LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

INSTITUT FÜR STATISTIK

Christian Seiler

Dynamic modelling of Nonresponse in Business Surveys

Technical Report Number 093, 2010
Department of Statistics
University of Munich

http://www.stat.uni-muenchen.de

Dynamic modelling of Nonresponse in Business Surveys

Christian Seiler
Ifo Institute for Economic Research
Poschingerstraße 5, 81679 Munich
Phone: +49(0)89/9224-1248, seiler@ifo.de

It is well-known that nonresponse affects the results of surveys and can even cause bias 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 paper is one of the first which analyses a large business survey on micro data level for unit nonresponse. The data base is the Ifo Business Tendency Survey, which was established in 1949 and has more than 5,000 responding firms each month. The panel structure allows to use statistical modelling including time-varying effects to check for the existence of a panel fatigue. The results show that there are huge differences in business characteristics such as size or subsector and that nonresponse is more frequent in economically good times.

JEL Code: C33, C44, C81, C83
Key words: Business survey, Logistic regression, Nonresponse, Panel survey, Varying-coefficient model

1 Introduction

Data collection is the essential tool and fundamental for all empirical studies. In the socioeconomic sciences the methods used are mostly surveys. These surveys often face the problem of nonresponse, i.e., partial or complete drop out of information. In contrast to one-time studies, in panel surveys nonresponse is much more problematic, because the same units are analysed over time. Since panel studies with sociological or economic background mostly base on household surveys, a large literature exists about techniques for reducing the effect of nonresponse. Only less is known about the processes and reasons for participation and responding behaviour in business surveys (Janik and Kohaut, 2009). Although individuals are questioned in the survey, they are representatives of an organisation, so that organisational relationships have to be considered (Tomaskovic-Devey et al., 1995).

This paper is one of the first that models unit nonresponse behaviour in business surveys on micro level. To this end we examine the Ifo Business Tendency Survey (Ifo BTS) for unit nonresponse. 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. Because the Ifo BTS is a survey performed since 1949 with more than 5,000 respondents each month, it provides a large amount of data with panel structure. For the nonresponse analysis in this paper, monthly data from 1994 to 2009 is used which lead to a total number of observations of more than one million. Former empirical studies on nonresponse mainly focussed on the aggregate response rate, in particular explaining effects of survey characteristics. Others like Harris-Kojetin and Tucker (1999) examine the relationship between political and economic indicators and nonresponse rates using ARMA models. Steel et al. (1996) show that, in general, results from macro level analysis cannot be transferred directly to the individual level, because the nonresponse rate is the aggregation of many individual decisions. In recent years more studies investigated nonresponse on micro data level, for example Lepkowski and Couper (2002), Kalsbeek et al. (2002), and Schräpler (2004). They all use (multivariate) logit or probit models for statistical modelling but do not include dynamic effects since most of them use one-time surveys.

For household panel studies, Laurie et al. (1999) argue that the main problem is the phenomenon of 'panel fatigue', i.e. the respondents may loose interest in taking part in the survey with running participation time. Including time-varying effects could help explain the reasons behind panel fatigue. However, as analysis incorporating time-effects need long panels for good parameter interpretation, such models related to nonresponse studies can be found rarely. For example, Hawkes and Plewis (2006) use dynamic models for analysing nonresponse in six successive cohort studies. In addition the panel has to be long enough for consistent parameter estimation. With our data, on the one hand we are able to investigate the estimation of time effects, such as panel fatigue, and on the other hand we can model other variables flexibly by using Generalised Additive Models with time-varying effects.

Our analysis shows that time-varying effects are strongly present in business surveys and that the responding behaviour depends to a greater extent on the characteristics of the participants and less on the survey questionnaire design. This paper is organised as follows: Section 2 gives an overview of the survey and introduces the data set in detail. The statistical methodology is presented in Section 3. Section 4 sums up the empirical findings and gives a short outlook.

