ECONSTOR
*Make Your Publications Visible.*

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Quiroz, Matias; Villani, Mattias

**Working Paper**

# Dynamic mixture-of-experts models for longitudinal and discrete-time survival data

Sveriges Riksbank Working Paper Series, No. 268

**Provided in Cooperation with:**
Central Bank of Sweden, Stockholm

*Suggested Citation:* Quiroz, Matias; Villani, Mattias (2013) : Dynamic mixture-of-experts models for longitudinal and discrete-time survival data, Sveriges Riksbank Working Paper Series, No. 268, Sveriges Riksbank, Stockholm

This Version is available at:
https://hdl.handle.net/10419/81873

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

SVERIGES RIKSBANK
WORKING PAPER SERIES 268

# DYNAMIC MIXTURE-OF-EXPERTS MODELS FOR LONGITUDINAL AND DISCRETE-TIME SURVIVAL DATA

*Matias Quiroz and Mattias Villani*

May 2013

WORKING PAPERS ARE OBTAINABLE FROM

Sveriges Riksbank • Information Riksbank • SE-103 37 Stockholm
Fax international: +46 8 787 05 26
Telephone international: +46 8 787 01 00
E-mail: info@riksbank.se

The Working Paper series presents reports on matters in
the sphere of activities of the Riksbank that are considered
to be of interest to a wider public.
The papers are to be regarded as reports on ongoing studies
and the authors will be pleased to receive comments.

The views expressed in Working Papers are solely
the responsibility of the authors and should not to be interpreted as
reflecting the views of the Executive Board of Sveriges Riksbank.

# DYNAMIC MIXTURE-OF-EXPERTS MODELS FOR LONGITUDINAL AND DISCRETE-TIME SURVIVAL DATA

MATIAS QUIROZ AND MATTIAS VILLANI

SVERIGES RIKSBANK WORKING PAPER SERIES

NO. 268

MAY 2013

ABSTRACT. We propose a general class of flexible models for longitudinal data with special emphasis on discrete-time survival data. The model is a finite mixture model where the subjects are allowed to move between components through time. The time-varying probabilities of component memberships is modeled as a function of subject-specific time-varying covariates. This allows for interesting within-subject dynamics and manageable computations even with a large number of subjects. Each parameter in the component densities and in the mixing function is connected to its own set of covariates through a link function. The models are estimated using a Bayesian approach via a highly efficient Markov Chain Monte Carlo (MCMC) algorithm with tailored proposals and variable selection in all set of covariates. The focus of the paper is on models for discrete-time survival data with an application to bankruptcy prediction for Swedish firms, using both exponential and Weibull mixture components. The dynamic mixture-of-experts models are shown to have an interesting interpretation and to dramatically improve the out-of-sample predictive density forecasts compared to models with time-invariant mixture probabilities.

KEYWORDS: Bayesian inference, Markov Chain Monte Carlo, Bayesian variable selection, Survival Analysis, Mixture-of-experts.

JEL Classification: C11, C41, D21, G33

Quiroz: *Research Division, Sveriges Riksbank, SE-103 37 Stockholm, Sweden* and *Department of Statistics, Stockholm University. E-mail: quiroz.matias@gmail.com.* Villani: *Division of Statistics, Department of Computer and Information Science of Statistics, Linkoping University.* E-mail: mattias.villani@liu.se. The views expressed in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Executive Board of Sveriges Riksbank.

## 1. Introduction

We propose a mixture model for flexible modeling of longitudinal data. Our model belongs to the mixture-of-expert type of models first proposed by Jacobs et al. (1991) and Jordan and Jacobs (1994). In particular, we extend the class of *Generalized Smooth Mixture* (GSM) models presented in Villani et al. (2009) and Villani et al. (2012) to a longitudinal data setting. Villani et al. (2012) generalizes the *Smoothly Mixing Regression* (SMR) model in Geweke and Keane (2007). The key features of our approach are: i) subjects are allowed to move between mixture components over time, ii) the within-subject dynamics is modeled by letting the component membership probabilities be functions of subject-specific time-varying covariates, and iii) an efficient Bayesian inference methodology using MCMC with variable selection.

Our methodology applies to essentially any mixture components, but most of the article will focus on mixture components for discrete-time survival (duration) data, in particular the exponential and Weibull components used in our application to firm bankruptcy. For a general introduction to survival models, see e.g. Miller et al. (1981). Ibrahim et al. (2005) provides an introduction and overview of Bayesian modeling in the field.

Finite mixtures are useful for modeling unobserved heterogeneity in many fields, see the examples in the referenced articles above, and Frühwirth-Schnatter (2006) for a general introduction to finite mixture models. Mixture models are also used in model based clustering. Given the longitudinal dimension of our data this paper is closely related to clustering panel data in form of relatively short time series. A recent survey on model based clustering of time series is given in Frühwirth-Schnatter (2011). A main difference between this literature and our approach is that we allow for the possibility of subjects to move across clusters over time.

Our main focus in this paper is on using dynamic mixture-of-experts models for analyzing survival data. One of the first regression models for survival data is the *Proportional Hazards*, or *Cox regression model* introduced in Cox (1972). The restrictiveness of the proportionality assumption and the inability to capture unobserved heterogeneity has lead researches to

develop more flexible models. Frailty models, which is a variant of a multiplicative random effects model, are much used in the literature. Introducing an unobserved random variable (the frailty) which acts multiplicative on the hazard can give a heterogeneous effect that vary across individuals or groups. A common practice is to assume a parametric distribution for the continuous frailty. Early references are Lancaster (1979) and Vaupel et al. (1979). Mosler (2003) surveys the theory and applications of these models in Econometrics. Recent work by Huynh and Voia (2009) assumes a finite mixture for the frailty distribution to capture a wide variety of functional shapes. Alternatively, finite mixture models offers a rich model class where some of the restrictive assumptions in the traditional survival models can be relaxed. McLachlan et al. (1994) provides a survey on the role of finite mixture models in survival analysis. Finite mixture of survival models are closely related to frailty models which is most easily seen when the distribution of frailty is discrete and finite. The intuitive interpretation of finite mixtures combined with the capability of modeling frailties makes it an interesting framework for analyzing complex data structures in survival analysis.

