Kauermann, Tutz:

Vanishing of Risk Factors for the Success and Survival of Newly Founded Companies

Sonderforschungsbereich 386, Paper 242 (2001)

Vanishing of Risk Factors for the Success and Survival of Newly Founded Companies

Göran Kauermann*
University of Glasgow

Gerhard Tutz†
Ludwig-Maximilians-Universtität München

15th May 2001

Abstract

The success of a newly founded company or small business depends on various initial risk factors or staring conditions, respectively, like e.g. the market the business aims for, the experience and the age of the founder, the preparation prior to the launch, the financial frame, the legal basis of the company and many others. These risk factors determine the chance of survival for the venture in the market. However, the effects of these risk factors often change with time. They may vanish or even increase with the time the company is in the market. In this paper we analyse the survival of 1123 newly founded companies in the state of Bavaria, Germany (see Brüderl, Preisendörfer & Ziegler, 1992). Our focus is thereby primarily on the investigation of time-variation of the initial factors for success. The time-variation is thereby tackled within the framework of varying coefficient models, as introduced by Hastie & Tibshirani (1993), where time modifies the effects of risk factors. An important issue in our analysis is the separation of risk factors which have time-varying effects from those which have time-constant effects. We make use of the Akaike criterion to separate these two types of factors.

*Department of Statistics , University Gardens, University of Glasgow, Glasgow G12 8QW, UK
†Department für Statistik, Akademiestraße 1, 80799 München, Germany

1 Introduction

In times of economic change many new firms come into the market, as e.g. in recent years start-up companies have become a talking point. It is of great interest for economists and financiers, as well for the company founders of course, to predict the success and the survival chances of the company. Of particular interest at the time-point of launching the business is thereby the investigation of available indicators of success, called starting conditions or risk factors which known at the beginning. These may be used in the prognosis of success or failure. Although a prognosis right from the start is important, one also often wants to know the chances of survival after some years in the marketplace. Then, the initial indicators which are most influential at the beginning, might loose their prognostic power as their effect varies with time. In this paper the objective is to investigate if and how the effects of starting conditions vary over the time a company is in the market.

We consider a sample of 1123 firms which have been founded during the years 1985 and 1986 in the state of Bavaria in the southern part of Germany. The dataset is available from the Central Archive for Empirical Social Research, University Cologne, Germany (`http://www.gesis.org/ZA/`). The original survey consists of 1849 firms but we pursue a complete case analysis here.

The data have been thoroughly investigated before in several papers by Brüderl, Preisendörfer & Ziegler (1992) or Brüderl (1995). The authors did not however investigate any time variation of effects. In general, the problem in the investigation of time variation in the effects of starting conditions is, that many influential variables are to be considered but possibly only a few of them have time-varying effects. It is therefore of particular interest to distinguish between variables which have time varying effects and variables which have time-constant effects. We do this by applying an Akaike criterion to different models for the data.

(Table 1 about here)

The variables considered in our analysis are given subsequently, see also Table 1. We consider variables describing the legal and financial frame of the company, as well as the number of employees and the length of preparation time for the business. Moreover, we investigate the effect of the target market for the business and the qualifications of the owner. In particular we include the following variables in our analysis.

- Legal basis of the business

Based on the German law we classify the business in small business ($legal1 = 1$, "*Kleingewerbe*"), partnership ($legal2 = 1$, "*Personengesellschaft*") and joint-stock companies ("*Kapitalgesellschaften*"), which is taken as reference category.

- Financing of the business

Companies without starting capital at the time-point of foundation are treated as reference category. For other companies we record whether they start with outside capital ($out.capital = 1$) or own company capital ($in.capital = 1$).

- Number of employees

We treat a single person business as reference category and distinguish between companies with 2 or more employees ($employ1 = 1$) respectively with 3 or more employees ($employ2 = 1$).

- Preparation of business

Companies were prepared by a planing process and/or a trial period or lead time.

We record whether these preparation periods were longer than 6 months ($plan = 1$ if business was planed $> 1/2$ year, $lead = 1$ if trial period $> 1/2$ year).

- Target market of business

The target market is described with with its price level ($price = 1$ if the target market has high price level, 0 otherwise), and whether the product offered or produced is already available in the market ($product = 1$ if product is in the market, 0 for new development). Moreover it is recorded whether the business aims for a national or international market ($national = 1$ if the target market is national, 0 for international) and whether the target market is a niche market with specialised customers ($customer = 1$ if target customer is not special, 0 for special customer).

- Founder of the business

The founder of the business is described with his/her gender ($gender = 1$ for males, 0 for females) and age ($age1 = 1$ for junior owner with $\leq$ 30 years of age, 0 otherwise; $age2$ =1 for senior owner older than 40 years, 0 otherwise). The education of the owner is characterised by its school and higher education degree ($school = 1$ for high-school degree, 0 otherwise; $training = 1$ for master degree including craftsman or university, 0 otherwise) Moreover the experience of the owner in the branch and area of business is considered ($branch = 1$ for expert of branch, 0 otherwise)

Since the exact date of failure is of minor importance, we consider discrete survival models where time is measured in months. Figures 1 and 2 show the size and the time-variation of the above effects. As can be seen from the plots, some effects clearly change with time, while others remain constant. We will apply an Akaike criteria to separate the two sets of covariables, i.e. those with time-varying and those with time-constant effects.

The paper is organised as follows. In Section 2 we introduce the model which yields the estimates shown in Figures 1 and 2. We introduce semiparametric model where some effects are time-varying and others or time-constant. Model selection based on the Akaike criteria is introduced to distinguish between risk factors with time varying and time constant effects, respectively. In Section 3 we apply the modelling approach and determine the factors influencing the success of new companies. Technical details as well as a description of the estimation routine is postponed to the Appendix.

