can occur. We analysed the effect on the indicators when the missing values are estimated by imputation in Chapter 5. Various imputation methods have been analysed to evaluate the data generating process to enable good predictions on microdata level. After the estimation of the missing values, the difference between the indicators which include the missing values and the original ones is small. Although a small bias pattern seems to prevail, the imputed indicators still receive a high correlation with the German GDP growth rates and also no significant reduction in forecasting performance was measured. This supports the usage of these macro level indicators. As this was a case study for German data, other tendency surveys maybe include other bias patterns which should be analysed. Also, the results show that for subareas, the differences between the indicators become higher. Therefore, indicators for countries with a smaller number of companies may be affected more by nonresponding units.

Another interesting research field is the extension of the data structures found in Chapter 5 to create synthetic data sets (SDS) for the Ifo Business Survey. SDS are artificially created data sets from imputation models where variables which are critical in terms of privacy policy, e.g. information about location, turnover or detailed business area. Their major advantage is the fact that they are not a random draw of the original data. They reflect the data structure to a certain extent but can be made freely available for potential data users. Therefore, SDS allow to get a deeper insight into the data structure than a pure random draw. Drechsler (2011b) gives an excellent overview on the usage of imputation models for synthetic data sets. Given the fact that most of the variables in the IBS have a 3-level Likert scale, the methods used in Chapter 5 can also be used to evaluate their predictive performance on these variables.

Appendix A

Appendix for Chapter 3

The following Figure A.1 shows the participation process in business surveys developed by Willimack et al. (2002). The results for the fixed effects model are listed in Tables 3.3 (model excluding weights) and A.2 (model including weights).

A.1 Conceptual particpation framework

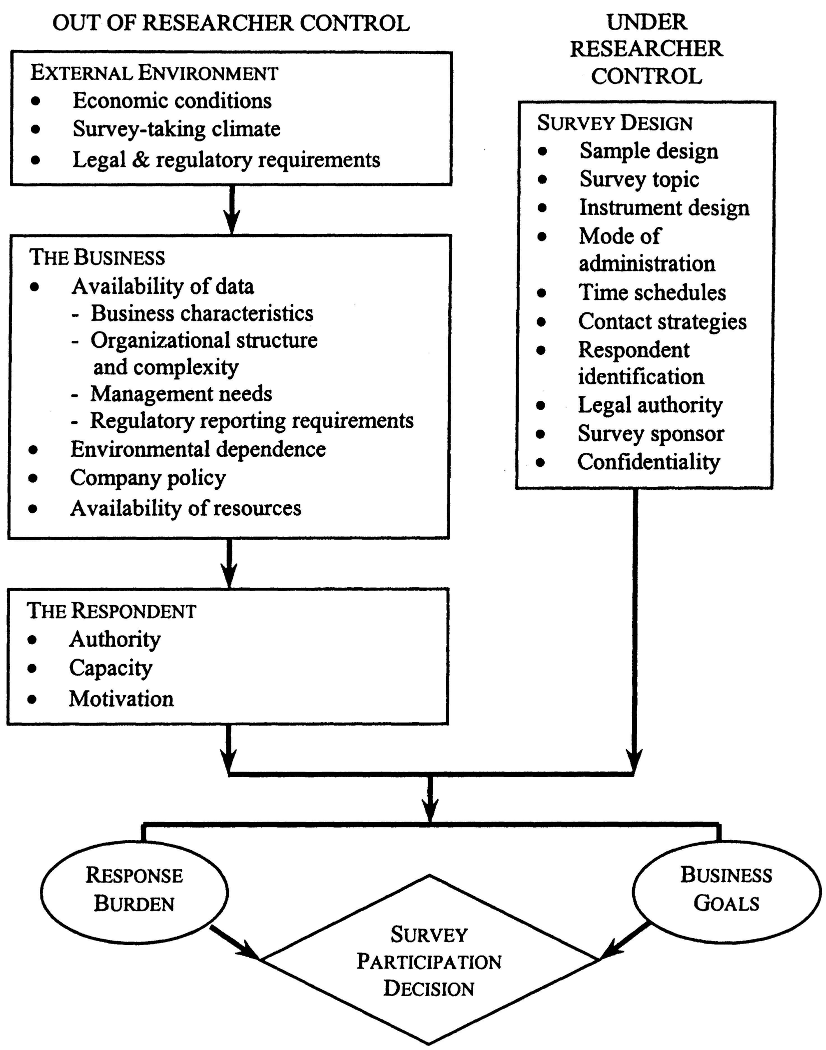

Figure A.1: Conceptual framework for the participation process in business surveys according to Willimack et al. (2002).

A.2 Results for the fixed effects model

Variable	Business Sit. coef.	Business Sit. p-value	GDP growth coef.	GDP growth p-value
Participation time	-0.001	0.000	-0.001	0.000
Calendar time	0.002	0.000	0.003	0.000
Calendar time × East	-0.011	0.000	-0.011	0.000
East	2.466	0.000	2.715	0.000
Cycle indicator	0.002	0.000	0.042	0.000
Additional survey	0.018	0.127	0.017	0.165
Number of questions	0.000	0.881	0.000	0.302
Short time schedule	0.091	0.000	0.091	0.000
Working days	-0.001	0.813	0.000	0.914
Vacation days	0.004	0.000	0.003	0.000
Size:				
Smallest	0.277	0.000	0.279	0.000
Small	0.096	0.000	0.095	0.000
Large	-0.117	0.000	-0.118	0.000
Largest	0.006	0.854	0.009	0.800
Subsector				
Food & tobacco	0.451	0.232	0.431	0.139
Textiles & textiles products	1.226	0.155	1.199	0.210
Wood	-0.856	0.022	-0.903	0.025
Pulp, paper, publishing & printing	-0.728	0.098	-0.764	0.075
Petroleum & chemical products	0.355	0.108	0.372	0.093
Rubber & plastic products	0.767	0.001	0.711	0.004
Other non-metallic mineral products	0.733	0.230	0.701	0.239
Basic metals & fabricated metal products	0.482	0.011	0.396	0.007
Machinery & equipment	-0.258	0.019	-0.397	0.000
Electrical & optical equipment	-0.179	0.158	-0.220	0.083
Transport equipment	-0.319	0.027	-0.407	0.004
Furniture & manufacture n.e.c.	1.326	0.000	1.174	0.000
Sale, maintenance and repair of motor vehicles	4.093	0.000	1.213	0.005
Wholesale trade	3.864	0.000	1.046	0.011
Retail trade	3.895	0.000	1.070	0.010

Table A.1: Estimation results of the unweighted FE model

	BUSINESS SIT.		GDP GROWTH	
VARIABLE	COEF.	P-VALUE	COEF.	P-VALUE
Participation time	-0.001	0.000	-0.001	0.000
Calendar time	0.003	0.000	0.003	0.002
Calendar time $\times$ East	-0.010	0.000	-0.010	0.000
East	1.322	0.000	1.746	0.000
Cycle indicator	0.001	0.002	0.035	0.004
Additional survey	0.011	0.000	0.014	0.000
Number of questions	0.004	0.872	0.004	0.776
Short time schedule	0.031	0.000	0.040	0.000
Working days	-0.007	0.854	-0.008	0.588
Vacation days	0.001	0.000	0.001	0.000
Size:				
Smallest	0.440	0.000	0.444	0.000
Small	0.151	0.001	0.153	0.000
Large	-0.092	0.025	-0.084	0.001
Largest	-0.033	0.745	-0.022	0.621
Subsector				
Food & tobacco	1.724	0.370	1.672	0.564
Textiles & textiles products	1.822	0.221	1.774	0.137
Wood	-1.556	0.088	-1.635	0.112
Pulp, paper, publishing & printing	-1.171	0.000	-1.190	0.000
Petroleum & chemical products	0.532	0.747	0.478	0.691
Rubber & plastic products	0.550	0.008	0.540	0.047
Other non-metallic mineral products	0.667	0.404	0.665	0.308
Basic metals & fabricated metal products	-0.694	0.009	-0.738	0.027
Machinery & equipment	-2.820	0.158	-2.988	0.000
Electrical & optical equipment	-1.145	0.027	-1.179	0.048
Transport equipment	-2.039	0.045	-2.024	0.059
Furniture & manufacture n.e.c.	-1.850	0.000	-1.884	0.000
Sale, maintenance and repair of motor vehicles	0.815	0.003	0.407	0.001
Wholesale trade	0.380	0.000	0.068	0.000
Retail trade	0.450	0.000	0.128	0.021

Table A.2: Estimation results of the weighted FE model

Appendix B

Appendix for Chapter 4

In this chapter, the calculation of the balance statistics is shown in Section B.1. Section B.2 includes the proof of $\rho_C^{obs} \neq 1$. In Section B.3, some theoretical notes on the correlation between the biased with the unbiased indicator are given. Section B.4 includes figures which display the effects of nonresponse patterns on the cycle functions $g_1(t), g_3(t)$ and $g_4(t)$. Finally, an outline on the effects of the nonresponse patterns on the forecasting performance is given in Section B.5.

B.1 Calculation of the balance statistics

As the Ifo index is based on balances and not on the calculation of the mean, we have to ensure that we obtain the same results. We define, as in Section 4.2.1, $+ \equiv 1$, $= \equiv 0$ and $- \equiv -1$. Then, let $B(t)$ be the balance statistics at time t, i.e.

$$B(t) = \frac{\sum_{i=1}^{n_t} I(s_{i,t} = 1) - \sum_{i=1}^{n_t} I(s_{i,t} = -1)}{n_t}$$

with the indicator function $I(\cdot)$. Then, the mean is defined as

$$\begin{aligned}
\mathrm{E}(B(t)) = \mathrm{E}_t(B) &= \mathrm{E}\left(\frac{\sum_{i=1}^{n_t} I(s_{i,t} = 1) - \sum_{i=1}^{n_t} I(s_{i,t} = -1)}{n_t}\right) \\
&= \frac{1}{n_t}\mathrm{E}\left(\sum_{i=1}^{n_t} I(s_{i,t} = 1) - \sum_{i=1}^{n_t} I(s_{i,t} = -1)\right) \\
&= \frac{1}{n_t}\sum_{i=1}^{n_t} \mathrm{E}[I(s_{i,t} = 1) - I(s_{i,t} = -1)] \\
&= \frac{1}{n_t} \cdot n_t \cdot [P(s_{i,t} = 1) - P(s_{i,t} = -1)] \\
&= P(s_{i,t} = 1) - P(s_{i,t} = -1) \\
&= \left[1 - \Phi\left(\frac{\tau^+ - g(t)}{\sigma}\right)\right] - \Phi\left(\frac{\tau^- - g(t)}{\sigma}\right) \\
&= \mathrm{E}_t(s).
\end{aligned}$$

B.2 Proof of $\rho_C^{obs} \neq 1$

In Section 4.2.3 we stated that $\rho_C^{obs} \neq 1$. The correlation between the observed indicator and the cycle function $\rho_C^{obs} = \rho(\mathrm{E}_t(s^{obs}), g(t)) = 1$ if

$$\begin{aligned}\mathrm{E}_t(s^{obs}) &= a \cdot g(t) + b \qquad \text{(B.1)}\\ \frac{1}{\bar{\pi}(t)} \cdot \left(\pi^+(t) \cdot \Phi^+(t) - \pi^-(t) \cdot \Phi^-(t)\right) &= a \cdot g(t) + b.\end{aligned}$$

Assuming $a \neq 0$, Equation (B.2) can be written as

$$\left(\frac{\pi^+(t)}{\bar{\pi}(t)} \cdot \Phi^+(t) = a^+ \cdot g(t) + b^+\right) \cap \left(\frac{\pi^-(t)}{\bar{\pi}(t)} \cdot \Phi^-(t) = a^- \cdot g(t) + b^-\right)$$

with $a^+, a^- \neq 0$. It is important to notice that a, a^+, a^-, b, b^+ and b^- have to be constant and therefore independent from t. Unfortunately, a perfect correlation of 1 can not be received for any NMAR process: As $\Phi^+(\cdot)$ and $\Phi^-(\cdot)$ are nonlinear if $\epsilon \not\sim U(c_l, c_u)$, linearity in $\Phi^s(\cdot)$ can only be obtained directly via the inverse of $\Phi^s(\cdot)$. As the $\pi^s(t)$ have a *multiplicative effect* on $\Phi^s(\cdot)$ in Equation (4.5), this way for correction is not possible. Another approach to obtain a linear relationship would be to constrain $\pi^+(t) = a \cdot g(t) + b$ or $\pi^-(t) = a \cdot g(t) + b$ and to oblige that $\Phi^+(t)$ *and* $\Phi^-(t)$ are reduced to time-invariant constants via division by $\bar{\pi}(t)$. But because we have stated above that $\pi^+(t)$ or $\pi^-(t)$ have to be a linear transformation of $g(t)$ and they are also included in $\bar{\pi}(t)$, $\Phi^+(t)$ and $\Phi^-(t)$ can not be changed simultaneously into time-invariant effects. Therefore, it is not possible to obtain a perfect correlation regardless of the type of $\pi^s(t)$. However, this fact does not exclude the possibility of receiving a higher correlation for a biased indicator $\mathrm{E}_t(s^{obs})$. In these cases, the functions $\frac{\pi^+(t)}{\bar{\pi}(t)}$ and $\frac{\pi^-(t)}{\bar{\pi}(t)}$ partially eliminate the bias of $\Phi(\cdot)$.

B.3 Correlation with the unbiased indicator

Similar to the statements mentioned in Sections 4.2.3 and B.2, we can derive conditions for a perfect correlation $\rho_E^{obs} = \rho(\mathrm{E}_t(s^{obs}), \mathrm{E}_t(s))$ between the observed mean function $\mathrm{E}_t(s^{obs})$ and the 'ideal' unbiased mean function $\mathrm{E}_t(s)$. The correlation $\rho(\mathrm{E}_t(s^{obs}), \mathrm{E}_t(s)) = 1$ holds if

$$\begin{aligned} \mathrm{E}_t(s^{obs}) &= a \cdot \mathrm{E}_t(s) + b \\ \frac{1}{\bar{\pi}(t)} \cdot \left(\pi^+(t) \cdot \Phi^+(t) - \pi^-(t) \cdot \Phi^-(t)\right) &= a \cdot [\Phi^+(t) - \Phi^-(t)] + b \\ \frac{\pi^+(t)}{\bar{\pi}(t)} \cdot \Phi^+(t) - \frac{\pi^-(t)}{\bar{\pi}(t)} \cdot \Phi^-(t) &= a \cdot \Phi^+(t) - a \cdot \Phi^-(t) + b. \end{aligned}$$

As before, a and b have to be time-independent. This is only the case, if

$$\pi^+(t) = \pi^-(t) \quad \text{and} \quad \frac{\pi^+(t)}{\bar{\pi}(t)} \perp t \quad \text{and} \quad \frac{\pi^-(t)}{\bar{\pi}(t)} \perp t.$$

Since $\bar{\pi}(t) = \sum_s \pi^s(t) \cdot \Phi^s(t)$ and $\sum_s \Phi^s(t) = 1, \forall t$, we can conclude that $\rho(\mathrm{E}_t(s^{obs}), \mathrm{E}_t(s)) = 1$ holds, if $\pi^+(t) = \pi^=(t) = \pi^-(t)$, i.e. nonresponse is MAR. This result is essential: It shows that any type of NMAR would decrease this type of correlation and therefore the linear correlation between $\mathrm{E}_t(s^{obs})$ and $\mathrm{E}_t(s)$.

B.4 Effects of bias patterns

The following figures show the effects of different bias patterns on the cycle function $g_1(t)$ (Figures B.1 and B.2), $g_3(t)$ (Figures B.3 and B.4) and $g_4(t)$ (Figures B.5 and B.6). In addition, the scatterplots for correlations $\tilde{\rho}_E$ and dispersions v are shown in Figure B.7.