2 The Ifo BTS Data

The Ifo Business Tendency Survey is a monthly panel survey that has been conducted by the Ifo Institute for Economic Research since 1949. The Ifo BTS collects data from German companies on different aspects of their business parameters, such as business situation, business expectations, demand situation or change in staff. For an overview of the collected variables, see Becker and Wohlrabe (2008), for more methodological background of the survey, see Goldrian (2007). The monthly data sets are available at the Economics & Business Data Center (EBDC)[1], a combined platform for empirical research in business administration and economics of the Ludwig Maximilian University of Munich (LMU) and the Ifo Institute for Economic Research.

A specificity of the survey is that a single firm can answer more than one questionaire if the company operates in various business areas. This applies particularly to larger companies. For each of these areas, the company is asked to fill in a separate questionaire which is normally done by different persons. For reasons of simplicity, in this paper each report is treated

[1]http://www.cesifo-group.de/ebdc and Hönig (2009)

as coming from a different company. Since the surveys for industry, construction and trade differ in questionnaire design and, thus, probably in the factors that influence the response behavior, they are analysed separately.

2.1 Variables

The dependent variable is the risk of nonresponse in the given month since taking part at the Ifo BTS. It can be supposed that when firms start participating the risk of the nonresponse is lower than later on, because the company agreed to take part in the survey and therefore showed interest. However, the risk may be reduced over time, when only reliably reporting firms participate. In addition, there are many other risk factors that may influence the response behaviour. These variables have been categorised according to the conceptual framework of Willimack et al. (2002). This framework distinguishes two major categories of variables: Firstly those factors which are under the control of the researcher, related to the survey design (time schedule, instrument design, etc.) and secondly those 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 4 are listed in Table 1 and 2.

2.1.1 Survey Design

Since 1949, the industrial sector questionnaire has undergone very minor changes in terms of number of questions. 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 European Union programme of business and consumer surveys. 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 the analysis a dummy variable for short time schedule is introduced, which indicates all Decembers since 2002. Actually, the number of days to answer the questionnaire would be interesting, but these data are only available since 2003. In order to avoid a large reduction of the data set, this information cannot be included into the analyses.

Besides 'classical' paper mail, an alternative response mode was offered to the respondents of the construction firms from June 2002 and for the industry and trade firms from July 2004 with replying via internet. Unfortunately, it is not recorded over the whole period of time which firms used the online questionnaire, so that a dummy variable is included for the months after its introduction. Therefore, it can only be analysed whether the *possibility for online-answering* influences the responding behaviour.

2.1.2 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 nonresponse 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 are taken in the survey: Industry and construction firms are categorised by the number of employees once a year whereas the trade companies by their annual sales volume. This information is only 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. Furthermore, we abstract differences in regional response behaviour, but account for differences between companies from the former Eastern and Western states. It can be assumed that there was a transition period when the Ifo BTS was established after the reunification of Germany in the states of the former GDR.

2.1.3 The Respondent

Tomaskovic-Devey et al. (1994) point out that the authority of the respondent is important for the answering behaviour. For the Ifo BTS, Characteristics of the respondent, such as gender, age and position in the company are not available, even not on a annual basis. Abberger et al. (2009) undertook a meta survey directed to this question in spring 2009 with respect to the trade firms. Since this was an one-time survey the data were not merged with the Ifo BTS 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 the capacity and motivation of the respondent.

2.1.4 External Enviroment

An external aspect of nonresponding behaviour are the economic conditions prevailing at the time of the survey. Harris-Kojetin and Tucker (1999) find lower cooperation in a population survey in periods of economically better times. As the Ifo BTS focuses on economic parameters of the companies, there is a variety of possible indicators for the current economic situation of the 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 exploited. The Ifo Institute computes business situation indicators for each (sub)subsector, so the indicators from the lowest available aggregation level are used as an approximation of the business situation of the single firm in the appropriate (sub)subsector. This approach is 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 responding 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 in the model.

Groves et al. (2004) mention, that the intensity of survey research can be a reason for nonresponse. The 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, we have 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.

2.1.5 Additional variables

Several studies find evidence for declining interest in survey participation over the last decades (for an overview see de Leeuw and de Heer, 2002). Brehm (1994) points 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 the 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 the responding behaviour between 1994 and 2009. To deal with the problem of different vacation and working days each month - which speaks to the number of available days to respond - one might consider including the months as dummy variables in the model. But because the vacation days differ significantly between the German states and for a better approximation, the relation $vacationdays/workingdays$ for the corresponding month is used.