2 Survival Model with Varying Coefficients

2.1 Discrete Survival Model

Let T_i be the (true) survival time of the i-th company. This may be unobserved as the company might be still in business. Instead, we observe the right censored version of survival time C_i and denote with δ_i the censoring indicator with

$$\delta_i \;=\; \begin{cases} 1 & T_i < C_i \\ 0 & T_i \geq C_i. \end{cases}$$

Thus $\delta_i = 1$ indicates that failure has been observed, i.e. the business got bankrupt, while $\delta_i = 0$ indicates that the company is still in the market at the end of the observation period. In the following, random censoring is assumed, which means that T_i and C_i are considered as independent. Survival is measured on a discrete scale like monthly intervals and the distribution of δ_i implicitly assumes that censoring takes place at the beginning of the interval which determines the discrete scale. For discrete time measurements the essential instrument to investigate survival is the discrete hazard function

$$\lambda(t) \quad = \quad P(T = t | T \geq t). \tag{1}$$

The hazard function gives the probability for failure in the next time period. We model $\lambda(t)$ to depend on covariates or risk factors, respectively, by assuming

$$\lambda(t|x) \quad = \quad h\{\beta_0(t) + \sum_{k=1}^{K} x_k \beta_k(t)\} \tag{2}$$

where $h(\cdot)$ is a known response function. A common choice for $h(\cdot)$ is the logit function which is used in the following. In (2), $\beta_0(t)$ is the smooth baseline hazard while $\beta_k(t)$ is the effect of the covariable x_k which is allowed to vary with time as smooth but unspecified functions, $k = 1, \ldots, K$. One may consider the variation of the effects to mirror the interaction between the covariates x_k and the survival time t. In general, the covariates x_k may also be time-varying itself, i.e. $x_k = x_{kt}$. For simplicity of notation and also for ease of interpretation we restrict the presentation in the following to covariates which are constant over time.

Model (2) provides a large amount of flexibility as seen from Figures 1 and 2. This is since the shape of the effects is unspecified at all. However, model (2) may not be parsimonious at the same time since some of the covariate effects appear to be constant over time, i.e. $\beta_k(t) \equiv \beta_k$. It seems therefore advisable to extend (2) by allowing for a mixture of time varying and time constant effects. This is achieved by the semiparametric model

$$\lambda(t|x) \quad = \quad h\{\beta_0(t) + \sum_{k \in \mathcal{V}} x_k \beta_k(t) + \sum_{k \in \mathcal{C}} x_k \beta_k\} \tag{3}$$

where $\mathcal{V}$ is the index set of covariables with time-varying effect and $\mathcal{C}$ is the index set of time-constant effect variables.

In the analysis of the data it is of particular interest to determine the sets $\mathcal{V}$ and $\mathcal{C}$ of time-varying and time-constant effects. We will make use of the Akaike criteria and select a flexible but still parsimonious model.

2.2 Model-selection

The selection of an appropriate model has in general to consider two components. First a bandwidth h used for smoothing has to be selected. Secondly the sets $\mathcal{V}$ and $\mathcal{C}$ of covariates with time varying and time constant effects have to be determined. For reasons of interpretation of the model, the second point of the model selection is more important than the first one and is therefore emphasized here. Let $\mathcal{M}$ be the model determined by specification of index sets $\mathcal{C}$ and $\mathcal{V}$ and let $\eta_{\mathcal{M}}$ denote the corresponding predictor, i.e. in model (3) one has $\eta_{\mathcal{M}} = \beta_0(t) + \sum_{k\in\mathcal{V}} x_k\beta_k(t) + \sum_{k\in\mathcal{C}} x_k\beta_k$. Correspondingly, $\widehat{\eta}_{\mathcal{M}}$ denotes the fitted predictor. The Akaike information for model $\mathcal{M}$ is then defined by

$$AIC(\mathcal{M}) \quad = \quad -2\mathbf{l}(\widehat{\eta}_{\mathcal{M}}) + 2\dim(\mathcal{M}) \tag{4}$$

where $\mathbf{l}(\cdot)$ is the likelihood evaluated at the predictor (see appendix for details) and $\dim(\mathcal{M})$ denotes the dimension or the degree of model $\mathcal{M}$. The intention is to select a model $\mathcal{M}$ which miniizes (4). In the time constant effects model $\mathcal{M}_{\mathcal{C}}$ with $\lambda(t|x)$ $= h\{\beta_0(t) + \sum_{k\in\mathcal{C}} x_k\beta_k\}$, i.e. the number of the elements in $\mathcal{C}$. In the solely smooth model $\mathcal{M}_{\mathcal{V}}$, as given in (2), we make use of approximations as derived in Kauermann & Tutz (2000) and define the dimension by the trace of a smoothing matrix. This means $\dim(\mathcal{M}_{\mathcal{V}}) = |\mathcal{V}|\mathrm{tr}(S_h)$ where S_h is the $T \times T$ dimensional smoothing matrix

$$S_{\mathcal{V},tj} \quad = \quad w_{tj}\tau_j \left(\sum_{s=1}^{T} w_{ts}\tau_s\right)^{-1} \tag{5}$$

where $w_{tj} = K\{(t-j)/h\}/K(0)$ with $K(\cdot)$ as unimodal, symmetrical kernel function, h as bandwidth and τ_j as the number of observations at risk at time j, i.e. $\tau_j =$