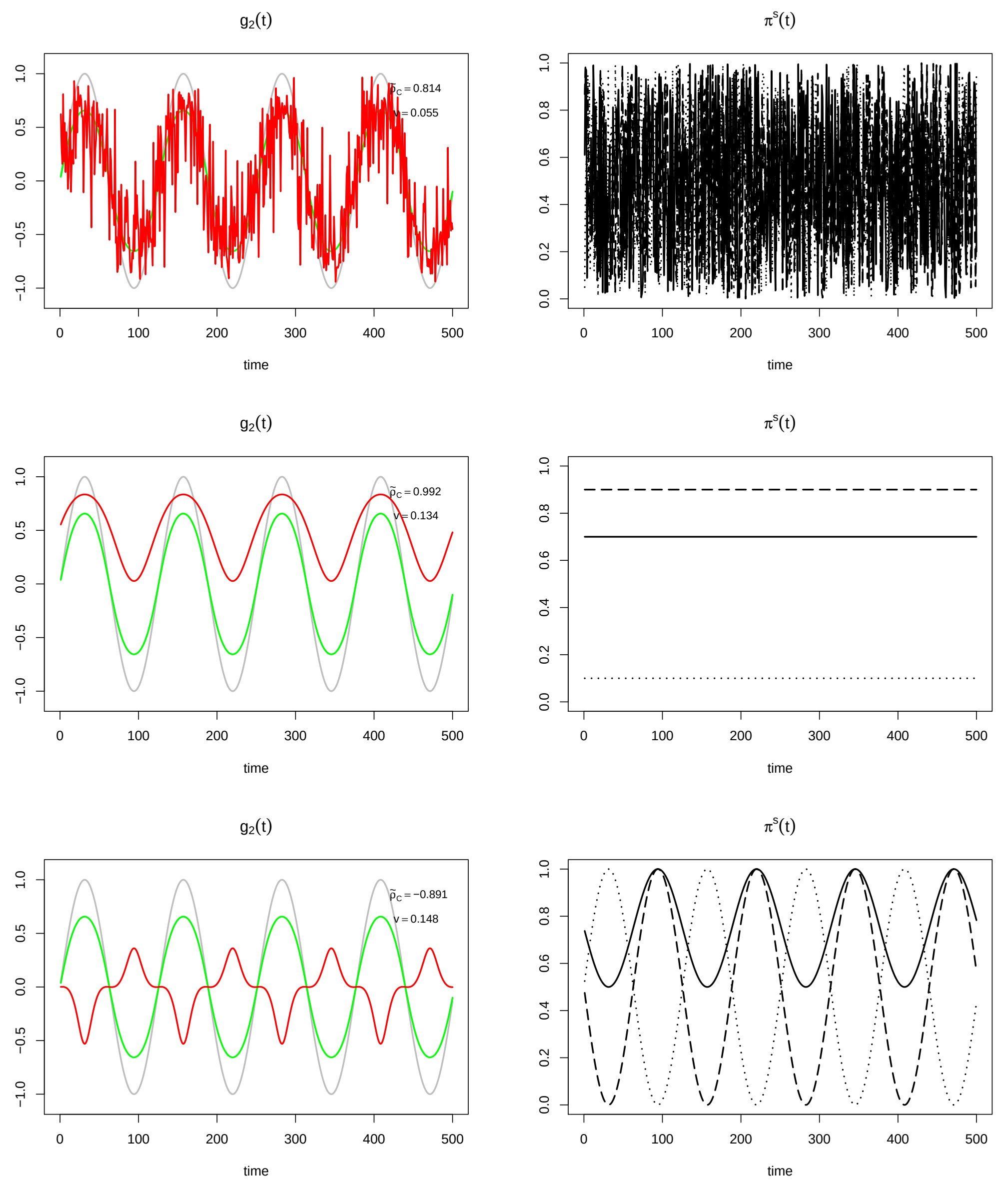

Figure B.1: Effects of bias patterns on cycle function $g_1(t)$ - part I. Left: cycle function $g_1(t)$ (—), unbiased mean function $\mathrm{E}_t(s)$ (—) and observed mean function $\mathrm{E}_t(s^{obs})$ (—). Right: acceptance rates $\pi^+(t)$ (- - -), $\pi^=(t)$ (—) and $\pi^-(t)$ ($\cdots$). From top to bottom: Random, fixed (with $\pi^+(t) = 0.9$, $\pi^=(t) = 0.7$ and $\pi^-(t) = 0.1$) and cycle-dependent (with $\pi^+(t) = 1 - \pi^C_{2,1}(t)$, $\pi^=(t) = 1 - \pi^C_{2,2}(t)$ and $\pi^-(t) = \pi^C_{2,1}(t)$) patterns.

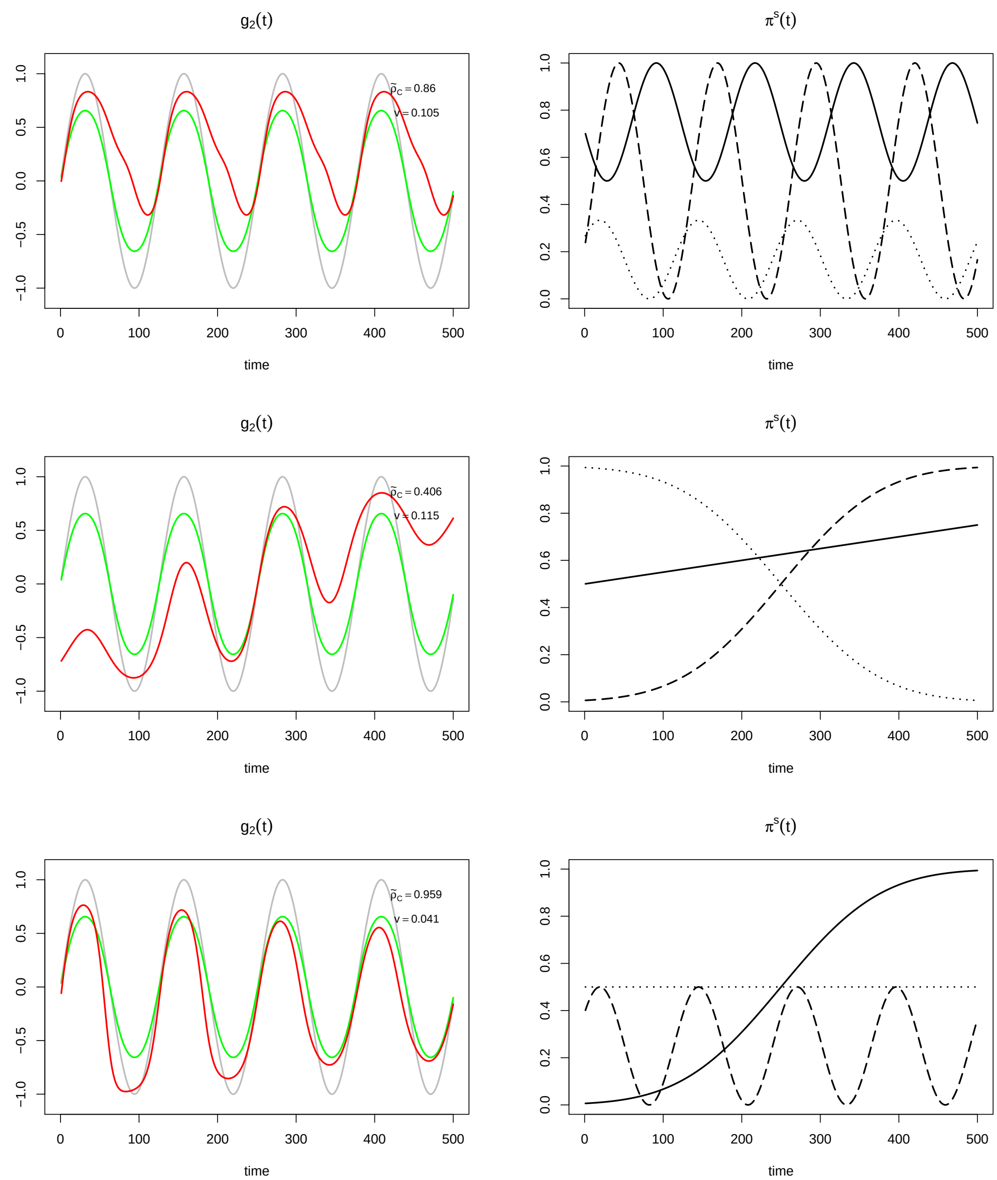

Figure B.2: Effects of bias patterns on cycle function $g_1(t)$ - part II. Left: cycle function $g_1(t)$ (—), unbiased mean function $\mathrm{E}_t(s)$ (—) and observed mean function $\mathrm{E}_t(s^{obs})$ (—). Right: acceptance rates $\pi^+(t)$ (- - -), $\pi^=(t)$ (—) and $\pi^-(t)$ ($\cdots$). From top to bottom: Cycle-shifted (with $\pi^+(t) = \pi^S_{2,1}(t-12)$, $\pi^=(t) = 1 - \pi^S_{2,2}(t+3)$ and $\pi^-(t) = \pi^S_{2,3}(t+12)$), monotone (with $\pi^+(t) = \pi^M_{100}(t)$, $\pi^=(t) = \pi^L_{2000}(t)$ and $\pi^-(t) = 1 - \pi^M_{100}(t)$) and mixed (with $\pi^+(t) = \pi^S_{2,2}(t+12)$, $\pi^=(t) = \pi^M_{100}(t)$ and $\pi^-(t) = 0.5$) patterns.

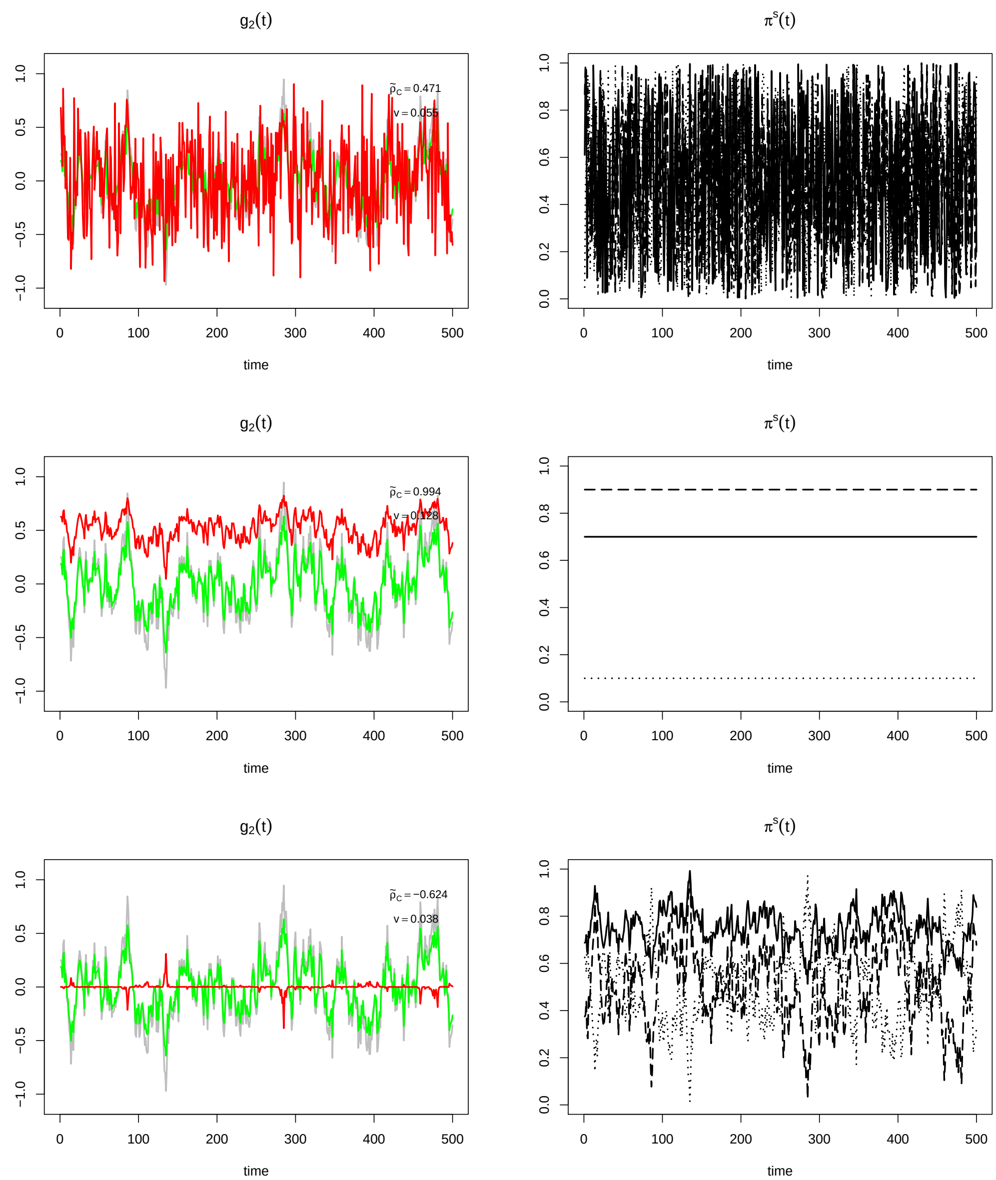

Figure B.3: Effects of bias patterns on cycle function $g_3(t)$ - part I. Left: cycle function $g_3(t)$ (—), unbiased mean function $\mathrm{E}_t(s)$ (—) and observed mean function $\mathrm{E}_t(s^{obs})$ (—). Right: acceptance rates $\pi^+(t)$ (- - -), $\pi^=(t)$ (—) and $\pi^-(t)$ $(\cdots)$. From top to bottom: Random, fixed (with $\pi^+(t) = 0.9$, $\pi^=(t) = 0.7$ and $\pi^-(t) = 0.1$) and cycle-dependent (with $\pi^+(t) = 1 - \pi^C_{2,1}(t)$, $\pi^=(t) = 1 - \pi^C_{2,2}(t)$ and $\pi^-(t) = \pi^C_{2,1}(t)$) patterns.

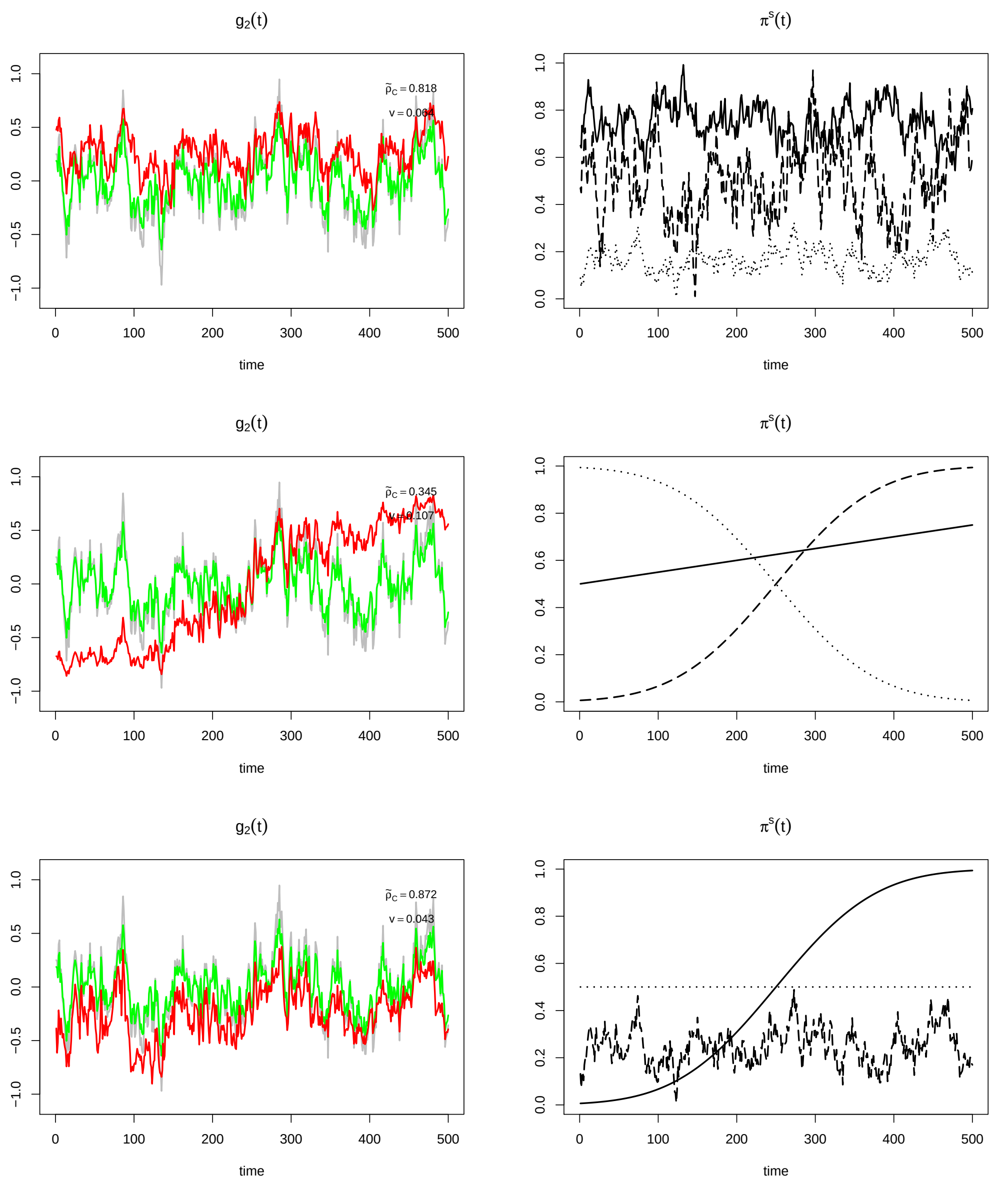

Figure B.4: Effects of bias patterns on cycle function $g_3(t)$ - part II. Left: cycle function $g_3(t)$ (—), unbiased mean function $\mathrm{E}_t(s)$ (—) and observed mean function $\mathrm{E}_t(s^{obs})$ (—). Right: acceptance rates $\pi^+(t)$ (- - -), $\pi^=(t)$ (—) and $\pi^-(t)$ ($\cdots$). From top to bottom: Cycle-shifted (with $\pi^+(t) = \pi^S_{2,1}(t-12)$, $\pi^=(t) = 1 - \pi^S_{2,2}(t+3)$ and $\pi^-(t) = \pi^S_{2,3}(t+12)$), monotone (with $\pi^+(t) = \pi^M_{100}(t)$, $\pi^=(t) = \pi^L_{2000}(t)$ and $\pi^-(t) = 1 - \pi^M_{100}(t)$) and mixed (with $\pi^+(t) = \pi^S_{2,2}(t+12)$, $\pi^=(t) = \pi^M_{100}(t)$ and $\pi^-(t) = 0.5$) patterns.