In most economics and social sciences applications, time is measured discretely (Allison, 1982). Examples include labor economics when studying the duration of individual unemployment measured e.g. in weeks (Carling et al., 1996), or educational research where the data is often recorded in school years (Singer and Willett, 1993). In our application we model time to bankruptcy (in years) for nascent firms. There are several advantages with the discrete time framework. One is that time dependent covariates easily can be incorporated. Another is that the proportionality assumption in Cox regression models can be relaxed by allowing the effect of a predictor to be different across time periods. Singer and Willett (1993) provides more arguments why one should consider discrete time survival models. Heterogeneity has not been explored as much in the discrete time framework. Notable exceptions are the continuous frailties in Xue and Brookmeyer (1997) and the finite mixture approach in Muthén and Masyn (2005).

Our article extends Muthén and Masyn (2005) in the following directions. First, we allow subjects to be classified to potentially different mixture components at each time

period (dynamic mixture) while Muthén and Masyn (2005) restrict each subject to belong to one and only one mixture component during its exposure time (static mixture). Second, we use the Bayesian paradigm and Markov Chain Monte Carlo (MCMC) to estimate the model. This allows us to use Bayesian variable selection to obtain model parsimony and give insights on importance of covariates in different parts of the model. Our approach can also be straightforwardly extended to include the general latent variable (factor analysis) part in Muthén and Masyn (2005).

This paper is organized as follows. Section 2 presents the longitudinal mixture-of-experts models in a general setting. In Section 3 we review necessary concepts of survival analysis in continuous and discrete time and introduce two models for the discrete survival time. The structure of the data and likelihood is particularly highlighted. Section 4 applies the framework in Section 2 to discrete-time survival models. Section 5 discusses the inference methodology. Priors are introduced in all parts of the model and the general MCMC algorithm with variable selection is presented. Section 6 illustrates the methodology by modeling the bankruptcy risk for nascent Swedish firms. Section 7 discusses future research and concludes.

## 2. MIXTURE-OF-EXPERTS MODELS FOR LONGITUDINAL DATA

In the standard non-longitudinal framework a smooth finite mixture density with $K$ components can be formulated as

$$p(y_i|x_i,\beta,\gamma) = \sum_{k=1}^{K} w_k(z_i|\gamma_k)p_k(y_i|x_i,\beta_k), \quad i=1,...,n, \tag{2.1}$$

where $w_k(z|\gamma_k)$ denotes the mixing probability and can be interpreted as the prior probability of belonging to the $k$th component density. Both $x$ and $z$ are vector of covariates which may be overlapping. When $w_k(z|\gamma_k)$ is the multinomial logit with $\gamma_1 = 0$ this is the Generalized Smooth Mixture (GSM) model in Villani et al. (2012). To simplify inference with the Gibbs sampler, augmented data $s_1, s_2, \ldots s_n$ is introduced so that $s_i = k$ means that the $i$th observation belongs to the $k$th component. The model in Equation (2.1) can

then be formulated as

$$\begin{aligned} y_i|(s_i = k, x_i, \beta_k) &\sim p_k(y_i|x_i, \beta_k) \\ P(s_i = k|z_i, \gamma_k) &= w_k(z|\gamma_k). \end{aligned}$$

To extend to a longitudinal mixture the following notation is introduced. Assume subject $i$ has been observed over $n_i$ time periods. Let $y_i = (y_{i1}, \ldots, y_{in_i})^T \in \mathbb{R}^{n_i \times 1}$, $x_i = (x_{i1}, \ldots, x_{in_i})^T \in \mathbb{R}^{n_i \times p_x}$ and $z_i = (z_{i1}, \ldots, z_{in_i})^T \in \mathbb{R}^{n_i \times p_z}$. Let $v_i \in \mathbb{R}^{p_v \times 1}$ denote the time-invariant predictors and $s_i \in \{1, \ldots, K\}^{n_i}$ where $s_{ij} = k$ if the subject belongs to component $k$ at time period $j$. The longitudinal dimension allows for two main specifications of $s$: $s_{ij} = k$ for all $j$ or $s_{ij} = k_j$ where $k_j \in \{1, 2 \ldots K\}$. We refer to the former as a *static mixture* and the latter as a *dynamic mixture*.

*Static mixture.* The static mixture model is a finite mixture on the joint distribution of $y_i$, i.e.

$$\begin{aligned} p(y_{i1}, y_{i2}, ..., y_{in_i}) &= \sum_{k=1}^{K} w_k p_k(y_{i1}, y_{i2}, ..., y_{in_i}) \\ &= \sum_{k=1}^{K} w_k p_k(y_{i1}) p_k(y_{i2}|y_{i1}) \cdots p_k(y_{in_i}|y_{i1}, y_{i2}, ..., y_{i(n_i-1)}) \end{aligned} \tag{2.2}$$

and the dependence on covariates and parameters is suppressed everywhere to save space. The covariates $x_{ij}$ and $v_i$ both enter in the component models, while the mixing function is only a function of $v_i$;it is clearly not possible to have time-varying covariates in the mixing function in a static mixture. The mixing probabilities are modeled with the multinomial logit

$$w_k(v_i) = \frac{\exp(v_i^T \gamma_k)}{\sum_{l=1}^{K} \exp(v_i^T \gamma_l)} \tag{2.3}$$

where $\gamma_k \in \mathbb{R}^{p_v \times 1}$ with $\gamma_1 = 0$ for identification. The model in Equation (2.2) expresses the *joint* distribution. It is straightforward to show that the density at period $t$ conditional on

previous values is given by

$$p(y_{it}|y_{i1},\ldots,y_{i(t-1)}) = \frac{\sum_{k=1}^{K} w_k p_k(y_{i1})\ldots p_k(y_{i(t-1)}|y_{i1},\ldots,y_{i(t-2)})p_k(y_{it}|y_{i1},\ldots,y_{i(t-1)})}{\sum_{k=1}^{K} w_k p_k(y_{i1})\ldots p_k(y_{i(t-1)}|y_{i1},\ldots,y_{i(t-2)})}. \tag{2.4}$$