$\sum_{i=1}^{n} \delta(C_i > j)$ with $\delta(\cdot)$ as Indicator function. Since $w_{tt} = 1$ one obtains

$$\dim(\mathcal{M}_{\mathcal{V}}) = |\mathcal{V}| \sum_{t=1}^{T} \tau_t \left(\sum_{s=1}^{T} w_{ts} \tau_s \right)^{-1} .$$

Note that $\dim(\mathcal{M}_{\mathcal{V}}) \geq |\mathcal{V}|$ where equality holds if bandwidth $h \to \infty$. In Kauermann & Tutz (2001) it is shown that if profile likelihood estimation is used in the semiparametric model $M_{\mathcal{V},\mathcal{C}}$ as given in (3), the smoothing step and the parametric fit are "locally orthogonal". This means that the degree of a semiparametric model can be defined by

$$\dim(\mathcal{M}_{\mathcal{V},\mathcal{C}}) = |\mathcal{V}| \text{tr}(S_h) + |\mathcal{C}|. \tag{6}$$

Note that $\dim(\mathcal{M}_{\mathcal{V}})$ and $\dim(\mathcal{M}_{\mathcal{C}})$ as given above result as special cases of (6). In the next session we use (4) to choose an appropriate model for the survival of companies.

3 Survival of Newly Founded Companies

We return to the data introduced in Section 1. As a first step, model (2) with all 18 covariates as described in Section 1 is fitted. (The smoothing parameter h was determined by minimizing the Akaike criterion yielding $h = 10$ for $K(\cdot)$ as the normal kernel) Model (2) allows all of the coefficients to be time-varying as seen from Figures 1 and 2. The baseline clearly shows a quadratic shape which indicates that the risk for getting bankrupt is increasing in the first 18 months and is slowly decreasing afterwards.

(Figures 1 and 2 about here)

We included the fitted time-constant effects resulting from model $\mathcal{M}_{\mathcal{C}}$ as horizontal lines. It is obvious that the effects given by the horizontal lines represent an average across the time-variation. For some covariates this seems adequate while for others time-variation is obvious. For example the time-variation in the legal basis is very close to a straight line, thus a time-constant effect seems to be adequate. In contrast, for variable *out.capital*, which indicates whether the company started with outside capital, a time-constant effect seems not appropriate as the effect is decreasing across time. Although the time-constant effect is negative and signals a slight reduction for the relative risk for companies going bankrupt if they have outside finance, the risk is actually not reduced in the start up period but it increases after about 2 years. This means that firms which were able to raise money before launching have a better chance of survival in the long run.

(Table 2 about here)

We now make use of the Akaike criterion to find an appropriate model for the data, i.e. to separate between time-constant and time-varying covariates. Table 2 shows a selection of models which have been fitted. From the top downwards the number of variables modelled with time-variation is successively reduced. The minimal model $\mathcal{M}_{\mathcal{C}}$ (model 5) is the model shown in Figures 1 and 2 as horizontal dashed lines. Clearly, based on the AIC criterion model 5 seems not adequate compared to other models. This indicates that there is time variation for some of the risk factors considered. The model with smallest AIC value is model 4 where only *out.capital*, *plan*, *price* and *age* are time-varying. The estimated varying coefficients are given in Figure 3 and fitted time constant-effects are listed in Table 3.

(Figure 3 and Table 3 about here)

It is seen from the effect of the variable *plan*, that if the launch of the new business was planed well in advance it reduces the hazard during the starting up period, but the effect vanishes with time and disappears after being in the market for about 3 years. The same holds for the variable *price*, meaning that a new business which aims for a high price market has a reduced risk of failure for the starting up period. This advantage however fades away after about three years. Finally, the age of the person who is running the business has a time-varying effect. If the person who is in charge of the business is young (*le* 30 years) the hazard is increased. This effect decreases with time even though it does not vanish.

Table 3 shows the time-constant effects some of which are not significant. Refitting the model with the significant variables provides similar parametric estimates as seen from Table 2 (model 4^*). The smooth estimates also resemble very much those from model 4 and are therefore not shown. The constant effects show that the legal form of the company is a time constant risk factor indicating an increased risk for small businesses and partnerships compared to joint stock companies. If a company is large enough at the launch to employ ≥ 3 people it proves to have better survival chances. Moreover, if the company was run in a trial period it reduces the risk of failure. Companies aiming for a national market only and companies serving a niche market with specialized customers also have better chances to remain in the market. Finally, prior knowledge of the branch is clearly an advantage for the company. All these effects remain constant over the time the company is in the market.

4 Discussion

The analysis shows that the effect of risk factors for the survival and success of newly founded companies can vary with the length of time that the business is in the market. This means that even though at the launch of the company some of the factors may be useful for predicting its probability of survival in the market, the predictive power can vary and decrease with the age of the company. On the other hand, the analysis also shows that not all risk factors have a time varying effects but some preserve a constant prognosis power even after years.