2.2 Descriptive analysis

Covering the period from 1994 to 2009, the total number of observations (including nonresponse) is 660,630 from 6,613 firms in industry (with an average nonresponse rate of 15.2%), 199,181 from 2,942 firms in construction (22.5%) and 273,873 from 4,151 firms in trade (22.5%). Table 1 gives an overview of all non-sector specific variables and Table 2 over the sector-specific. For the empirical analysis the medium categories for the companies' size and the first category of the subsectors are chosen as reference categories.

CATEGORY	VARIABLE	ABBREVIATION	CODING	% IND	% CON	% TRA
Survey design	Online questionnaire	$online$	1 since 06/2002 (con) / since 07/2004 (ind, tra)	30.6	43.9	35.0
	Number of questions	$questions$		-	-	-
	Short time schedule in December	$short\ ts$	1 for short time schedule	3.8	3.8	4.1
Business	Location	$east$	1 for firms located in Eastern Germany	22.7	33.5	24.2
	Size of the company	$size$			(see table 2)	
	Subsector of the company	$subsector$			(see table 2)	
External Enviroment	Business situation index in the (sub)sector	$business\ situation, bs$		-	-	-
	Received an additional survey by Ifo	$add\ survey$	1 for additional survey	10.0	4.0	5.9
Additional variables	Calendar time	$calendar\ time, ct$		-	-	-
	Vacation days in the federal state	$vacation\ days$		-	-	-
	Working days in the federal state	$working\ days$		-	-	-

Table 1: Description and coding of non-sector specific variables

SECTOR	VARIABLE	ABBREVIATION	CODING	%
Industry	Size of the company (no. employees)	$size$	1 for < 50 employees (smallest)	27.1
			2 for 50-199 employees (small)	31.5
			3 for 200-499 employees (medium)	19.1
			4 for 500-999 employees (large)	9.8
			5 for >1,000 employees (largest)	12.6
	Subsector	$subsector$	1 for food & tobacco	7.7
			2 for textiles, textiles products & leather	7.1
			3 for wood	4.1
			4 for pulp, paper, publishing & printing	15.1
			5 for petroleum & chemical products	5.2
			6 for rubber & plastic products	6.7
			7 for other non-metallic mineral products	6.3
			8 for basic metals & fabricated metal products	12.2
			9 for machinery & equipment	15.8
			10 for electrical & optical equipment	12.1
			11 for transport equipment	2.8
			12 for furniture & manufacture n.e.c.	4.9
Construction	Size of the company (no. employees)	$size$	1 for < 100 employees (smallest)	54.9
			2 for 100-199 employees (small)	23.4
			3 for 200-499 employees (medium)	12.2
			4 for 500-999 employees (large)	5.9
			5 for >1,000 employees (largest)	3.6
Trade	Size of the company (annual sales volume)	$size$	1 for < 1.0 million (smallest)	25.4
			2 for 1.0-5.0 million (small)	28.6
			3 for 5.0-12.5 million (medium)	18.6
			4 for 12.5-50.0 million (large)	20.4
			5 for > 50.0 million (largest)	7.1
	Subsector	$subsector$	1 for motor trade	10.5
			2 for wholes. trade of agricultural mat., food, beverages & tobacco	5.7
			3 for wholes. trade of household goods	14.3
			4 for wholes. trade of non-agricultural products	22.8
			5 for wholes. trade of machinery, equipment & supplies	4.7
			6 for retail trade	42.1

Table 2: Description and coding of sector specific variables

3 Methodology

All variables presented in Section 2 have a panel structure, so the data set has the form $(y_{it_i}, x_{it_i}), i = 1, \ldots, n_{sector}$ and $t_i = 1, \ldots, T_i$. Given that the dependent variable is an 1/0-dummy, $y_{it_i} = 1$ if company i did not answer the questionnaire in the t_i-th month of survey participation and $y_{it_i} = 0$ if it was observed in the data. In a non-dynamic approach, one would estimate a model specified as