In contrast to the *dynamic mixture* this model lacks the nice interpretation that the conditional distribution at any given time period is a mixture of conditional distributions. The latent variable formulation of the model in Equation (2.2) is

$$\begin{aligned} y_i|s_i=k,x_i &\sim p_k(y_{i1})\ldots \quad p_k(y_{i(n_i-1)}|y_{i1},\ldots,y_{i(n_i-2)})p_k(y_{in_i}|y_{i1},\ldots,y_{i(n_i-1)}) \\ P(s_i=k|v_i) &= \frac{\exp(v_i^T\gamma_k)}{\sum_{l=1}^{K}\exp(v_i^T\gamma_l)}. \end{aligned} \tag{2.5}$$

The assumption of this model is that given the component membership, responses are only associated with the predictors in $x_i$.

*Dynamic mixture.* Restricting a subject to a single component over time may not be realistic in some situations because individual behavior may not be homogeneous over time. To exemplify, consider the modeling of firm bankruptcy. If the components can be interpreted as high versus low risk for bankruptcy, the assumption of being constantly a risky or a safe firm is unrealistic. The economic surrounding and individual financial variables do change over time which is likely to make the firm more or less risky. It is therefore sometimes useful to let subjects be classified at each time period.

The obvious approach to a dynamic mixture is to let $s_i = \{s_{i1},\ldots,s_{in_i}\}$ follow a (hidden) Markov chain, see Baum and Petrie (1966) and Kim and Nelson (2003). The posterior sampling of $s_i$ is then performed sequentially from the conditional distribution at each time point using e.g. Sequential Monte Carlo (SMC) (Doucet et al., 2000). Such an approach is computationally infeasible in many longitudinal applications since the SMC would have to be performed for each of the subjects, which is clearly not an option in our application to firm bankruptcy. An alternative approach is to sample directly from the joint distribution for

each $s_i$ sequence (Franzén, 2008), but the sample space of $s_i$ grows dramatically with $K$ and the number of time periods so $P(s_i|y_i, x_i, z_i)$ quickly becomes computationally intractable.

To overcome these problems we suggest the following approach. Let $\{s_{i1}, \dots s_{in_i}\}$ be an independent sequence *conditional* on the path of time-varying covariates $z_i$, i.e.

$$P(s_{i1} = k_1, \dots s_{in_i} = k_{n_i}|z_{i1}, \dots z_{in_i}) = P(s_{i1} = k_1|z_{i1}) \dots P(s_{in_i} = k_{n_i}|z_{in_i}) \tag{2.6}$$

for $1 \leq k_i \leq K$. The temporal dependence of $s_i$ is thus induced by the covariate path itself. The strength of this approach is that given the covariates (and other model parameters) the component allocations can be sampled independently for all observations and time periods in the Gibbs sampler, see Section 5.2.

The dynamic mixture model is formulated as

$$p(y_{i1}, \dots, y_{i(n_i-1)}, y_{in_i}) = p(y_{i1}) \dots p(y_{i(n_i-1)}|y_{i1}, \dots, y_{i(n_i-2)}) p(y_{in_i}|y_{i1}, \dots, y_{i(n_i-1)}),$$

where

$$p(y_{it}|y_{i1}, \dots, y_{i(t-1)}) = \sum_{k=1}^{K} w_{it}^k p_k(y_{it}|y_{i1}, \dots, y_{i(t-1)}),\ t = 1, ..., n_i \tag{2.7}$$

and

$$w_{ij}^k = \frac{\exp(z_{ij}^T \gamma_k)}{\sum_{l=1}^{K} \exp(z_{ij}^T \gamma_l)}, \tag{2.8}$$

$\gamma_k \in \mathbb{R}^{p_z \times 1}$ with $\gamma_1 = 0$ for identification. Time invariant predictors $v_i$ might also be included in $z_i$. Persistence in component allocations over time can be achieved by defining $z_{ij}$ as an exponential moving average of the time dependent covariates $x_{ij}$

$$z_{it} = \alpha x_{it} + (1 - \alpha) z_{i(t-1)},$$

and $z_{i1} = x_{i1}$, where $0 \leq \alpha \leq 1$, and $\alpha = 1$ corresponds to no smoothing. Persistence prevents a sudden change in the explanatory variables to trigger an immediate reallocation

of the subject; a sudden decrease in a firm's profits may not immediately make it a high risk firm, but several consecutive years of losses might.

The latent variable formulation of Equation (2.7) is

$$\begin{aligned} y_{ij}|s_{ij}=k,x_{ij} &\sim p_k(y_{ij}|y_{i1},\ldots,y_{i(j-1)}) \\ P(s_{ij}=k|z_{ij}) &= \frac{\exp(z_{ij}^T\gamma_k)}{\sum_{l=1}^K \exp(z_{ij}^T\gamma_l)}. \end{aligned} \tag{2.9}$$

The dynamic mixture has the interpretation that the conditional density at a given time period is a mixture of densities, as can be seen from Equation (2.7). The assumption of this model is that given the component allocation at each time-period, responses are only associated with the $x$ predictors at the given time-point.