A Technical Details

A.1 Varying Coefficient Model

Local Likelihood

We describe briefly how to fit model (2) by local likelihood. Let observations be given by (t_i, δ_i, x_i), $i = 1, \ldots, n$, where time $t_i = \min(T_i, C_i)$ is the minimum of failure time T_i and censoring time C_i. Assuming random censoring, the likelihood contribution of observation i is given by

$$l_i \;=\; c_i \lambda(t_i|x_i)^{\delta_i} \prod_{j=1}^{t_i - 1} \{1 - \lambda(j|x_i)\} \tag{7}$$

where the constant $c_i = P(C_i > t_i)^{\delta_i} P(C_i = t_i)^{1-\delta_i}$ is considered as non-informative and is therefore omitted in the following (see e.g. Fahrmeir & Tutz, 2001, chapter 9). For ease of presentation it is useful to rearrange the data by using two types of indicator variables. The variable

$$y_{it} \;=\; \begin{cases} 1 & \text{if } t_i = t, \delta_i = 1 \\ 0 & \text{otherwise} \end{cases}$$

indicates failure at time point t while

$$\delta_{it} \;=\; \begin{cases} 1 & \text{if } t \le t_i \\ 0 & \text{otherwise} \end{cases}$$

indicates whether the firm belongs to the risk set at time point t. Note that by construction $\delta_{it} = 1$ implies $t_i \geq t$. It is easily seen that based on (7) the log likelihood contribution for the ith firm has the form $\sum_{t=1}^{T} \delta_{it} l_{it}\{\beta(t)\}$ with T as maximal survival time and

$$l_{it}\{\beta(t)\} \quad = \quad \delta_{it}\left[y_{it}\log\{\lambda(t,x_i)\} + (1-y_{it})\log\{1-\lambda(t,x_i)\}\right]. \tag{8}$$

The likelihood term given in (8) corresponds to the likelihood of a binomial distribution (see e.g. Fahrmeir & Tutz, 2001, chapter 9). The overall likelihood is now given by

$$\mathbf{l}(\boldsymbol{\beta}) \quad = \quad \sum_{i=1}^{n}\sum_{t=1}^{T} \delta_{it} l_{it}\{\beta(t)\} \tag{9}$$

which has to be maximized with respect to $\boldsymbol{\beta} = \{\beta(t),\ t = 1,\ldots,T\}$. Direct maximization of (9) will not necessarily yield smooth estimates, since no smoothness restrictions are imposed on $\beta(t)$. The suggestion is therefore to maximize a local version of (9) in order to achieve smoothness. For estimation of $\beta(t_0)$ at time point t_0 we introduce the kernel weights $w_{0t} = K\{(t_0 - t)/h\}$ with $K(\cdot)$ being a unimodal kernel function and h denoting the bandwidth. Incorporating the weights into the likelihood yields the local likelihood function

$$\sum_{i=1}^{n}\sum_{t=1}^{T} w_{0t}\delta_{it} l_{it}(\beta). \tag{10}$$

Maximization of (10) provides the local maximum likelihood estimate of $\beta(t_0)$. Repetition of the maximization for $t_0 = 1,\ldots,T$ finally yields a smooth estimates $\widehat{\beta}(1),\ldots,\widehat{\beta}(T)$ with the amount of smoothness corresponding to the bandwidth h.

Properties of the smooth estimate

The following derivation is basically as in Kauermann & Tutz (2000) (see also Fan, Farmen & Gijbels, 1998). The difference occurring here is due to the censor variables δ_{it} in (9) which are random numbers. Let $l_{\eta,it}$ denotes the score contribution evaluated at the true parameter. This implies $E(\delta_{it} l_{\eta,it}) = 0$ which follows from

$$E(\delta_{it} l_{\eta,it}) \;=\; P(\delta_{it}=1)E(l_{\eta,it}|\delta_{it}=1) = P(\delta_{it}=1)\{E(y_{it}|\delta_{it}=1) - \lambda_{it}(t)\} = 0$$

Moreover, $\sum_{t=1}^{T} \delta_{it} F_{it}$ serves as unbiased estimate for $E(\sum_{t=1}^{T} \delta_{it} F_{it})$. Finally one finds for mixed moments

$$\begin{aligned} E\{\sum_{t=1}^{T}\sum_{s=1}^{T} \delta_{it}\delta_{is} l_{\eta,it} l_{\eta,is}\} &= \sum_{t=1}^{T} P(\delta_{it}=1)E\{l_{\eta,it} l_{\eta,it}|\delta_{it}=1\} \\ &\quad +2\sum_{t=1}^{T}\sum_{s>t}^{T} P(\delta_{it}=1,\delta_{is}=1)E\{l_{\eta,it} l_{\eta,is}|\delta_{it}=1,\delta_{is}=1\} \quad (11) \\ &= \sum_{t=1}^{T} P(\delta_{it}=1)F_{it}, \quad (12) \end{aligned}$$

where (12) follows from (11) since for $s > t$ one has $\delta_{is} = 1 \Rightarrow \delta_{it} = 1$ and $y_{it} = 0$ such that $E(l_{\eta,it} l_{\eta,is}|\delta_{is}=1) = -\lambda_{it}(t)E(l_{\eta,is}|\delta_{is}=1) = 0$. These results secure that the randomness of the censor variable is not disturbing in the derivation of what follows. From (10) we get by expansion

$$\begin{aligned} 0 &= \sum_{i=1}^{n}\sum_{t=1}^{T} w_{0t}\delta_{it} X_i^T l_{\eta,it}\{\widehat{\beta}(t_0)\} \\ &= \sum_{i=1}^{n}\sum_{t=1}^{T} w_{0t}\delta_{it} X_i^T l_{\eta,it} - \sum_{i=1}^{n}\sum_{t=1}^{T} w_{0t}\delta_{it} X_i^T F_{it} X_i\{\widehat{\beta}(t_0) - \beta(t)\} + \ldots \\ \Leftrightarrow \{\widehat{\beta}(t_0) - \beta(t_0)\} &= \{\sum_{i=1}^{n}\sum_{t=1}^{T} w_{0t}\delta_{it} X_i^T F_{it} X_i\}^{-1} \\ &\quad \times \left[\sum_{i=1}^{n}\sum_{t=1}^{T} w_{0t}\delta_{it}[X_i^T l_{\eta,it} + X_i^T F_{it} X_i\{\beta(t) - \beta(t_0)\}]\right] + \ldots. \end{aligned}$$