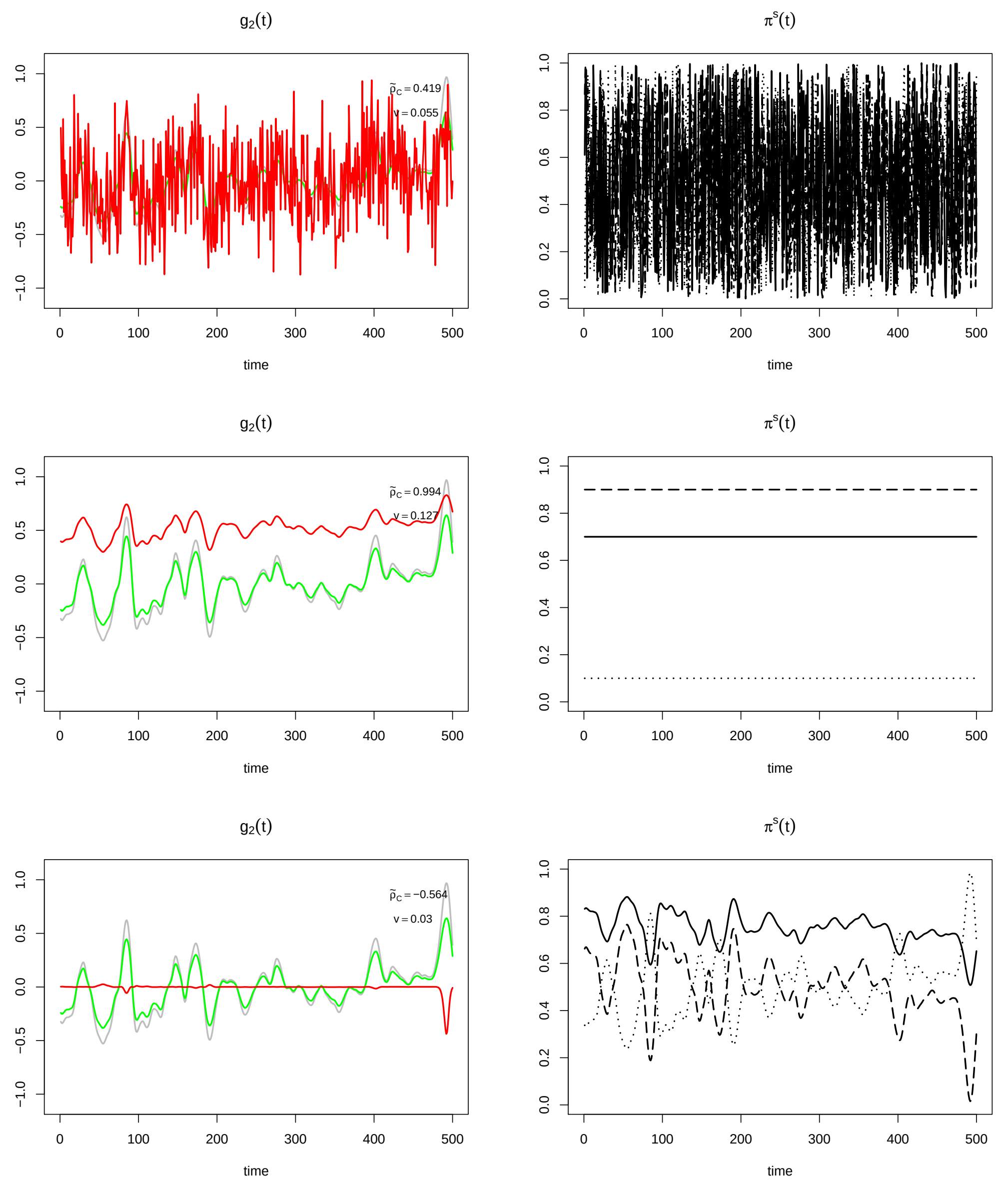

Figure B.5: Effects of bias patterns on cycle function $g_4(t)$ - part I. Left: cycle function $g_4(t)$ (—), unbiased mean function $\mathrm{E}_t(s)$ (—) and observed mean function $\mathrm{E}_t(s^{obs})$ (—). Right: acceptance rates $\pi^+(t)$ (- - -), $\pi^=(t)$ (—) and $\pi^-(t)$ ($\cdots$). From top to bottom: Random, fixed (with $\pi^+(t) = 0.9$, $\pi^=(t) = 0.7$ and $\pi^-(t) = 0.1$) and cycle-dependent (with $\pi^+(t) = 1 - \pi^C_{2,1}(t)$, $\pi^=(t) = 1 - \pi^C_{2,2}(t)$ and $\pi^-(t) = \pi^C_{2,1}(t)$) patterns.

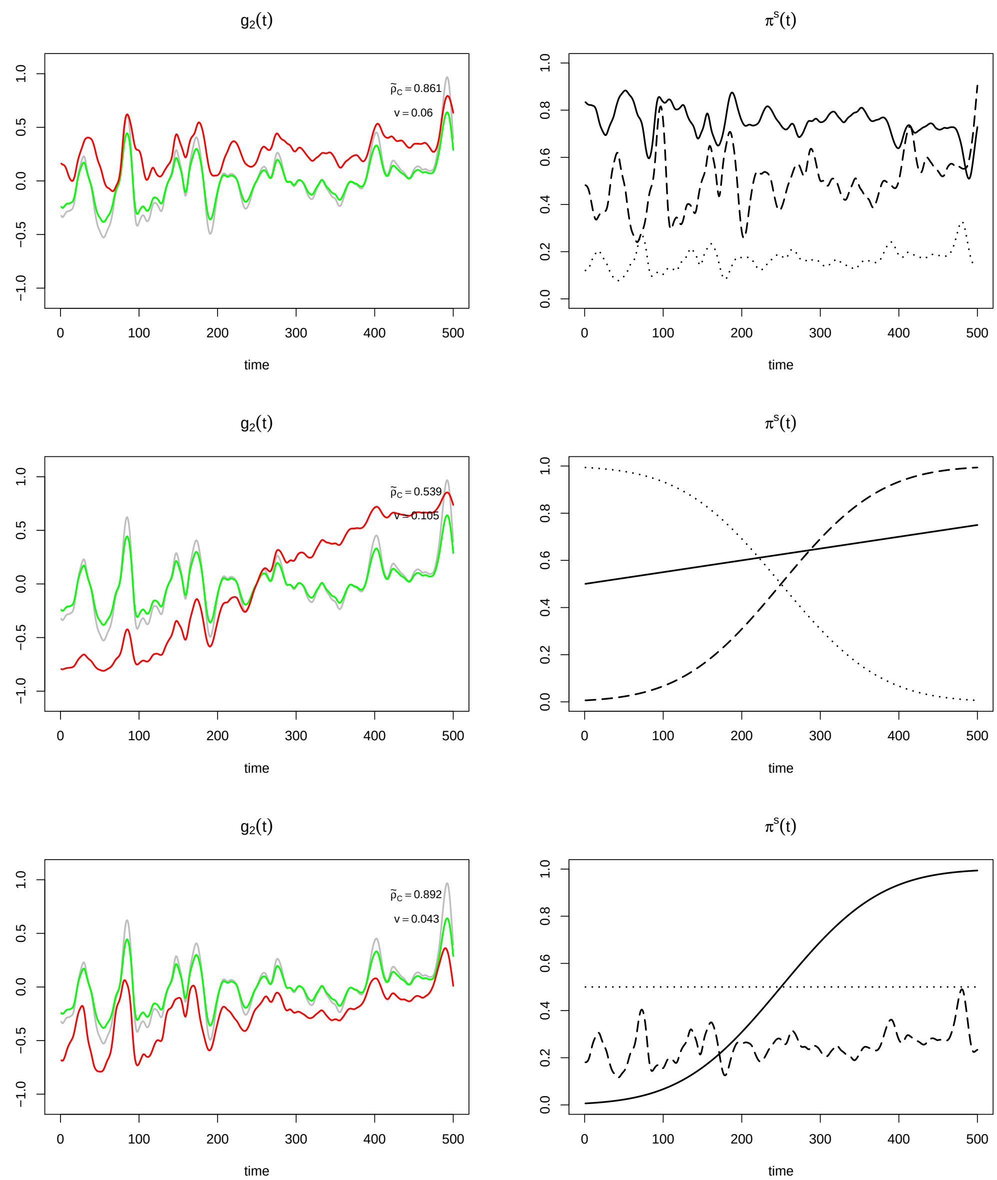

Figure B.6: Effects of bias patterns on cycle function $g_4(t)$ - part II. Left: cycle function $g_4(t)$ (—), unbiased mean function $\mathrm{E}_t(s)$ (—) and observed mean function $\mathrm{E}_t(s^{obs})$ (—). Right: acceptance rates $\pi^+(t)$ (- - -), $\pi^=(t)$ (—) and $\pi^-(t)$ ($\cdots$). From top to bottom: Cycle-shifted (with $\pi^+(t) = \pi^S_{2,1}(t-12)$, $\pi^=(t) = 1 - \pi^S_{2,2}(t+3)$ and $\pi^-(t) = \pi^S_{2,3}(t+12)$), monotone (with $\pi^+(t) = \pi^M_{100}(t)$, $\pi^=(t) = \pi^L_{2000}(t)$ and $\pi^-(t) = 1 - \pi^M_{100}(t)$) and mixed (with $\pi^+(t) = \pi^S_{2,2}(t+12)$, $\pi^=(t) = \pi^M_{100}(t)$ and $\pi^-(t) = 0.5$) patterns.

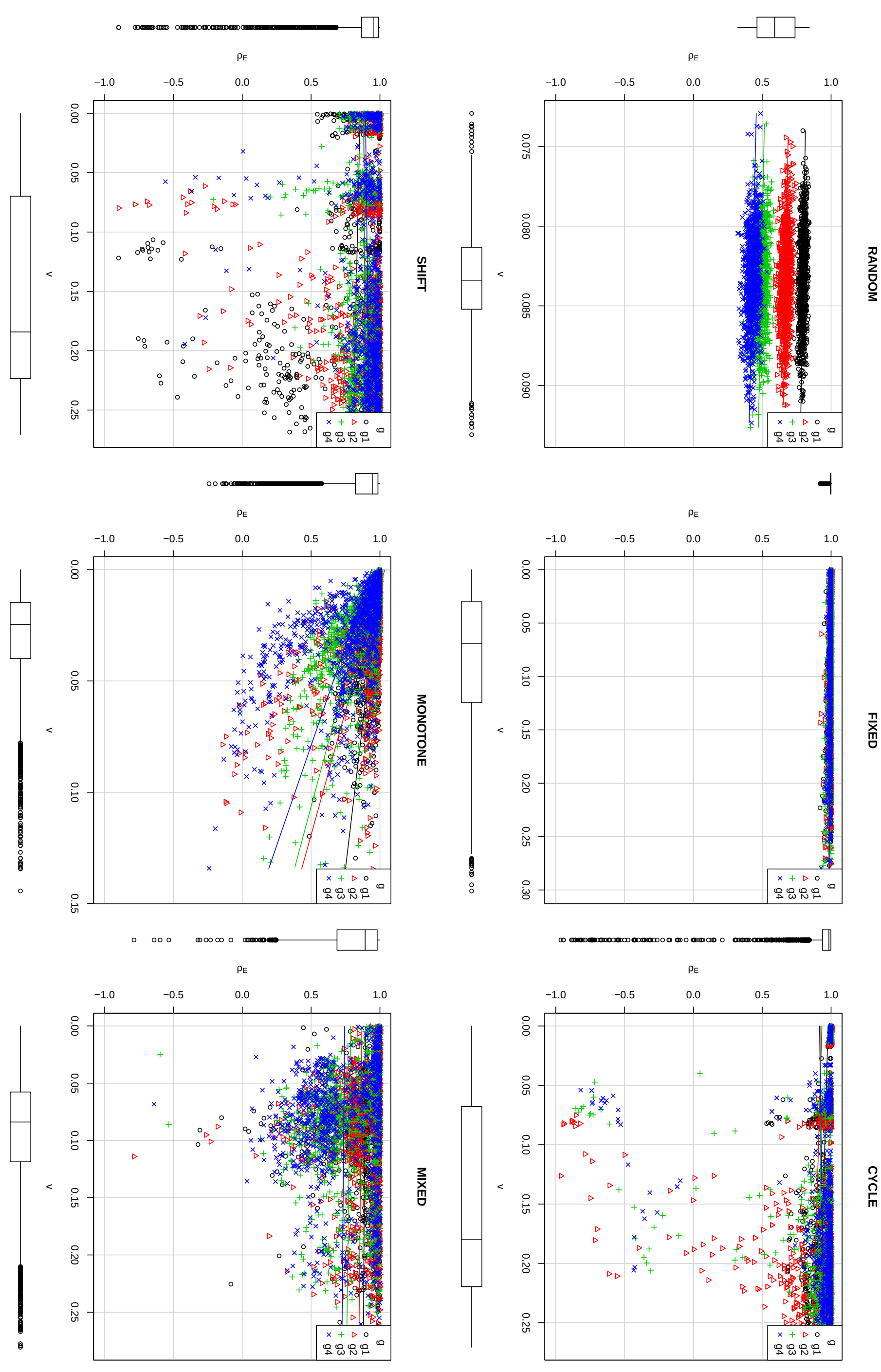

Figure B.7: Scatterplots for correlations $\tilde{\rho}_E$ and dispersions v for 6 different types of acceptance rates $\pi^s(t)$ as defined in Section 4.3.3.

B.5 Effects of nonresponse patterns on forecast performance

B.5.1 Definitions

In this section, it is analysed how nonresponse patterns might affect the forecasting performance of balance statistics indicators. In contrast to the correlation analysis in Section 4.3.3, we have to make some adjustments in order to enable sensible forecasting comparisons. First, we have to notice that the cycle functions $g_1(t)$, $g_2(t)$ and $g_4(t)$ display very ideal cases of a business cycle which is not actually observed in reality. Second, empirical analyses have shown that these smooth functions are not suitable for a forecasting comparison. We therefore focus on AR-processes as function $g_3(t)$ which come closer to real-life time series. To obtain more detailed results, we define cycle functions $g(t)$ based on AR(1)-processes

$$g_{\beta^{AR}}(t) \quad := \quad \beta^{AR} \cdot g_{\beta^{AR}}(t-1) + u_t$$

with 11 different auto-correlation factors $\beta^{AR} \in \{0.1, 0.2, 0.3, \ldots, 0.9, 0.95, 0.99\}$ in order to reflect a wide range of time series from very noisy ($g_{0.1}(t)$) to very smooth ($g_{0.99}(t)$). To stabilise our analysis further, we repeat this step 50 times, i.e. we create 50 different cycle functions for each β^{AR}-setting which leads to 11 $\times$ 50 = 550 cycle functions. Figure B.8 shows the 11 time series for one replication. As in Section 4.3.1, we standardise all of these functions to $g_{\beta^{AR}}(t) \in [-1, 1]$ and set the thresholds τ^s to $\tau^+ = 1/3$ and $\tau^- = -1/3$. For each function $g_{\beta^{AR}}(t)$ and NMAR pattern, we draw $z = 1000$ response functions $\pi^s(t)$ as in Section 4.3.2.

Due to the fact that observed indicator $\mathrm{E}_t(s^{obs})$ was defined in Equation (4.5) as an expected value which is not affected by any error term, we have to change the simulation setup compared with those in Section 4.3. First, the observation probabilities $\frac{\pi^s(t)}{\bar{\pi}(t)} \cdot \Phi^s(t)$ for the three states are calculated. Then, $n \in \{500, 1000, 2000, 5000\}$ observations are drawn from this distribution to analyse the effect of different number of observations. Finally, the balance statistics

$$B(t) = \frac{\sum_{i=1}^{n_t} I(s_{i,t} = 1) - \sum_{i=1}^{n_t} I(s_{i,t} = -1)}{n_t}$$

is calculated as in defined in Section B.1. As we are interested in a comparison of the unbiased with a biased indicator, we also calculate an unbiased balance statistics indicator.

To sum up, the simulation setup to generate the balance statistics indicators is as follows:

Step 1: Draw 11 different AR(1)-processes with an auto-correlation factor $\beta^{AR} \in \{0.1, 0.2, \ldots, 0.9, 0.95, 0.99\}$ which define $g_{\beta^{AR}}(t)$ and calculate the unbiased balance statistics indicator $B_{\beta^{AR}}^{unbiased}(t)$.

Step 2: For each $g_{\beta^{AR}}(t)$, each NMAR pattern (fixed, monotone, cycle-dependent, cycle-shifted and mixed) and each draw 1000 acceptance rates functions $\pi^s(t)$ for each state s.

Step 3: Calculate the observation probabilities $\frac{\pi^s(t)}{\bar{\pi}(t)} \cdot \Phi^s(t)$ for each state s and time t and draw n observations from this distribution.

Step 4: Calculate the biased balance statistics indicator $B_{\beta^{AR}}^{biased}(t)$.

Step 5: Repeat steps 1-4 50 times.

Step 6: Repeat steps 1-5 4 times for different number of observations $n \in \{500, 1000, 2000, 5000\}$.