## 3. Survival analysis

This section gives a brief review of concepts in continuous time survival analysis and some background on the extension to discrete time models emphasizing the construction of the likelihood.

### 3.1. **Continuous time framework.**

Let the random variable $T^c$ denote the time to some unrepeatable event. Assume for the moment that $T^c$ is continuous with sample space $\{t : t \geq 0\}$ and has a distribution $f(t|\lambda)$ parametrized by some parameter vector $\lambda$. Subject $i$ is right-censored if the event has not been observed before the *censoring time* $T^*$, and no further information about subject $i$ is available after $T^*$. $T^*$ is typically the end of the study, but some subject may leave the study early so $T^*$ may vary across subject. Introducing $c_i = 1$ if the $i$th subject is censored at time $T_i^*$ and $c_i = 0$ otherwise, the likelihood decomposes as

$$L(\lambda) = \prod_{i=1}^n f(t_i|\lambda)^{1-c_i} S(T_i^*|\lambda)^{c_i}. \tag{3.1}$$

where $S(t|\lambda) = P(T^c > t|\lambda)$ is the survival function. A common way of representing the distribution of $T^c$ is through the hazard function

$$\begin{aligned} h(t|\lambda) &= \lim_{\Delta t \to 0} \frac{P(t \leq T^c < t + \Delta t | T^c \geq t, \lambda)}{\Delta t}, \\ &= \frac{f(t|\lambda)}{S(t|\lambda)}. \end{aligned} \tag{3.2}$$

which is the instantaneous rate of experiencing the event given that it has not been experienced yet. The survival function relates to the hazard function through

$$S(t|\lambda) = \exp(-\int_0^t h(u|\lambda)\mathrm{d}u). \tag{3.3}$$

The extension to regression is made by including dependence of the distribution of $T^c$ on covariates. As an example the Proportional Hazards model Cox (1972) is obtained by $h(t|\lambda, x) = h_0(t)\exp(x^T\beta)$ where $x$ is the covariate vector, $\beta$ the regression parameters, and $h_0(t)$ is the baseline hazard.

3.2. **Discrete time framework.** Survival data are often observed in discrete time, for example monthly or yearly, see e.g. Allison (1982) and Singer and Willett (1993). Assume that a study is observed over $J$ periods which can be divided as $(0, t_1], (t_1, t_2], \ldots (t_{J-1}, t_J]$. Let $T^c$ denote the continuous random variable introduced in Section 3.1. Let $T \in \{1, 2, ...\}$ be the discrete random variable recording the time period where the event occurs, i.e. $T = j$ if $T^c \in (t_{j-1}, t_j]$. It is convenient to express the joint likelihood of the data in terms of the hazard, which in discrete time is a probability $h_j = P(T = j|T \geq j)$. It will later be useful to express the hazard in terms of the survival function $S(a) = P(T > a)$

$$h_j = \frac{P(T > j-1) - P(T > j)}{P(T > j-1)} = 1 - \frac{S(j)}{S(j-1)}, \tag{3.4}$$

where $S(a)$ may be computed by Equation (3.3) with $h(t|\lambda)$ being interpreted as the hazard *rate* of the continuous random variable underlying the discrete time variable $T$. Furthermore

the likelihood terms are expressed in form of hazards as

$$P(T = j) = \left( \prod_{k=1}^{j-1} (1 - h_k) \right) h_j, \tag{3.5}$$

if the subject experienced the event in time period $j$ and

$$P(T > j) = \prod_{k=1}^{j} (1 - h_k) \tag{3.6}$$

if the subject left the sample after time period $j$ without experiencing the event. Let the $i$th subjects' hazard probability at period $j$ be denoted $h_{ij}(x_{ij})$. Assuming $n$ independent subjects, the likelihood is expressed as

$$L = \prod_{i=1}^{n} \prod_{j=1}^{n_i} h(x_{ij})^{y_{ij}} (1 - h(x_{ij}))^{1-y_{ij}}, \tag{3.7}$$

where

$$y_{ij} = \begin{cases} 0 \text{ if subject i does not experience the event at period j,} \\ 1 \text{ if subject i does experience the event at period j.} \end{cases}$$

Singer and Willett (1993) and Shumway (2001) note that this likelihood has the same form as regression for binary data with $h^{-1}$ as the link function. Note that the recording of data is done through the binary representation of the response - if subject $i$ has $n_i$ periods then the observation is recorded as a sequence $\{y_{ij}, x_{ij}\}_{j=1}^{n_i}$. Whenever the subject is censored $y$ consist of only zeros, and if the event is experienced the sequence is terminated by 1 at the time period where the event took place.

3.3. **Survival models in discrete time.** In this paper two different models are considered; a single parameter and a two-parameter model. The first, the *exponential* model, is derived by assuming that $T^c \in Exp(\lambda)$ which gives $h(t|\lambda) = \lambda$. Since $\lambda > 0$ the link $g(\lambda) = \log(\lambda)$ is suitable. Then

$$h(x_{ij}) = 1 - \exp(-\exp(\alpha + x_{ij}^T \beta)(t_{ij} - t_{i(j-1)})).$$

The second, the *Weibull* model, is derived by assuming the Weibull distribution for $T^c$, parametrized by $f(t|\lambda,\rho) = \rho\lambda t^{\rho-1}\exp(-\lambda t^\rho)$ which implies $S(t|\lambda,\rho) = \exp(-\lambda t^\rho)$ and therefore

$$h(\lambda_{ij},\rho_{ij}) \;=\; 1-\exp(-\lambda_{ij}(t_{ij}^{\rho_{ij}} - t_{i(j-1)}^{\rho_{ij}})).$$

Because both $\lambda$ and $\rho$ are positive the dependence on the covariates are modeled through

$$\begin{aligned}\log(\lambda_{ij}) &= \alpha_\lambda + x_{\lambda_{ij}}^T\beta_\lambda \\ \log(\rho_{ij}) &= \alpha_\rho + x_{\rho_{ij}}^T\beta_\rho.\end{aligned}$$

Note that the hazard probability of the Weibull model depends on the level of $t$, while it only depends on the difference in $t$ in the exponential model. Both these models can easily be extended to allow for an effect $\eta_j$ to model flexible hazards, see our application in Section 6.