$$g(\pi_{it_i}) = \eta_{it_i} = \beta_0 + x_{it_i}\beta \tag{1}$$

with $\pi_{it_i} = E(y_{it_i})$ and an appropriate link function $g(\cdot)$, such as logit or probit. For simplicity i is not further illustrated. To model time-varying effects, however, (1) is extended to a varying-coefficient model in line with Hastie and Tibshirani (1993). The explanatory variables will be separated in the first step into two groups: Let $x_{\mathcal{V}}$ be the set of variables to be estimated time-varying and $x_{\mathcal{C}}$ all variables estimated parametrically with time-constant effects. Then (1) yields to

$$\eta_t = \beta_0 + x_{\mathcal{C}}\beta_{\mathcal{C}} + x_{\mathcal{V}}\beta_{\mathcal{V}}(t_{\mathcal{V}}) \tag{2}$$

where $\beta_{\mathcal{C}} = (\beta_{\mathcal{C}_1}, \ldots, \beta_{\mathcal{C}_p})'$, $\dim(x_{\mathcal{C}}) = p$, and $\beta_{\mathcal{V}}(t_{\mathcal{V}}) = (\beta_{\mathcal{V}_1}(t_{\mathcal{V}_1}), \ldots, \beta_{\mathcal{V}_q}(t_{\mathcal{V}_q}))'$, $\dim(x_{\mathcal{V}}) = q$, with $t_{\mathcal{V}_j}, j = 1, \ldots, q$, as time-varying effect modifier. The functions in $\beta_{\mathcal{V}}(t_{\mathcal{V}})$ can also be interpreted as semi-parametric terms reflecting an interaction between the effect modifiers $t_{\mathcal{V}}$ and $x_{\mathcal{V}}$-covariates. Note that each $\beta_{\mathcal{V}_j}(t_{\mathcal{V}_j})$ can be modified by any $t_{\mathcal{V}_j}$ and the effect varies smoothly over it. This allows us to model time-varying effects over the two different time dimensions in the data set, the participation month t and $calendar\ time$.

Although model (2) includes flexible modelling, the covariates $x_{\mathcal{C}}$ have still a linear effect on η_t. In particular, this is rather restrictive for metrically scaled variables. Therefore, let $x_{\mathcal{N}}, \dim(x_{\mathcal{N}}) = r$, be the set of all variables to model nonparametrically, then (2) is extended to

$$\eta_t = \beta_0 + x_{\mathcal{C}}\beta_{\mathcal{C}} + x_{\mathcal{V}}\beta_{\mathcal{V}}(t_{\mathcal{V}}) + \sum_{k=1}^{r} f_{(k)}(x_{\mathcal{N}_k}) \tag{3}$$

with unspecified smooth functions $f_{(k)}(\cdot), k = 1, \ldots, r$, leading to a *Generalised Additive Model* (GAM) with $x_t \equiv (x_{\mathcal{C}}, x_{\mathcal{V}}, x_{\mathcal{N}}), \dim(x_t) = p + q + r$. It is also possible to modify nonparametrically estimated variables with time-varying coefficients, but this will not be covered in this paper since all of these variables vary evenly over time. For further information of this type of modelling see Tutz and Binder (2004). Finally, this model class provides a very flexible kind of estimation but inherent in it are some identification restrictions to be made which have an impact on the interpretation of the results. An outline is given in the appendix.

The estimation of the smooth functions $\beta_{\mathcal{V}_j}(t_{\mathcal{V}_j})$ as well as $f_{(k)}(x_{\mathcal{N}_k})$ is based on penalised basis functions (P-splines) which were introduced by Eilers and Marx (1996). Define $h(\cdot)$ as a function to be estimated, then it can be written as

$$h(x) = \sum_{s=1}^{M} \alpha_s B_s(x)$$

with $B_s(\cdot)$ being one of the M basis functions, connected at the knots κ_s. The fitted function is sensitive to the number of basis functions as well as the number and location of the knots κ_s. Depending on the degree l of $B_s(x)$ the B-spline functions only have positive values on the interval based on $l+2$ adjacent knots. To avoid overfitting, second order penalty terms of form $\lambda \int (f''(z))^2 dz$ are used. For more information on the methodological and computational implementation see Wood (2006).