Taking expectation yields in first order approximation

$$\begin{aligned} E\{\widehat{\beta}(t_0) - \beta(t_0)\} &= \sum_{i=1}^{n} E\left[\{\sum_{j\neq i}^{n}\sum_{t=1}^{T} w_{0t}\delta_{jt} X_j^T F_{jt} X_j\}^{-1} \sum_{t=1}^{T} w_{0t}\delta_{it} X_i^T F_{it} X_i\{\beta(t) - \beta(t_0)\}\right] + \ldots \\ &= \left\{\sum_{i=1}^{n}\sum_{t=1}^{T} w_{0t} P(\delta_{it}=1) X_i^T F_{it} X_i\right\}^{-1} \end{aligned}$$

$$\times \left[\sum_{t=1}^{n} w_{0t} P(\delta_{it}=1) X_i^T F_{it} X_i \{\beta(t)-\beta(t_0)\}\right] + \dots \tag{13}$$

Moreover, for the variance one finds by making use of (12)

$$\begin{aligned} Var\{(\widehat{\beta}(t_0)\} &= E\Big[\{\sum_{i=1}^{n}\sum_{t=1}^{T} w_{0t}\delta_{it} X_i^T F_{it} X_i\}^{-1} \\ &\quad \times \{\sum_{t=1}^{T} w_{0t}^2 \delta_{it}^2 X_i^T l_{\eta,it}^2 X_i\}\{\sum_{i=1}^{n}\sum_{t=1}^{T} w_{0t}\delta_{it} X_{it}^T F_{it} X_{it}\}^{-1}\Big] \\ &\approx \left\{\sum_{i=1}^{n}\sum_{t=1}^{T} w_{0t}^2 P(\delta_{it}=1)\right\} \Big/ \left\{\sum_{i=1}^{n}\sum_{t=1}^{T} w_{0t} P(\delta_{it}=1)\right\} \\ &\quad \times \left\{\sum_{i=1}^{T}\sum_{t=1}^{T} w_{0t} P(\delta_{it}=1) X_{it}^T F_{it} X_{it}\right\}^{-1} \end{aligned} \tag{14}$$

Using standard smoothing results one finds that the bias has order $O(h^2)$ while the variance has order $O(n^{-1})$. It should be noted that the order of the variance is different from what is typically met for smooth estimates (see e.g. Fan & Gijbels, 1996). The reason for this is that the asymptotic consideration is not based on the assumption that t is getting infinitely dense. Instead, in accordance with the conditions met in discrete survival, it is assumed that t is measured on a grid while the sample size n is growing.

A.2 Semiparametric Model

Estimation in the semiparametric model (3) is slightly more complicated. We make use of a combination of local and profile likelihood, as generally suggested by Severini & Wong (1992), see also Kauermann & Tutz (2001). For normal response models the resulting estimate is also known as the Speckman estimate (Speckman, 1988). The basic idea is to apply a backfitting type algorithm, where the covariates for the parametric component are locally balanced in each step. The welcome property of this approach is that estimates of the parametric components achieve the standard parametric precision, i.e. they possess $\sqrt{n}$ convergence.

Let $\eta_{\mathcal{C},i} = \sum_{k \in \mathcal{C}} x_{k,i}\beta_k$ be the predictor of the parametric component. We assume for the moment that $\eta_{\mathcal{C}}$ is fixed and known. The smooth estimate then results by maximizing the local likelihood as in (10) but now with given offset $\eta_{\mathcal{C},i}$, i.e.

$$\sum_{i=1}^{n}\sum_{t=1}^{T} w_{0t}\delta_{it} l_{it}\{X_{\mathcal{V},i}\beta_{\mathcal{V}}(t_0) + \eta_{\mathcal{C},i}\} \tag{15}$$

where $X_{\mathcal{V},i} = (x_{ki}, k \in \mathcal{V})$ is the design (row) vector built from the intercept and the covariates and $\beta_{\mathcal{V}}(t_0)$ as corresponding parameter vector. Note that we slightly changed the notation here by writing the likelihood contributions $l_{it}(\eta)$ with the predictor as argument. Even though this is unusual it will be helpful in the following for notational reasons. Maximizing (15) with respect to $\beta_{\mathcal{V}}(t_0)$ is done by solving the local score equation

$$0 = \sum_{i=1}^{n}\sum_{t=1}^{T} w_{0t} X_{\mathcal{V},i}^T \delta_{it} l_{\eta,it}\{X_{\mathcal{V},i}\widehat{\beta}_{\mathcal{V}}(t_0) + \eta_{\mathcal{C},i}\}. \tag{16}$$

The resulting estimates $\widehat{\beta}_{\mathcal{V}}(t_0)$, $t_0 = 1, \ldots, T$ are inserted in the likelihood for $\beta_{\mathcal{C}}$, yielding the profile likelihood

$$\sum_{i=1}^{n}\sum_{t=1}^{T} \delta_{it} l_{it}(X_{\mathcal{V},i}\widehat{\beta}_{\mathcal{V}}(t) + X_{\mathcal{C},i}\beta_{\mathcal{C}}). \tag{17}$$