## 4. Mixture-of-experts model for survival data

This section presents the survival models that will be used as the components in the finite mixture. The dynamic and static mixture of survival models are presented using results in Section 2.

### 4.1. **Component models.**

We characterize the distribution by the hazard probability. The hazard probability will depend on a set of model parameters $\phi_1,\ldots,\phi_L$. As in Villani et al. (2012) each parameter depends on a set of predictors through link functions $g_l(\phi_l) = x_l^T\beta_l$ The expression of the likelihood of a given component is

$$L(\beta_1,...,\beta_L) \;=\; \prod_{i=1}^{n}\prod_{j=1}^{n_i} h(x_{ij}|\phi_1,\ldots\phi_L)^{y_{ij}}(1-h(x_{ij}|\phi_1,\ldots\phi_L))^{1-y_{ij}} \tag{4.1}$$

where $\phi_l = g_l^{-1}(x_l^T\beta_l)$.

4.2. **Smooth mixtures of survival models.** Section 2 presents the dynamic and static mixture in a general setting. Here it is restricted to discrete time survival data which is recorded as the binary vector $y_i = \{0,1\}^{n_i}$ for the $i$th subject.

*Static mixture.* The general expression for this model is given in Equation (2.2). This is the latent class model considered in Muthén and Masyn (2005), but without the general latent variable part and not restricted to the logit hazard model. The interpretation is that the mixture is on the joint distribution of $y_i$, i.e.

$$\begin{aligned} p(y_{i1}=0,\ldots,y_{i(n_i-1)}=0,y_{in_i}=c_i) &= \sum_{k=1}^{K} w_k p_k(y_{i1}=0,\ldots,y_{i(n_i-1)}=0,y_{in_i}=c_i) \\ &= \sum_{k=1}^{K} w_k \left(\prod_{j=1}^{n_i-c_i}(1-h_{ij}^k)\right)(h_{in_i}^k)^{c_i} \end{aligned} \tag{4.2}$$

where $c_i \in \{0,1\}$ is the censor indicator and the dependence on covariates and parameters is suppressed everywhere to save space. In the component model $h_{ij}^k = h^k(x_{ij}, v_i)$ while the mixing function $w_k = w_k(v_i)$. The mixing probabilities are modeled with the multinomial logit as in Equation (2.3).

The hazard probability at period $t$ is the equivalent of the conditional density in Equation (2.4), i.e.

$$p(y_{it}=1|y_{i(t-1)}=0) = \frac{\sum_{k=1}^{K} w_k \left(\prod_{j=1}^{t-1}(1-h_{ij}^k)\right) h_{it}^k}{\sum_{k=1}^{K} w_k \prod_{j=1}^{t-1}(1-h_{ij}^k)}. \tag{4.3}$$

Note that the overall hazard at any given time period is not a mixture of hazard probabilities.

The marginal effect of a covariate $x_t$ (which does not enter in the mixing function) is easily computed as

$$\frac{d}{dx_t}h_t(x_t) = \frac{\sum_{k=1}^{K} w_k \prod_{j=1}^{t-1}(1-h_{ij}^k)\left(\frac{d}{dx_t}h_t^k(x_t)\right)}{\sum_{k=1}^{K} w_k \prod_{j=1}^{t-1}(1-h_{ij}^k)}.$$

The latent variable formulation of the model in Equation (4.2) is

$$\begin{aligned} y_i|s_i = k, \{x_{ij}\}_{j=1}^{n_i} &\sim \left(\prod_{j=1}^{n_i - c_i} (1 - h_{ij}^k(x_{ij}))\right) (h_{in_i}^k(x_{ij}))^{c_i} \\ P(s_i = k|v_i) &= \frac{\exp(v_i^T \gamma_k)}{\sum_{l=1}^K \exp(v_i^T \gamma_l)}. \end{aligned} \tag{4.4}$$

*Dynamic mixture.* The general dynamic mixture model in Equation (2.7) can be formulated in terms of hazards as

$$\begin{aligned} p(y_{i1} = 0, \ldots, y_{in_i} = c_i) &= p(y_{i1} = 0) \ldots p(y_{in_i} = c_i | y_{i(n_i - 1)} = 0) \\ &= \left(\sum_{k=1}^K w_{i1}^k (1 - h_{i1}^k)\right) \ldots \left(\sum_{k=1}^K w_{in_i}^k (h_{in_i}^k)^{c_i} (1 - h_{in_i}^k)^{1 - c_i}\right) \end{aligned} \tag{4.5}$$

where $h_{ij}^k = h^k(x_{ij}, v_i)$ and $w_{ij}^k$ follows the multinomial model in Equation (2.8). The dynamic mixture thus has the interpretation that the hazard at a given time period is a mixture of hazards.