4 Results and Discussion

All variables described in Section 2 and listed in Tables 1 and 2 are potential factors that may influence the responding behaviour. They enter the model as follows:

$$\begin{aligned}\eta_t &= \beta_0 + \beta_t(t) + \beta_{ct}(calendar\ time) + east\ \beta_{east}(calendar\ time)\\ &\quad + size\ \beta_{size} + online\ \beta_{online} + subsector\ \beta_{subsector}\\ &\quad + short\ ts\ \beta_{short\ ts} + \frac{vacation\ days}{working\ days}\ \beta_{days} + add\ survey\ \beta_{add\ survey}\\ &\quad + questions\ \beta_{questions} + f_{bs}(business\ situation)\end{aligned}$$

with a logit link function. Note that β_{size} and $\beta_{subsector}$ are vectors and $\beta_{subsector}$ is excluded from the model for the construction companies since this information is not available. $\beta_{east}(calendar\ time)$ is estimated as time-varying since it is assumed that the responding behavior differs between Eastern and Western firms over time. The adjusted R^2 for this model is 0.15 (industry), 0.06 (construction) and 0.05 (trade).

4.1 Interpretation of the results

For the interpretation of the results, all parametrically estimated variables will be discussed first. The estimates can be found in Table 3. Subsequently, all nonparametrically estimated variables will be interpreted. The fitted functions are displayed in Figure 1. Since a logit function was used, an estimate γ increases the odds for nonresponse by the factor $\exp(\gamma)$.

The estimates of the intercepts in Table 3 show that the industry firms tend less to nonresponse than the construction and trade firms. Besides, for the industry sector there exist major differences between the subsectors with coefficients ranging from -0.2 to -0.6. Since these are aggregated, this may be a potential source of distortion in cases of significantly different economic development. The introduction of the online survey reduced the probability for nonresponse strongly in construction, but has led to a higher probability in the industrial and trade sector. Kwak and Radler (2002) find that online surveys lead to higher nonresponse rates, but they show that their advantages results in a faster response speed. This is probably due to the fact that the companies feel more obliged to respond when they receive a paper questionnaire. However, as discussed in Section 2.1.1, this variable is perhaps not very accurate because information on micro level does not enter the model. Next, the short time schedule in December has a negative impact on the responding behaviour in all sectors except of trade. This may be because the trade sector shows a higher willingness to answer in the Advent season, which is the most important time of the year for their business. However, the effect is close to zero and not significant. Sending an additional survey to the firms also affects the responding behaviour for construction and trade in a negative way, so that an excessive questioning of participants should be avoided. For example, sending an additional survey in trade raises the odds for nonresponse by the factor $\exp(0.0586) = 1.06$. For the industrial sector we find a positive effect, but this is not significant at the 90%-level. The number of questions on the questionnaire seems to have only a small impact on the responding behaviour. Overall, the effects are very close to zero. Unsurprisingly, with rising proportion of vacation to working days in a given month, the firms tend more to nonresponse. It can be assumed that the respondent is more likely not in office and therefore has less time to fill the questionnaire. The responding behaviour also varies for different business' sizes: Basically, larger firms are more likely to respond than smaller ones. This effect is pretty stable across all sectors. Although organisational performance generally rises with the size of the company, suggesting that they may benefit more from the survey results than the smaller firms and are therefore more willing to respond regularly.

For the interpretation of the nonparametric and the time-varying coefficient terms, identification restrictions mentioned in Section 3 and outlined in the appendix have to be considered. The first row in Figure 1 shows the effect of $\beta(t)$. As can be seen the firms are more likely to nonresponse when they become participants. With proceeding participation time, the responding behaviour increases in general. For example, in all sectors we estimated in first participation month an effect of about $\beta_t(t = 1) = 0.5$. So, the odds of nonresponse increase by the factor $\exp(0.5) = 1.65$. After 700 months of participation, in industry the effect declines to about $\exp(-0.5) = 0.61$. Row