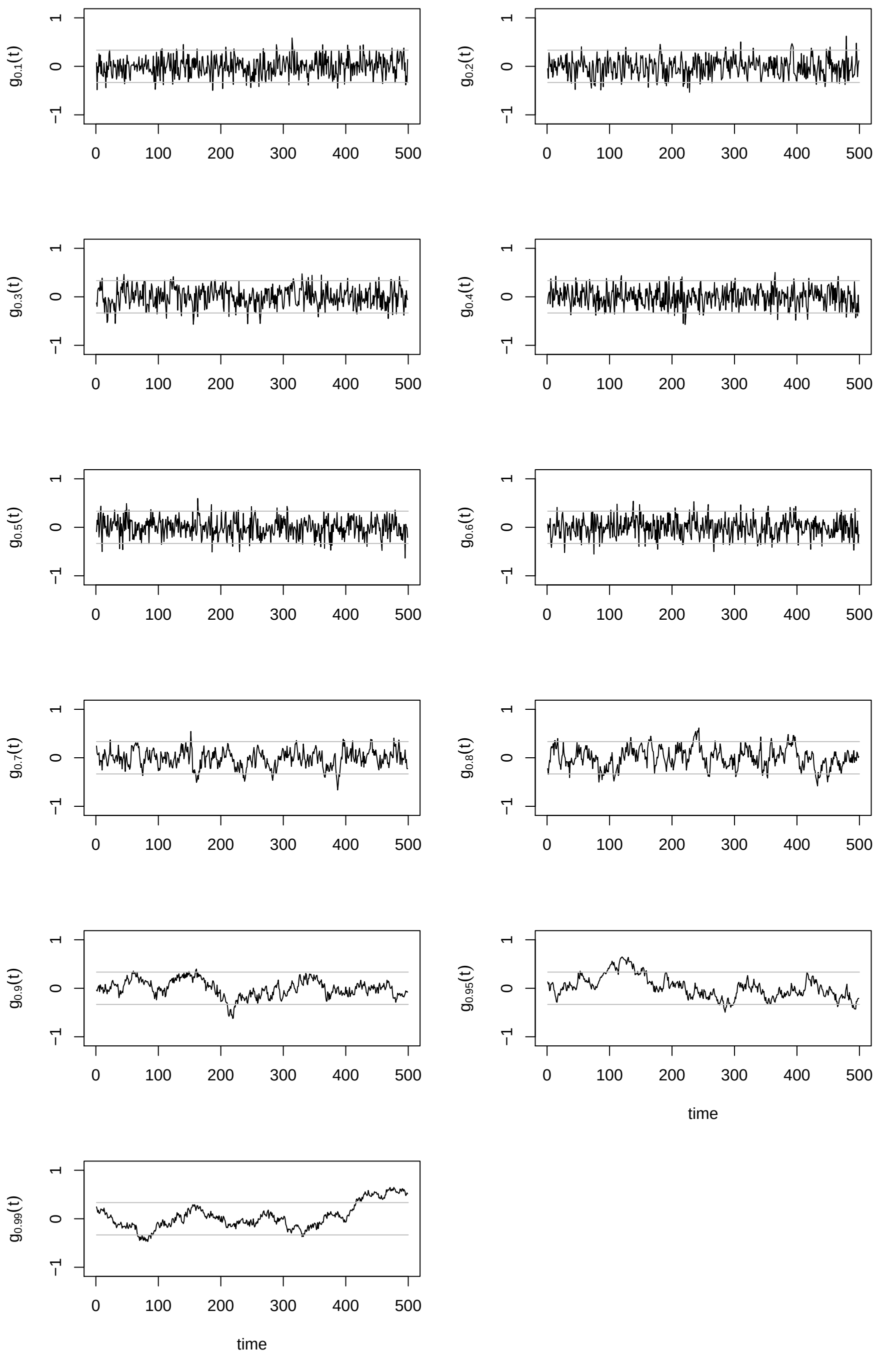

Figure B.8: Cycle functions $g_{\beta^{AR}}(t)$ (in black) for the 11 parameters β^{AR} from the first replication with fixed thresholds τ^s (in grey).

B.5.2 Forecast model

Due to the fact that balance statistics indicators $B(t)$ do not inherent any kind of leading information[1], we assume that the t-th value of the balance statistics indicator $B(t)$ (e.g. the Ifo Business Climate Index) is published earlier than the t-th value of $g(t)$ (e.g the industrial production). We therefore use $B(t)$ in order to forecast the t-th value of $g(t)$. This procedure is known as *nowcasting* in economic literature.

As we are not interested in the correct selection of leads or different variables with might explain our dependent variable best, perform a simple linear model

$$g_{\beta^{AR}}(t) = \alpha + \beta \cdot B_{\beta^{AR}}(t). \tag{B.2}$$

Then, the estimated forecast turns to

$$\hat{g}_{\beta^{AR}}(t+1) = \hat{\alpha} + \hat{\beta} \cdot B_{\beta^{AR}}(t+1).$$

As $t = 1, \ldots, 500$, we perform our first forecast at $t = 451$ and proceed to $T = 500$. Then, we calculate the RMSE ratio between the forecast error of the biased and the unbiased balance statistics indicator.

Empirical analyses showed that a direct forecast of $g_{\beta^{AR}}(t)$ leads to enormous RMSE ratios (e.g. >80 on average for the random pattern) which clearly favour the unbiased indicator. This results from the fact that the unbiased indicator approximates the cycle function $g_{\beta^{AR}}(t)$ very fast with respect to the number of respondents. As this is very unrealistic to appear in a real-life situation, we assume that the dependent cycle function $g_{\beta^{AR}}(t)$ in Equation (B.2) is affected by an error term $\epsilon \sim N(0, \sigma^2)$. This seems to be plausible as also official GDP growth rates can not be estimated without a measurement error. To show how strongly this fact affects the RMSE ratios, we set $\sigma^2 \in \{0.1, 0.25, 0.5\}$.

[1] If, for example, the balance statistics indicator $B(t)$ would have lead of 3 months with respect to the cycle function $g(t)$, the observation probabilities would turn to $\frac{\pi^s(t)}{\bar{\pi}(t)} \cdot \Phi^s(t+3)$.

B.5.3 Results

In Figures B.9 - B.12 the boxplots of the RMSE ratios are drawn according to the NMAR patterns and different values of the measurement error variance σ^2. I can be seen that the number of observations n does not seem to influence the results strongly. They remain similar in all cases of n. In general, the results are strongly affected by the assumed measurement error variance σ^2. For $\sigma^2 = 0.1$, the RMSE ratios are higher than 2 for nearly all NMAR patterns and auto-correlation factors β^{AR}.[2] This parameter has a strong effect on the RMSE of the unbiased indicator and therefore defines the level to be reached of the biased indicators. For $\sigma^2 = 0.25$ and 0.5, the RMSE ratios decrease to a more realistic range. For the different NMAR patterns, it can be seen that random, monotone and mixed patterns affect the forecasting performance of the biased indicators stronger. This result is similar to correlation analysis in Section 4.3.5, Figure 4.2. Fixed, cycle-dependent and cycle-shifted seem to affect the forecasting accuracy less. With respect to the auto-correlation factors, higher RMSE ratios were recorded with higher β^{AR}. As the time series are smoother in cases of high values of β^{AR}, the NMAR patterns seem to introduce a stronger decrease in forecasting performance as the series it self includes less uncertainty.

[2]Notice that the y-axis are of different scale for σ^2.

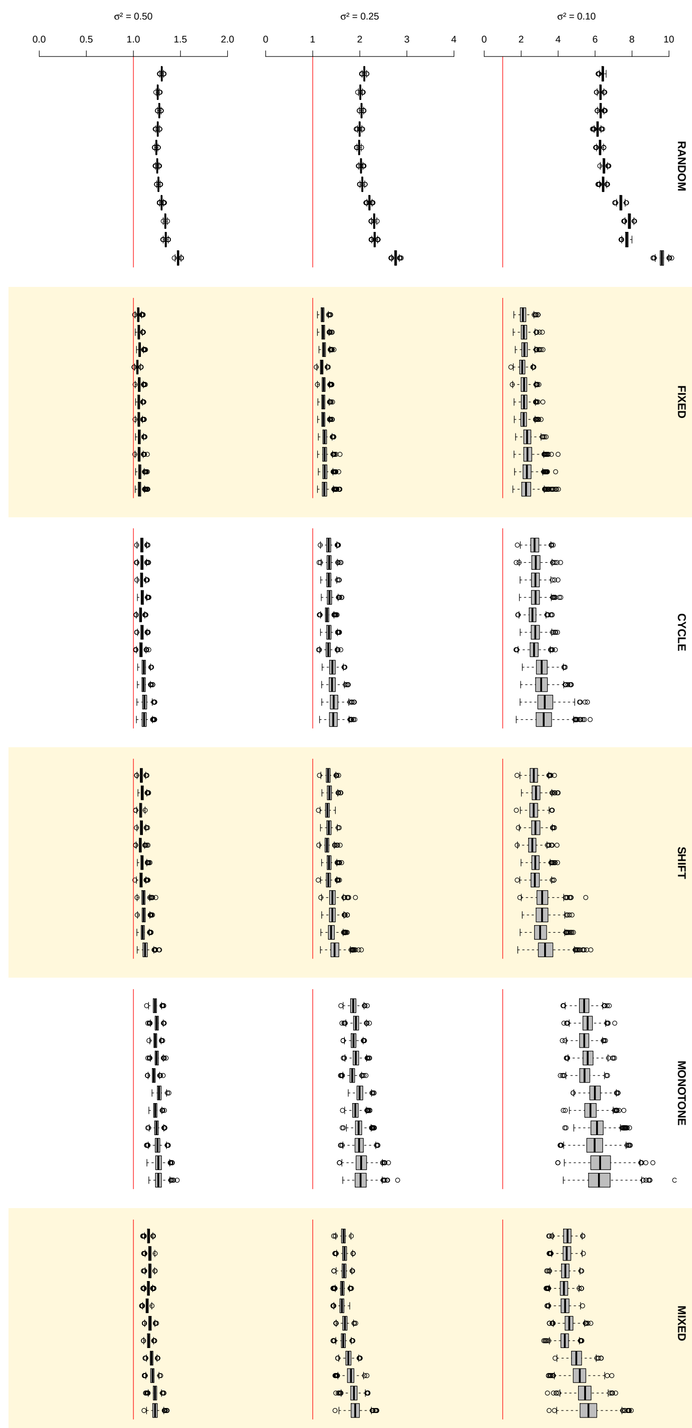

Figure B.9: Boxplots for RMSE ratios according to the six main NMAR patterns (columns) and three different values of σ^2 (rows) for $n = 500$ observations. In each cell, a boxplot for each of the 11 different β^{AR} is drawn (from left to right: $0.1, 0.2, \ldots, 0.9, 0.95, 0.99$).

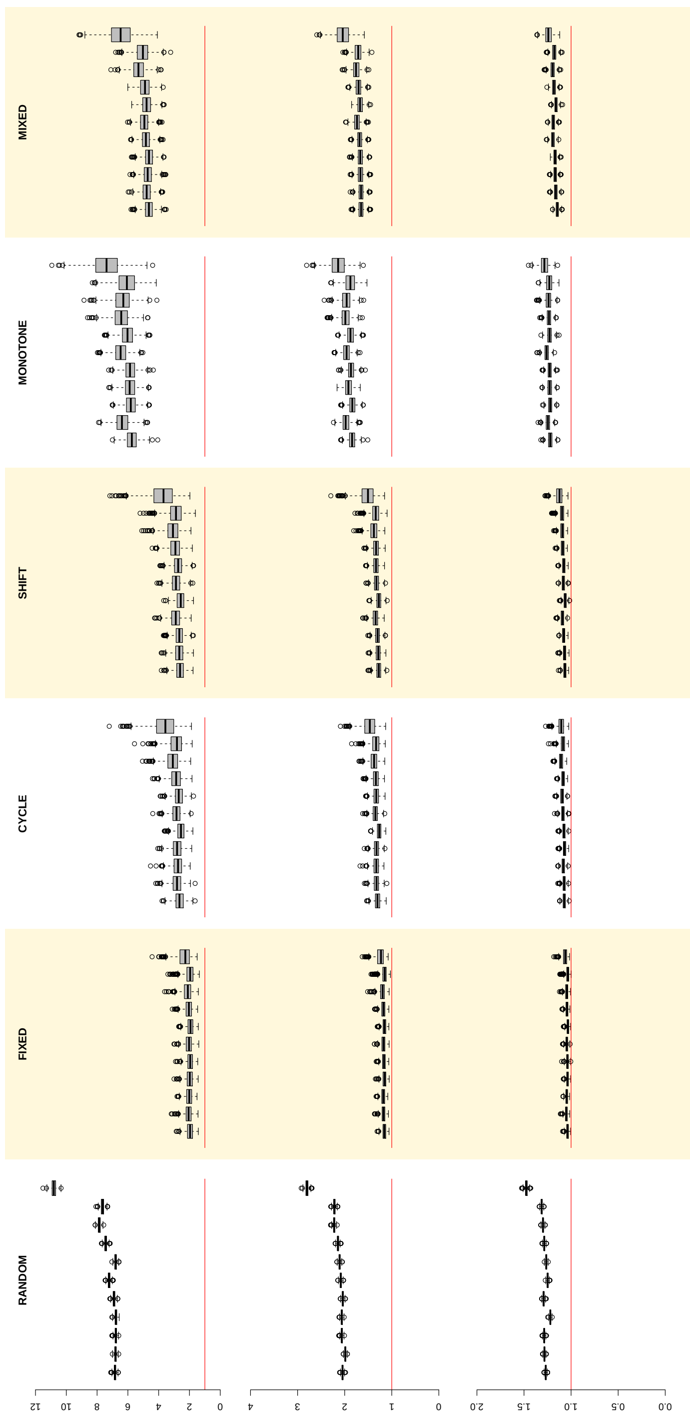

Figure B.10: Boxplots for RMSE ratios according to the six main NMAR patterns (columns) and three different values of σ^2 (rows) for $n = 1000$ observations. In each cell, a boxplot for each of the 11 different β^{AR} is drawn (from left to right: $0.1, 0.2, \ldots, 0.9, 0.95, 0.99$).

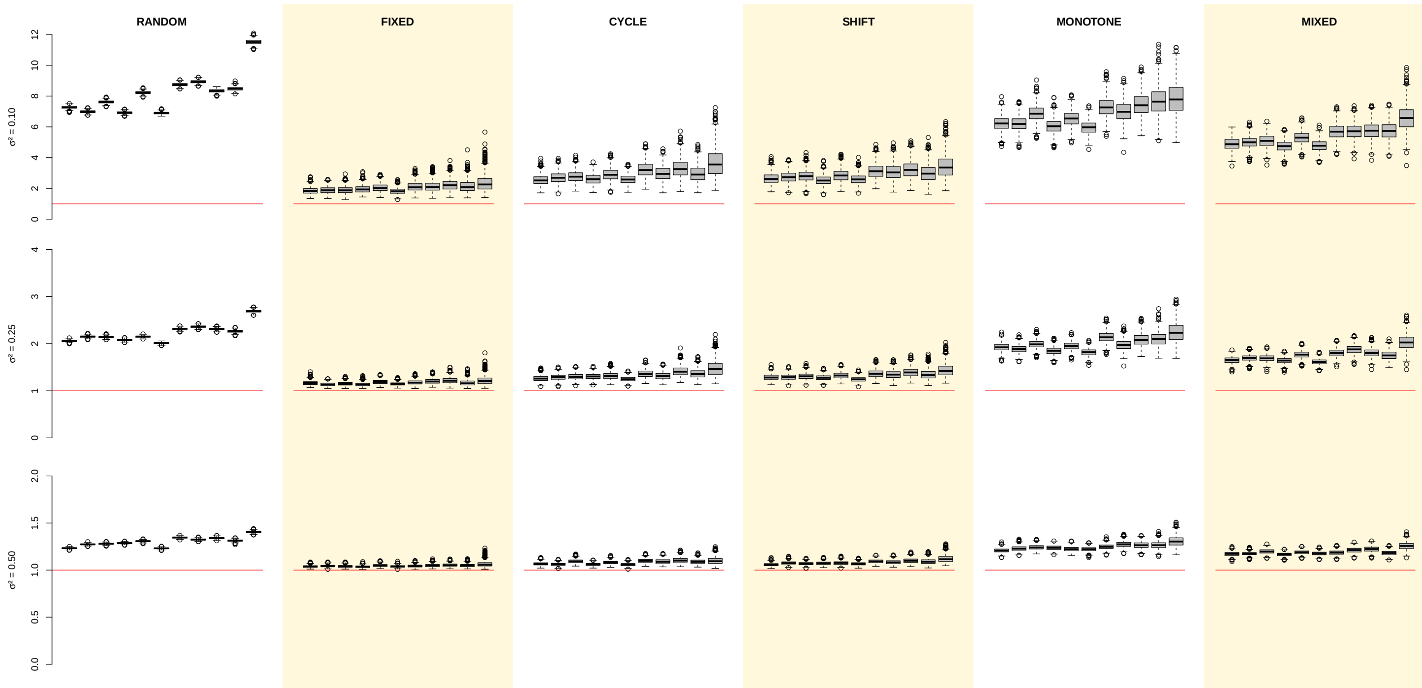

Figure B.11: Boxplots for RMSE ratios according to the six main NMAR patterns (columns) and three different values of σ^2 (rows) for $n = 2000$ observations. In each cell, a boxplot for each of the 11 different β^{AR} is drawn (from left to right: $0.1, 0.2, \ldots, 0.9, 0.95, 0.99$).

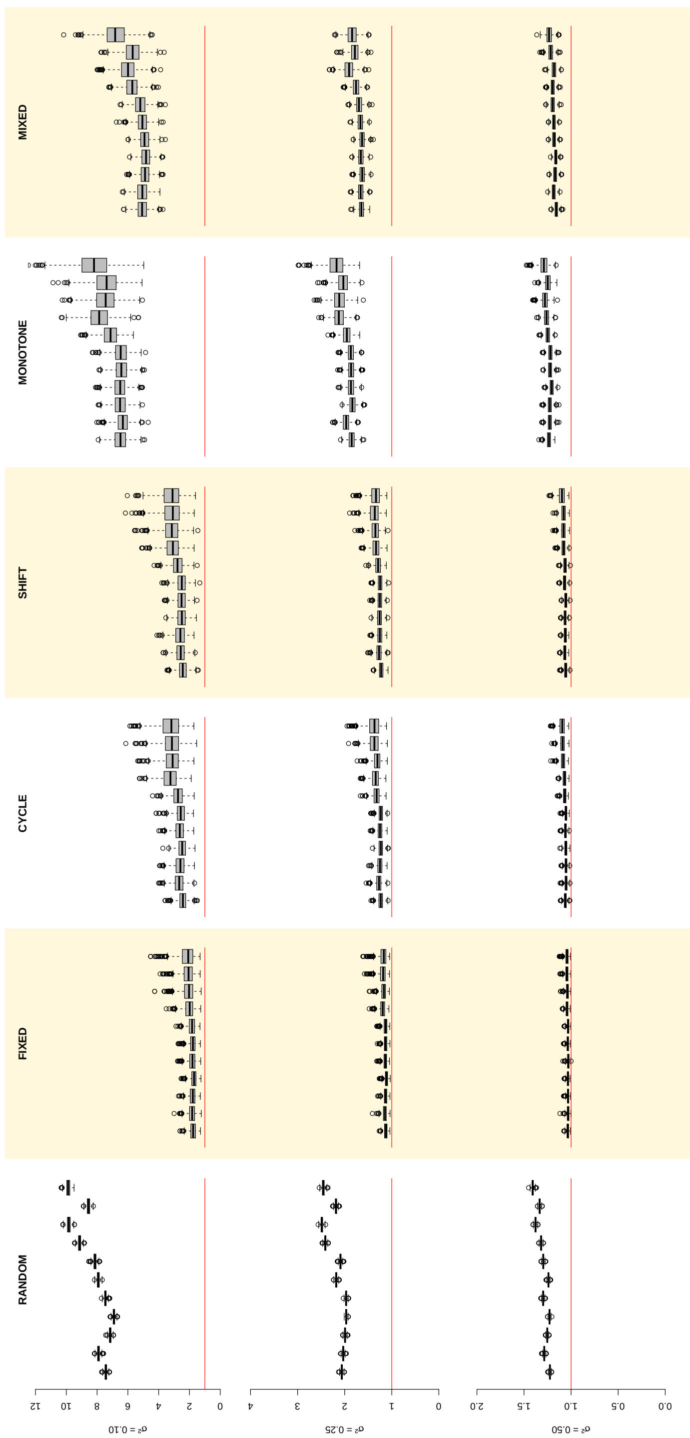

Figure B.12: Boxplots for RMSE ratios according to the six main NMAR patterns (columns) and three different values of σ^2 (rows) for $n = 5000$ observations. In each cell, a boxplot for each of the 11 different β^{AR} is drawn (from left to right: $0.1, 0.2, \ldots, 0.9, 0.95, 0.99$).