The marginal effect of covariate $x_t$, on the hazard when $z_t = \alpha x_t + (1 - \alpha) f(x_{t-1}, \ldots x_1)$, is given by

$$\frac{d}{dx_t} h_t(x_t) = \sum_{k=1}^K \alpha \frac{d}{dz_t} \left(w_k(z_t)\right) h_t^k(x_t) + w_k(z_t) \frac{d}{dx_t} h_t^k(x_t),$$

where the derivative of the multinomial logit is

$$\frac{d}{dz_t} \left(w_k(z_t)\right) = w_k(z_t) \left[\gamma_k - \sum_{l=1}^K w_k(z_t) \gamma_l\right].$$

The latent variable formulation of Equation (4.5) is

$$\begin{aligned} y_{ij}|s_{ij} = k, x_{ij} &\sim \begin{cases} 1 - h_{ij}^k(x_{ij}), & y_{ij} = 0 \\ h_{ij}^k(x_{ij}), & y_{ij} = 1 \end{cases} \\ P(s_{ij} = k|z_{ij}) &= \frac{\exp(z_{ij}^T \gamma_k)}{\sum_{l=1}^K \exp(z_{ij}^T \gamma_l)}. \end{aligned} \tag{4.6}$$

## 5. Inference

We adopt a Bayesian approach to inference and use a Metropolis-within-Gibbs sampler with variable selection to sample from the posterior distribution. The sampler utilizes the gradient and Hessian of the full conditional posterior to construct tailored proposals.

This section is organized as follows. First, prior distributions are introduced in all parts of the model. These priors are simple and the user only needs to specify prior beliefs about scalar parameters. Then the general MCMC scheme is illustrated, followed by a section describing the algorithm that construct tailored proposals for efficient inference. Finally, the method for choosing the number of components is explained.

### 5.1. **Prior Elicitation.**

5.1.1. *Components.* We use the prior construction initially developed in Ntzoufras et al. (2003) for the Generalized Linear Model (GLM) and subsequently refined and extended in Villani et al. (2012) to GSM models. Assume a component model with a single model parameter $\lambda$ and a link function $g$ such that $g(\lambda) = \alpha_\lambda + x^T\beta_\lambda$. We first discuss the prior on the intercept. Start by standardizing the covariates to have mean zero and unit standard deviation. The intercept $\alpha_\lambda$ is then $g(\lambda)$ at the mean of the original covariates. Assume that $\alpha_\lambda \sim N(m_\lambda, s_\lambda^2)$ and the task is to find $m_\lambda$ and $s_\lambda^2$ by eliciting a suitable prior on the model parameter $\lambda$ with mean and variance specified by the user, say $E(\lambda) = m_\lambda^*$ and $V(\lambda) = s_\lambda^{*2}$. In the simplest example, with the identity link, $\lambda \sim N(m_\lambda^*, s_\lambda^{*2})$ transforms directly to $\alpha_\lambda \sim N(m_\lambda, s_\lambda^2)$ with $m_\lambda = m_\lambda^*$ and $s_\lambda^2 = s_\lambda^{*2}$. In the case with a log-link, a suitable prior on $\lambda$ is the log-normal density with mean $m_\lambda^*$ and variance $s_\lambda^{*2}$ which transforms to $\alpha_\lambda \sim N(m_\lambda, s_\lambda^2)$ with $s_\lambda^2 = \log\left[\left(\frac{s_\lambda^*}{m_\lambda^*}\right)^2 + 1\right]$ and $m_\lambda = \log(m_\lambda^*) - s_\lambda^2/2$. For more complicated links it is a simple exercise to derive the implied distribution of $\alpha_\lambda$ and then use numerical optimization methods to find the parameters in the normal prior, $m_\lambda$ and $s_\lambda$, that minimizes the Kullback-Leibler distance between the distribution of $\lambda$ and $\alpha$.

The regression coefficients in $\beta_\lambda$ are assumed to be a priori independent of $\alpha_\lambda$ with $\beta_\lambda \sim N(0, c_\lambda\Sigma_\lambda)$. Here $\Sigma_\lambda = (W^T\hat{D}_\lambda W)^{-1}$, where $W$ is the matrix of covariates excluding the

intercept and $\hat{D}_\lambda$ is the conditional Fisher information for $\lambda$ evaluated at the prior modes of $\alpha_\lambda$ and $\beta_\lambda$, which is the vector $\hat{\beta}_\lambda = (m_\lambda, \mathbf{0}^T)^T$. Thus $\hat{D}_\lambda$ depends only on the constant $m_\lambda$. The conditional Fisher information for $\lambda = (\lambda_1, \ldots \lambda_n)^T$ is a diagonal matrix with elements

$$-E\left[\frac{\partial^2 \log p(y_i|\lambda_i)}{\partial \lambda_i^2}\right] g'_\lambda(\lambda_i)^{-2}.$$

Setting $c_\lambda = n$ gives a unit information prior, i.e. a prior that carries the information equivalent to a single subject from the model. For the models in our framework $\hat{D}_\lambda$ can not be obtained analytically but is easily computed by simulation. It is straightforward to extend the argument to elicit priors for more than one model parameter. For details and examples see Villani et al. (2012).

We allow for variable selection in all covariate sets in the model. For a given component let the indicator variable $\mathcal{I} = \{I_1, \ldots I_{p_x}\}$ be defined such that $I_j = 0$ means that the $j$th element in $\beta$ is zero and the corresponding covariate drops out. Let $\beta_\mathcal{I}$ be the vector of non-zero coefficients, and for any $\mathcal{I}$ let $\mathcal{I}^c$ denote its complement. We make the assumption that the intercept is always in the model. Let $\beta \sim N(0, c\Sigma)$ as discussed above for the regression coefficients. Conditioning on the variables that are in the model we obtain

$$\beta_\mathcal{I}|\mathcal{I} \;\sim\; N\left[0, c(\Sigma_{\mathcal{I},\mathcal{I}} - \Sigma_{\mathcal{I},\mathcal{I}^c}\Sigma^{-1}_{\mathcal{I}^c,\mathcal{I}^c}\Sigma^T_{\mathcal{I}^c,\mathcal{I}})\right]$$

and $\beta_{\mathcal{I}^c}|\mathcal{I}$ is identically zero.

5.1.2. *Mixing function.* For the vector $\gamma = (\gamma_2^T, \ldots \gamma_K^T)^T$ (recall that $\gamma_1 = 0$) we assume $\gamma \sim N(0, c_\gamma I)$. It is also possible to use a prior with non-diagonal structure as above but this is not pursued here. Variable selection is done similarly as above by introducing the indicator $\mathcal{I}_Z$ for $\gamma$.