2 displays the effect of $\beta_{ct}(calendar\ time)$. It can be seen that the willingness to participate over time in the last 16 years was subject to only minor fluctuations. A general trend towards a lower willingness to participate is not visible. Row 3 displayes the differences between Eastern and Western firms with running calendar time $\beta_{east}(calendar\ time)$. For all three sectors, the difference in answering behaviour dropped and at the end of 2009 nearly no significant difference can be observed. This could be due to a transition period when a existing panel is introduced into a new region has to become established. In row 4, the effect of business situation $f_{bs}(bs)$ is shown. For reasons of comparability, these indicators were centered to their long-running mean. A higher value indicates a better business situation. For industry and trade, a differently pronounced U-shape function was estimated, while in the construction sector a more linear relationship can be found. All of them also differ in the magnitude of the effect. It can be seen that in particular in economically good times the firms tend more to nonresponse which verifies the result of Harris-Kojetin and Tucker (1999). This is presumably due to the fact that in boom times the companies have less time to answer the questionnaire because of many orders. This can, but not has to, be a possible source of bias, since potential positive replies are missing.

Our results differ from those in Janik and Kohaut (2009), who also examine the response behaviour of German companies, but do not model dynamics since they use only the 2006 data from the IAB Establishment Panel. For this reason, their analysis do not contain dynamically modelled effects. The only dynamic component is the participation time in years of each company, where a declining trend was found. This result is consistent with the analysis presented here. However, we can show that this effect is clearly not linear with rising participation time.

4.2 Summary and Outlook

This paper models unit nonresponse behaviour in a business panel survey with varying-coefficient models including additive effects. The analysis shows that the risk of nonresponse decreases over participation time. A panel fatigue in the sense of an increasing nonresponse behaviour with running participation time is not present. Considering the framework of Willimack et al. (2002) and the magnitudes of the estimated effects, the main reasons for different responding 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, play a minor role in the participation process. After controlling for these survey methodologic related effects, the willingness to participate also depends to a small extent on the business situation. In particular, in economically good times the companies respond less often. Since the Ifo BTS focusses 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 low distortion, imputation methods can be used for analysing these effects by developing a consistent estimation for the missing data and recalculating the survey results. Using these methods can analyse how much the bias is and how a consistent and economically motivated estimation of the missing values can be constructed. Since the data is in a high frequency, the panel structure can be used.

Acknowledgements

The author thanks Teresa Buchen, Kai Carstensen, Oliver Falck, Christian Heumann, Stefan Kipar, Gerhard Tutz, Joachim Winter and Klaus Wohlrabe for their useful comments that have helped to improve this paper. The author also thanks Heike Mittelmeier and Wolfgang Ruppert for their useful help with the data sets.

	Industry		Construction		Trade	
Variable	coef.	p-value	coef.	p-value	coef.	p-value
Intercept	-2.1817	0.000	-1.2893	0.000	-1.3479	0.000
Additional survey	-0.0212	0.102	0.0611	0.029	0.0586	0.005
Number of questions	0.0006	0.495	-0.0006	0.004	-0.0127	0.000
Online	0.3400	0.000	-0.0942	0.015	0.0641	0.056
Short time schedule	0.0892	0.000	0.0869	0.004	-0.0144	0.564
Vacation days/Working days	0.0655	0.000	0.0314	0.121	0.0232	0.175
Size:						
Smallest	-0.1118	0.000	0.1282	0.000	0.2268	0.000
Small	0.1946	0.000	0.0139	0.490	0.0502	0.001
Large	-0.2224	0.000	-1.0753	0.000	-0.1725	0.000
Largest	-0.4702	0.000	-1.1964	0.000	-0.0821	0.000
Subsector (industry):						
Textiles, textiles products	-0.5224	0.000	-	-	-	-
Wood	-0.5727	0.000	-	-	-	-
Pulp, paper, publishing & printing	-0.2241	0.000	-	-	-	-
Petroleum & chemical products	-0.4402	0.000	-	-	-	-
Rubber & plastic products	-0.3105	0.000	-	-	-	-
Other non-metallic mineral products	-0.6415	0.000	-	-	-	-
Basic metals & fabricated metal products	-0.3869	0.000	-	-	-	-
Machinery & equipment	-0.4280	0.000	-	-	-	-
Electrical & optical equipment	-0.4922	0.000	-	-	-	-
Transport equipment	-0.2275	0.000	-	-	-	-
Furniture & manufacture n.e.c.	-0.4484	0.000	-	-	-	-
Subsector (trade):						
Wholes. trade of agr. materials, food, beverages & tobacco	-	-	-	-	-0.1714	0.000
Wholes. trade of household goods	-	-	-	-	-0.0070	0.719
Wholes. trade of non-agricultural products	-	-	-	-	-0.1600	0.000
Wholes. trade of machinery, equipment & supplies	-	-	-	-	0.0517	0.048
Retail trade	-	-	-	-	-0.0058	0.727