Appendix C

Appendix for Chapter 5

In this chapter, the aggregation scheme of the Ifo index is shown in Section C.1. Table C.1 in Section C.2 shows the covariates $\boldsymbol{x}_{t-1}^{sec}$ of the sector-specific regression-based imputation models from Equation (5.6). In Section C.3 the non-standardised and in Section C.4 the standardised imputed indicators are shown.

C.1 Aggregation scheme of the Ifo index

As mentioned in Section 5.2, each company can give three possible replies to both, business situation and business expectations: a positive, a unchanged and an negative reply. The microdata aggregation has a hierarchical structure according to the Classification of Economic Activities (edition 2008) from the German Federal Statistical Office (Destatis, 2008). The lowest aggregation level evaluated in the Ifo Business Survey is IV in industry, VI in trade and V in construction (the Ifo index is level 0). We show the aggregation process based on the industry sector. For example, a company from manufacture of metal forming machinery (level IV) is part of manufacture of metal forming machinery and machine tools (level III) and part of manufacture of machinery and equipment n.e.c. (level II) and part of the industry sector (level I). There are 32 different groups on level II, 119 on level III, 171 on level IV, 24 on level V and 4 on level VI[1]. This process is equal for all variables measured on a 3-level Likert scale and all points in time.

We define $SEC_u^{IV}, u = 1, \ldots, U = 112$ as the u-th subsector on aggregation level IV. Each company can be clearly classified to one of these subsectors. We first count all weighted positive, unchanged and negative replies $\widetilde{n}^{+}_{t,SEC_u^{IV}}, \widetilde{n}^{=}_{t,SEC_u^{IV}}$ and $\widetilde{n}^{-}_{t,SEC_u^{IV}}$ in each subsector SEC_u^{IV}. The answers are weighted by the companies size, e.g. the answers of an industry firm with more than 500 employees get a weight of 15 whereas the answers of a company with less than 10 employees gets a weight of 1. Then, the number of weighted replies are scaled to the unit interval by dividing by $\widetilde{n}_{t,SEC_u^{IV}} = \widetilde{n}^{+}_{t,SEC_u^{IV}} + \widetilde{n}^{=}_{t,SEC_u^{IV}} + \widetilde{n}^{-}_{t,SEC_u^{IV}}$, so that

$$n^{\cdot}_{t,SEC_u^{IV}} = \frac{\widetilde{n}^{\cdot}_{t,SEC_u^{IV}}}{\widetilde{n}_{t,SEC_u^{IV}}}, \qquad n^{\cdot}_{t,SEC_u^{IV}} \in [0,1].$$

Then, the balance of u-th subsector SEC_u^{IV} is defined as

$$B_{SEC_u^{IV}}(t) = (n^{+}_{t,SEC_u^{IV}} - n^{-}_{t,SEC_u^{IV}}), \qquad B_{SEC_u^{IV}}(t) \in [-1,1], \tag{C.1}$$

[1]Not all subgroups which occur in the German Classification of Economic Activities are calculated by the Ifo Institute, in particular no value for the whole trade sector is calculated.

i.e. to substract the fraction of negative from the fraction of positive replies. To calculate the balances for the next higher aggregation level, the fractions $n^{+}_{t,SEC^{IV}_u}, n^{=}_{t,SEC^{IV}_u}$ and n^{-}_{t,SEC^{IV}_u} of replies are weighted, i.e.

$$n^{\cdot}_{t,SEC^{III}_v} = (n^{\cdot}_{t,SEC^{IV}_1}, \ldots, n^{\cdot}_{t,SEC^{IV}_u})'\omega_{SEC^{III}_v}$$

with $\omega_{SEC^{III}_v} = (\omega_{1,v}, \ldots, \omega_{U,v})', \omega_{u,v} \in [0,1], \sum_{u=1}^{U} \omega_{u,v} = 1$. Note that only $\omega_{u,v} > 0$ if $SEC^{IV}_u \in SEC^{III}_v$, i.e. if SEC^{IV}_u is subsector of SEC^{III}_v. The balances $B_{SEC^{III}_v}(t)$ are just as calculated as in equation (C.1). The aggregation to level II, I and 0 is also carried out as described above. The index' value is obtained by scaling the balances to the average of the year 2005.

C.2 Covariates for regression-based imputation

Industry	stock of inventories
	orders vs. previous months
	orders (appraisal)
	prices vs. previous months
	expected production
	expected domestic prices
	expected export trade
	expected commercial operations
	foreign orders (appraisal)
Construction	construction activity vs. previous month
	construction activity in 3 months
	constraints
	constraints: lack of manpower
	constraints: lack of material
	constraints: weather conditions
	constraints: financing
	constraints: other reasons
	orders vs. previous month
	orders (appraisal)
	range of orders in months
	prices vs. previous month
	prime costs covering
	expected prices
	expected employees
	industrial worker
	employee
	status of employee's illness in %
Trade	business volume vs. previous year
	feedstock (appraisal)
	prices vs. previous month
	expected prices
	orders vs. previous year

Table C.1: Covariates included in $\boldsymbol{x}_{t-1}^{sec}$ for the different sector models

C.3 Results for non-standardised indicators

The following figures show the non-standardised original and imputed Ifo indicators for the German economy (Figure 5.2), the manufacturing (Figure C.1) and the construction sector (Figure C.2) as well as the Ifo indicators for retail (Figure C.3) and wholesale trade (Figure C.4).

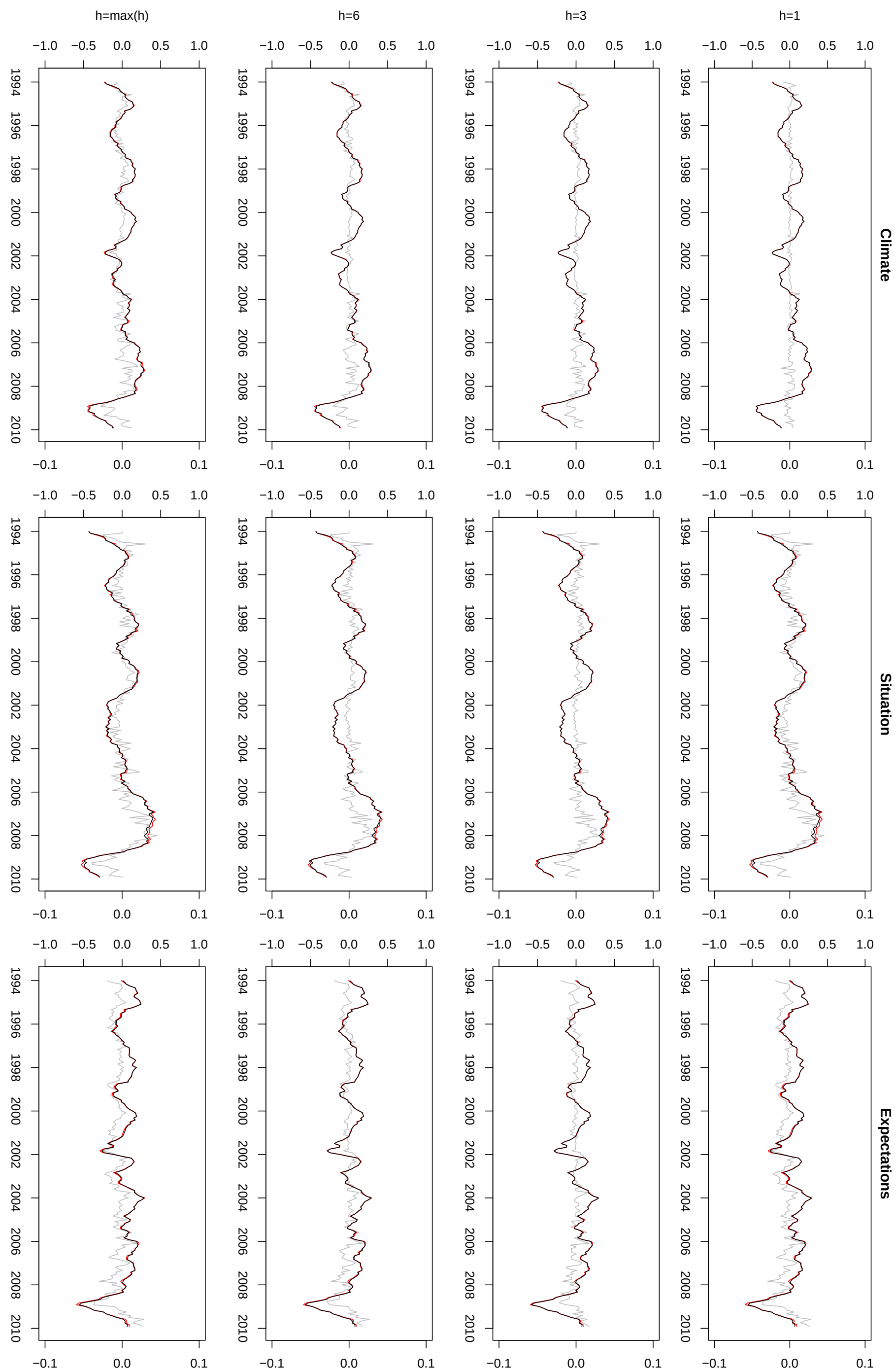

Figure C.1: Original (black) and imputed (red) *non-standardised* Ifo indicators for the manufacturing sector and their difference (grey, right scale)

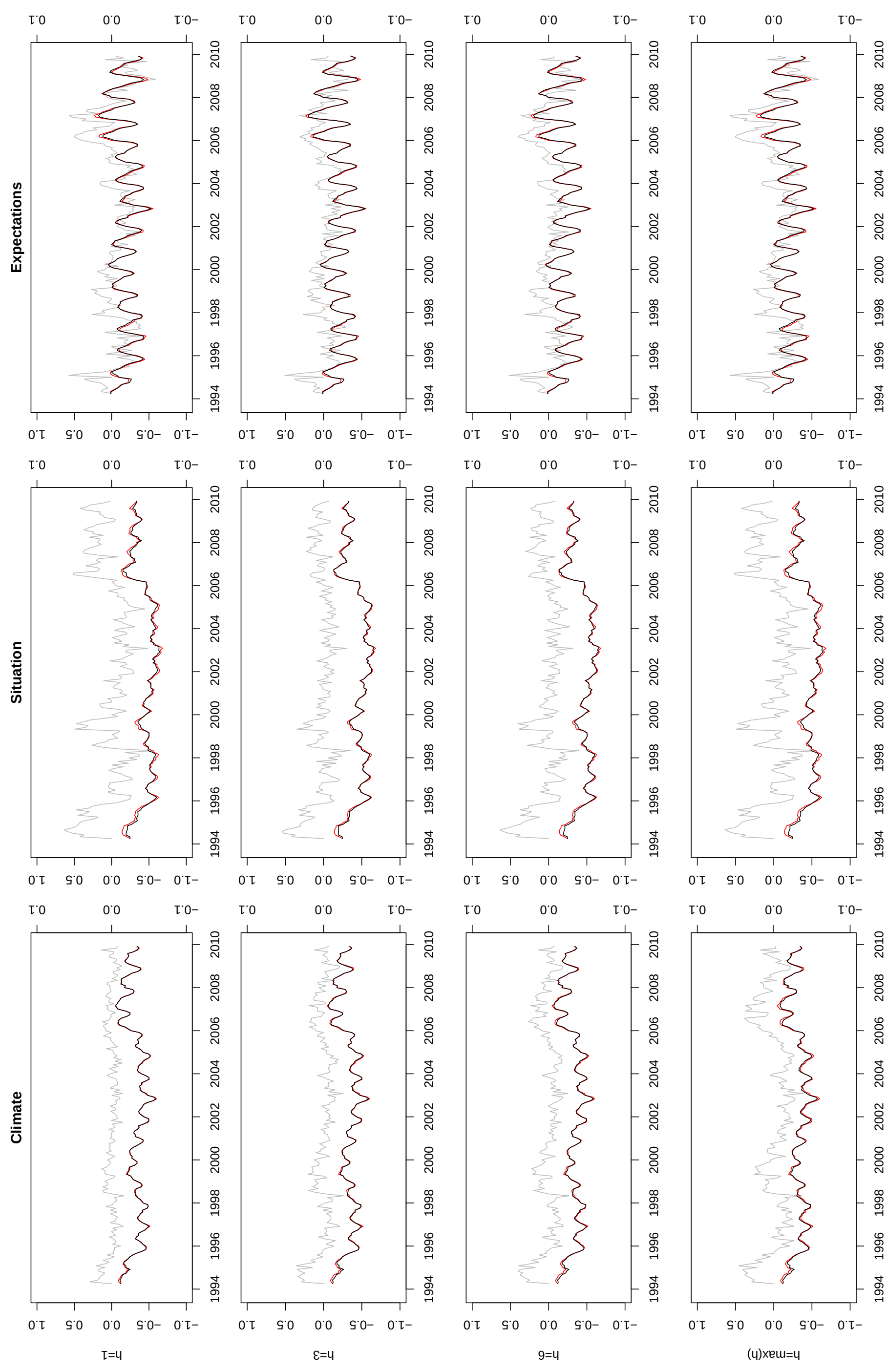

Figure C.2: Original (black) and imputed (red) *non-standardised* Ifo indicators for the construction sector and their difference (grey, right scale)

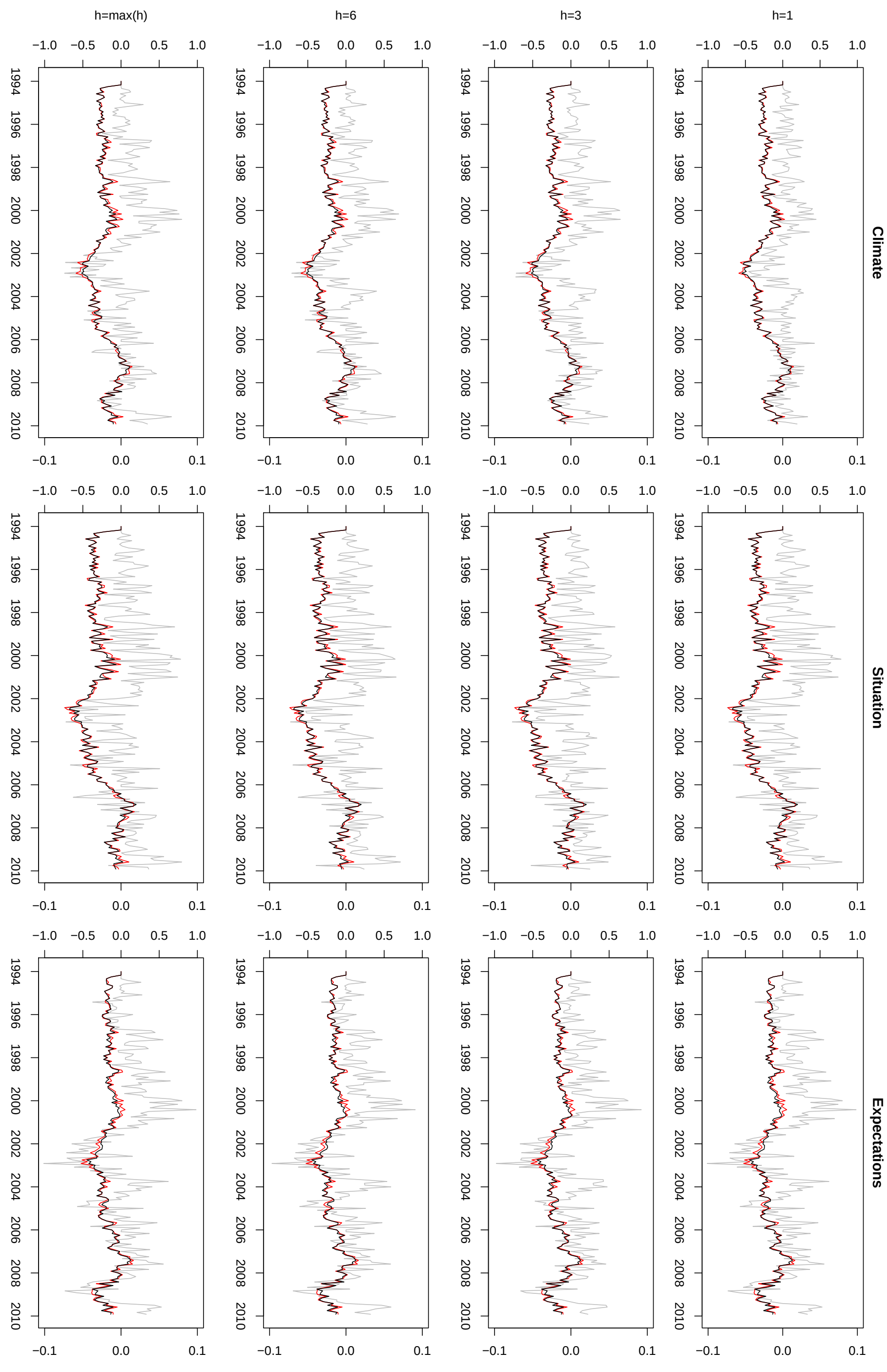

Figure C.3: Original (black) and imputed (red) *non-standardised* Ifo indicators for retail trade and their difference (grey, right scale)