Due to (16), the smooth estimates $\widehat{\beta}_{\mathcal{V}}(t)$ depend on the fixed parameters $\beta_{\mathcal{C}}$, which however is suppressed in the notation. Differentiation of the profile likelihood (17) has to account for this dependence and therefore yields the profile score equation

$$0 = \sum_{i=1}^{n}\sum_{t=1}^{T} \left(X_{\mathcal{C},i}^T + \frac{\partial X_{\mathcal{V},i}\widehat{\beta}_{\mathcal{V}}(t)}{\partial \beta_{\mathcal{C}}} \right) \delta_{it} l_{\eta,it}\{X_{\mathcal{V},i}\widehat{\beta}_{\mathcal{V}}(t) + X_{\mathcal{C},i}\widehat{\beta}_{\mathcal{C}}\}. \tag{18}$$

The derivative $\partial\widehat{\beta}_{\mathcal{V}}(t)/\partial\beta_{\mathcal{C}}$ can now be calculated from differentiating (16) with respect to $\beta_{\mathcal{C}}$. This provides in first order approximation

$$\frac{\partial\widehat{\beta}_{\mathcal{V}}(t_0)}{\partial\beta_{\mathcal{C}}^T} = -\left(\sum_{i=1}^{n}\sum_{t=1}^{T} w_{0t}\delta_{it} X_{\mathcal{V},t}^T F_{it} X_{\mathcal{V},t}\right)^{-1} \left(\sum_{i=1}^{n}\sum_{t=1}^{T} w_{0t}\delta_{it} X_{\mathcal{V},t}^T F_{it} X_{\mathcal{C},t}\right). \tag{19}$$

This derivative can be seen as a smoothing of the covariates $X_{\mathcal{C}}$. To see this let $X_{\mathcal{V}} = 1$, for simplicity, which yields

$$\frac{\partial \widehat{\beta}_{\mathcal{V}}(t_0)}{\partial \beta_{\mathcal{C}}^T} = -(\sum_{i=1}^{n} w_{0t}\delta_{it}F_{it}X_{\mathcal{C},i})/(\sum_{i=1}^{n}\sum_{t=1}^{T} w_{0t}\delta_{it}F_{it}).$$

It is shown in Severini & Wong (1992) for normal smoothing models and in Hunsberger (1995) for general smoothing models that estimates resulting from solving the profile score function are more efficient than estimates ignoring (19), i.e. simple backfitting estimates where iteratively the parametric and the smooth part of the model are fitted. The same result holds in this setting, since the random variables δ_{it} do not disturb the asymptotic behavior. In particular one finds

$$Var(\widehat{\beta}_{\mathcal{C}}) = \left(\sum_{i=1}^{n}\sum_{t=1}^{T} \pi it \tilde{X}_{\mathcal{C},i}^T F_{it} \tilde{X}_{\mathcal{C},i}\right)^{-1}$$

with $\tilde{X}_{\mathcal{C},i} = X_{\mathcal{C},i}^T + \partial X_{\mathcal{V},i}\widehat{\beta}_{\mathcal{V}}(t)/(\partial\beta_{\mathcal{C}})$, while the smooth estimates fulfill (13) and (14) but with X_i replaced by $X_{\mathcal{V},i}$ and F_{it} evaluated at $X_{\mathcal{V},i}\beta_{\mathcal{V}}(t,i) + X_{\mathcal{C},i}\beta_{\mathcal{C}}$.

References

Brüderl, J. (1995). Survival and growth of newly founded firms. Technical Report, Institute of Sociology, Ludwig-Maximilians-Universität München. Preliminary Version.

Brüderl, J., Preisendörfer, P., and Ziegler, R. (1992). Survival chances of newly founded business organizations. *American Sociological Review* **57**, 227–242.

Fahrmeir, L. and Tutz, G. (2001). *Multivariate Statistical Modelling Based on Generalized Linear Models* (2 ed.). New York: Springer Verlag.

Fan, J., Farmen, M., and Gijbels, I. (1998). Local maximum likelihood estimation and inference. *Journal of the Royal Statistical Society, Series B* **B 60**, 591–608.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications.* London: Chapman & Hall.

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

Hunsberger, S. (1995). Semiparametric regression in likelihood based models. *J. Amer. Statist. Assoc.* **89**, 1354–1365.

Kauermann, G. and Tutz, G. (2000). Local likelihood estimation in varying-coefficient models including additive bias correction. *Journal of Nonparametric Statistics* **12**, 343–371.

Kauermann, G. and Tutz, G. (2001). Testing generalized linear and semiparametric models against smooth alternatives. *Journal of the Royal Statistical Society, Series B* **63**, 147 – 166.

Severini, T. A. and Wong, W. H. (1992). Profile likelihood and conditionally parametric models. *Ann. Statist.* **20**, 1768–1802.

Speckman, P. (1988). Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society, Series B* **50**, 413–436.

Aspect	Variable	Coding
legal basis	*legal*1	= 1 for small business
	*legal*2	= 1 for partnerships
financing	*out.capital*	= 1 for outside capital > 0
	in.capital	= 1 for company capital > 0
employees	*employ*1	= 1 if $\geq$ 2 employees
	*employ*2	= 1 if $\geq$ 3 employees
preparation	*plan*	= 1 if business was planed > 1/2 year
	lead	= 1 if business had a lead of > 1/2 year
target market	*product*	= 1 if offered product is already on the market
		= 0 for new development
	price	= 1 if target market has high price level
	national	= 1 if target market is national
	customer	= 1 if target customers are not special
		= 0 for special customers
owner	*gender*	= 1 for male
		= 0 for female
	training	= 1 for master degree (craftsman or university)
	school	= 1 for high-school degree
	branch	= 1 for expert of branch and area of business
	*age*1	= 1 for junior owner, i.e. age $\leq$ 30 years
	*age*2	= 1 for senior owner, i.e. age > 40 years