Table 3: Estimates of the parametric terms

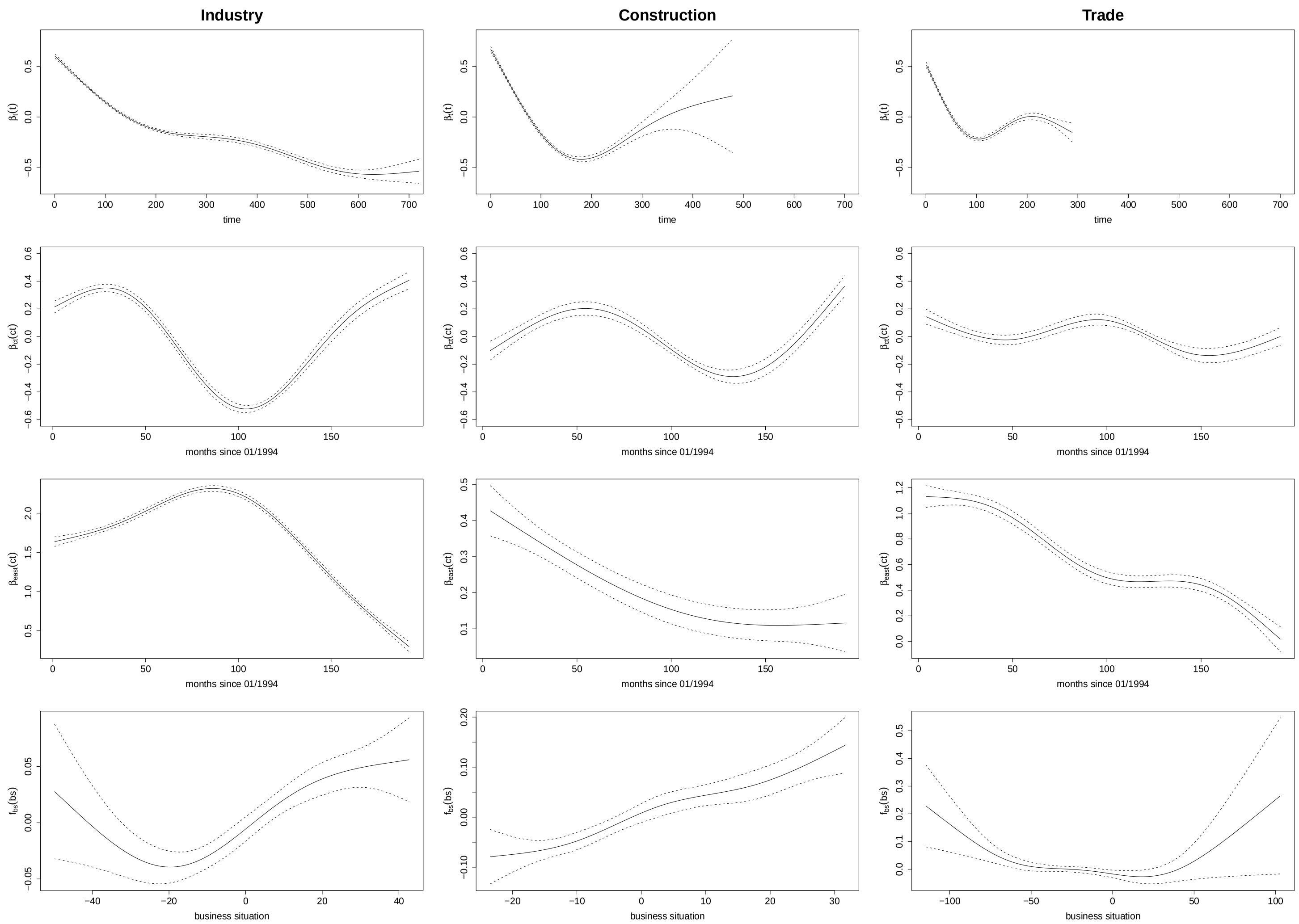

Figure 1: Nonparametric functions as noted in the model. Row 1 shows the effect of participation time on non-response. Row 2 displays the time-effect since 01/1994, row 3 the additional time-effect for the Eastern German firms. The effect of the aggregated business situation value of the responding firms is shown in row 4.