5.1.3. *Variable selection indicators.* For both the component and the mixing parts of the model the indicators are assumed to be a priori independent and Bernoulli distributed, i.e $P(I_i = 1) = \pi$, $0 \leq \pi \leq 1$ and $\pi$ is allowed to be different for each model parameter. It is

straightforward to let $\pi$ be unknown and estimate it in a separate updating step as in Kohn et al. (2001).

5.2. **General MCMC scheme.** Villani et al. (2009) experimented with different algorithms for finite mixture models in a related setting. Their preferred algorithm is the one used in this paper. The algorithm is a Metropolis within Gibbs sampler that draws the regression parameters and variable selection indicators jointly. Assume a component density with $L$ different model parameters and $K$ components. The following three blocks are sampled

(1) $s$

(2) $\gamma, \mathcal{I}_Z$

(3) $\{(\beta_1, \mathcal{I}_1), \ldots, (\beta_L, \mathcal{I}_L)\}_{k=1}^K$.

How to sample $s$ depends if it is a static or dynamic mixture. For the static mixture

$$P(s_i = k|x_i, v_i, y_i) \propto \left(\prod_{j=1}^{n_i - c_i} (1 - h^k(x_{ij}))\right) (h^k(x_{in_i}))^{c_i} \frac{\exp(v_i^T \gamma_k)}{\sum_{l=1}^K \exp(v_i^T \gamma_l)} \tag{5.1}$$

independently for $i = 1, \ldots, N$. For the dynamic mixture, the full conditional of $s_{ij}$ is independent of all other $s_{ij}$, $i = 1, ..., n$ and $j = 1, ..., n_i$, and is of the form

$$P(s_{ij} = k|x_i, z_i, y_i) \propto \begin{cases} h_{ij}^k \frac{\exp(z_{ij}^T \gamma_k)}{\sum_{l=1}^K \exp(z_{ij}^T \gamma_l)} & \text{if } c_i = 1 \text{ and } j = n_i \\ (1 - h_{ij}^k) \frac{\exp(z_{ij}^T \gamma_k)}{\sum_{l=1}^K \exp(z_{ij}^T \gamma_l)} & \text{otherwise.} \end{cases} \tag{5.2}$$

Note that this allows us to sample $s_{ij}$ independently for all $i$ and $j$ so this updating step is very fast in comparison with Markov models of $s_{ij}$.

Conditional on $s$, Step 2 is a multinomial logistic regression with variable selection. It is possible to apply a generalization of the algorithm described in the next section to handle this updating step efficiently, see Villani et al. (2009) for details.

5.3. **Variable-dimension finite step Newton proposals.** This section presents how to construct the tailored proposals for any component model presented in Section 4.1 based on the algorithm in Villani et al. (2009) and Villani et al. (2012), which generalizes earlier algorithms in Gamerman (1997), Qi and Minka (2002) and Nott and Leonte (2004). For

clarification the algorithm is first presented in the case with no variable selection and then extended. The only requirement is that the likelihood part of the posterior can be factorized as

$$p(\beta|y) = \prod_{i=1}^{N} p(y_i|\phi_i)p(\beta) \tag{5.3}$$

where $\phi_i = g^{-1}(x_i^T\beta)$. Note that there can be more than one model parameter and then $p(\beta|y)$ is a full conditional posterior distribution and the algorithm can be used as a Metropolis-within-Gibbs step. After a proper relabeling of the product in the likelihood in Equation (4.1) it has the same form as the likelihood part in Equation (5.3). The proposal distribution is tailored using an approximate posterior mode and the curvature around that mode. The approximate mode is found by taking a few steps with Newton's algorithm. To implement the algorithm we need the following results from Lemma 1 in Villani et al. (2012)

$$\frac{\partial \log p(y|\beta)}{\partial \beta} = X^T\tilde{g} \tag{5.4}$$

where $X$ is the covariate matrix, $\tilde{g} = (\tilde{g}_1, \ldots, \tilde{g}_n)^T$,

$$\tilde{g}_i = \frac{\partial \log p(y|\phi_i)}{\partial \phi_i} g'(\phi_i)^{-1}.$$

The outer-product approximation of the Hessian is

$$\frac{\partial^2 \log p(\beta|y)}{\partial \beta \partial \beta^T} \approx X^T D X, \tag{5.5}$$

where $D = \text{diag}(\tilde{g}_i^2)$. Villani et al. (2012) also derives expression for the exact Hessian but we have found the outer-product approximation to be more numerically stable for our problem. Note that the Lemma only requires derivatives for the scalar parameters of the log-likelihood which makes it possible to write general computer code and the user is only required to compute analytical derivatives for one-dimensional quantities. Newton's algorithm is

$$\beta_{r+1} = \beta_r - A_r^{-1}s_r,\ r = 0, \ldots, R \tag{5.6}$$

where $s_r$ and $A_r$ is the gradient and Hessian of the log posterior, respectively. Using the results above we have

$$\begin{aligned} s_r &= X^T\tilde{g} + \frac{\partial \log p(\beta)}{\partial \beta} \\ A_r &= X^T DX + \frac{\partial^2 \log p(\beta)}{\partial \beta \partial \beta^T}. \end{aligned}$$

Start with $\beta_0 = \beta_c$ and let $\hat{\beta}$ be the vector obtained after $R$ Newton steps. This is not necessarily the mode but is often close because the previously accepted draw is used as initial value. Setting $R = 1, 2$ or 3 is usually sufficient. Let $\beta_c \in \mathbb{R}^{p_x \times 1}$ denote the current and $\beta_p \in \mathbb{R}^{p_x \times 1}$ the proposed posterior draw. The proposal distribution is a multivariate t-distribution with $\nu \geq 2$ degrees of freedom, i.e

$$\beta_p|\beta_c \sim t_\nu \left[\hat{\beta}, -\left(\frac{\partial^2 \log p(\beta|y)}{\partial \beta \partial \beta^T}\right)\Big|_{\beta=\hat{\beta}}\right].$$