Table 1: Description and coding of variables

Number	Varying Coefficients	log-likelihood	degree of model	AIC-criterion
1	all	-1652	66.9	3437
2	*baseline, legal1+2 , out.capital, in.capital plan, lead, price, national, customer, gender, branch, age1+2*	-1658	54.3	3425
3	*baseline, out.capital, plan, price, national, gender, branch, age1+2*	-1665	43.1	3416
4	*baseline, out.capital, plan, price, age1*	-1676	31.2	3414
5	*baseline*	-1692	21.5	3428
4*	*baseline, out.capital, plan, price, age1*	-1680	24.2	3410

Table 2: Likelihood and Akaike criterion for various models.

	model 4		model 4*	
Effect	Estimate	Stud. Value	Estimate	Stud. Value
legal1	1.11	5.40	1.13	6.17
legal2	0.98	5.02	0.98	5.51
in.capital	-0.11	-1.42		
employ1	-0.14	-1.09		
employ2	-0.51	-2.88	-0.57	-2.50
lead	-0.89	-2.72	-0.78	-3.72
product	0.20	1.58		
national	-0.27	-2.40	-0.30	-2.74
customer	0.25	2.26	0.27	2.50
gender	0.07	0.55		
age2	-0.06	-0.44		
school	-0.17	-1.36		
training	0.01	0.06		
branch	-0.53	-4.91	-0.47	-4.53

Table 3: Parameter estimates, standard deviation and studentized value for model 4 and model 4^*

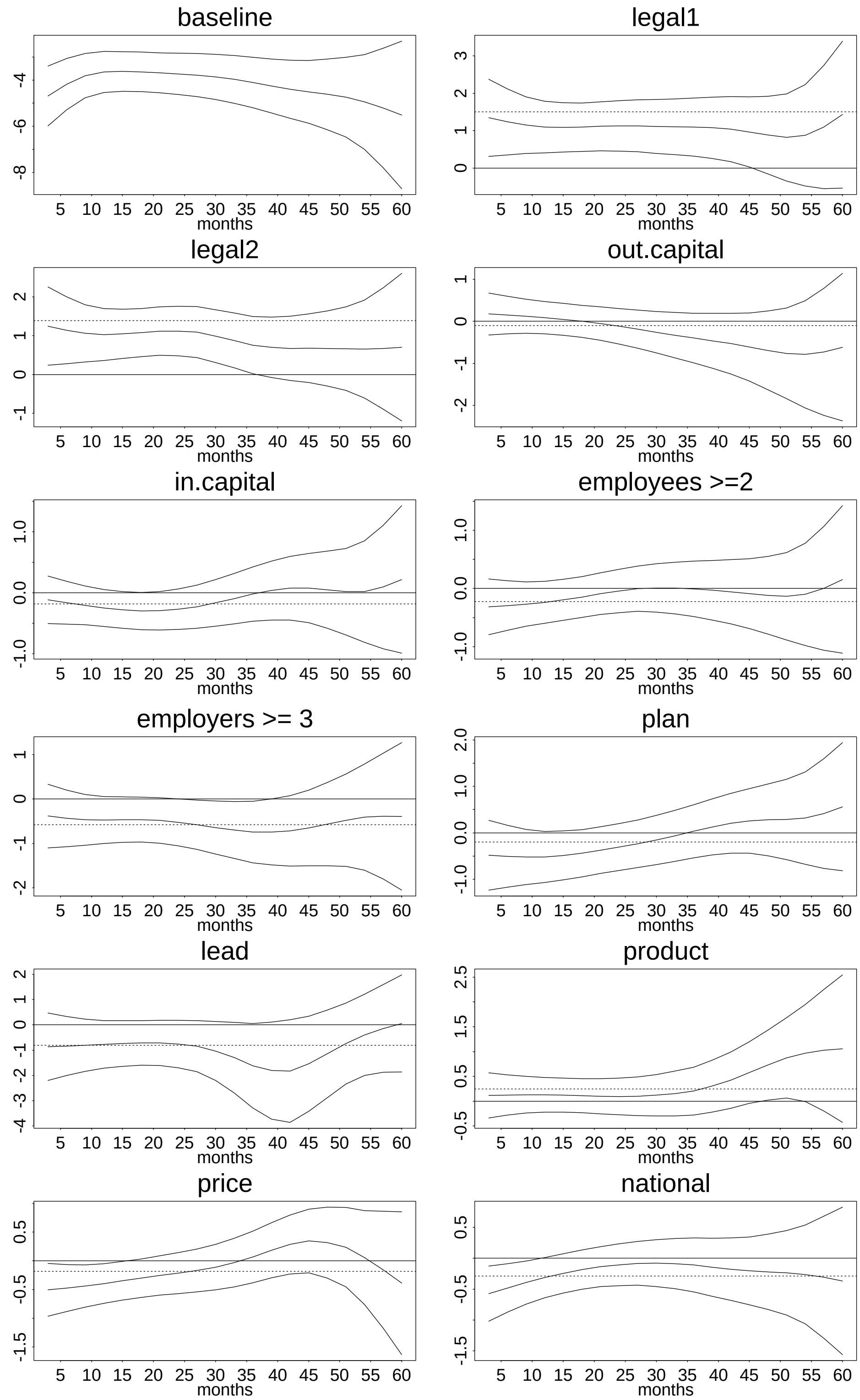

Figure 1: Varying coefficients for model 1 (first part). Solid horizontal line shows zero effect as reference, dashed horizontal line gives the estimated time-constant effect.