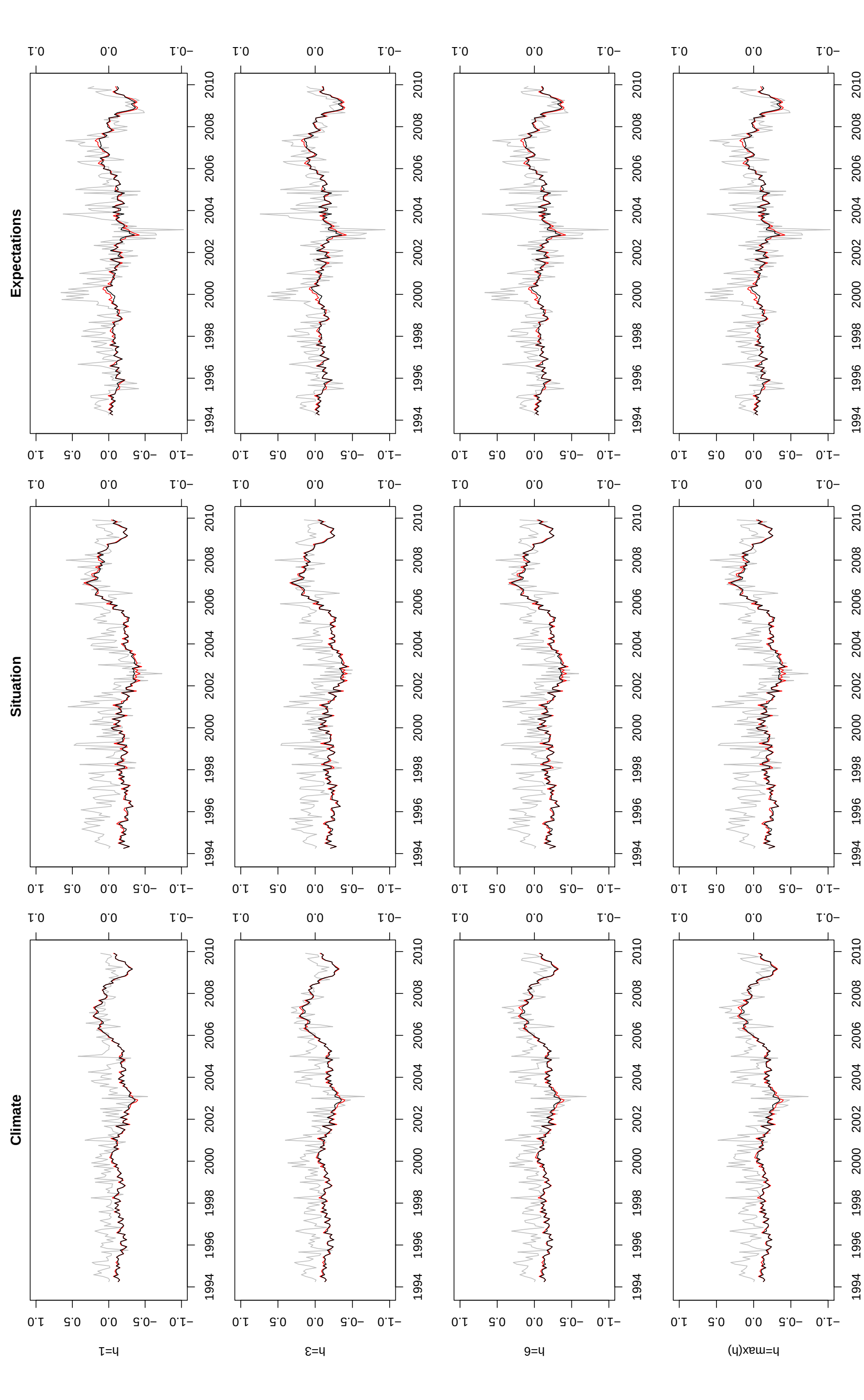

Figure C.4: Original (black) and imputed (red) *non-standardised* Ifo indicators for whole sale trade and their difference (grey, right scale)

C.4 Results for standardised indicators

The following figures show the standardised original and imputed Ifo indicators for the German economy (Figure C.5), the manufacturing (Figure C.6) and the construction sector (Figure C.7) as well as the Ifo indicators for retail (Figure C.8) and wholesale trade (Figure C.9). In addition, Figures C.10 and C.11 display the boxplots for the distribution of ρ_{GDP} for the standardised original and imputed indicators.

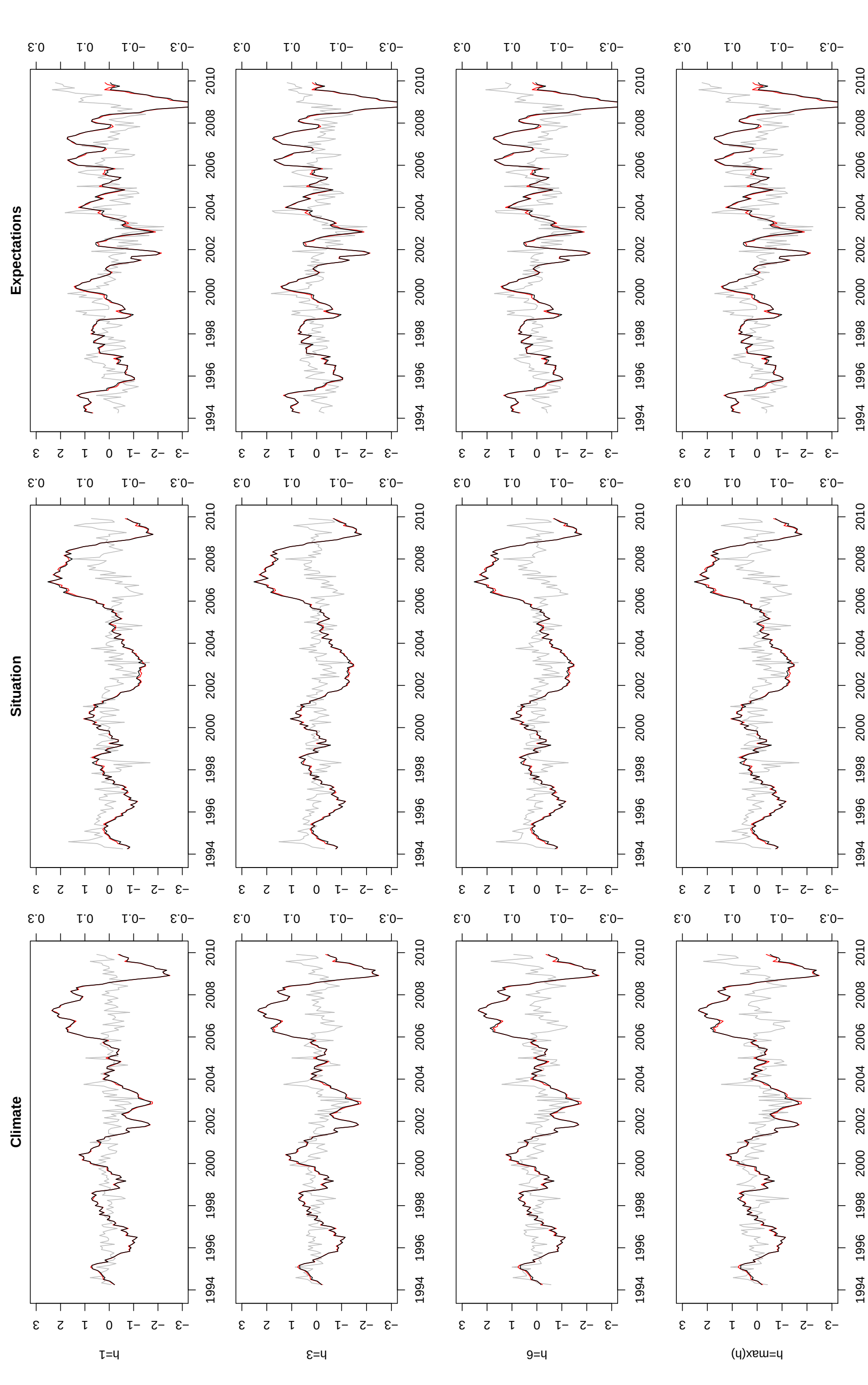

Figure C.5: Original (black) and imputed (red) *standardised* Ifo indicators for the German economy and their difference (grey, right scale)

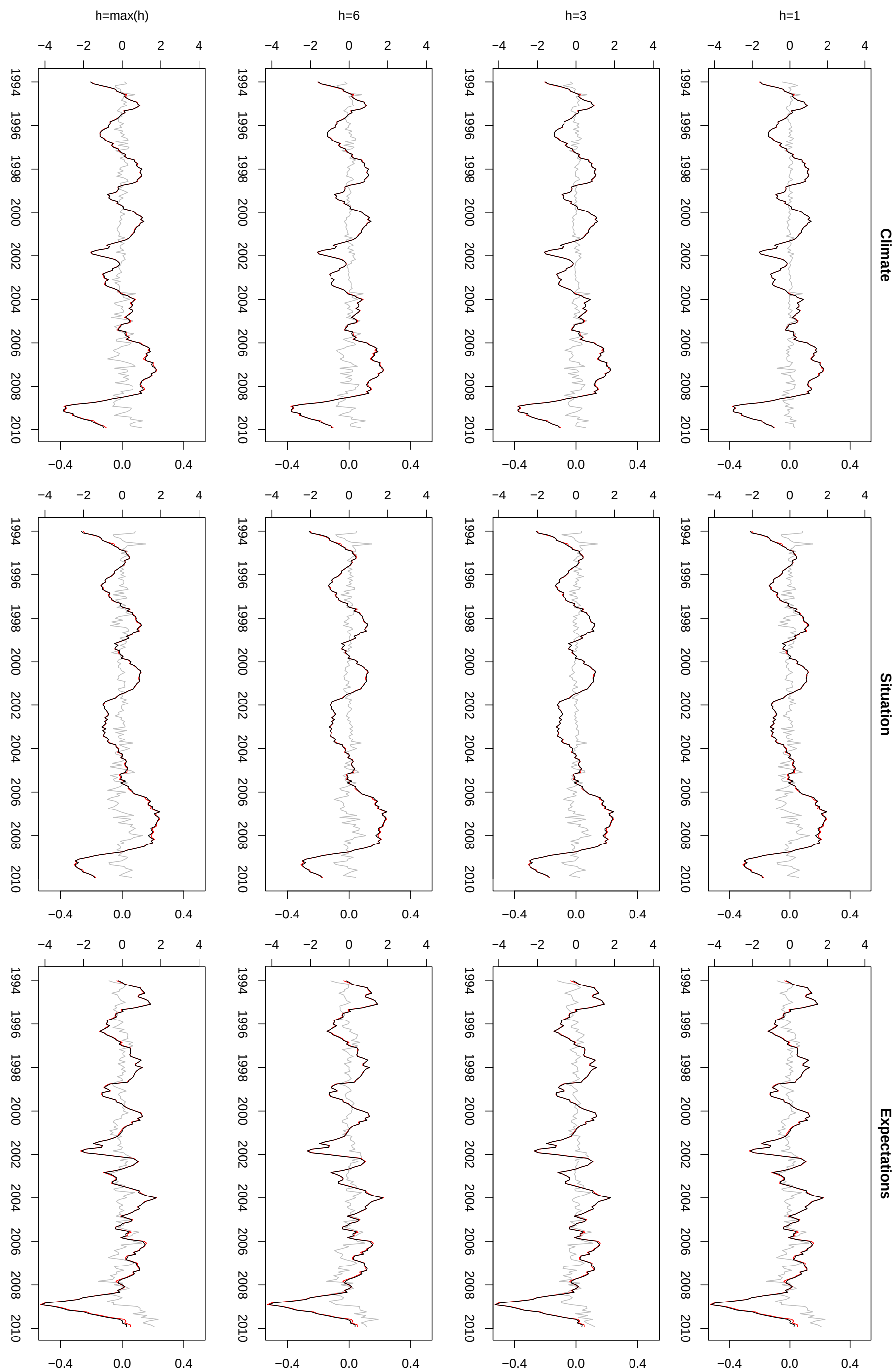

Figure C.6: Original (black) and imputed (red) *standardised* Ifo indicators for the manufacturing sector and their difference (grey, right scale)

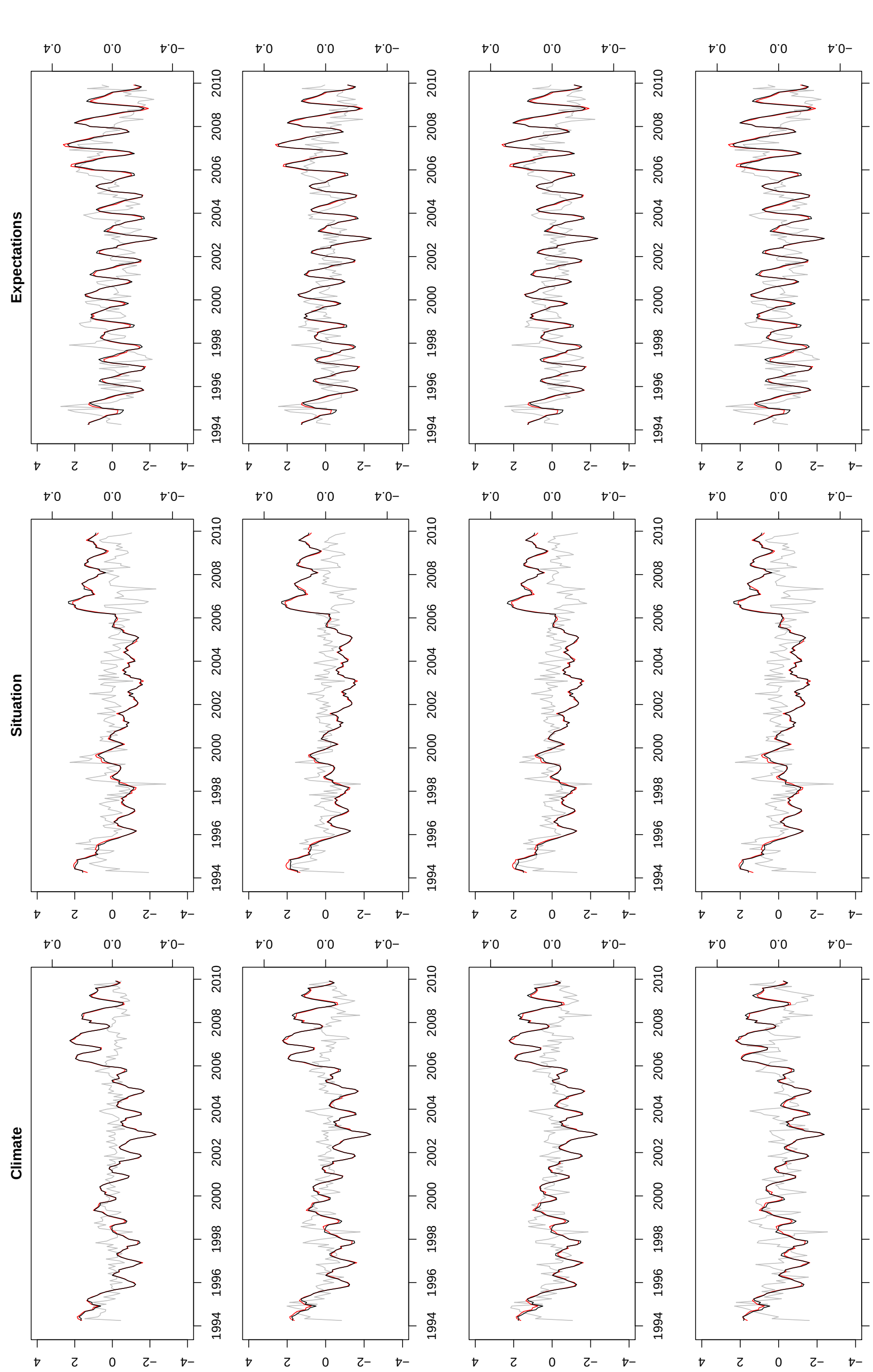

Figure C.7: Original (black) and imputed (red) *standardised* Ifo indicators for the construction sector and their difference (grey, right scale)