A Identification restrictions in GAMs

As mentioned in section 3, using GAMs come along with some identification restrictions. Suppose you have functions $f_1(z_1)$ and $f_2(z_2)$. Now z_1 and z_2 are transformed to $\tilde{z}_1 = z_1 + c$ and $\tilde{z}_2 = z_2 - c, c \neq 0$. Then the sum of the functions

$$\tilde{f}_1(\tilde{z}_1) + \tilde{f}_2(\tilde{z}_2) = f_1(z_1) + c + f_2(z_2) - c$$

stays unchanged, i.e. the effect of the predictors does not change. Although the shape of the functions is the same, the functions are not identifiable since the level has to be fixed. This is normally done by centering all functions around 0. The same problem arises when one variable is modified by another. For example, suppose you have one dummy variable z and an effect modifier t. If you estimate the model

$$y = \beta_0 + \beta_1 z + \beta(t) + \beta_z(t) z + \epsilon$$

then $\beta(t)$ can be interpreted as the effect of t in the reference category whereas $\beta(t) + \beta_1 z + \beta_z(t) z$ is the effect of t if $z = 1$. In this setting, $\beta_z(t)$ is scaled around 0. Therefore, if this is not required for reasons of interpretation, excluding the main effect β_1, $\beta_z(t)$ is no longer forced to be scaled around 0.

References

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.

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.

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.

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.

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing using B-splines and penalized likelihood. *Statistical Science*, 11:89–121.

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

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.

Hastie, T. and Tibshirani, R. (1993). Varying-coefficient Models. *Journal of the Royal Statistical Society, Series B*, 55(4):757–796.

Hawkes, D. and Plewis, I. (2006). Modelling non-response in the National Child Development Study. *Journal of the Royal Statistical Society, Series A*, 169(3):479–491.

Hönig, A. (2009). The New EBDC Dataset: An Innovative Combination of Survey and Financial Statement Data. *CESifo Forum*, 10(4):62–63.

Janik, F. and Kohaut, S. (2009). Why Don't They Answer? – Unit Non-Response in the IAB Establishment Panel. FDZ Methodenreport 7/2009, Bundesagentur für Arbeit.

Kalsbeek, W. D., Yang, J., and Agans, R. P. (2002). Predictors of nonresponse in a longitudinal survey of adolescents. In *ASA Proceedings of the Joint Statistical Meetings*, pages 1740–1745.

Kwak, N. and Radler, B. (2002). A Comparison Between Mail and Web Surveys: Response Pattern, Respondent Profile, and Data Quality. *Journal of Official Statistics*, 18(2):257–273.

Laurie, H., Smith, R., and Scott, L. (1999). Strategies for Reducing Non-response in a Longitudinal Panel Survey. *Journal of Official Statistics*, 15(2):269–282.

Lepkowski, J. M. and Couper, M. P. (2002). Nonresponse in the Second Wave of Longitudinal Household Surveys. In Groves, R. M., Dillman, D. A., Eltinge, J. L., and Little, R. J., editors, *Survey Nonresponse*, chapter 17, pages 259–272. Wiley.

Schräpler, J. P. (2004). Respondent Behavior in Panel Studies - A Case Study for Income-Nonresponse by means of the German Socio-Economic Panel (SOEP). *Sociological Methods and Research*, 33(1):118–156.

Steel, D. G., Holt, D., and Tranmer, M. (1996). Making Unit-Level Inferences from Aggregated Data. *Survey Methodology*, 22:3–15.

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.

Tutz, G. and Binder, H. (2004). Flexible modelling of discrete failure time including time-varying smooth effects. *Statistics in Medicine*, 23:2445–2461.

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.

Wood, S. N. (2006). *Generalized Additive Models: An Introduction with R*. Chapman and Hall, CRC.