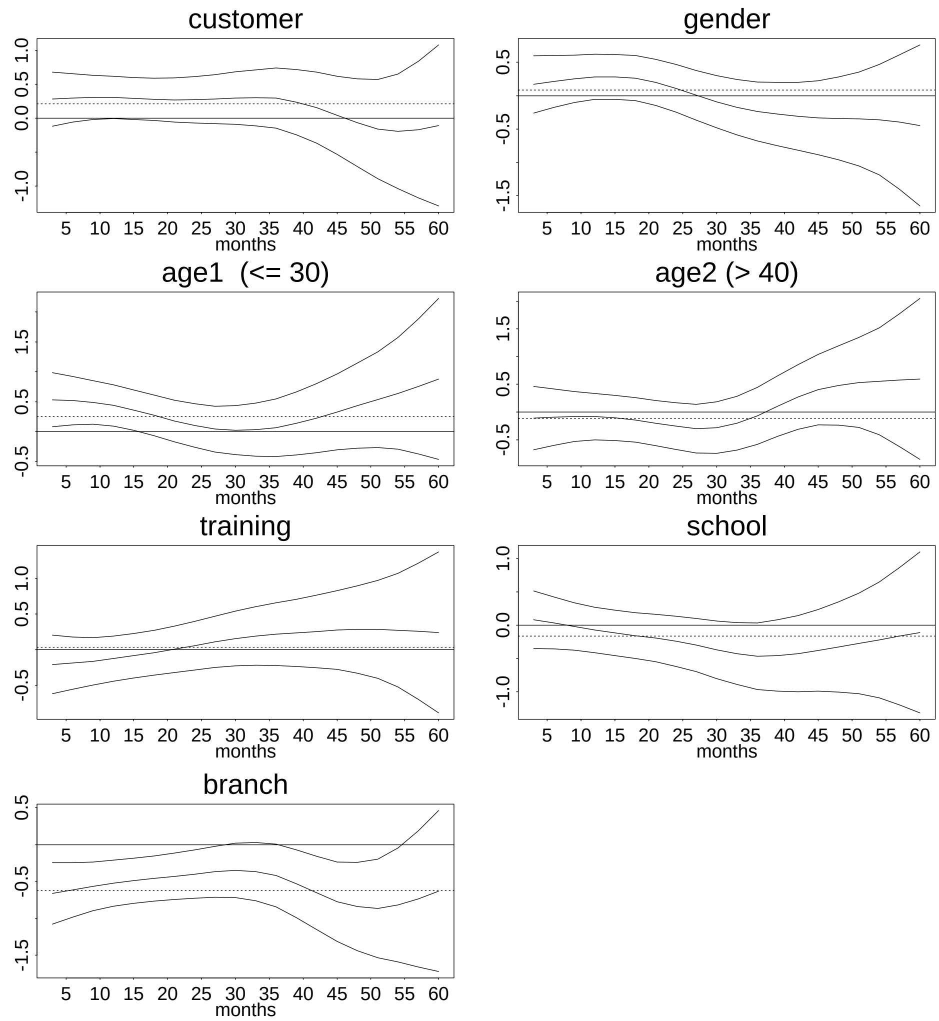

Figure 2: Varying coefficients for model 1 (second part). Solid horizontal line shows zero effect as reference, dashed horizontal line gives the estimated time-constant effect.

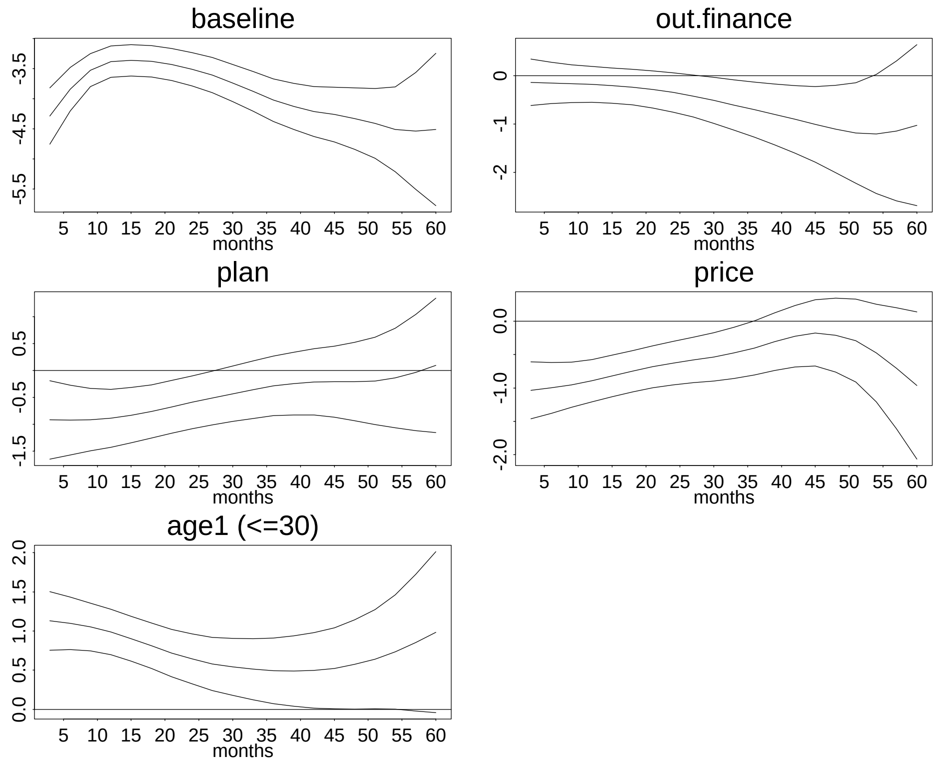

Figure 3: Varying Coefficients for Model 3