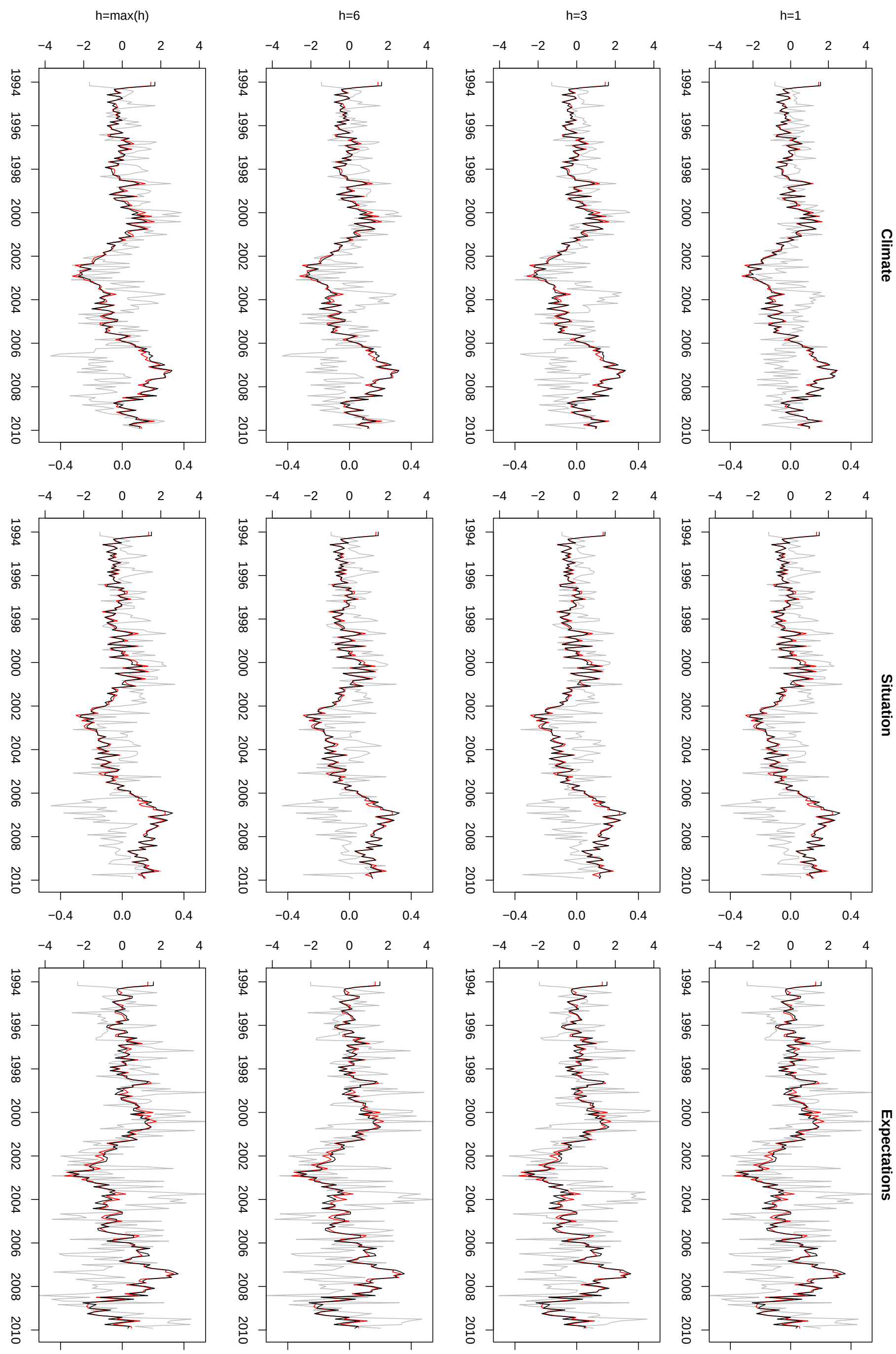

Figure C.8: Original (black) and imputed (red) *standardised* Ifo indicators for retail sale trade and their difference (grey, right scale)

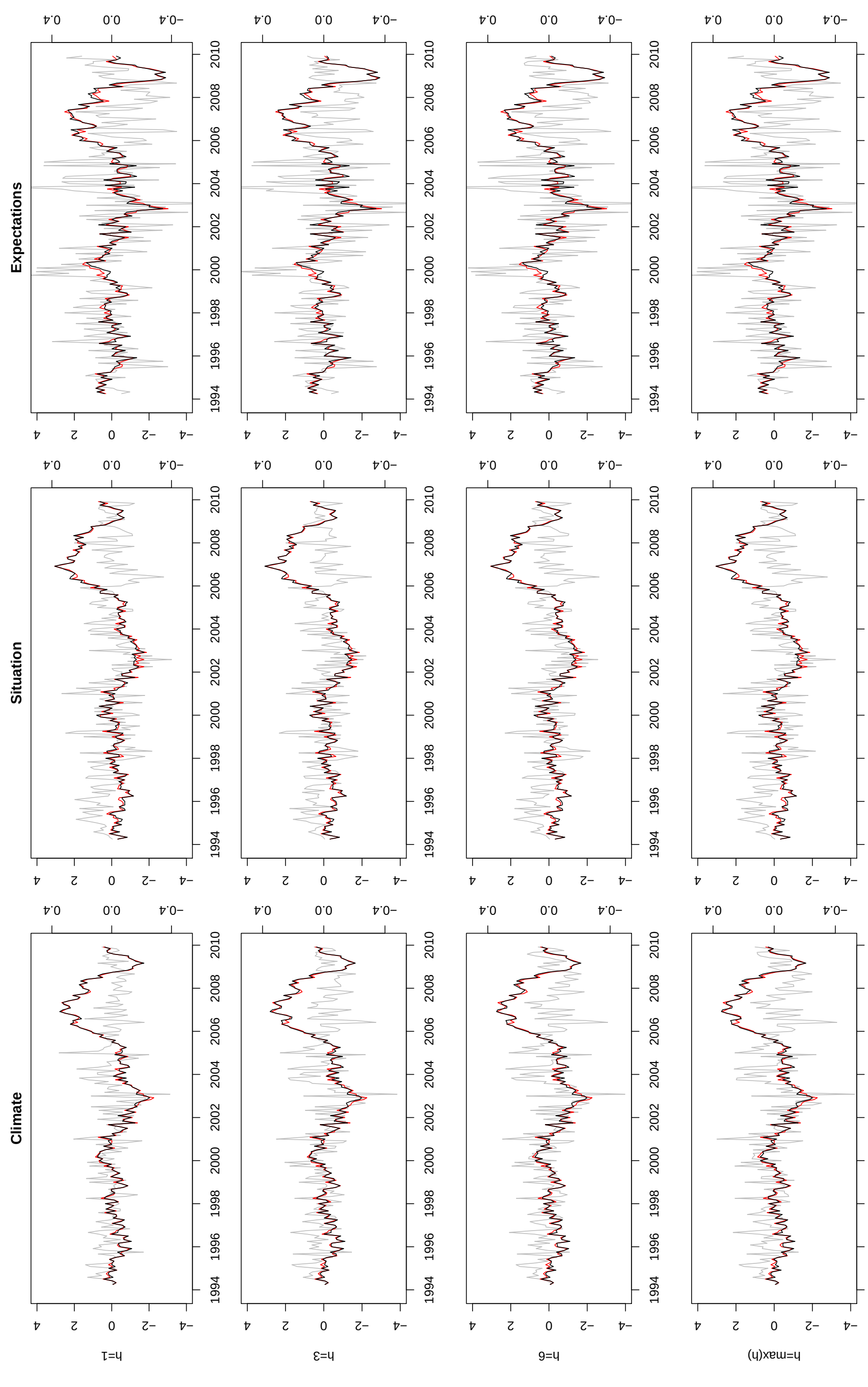

Figure C.9: Original (black) and imputed (red) *standardised* Ifo indicators for whole sale trade and their difference (grey, right scale)

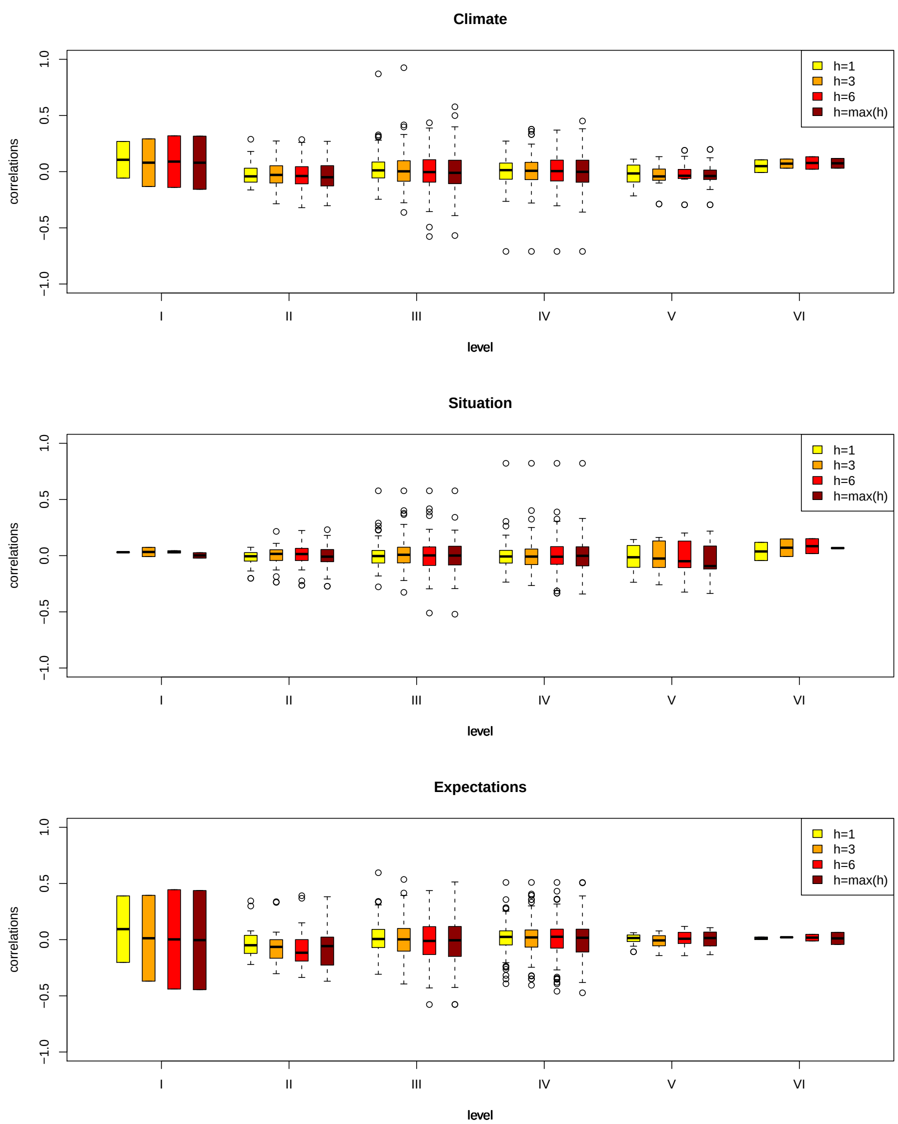

Figure C.10: Boxplots for the distribution of ρ_{GDP} for the *standardised* original and imputed indicators, different aggregation levels and horizons h.

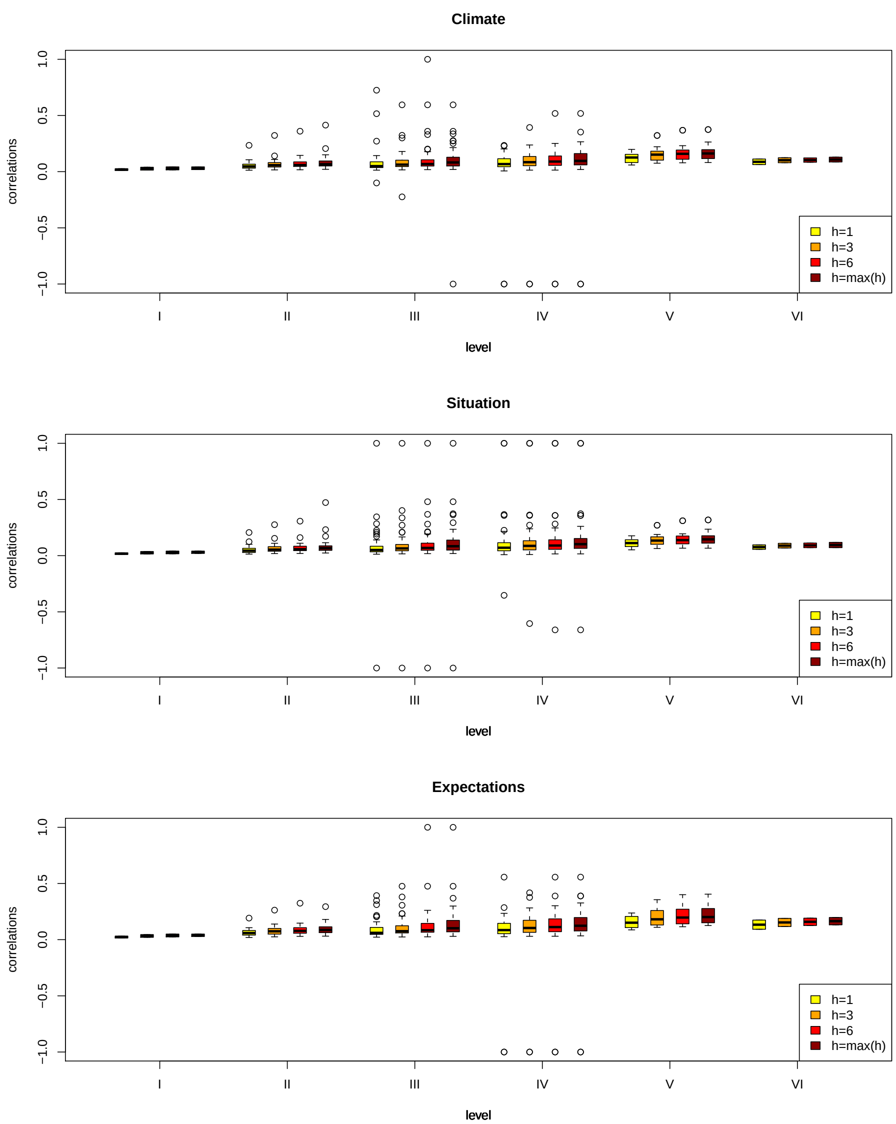

Figure C.11: Boxplots for the distribution of ρ_{IND} for the *standardised* original and imputed indicators, different aggregation levels and horizons h.

Bibliography

Abberger, K., Birnbrich, M., and Seiler, C. (2009). Der 'Test des Tests' im Handel - Eine Metaumfrage zum ifo Konjunkturtest. *ifo Schnelldienst*, 62(21):34–41.

Abberger, K. and Wohlrabe, K. (2006). Einige Prognoseeigenschaften des ifo Geschäftsklimas - Ein Überblick über die neuere wissenschaftliche Literatur. *ifo Schnelldienst*, 59(22):19–26.

Adebayo, S. A., Fahrmeir, L., Seiler, C., and Heumann, C. (2011). Geoadditive latent variable modelling of count data on multiple sexual partnering in Nigeria. *Biometrics*, 67(2):19–26.

Allen, R., Huggins, V. J., and Killion, R. A. (1995). The Evolution of Agricultural Data Collection in the United States. In Cox, B. G., Binder, D. A., Chinnappa, B. N., Christianson, A., Colledge, M. J., and Kott, P. S., editors, *Business Survey Methods*, chapter 31, pages 609–632. Wiley.

Anderson, O. (1951). Konjunkturtest und Statistik. *Allgemeines Statistisches Archiv*, 35:209–220.

Anderson, O. (1952). The business test of the IFO-Institute for Economic Research. *Revue del'Institute International de Statistique*, 20:1–17.

Ang, A., Bekaert, G., and Wei, M. (2007). Do macro variables, asset markets, or surveys forecast inflation better? *Journal of Monetary Economics*, 54(4):1163–1212.

Bachmann, R., Elstner, S., and Sims, E. R. (2012). Uncertainty and economic activity: Evidence from business survey data. *American Economic Journal – Macroeconomics*. to appear.

Bartholomew, D. J. and Knott, M. (1999). *Latent variable models and factor analysis*. Charles Griffin.

Becker, S. O. and Wohlrabe, K. (2008). Micro Data at the Ifo Institute for Economic Research - The "Ifo Business Survey", Usage and Access. *Journal of Applied Social Science Studies*, 128(2):307–319.

Bethlehem, J. G. (1988). Reduction of nonresponse bias through regression estimation. *Journal of Official Statistics*, 4(3):251—-260.

Brehm, J. (1994). Stubbing out Toes for a Foot in the Door? Prior Contacts, Incentives and Survey Response. *International Journal of Public Opinion Research*, 6(1):45–63.

Burns, A. F. and Mitchell, W. C. (1946). *Measuring Business Cycles*. Number 46-1 in NBER Books. National Bureau of Economic Research, Inc.

Cameron, A. C. and Trivedi, P. K. (2005). *Microeconometrics. Methods and Applications.* Cambridge University Press.

Carlson, J. A. and Parkin, M. (1975). Inflation expectations. *Economica*, 42:123–138.

Chen, J. and Shao, J. (2000). Nearest Neighbor Imputation for Survey Data. *Journal of Official Statistics*, 16(2):113—-131.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20:37–46.

Cook, R. J., Zeng, L., and Yi, G. Y. (2004). Marginal Analysis of Incomplete Longitudinal Binary Data: A Cautionary Note on LOCF Imputation. *Biometrics*, 60(3):820–828.

de Leeuw, E. and de Heer, W. (2002). Trends in Household Survey Nonresponse: A Longitudinal and International Comparison. In Groves, R. M., Dillman, D. A., Eltinge, J. L., and Little, R. J., editors, *Survey Nonresponse*, chapter 3, pages 41–54. Wiley.

Dempster, A. P. and Rubin, D. B. (1983). Introduction. In Madow, W. G., Olkin, I., and Rubin, D. B., editors, *Incomplete Data in Sample Surveys*, pages 3–10. Academic Press.

Destatis (2008). *Classification of Economic Activities, Edition 2008*. Federal Statistical Office of Germany.

Dillman, D. A., Eltinge, J. L., Groves, R. M., and Little, R. J. (2002). Survey Nonresponse in Design, Data Collection, and Analysis. In Groves, R. M., Dillman, D. A., Eltinge, J. L., and Little, R. J., editors, *Survey Nonresponse*, chapter 1, pages 3–26. Wiley.

Drechsel, K. and Scheufele, R. (2010). Should We Trust in Leading Indicators? Evidence from the Recent Recession. IWH-Diskussionspapiere 10/2010, Institut für Wirtschaftsforschung Halle.

Drechsler, J. (2011a). Multiple imputation in practice—a case study using a complex German establishment survey. *AStA Advances in Statistical Analysis*, 95:1–26.

Drechsler, J. (2011b). *Synthetic Datasets for Statistical Disclosure Control: Theory and Implementation*. Springer.

Ekholm, A. and Laaksonen, S. (1991). Weighting via response modelling in the finnish household budget survey. *Journal of Official Statistics*, 7(3):325—338.

Engels, J. M. and Diehr, P. (2003). Imputation of missing longitudinal data: a comprison of methods. *Journal of Clinical Epidemiology*, 56:968–976.

European Union (2006). Joint Harmonised EU Programme of Business and Consumer Surveys. *Official Journal of the European Union*, 49(C 245):5–8.

Finch, W. H. (2010). Imputation Methods for Missing Categorical Questionnaire Data: A Comparison of Approaches. *Journal of Data Science*, 8:361–378.

Gayer, C. (2005). Forecast Evaluation of European Commission Survey Indicators. *Journal of Business Cycle Measurement and Analysis*, 7(2):157—-183.

Giacomini, R. and White, H. (2006). Tests of Conditional Predictive Abilitys. *Econometrica*, 74(6):1545–1578.

Goldrian, G., editor (2007). *Handbook of Survey-Based Business Cycle Analysis*. Edward Elgar Publishing.

Graham, J. W., Olchowski, A. E., and Gilreath, T. D. (2007). How Many Imputations are Really Needed? Some Practical Clarifications of Multiple Imputation Theory. *Preventative Science*, 8:208–213.

Groves, R. M. and Couper, M. P. (1998). *Nonresponse in Household Interview Surveys*. Wiley and Sons.

Groves, R. M., Dillman, D. A., Eltinge, J. L., and Little, R. J., editors (2002). *Survey Nonresponse*. Wiley.

Groves, R. M., Fowler, J. F., Couper, M. P., Lepkowski, J. M., Singer, E., and Tourangeau, R. (2004). *Survey Methodology*. Wiley and Sons.

Harris-Kojetin, B. and Tucker, C. (1999). Exploring the Relation of Economical and Political Conditions with Refusal Rates to a Government Survey. *Journal of Official Statistics*, 15(2):167–184.

Hartmann, J. and Kohaut, S. (2000). Analysen zu Ausfällen im IAB-Betriebspanel. *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung*, 33(4):609–618.

Heagerty, P. and Zeger, S. (1996). Marginal regression models for clustered ordinal measurements. *Journal of the American Statistical Association*, 91:1024—1036.

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1):153—161.

Hodrick, R. and Prescott, E. (1997). Postwar US business cycles: an empirical investigation. *Journal of Money, Credit and Banking*, 29(1):10–16.

Honaker, J. and King, G. (2010). What to do About Missing Values in Time Series Cross-Section Data. *American Journal of Political Science*, 54(2):561–581.

Honaker, J., King, G., and Blackwell, M. (2012). *Amelia II: A Program for Missing Data*. R package version 1.6.3.

Horton, N. J. and Lipsitz, S. R. (2001). Multiple imputation in practice: Comparison of software packages for regression models with missing variables. *The American Statistician*, 55(3):244–254.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 221–233.

Janik, F. and Kohaut, S. (2012). Why don't they answer? - Unit nonresponse in the IAB Establishment Panel. *Quality & Quantity*, 46(3):917–934.

Jennrich, R. and Schluchter, M. (1986). Incomplete repeated-measures models with structured covariance matrices. *Biometrics*, 42:805–820.

Khan, H. R. and Shaw, E. (2012). *imputeYn: Imputing the last largest censored datum under weighted least squares*. R package version 1.0.

Kholodilin, K. A. and Siliverstovs, S. (2006). On the Forecasting Properties of the Alternative Leading Indicators for the German GDP: Recent Evidence. *Journal of Economics and Statistics*, 226(3):234–259.

King, G., Honaker, J., Joseph, A., and Scheve, K. (2001). Analyzing incomplete political science data: An alternative algorithm for multiple imputation. *American Political Science Review*, 95(1):49–69.

Kleinke, K., Stemmler, M., Reinecke, J., and Lösel, F. (2011). Efficient ways to impute incomplete panel data. *AStA Advances in Statistical Analysis*, 95(4):351–373.

Knöbl, A. (1974). Price expectations and actual price behaviour in Germany. IMF Staff Papers 21.

Korotayev, A. V. and Tsirel, S. V. (2010). A Spectral Analysis of World GDP Dynamics: Kondratieff Waves, Kuznets Swings, Juglar and Kitchin Cycles in Global Economic Development, and the 2008–2009 Economic Crisis. *Structure and Dynamics*, 4(1):3–57.

Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174.

Langelütke, H. and Marquardt, W. (1951). Das Konjunkturtest-Verfahren. *Allgemeines Statistisches Archiv*, 35:189–208.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22.

Little, R. J. A. (1992). Regression with missing X's: A review. *Journal of the American Statistical Association*, 87(420):1227–1237.

Little, R. J. A. and Rubin, D. (2002). *Statistical Analysis with Missing Data*. Wiley.

Little, R. J. A. and Su, H. L. (1989). Item Nonresponse in Panel Surveys. pages 400–425. Wiley.

Mander, A. (2003). WHOTDECK: Stata module to perform multiple imputation using the Approximate Bayesian Bootstrap with weights.

Manski, C. (2003). *Partial Identification of Probability Distributions*. Springer.

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society, Series B*, 42:109—142.

Namkung, J., Hwang, T., Kwon, M., Yi, S., and Chung, W. (2011). *imputeMDR: The Multifactor Dimensionality Reduction (MDR) Analysis for Imcomplete Data*. R package version 1.1.1.

Nardo, M. (2003). The quantification of qualitative survey data: A critical assessment. *Journal of Economic Surveys*, 17(5):645–668.

OECD (2003). *Business Tendency Surveys - A Handbook*. Organisation for Economic Co-operation and Development.

Pesaran, M. H. and Timmermann, A. (2009). Testing Dependence Among Serially Correlated Multicategory Variables. *Journal of the American Statistical Association*, 104(485):325–337.

Puhani, P. (2000). The Heckman Correction for Sample Selection and Its Critique. *Journal of Economic Surveys*, 14:53–68.

Raghunathan, T. E., S. P. and van Hoewyk, J. (2002). IVEware: Imputation and variance estimation software.

Robinzonov, N. and Wohlrabe, K. (2010). Freedom of Choice in Macroeconomic Forecasting. *CESifo Economic Studies*, 56(2):192–220.

Rottmann, H. and Wollmershäuser, T. (2013). A Micro Data Approach to the Identification of Credit Crunches. *Applied Economics*, 45(17):2423–2441.

Royston, P. (2005). Multiple imputation of missing values: Update of ice. *The Stata Journal*, 5:527–536.

Royston, P. (2011). STSURVIMPUTE: Stata module for flexible imputation of censored survival data. Statistical Software Components, Boston College Department of Economics.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. Wiley.

Saha, C. and Jones, M. P. (2009). Bias in the last observation carried forward method under informative dropout. *Journal of Statistical Planning and Inference*, 139(2):246–255.

SAS-Institute (2010). *SAS/STAT® 9.22 User's Guide - The MI Procedure*.

Schafer, J. (1997). *Analysis of Incomplete Multivariate Data*. Chapman and Hall.

Schnabel, A. (1997). Teilnahmeverhalten bei Unternehmensbefragungen. *Arbeit*, 6(2):154—-172.

Schunk, D. (2008). A Markov chain Monte Carlo algorithm for multiple imputation in large surveys. *AStA Advances in Statistical Analysis*, 92(1):101–114.

Seiler, C. (2010). Dynamic Modelling of Nonresponse in Business Surveys. Ifo Working Paper 93, Ifo Institute.

Seiler, C. (2012). On the Robustness of the Balance Statistics with respect to Nonresponse. Ifo Working Paper 126, Ifo Institute.

Seiler, C. and Heumann, C. (2012). Microdata Imputation and Macrodata Implications - Evidence from the Ifo Business Survey. Technical report, Department of Statistics, University of Munich.

Shumway, R. H. and Stoffer, D. S. (1982). An approach to time series smoothing and forecasting using the EM algorithm. *Journal of Time Series Analysis*, 3:253–264.

Stangl, A. (2007). European data watch: Ifo world economic survey micro data. *Journal of Applied Social Science Studies*, 127(3):487–496.

Stangl, A. (2009). *Essays on the Measurement of Economic Expectations*. PhD thesis, Ludwig-Maximilians-Universität München.

STATA-Corporation (2009). *Stata 11 Multiple-Imputation Reference Manual*.

Su, Y.-S., Gelman, A., Hill, J., and Yajima, M. (2011). Multiple Imputation with Diagnostics (mi) in R: Opening Windows into the Black Box. *Journal of Statistical Software*, 45(2):1–31.

Theil, H. (1952). On the shape of economic microvariables and the Munich business test. *Revue del'Institute International de Statistique*, 20:105–120.

Theil, H. (1955). Recent experiences with the Munich business test: An expository article. *Econometrica*, 23:184—192.

Tomaskovic-Devey, D., Leiter, J., and Thompson, S. (1994). Organizational Survey Nonresponse. *Administrative Science Quarterly*, 39:439–457.

Tomaskovic-Devey, D., Leiter, J., and Thompson, S. (1995). Item Nonresponse in Organizational Surveys. *Sociological Methodology*, 25:77–100.

Tourangeau, R., Rips, L. J., and Rasinksi, K., editors (2000). *The Psychology of Survey Response*. Cambridge University Press.

van Buuren, S. and Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, 45(3):1–67.

Watson, N. and Starick, R. (2011). Evaluation of Alternative Income Imputation Methods for a Longitudinal Survey. *Journal of Official Statistics*, 27(4):693–715.

Wedderburn, R. W. M. (1974). Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method. *Biometrika*, 61:439–447.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50:1–26.

Willimack, D. K. and Nichols, E. (2010). A Hybrid Response Process Model for Business Surveys. *Journal of Official Statistics*, 26(1):3–24.

Willimack, D. K., Nichols, E., and Sudman, S. (2002). Understanding Unit and Item Nonresponse in Business Surveys. In Groves, R. M., Dillman, D. A., Eltinge, J. L., and Little, R. J., editors, *Survey Nonresponse*, chapter 14, pages 213–227. Wiley.

Wollmershäuser, T. and Henzel, S. (2005). Quantifying Inflation Expectations with the Carlson-Parkin Method: A Survey-based Determination of the Just Noticeable Difference. *Journal of Business Cycle Measurement and Analysis*, 2(3):321–352.

Wong, J. (2011). *imputation*. R package version 1.3.

Woolley, S. B., Cardoni, A. A., and Goethe, J. W. (2009). Last-observation-carried-forward imputation method in clinical efficacy trials: review of 352 antidepressant studies. *Pharmacotherapy*, 29(12):1408–1416.

Worton, D. A. and Platek, R. (1995). A History of Business Surveys at Statistics Canada: From the Era of the Gifted Amateur to That of Scientific Methodology. In Cox, B. G., Binder, D. A., Chinnappa, B. N., Christianson, A., Colledge, M. J., and Kott, P. S., editors, *Business Survey Methods*, chapter 32, pages 633–654. Wiley.

Zorn, C. J. W. (2001). Generalized Estimating Equation Models for Correlated Data: A Review With Applications. *American Journal of Political Science*, 45(2):470–490.

Curriculum vitae

Personal details

Name	Christian Seiler
Date of birth	11th May 1983
Place of birth	Sindelfingen, Germany
Citizenship	German

Education

since 07/2009	Doctorate in Statistics (Ludwig-Maximilians-University, Munich)
04/2003–09/2008	Diplom-Statistiker (Ludwig-Maximilians-University, Munich)
1993–2002	Abitur (Wilhelm-Hausenstein-Gymnasium, Munich)

Professional experience

since 09/2008	Survey specialist (Ifo Institute for Economic Research, Munich)
04/2007–08/2008	Student assistant (Ludwig-Maximilians-University, Munich)
10/2004–03/2007	Student assistant (Fraunhofer Society, Munich)

Scholarships, awards and prizes

09/2012	Isaac Kerstenetzsky Young Economist Award
07/2012	Prize for outstanding success in the management of contract studies

Publications

Articles in refereed journals (4)

- C. Seiler and K. Wohlrabe (2013), "Archetypal Scientists", ***Journal of Informetrics*** 7(2), 345-356.
- C. Seiler (2012), "The data sets of the LMU-ifo Economics & Business Data Center - A guide for researchers", ***Journal of Applied Social Science Studies*** 132(4), (in press).
- C. Seiler and K. Wohlrabe (2012), "Ranking economists based on many indicators: An alternative approach using RePEc data", ***Journal of Informetrics*** 6(3), 389-402.
- S. B. Adebayo, L. Fahrmeir, C. Seiler and C. Heumann (2011), "Geoadditive latent variable modelling of count data on multiple sexual partnering in Nigeria", ***Biometrics*** 67(2), 620-628.

Working papers & technical reports (9, selection)

- C. Seiler (2012), "Modelling Unit-Nonresponse in the Ifo Business Survey", *mimeo* (submitted to ***AStA Wirtschafts- und Sozialstatistisches Archiv***).
- C. Seiler (2012), "On the Robustness of the Balance Statistics with respect to Nonresponse", *Ifo Working Paper 125* (submitted to ***Journal of Business Cycle Measurement and Analysis***).
- C. Seiler and C. Heumann (2012), "Microdata Imputations and Macrodata Implications: Evidence from the Ifo Business Survey", Department of Statistics, *Technical Report 119* (submitted to ***Economic Modelling***).

Monographs (1)

- K. Abberger, S. Sauer and C. Seiler (2011), "Der Test des Tests im ifo Konjunkturtest Handel", ***ifo Forschungsberichte*** 52.

Papers in academic volumes (1)

- C. Strobl, C. Dittrich, S. Hackensperger, C. Seiler and F. Leisch (2010), "Measurement and Predictors of a Negative Attitude towards Statistics in LMU Students", in: Kneib, T. und Tutz, G. (eds.), ***Statistical Modelling and Regression Structures, Festschrift in Honour of Ludwig Fahrmeir***, Physica-Verlag, Berlin, 217-230.

Articles in non-refereed journals (11)

- C. Seiler and K. Wohlrabe (2012), "Archetypal Analysis: Ein neuer Ansatz zur Klassifizierung von Wissenschaftlern", ***ifo Schnelldienst*** 65(22), 7-12.

- C. Seiler (2012), "Zur Robustheit des ifo Geschäftsklimaindikators in Bezug auf fehlende Werte", ***ifo Schnelldienst*** 65(17), 19-22.

- C. Seiler and K. Wohlrabe (2010), "A Critique of the 2009 Global "Go-To Think Tanks" Ranking", ***CESifo DICE Report*** 8(2), 60-63.

- C. Seiler and K. Wohlrabe (2010), "Eine Kritik des »Global Go-To Think Tanks«-Rankings 2009", ***ifo Schnelldienst*** 63(11), 46-48.

- C. Seiler and K. Wohlrabe (2010), "RePEc - an Independent Platform for Measuring Output in Economics", ***CESifo Forum*** 11(4), 72-77.

- C. Seiler and K. Wohlrabe (2010), "RePEc - Eine unabhängige Plattform zur wirtschaftswissenschaftlichen Output-Messung", ***ifo Schnelldienst*** 63(7), 43-48.

- C. Seiler and K. Wohlrabe (2010), "Eine Anmerkung zur Zeitschriftengewichtung im Handelsblatt-Ranking 2010", ***ifo Schnelldienst*** 63(6), 38-39.

- B. Schirwitz, C. Seiler and K. Wohlrabe (2009), "Regionale Konjunkturzyklen in Deutschland - Teil I: Die Datenlage", ***ifo Schnelldienst*** 62(13), 18-24.

- B. Schirwitz, C. Seiler and K. Wohlrabe (2009), "Regionale Konjunkturzyklen in Deutschland - Teil II: Die Zyklendatierung", ***ifo Schnelldienst*** 62(14), 24-31.
- B. Schirwitz, C. Seiler and K. Wohlrabe (2009), "Regionale Konjunkturzyklen in Deutschland - Teil III: Konvergenz", ***ifo Schnelldienst*** 62(15), 23-32.
- K. Abberger, M. Birnbrich and C. Seiler (2009), "Der 'Test des Tests' im Handel - eine Metaumfrage zum ifo Konjunkturtest", ***ifo Schnelldienst*** 62(21), 34-41.

Presentations at international conferences

- "On the Robustness of the Balance Statistics with respect to Nonresponse", 31st CIRET Conference, 05.-08.09.2012, Vienna
- "Microdata Imputations and Macrodata Implications: Evidence from the Ifo Business Survey", 18th International Panel Data Conference, 05.-06.07.2012, Paris
- "Microdata Imputations and Macrodata Implications: Evidence from the Ifo Business Survey", 2011 German Statistical Week, 20.-23.09.2011, Leipzig
- "Microdata Imputations and Macrodata Implications: Evidence from the Ifo Business Survey", 2011 Joint Statistical Meetings, 30.07.-04.08.2011, Miami
- "Dynamic Modelling of Nonresponse in Business Surveys", 4th Conference of the European Survey Research Association, 18.-22.07.2011, Lausanne

Hiermit versichere ich, dass ich diese Dissertation selbstständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe.

München, den 20.02.2013

Christian Seiler