To extend the algorithm to variable selection the pair $(\beta, \mathcal{I})$ is proposed jointly conditional on the previously accepted parameter and indicator. This proposal can be factorized as

$$J(\beta_p, \mathcal{I}_p|\beta_c, \mathcal{I}_c) = J_1(\beta_p|\mathcal{I}_p, \beta_c)J_2(\mathcal{I}_p|\beta_c, \mathcal{I}_c) \tag{5.7}$$

$J_1$ is a generalization of the proposal for $\beta_p$ above and $J_2$ is the proposal for the indicators. Consider first the $\beta$ proposal. Since $\beta_c$ and $\beta_p$ may be of different dimensions we use the following generalized Newton algorithm from Villani et al. (2012)

$$\beta_{r+1} = A_r^{-1}(B_r\beta_r - s_r),\ r = 0, \ldots, R \tag{5.8}$$

where

$$\begin{aligned} s_r &= X_{r+1}^T\tilde{g} + \frac{\partial \log p(\beta)}{\partial \beta} \\ A_r &= X_{r+1}^T DX_{r+1} + \frac{\partial^2 \log p(\beta)}{\partial \beta \partial \beta^T} \\ B_r &= X_{r+1}^T DX_r + \frac{\partial^2 \log p(\beta)}{\partial \beta \partial \beta^T}, \end{aligned}$$

where $X_r$ is the matrix with columns corresponding to the non-zero coefficients in $\beta_r$, and the likelihood part of the expressions are evaluated at $\beta = \beta_r$. The prior parts are evaluated at the entire vector $\beta$ (including the zero parameters) and then the sub-vector conformable with $\beta_{r+1}$ is extracted from the result. Note that after the first step the parameter no longer changes dimension and the generalized Newton algorithm reduces to the usual Newton algorithm.

Following Villani et al. (2009) and Villani et al. (2012) we choose a simple proposal of $\mathcal{I}$ where a subset of the indicators is randomly selected and a change of the selected indicators is proposed, one variable at a time.

With these proposals the acceptance probability in the Metropolis Hastings algorithm is

$$\alpha[(\beta_c, \mathcal{I}_c) \to (\beta_p, \mathcal{I}_p)] = \min\left(1, \frac{p(y|\beta_p, \mathcal{I}_p)p(\beta_p|\mathcal{I}_p)p(\mathcal{I}_p)/J_1(\beta_p|\mathcal{I}_p, \beta_c)J_2(\mathcal{I}_p|\beta_c, \mathcal{I}_c)}{p(y|\beta_c, \mathcal{I}_c)p(\beta_c|\mathcal{I}_c)p(\mathcal{I}_c)/J_1(\beta_c|\mathcal{I}_c, \beta_p)J_2(\mathcal{I}_c|\beta_p, \mathcal{I}_p)}\right).$$

The proposal density for $\beta$ at the proposed point $J_1(\beta_p|\mathcal{I}_p, \beta_c)$ is the multivariate t-density with mode $\hat{\beta}$ and covariance matrix evaluated at $\hat{\beta}$, where $\hat{\beta}$ is obtained by iterating Equation (5.8) with $\beta_0 = \beta_c$. The proposal density at the current point $J_1(\beta_c|\mathcal{I}_c, \beta_p)$ is also a multivariate t-density but with mode $\tilde{\beta}$ and covariance matrix evaluated at $\tilde{\beta}$, obtained from the same iteration scheme but this time from initial value $\beta_0 = \beta_p$. The proposal density for $\mathcal{I}$ at the current and proposed is the same for this simple proposal.

It is well-known that finite mixtures have identification problems because the likelihood is invariant with respect to permutations of the components. This is referred to as the label switching problem, see Frühwirth-Schnatter (2006) and Jasra et al. (2005). When estimating the predictive density this is not a problem (Geweke, 2007) but if the model is used for model based clustering one needs to proceed with caution. Plotting the MCMC samples may reveal if there was a problem with switching labels. Order conditions on the parameter space may be imposed to avoid the identification problem, see Jasra et al. (2005).

5.4. **Selecting number of components.** The key quantity for selecting models in the Bayesian framework is the marginal likelihood which allows to compute Bayes factors and determine the plausibility of one model against another. However, the marginal likelihood may be sensitive to the choice of prior distribution, especially when the prior information is vague. For a general discussion see Kass (1993) and Richardson and Green (2002) in the context of mixture models.

Following Geweke and Keane (2007) and Villani et al. (2009) we therefore choose models based on the log predictive score (LPS). The LPS removes most of the dependence on the prior by sacrificing a subset of the data to train the prior to get a posterior based on the training data. If $y_{test}$ denotes the test data and $y_{train}$ the training data then the LPS is

$$p(y_{test}|y_{train}) = \int p(y_{test}|\theta)p(\theta|y_{train})\mathrm{d}\theta$$

if the test and training data are independent conditional on $\theta$, which is the case in our longitudinal setting since the entire time series for a single subject belongs to either the test or training set. To deal with the arbitrary division into training and test data, a cross validated version of the LPS is used

$$LPS = \frac{1}{B}\sum_{b=1}^{B}\log p(\tilde{y}_b|\tilde{y}_{-b}, x),$$

where $\tilde{y}_b$ is the test data in the $b$th test sample and $\tilde{y}_{-b}$ denotes the training data. Since subjects are independent conditional on the parameters

$$p(\tilde{y}_b|\tilde{y}_{-b}, x) = \int \prod_{i\in\tau_b} p(y_i|\theta, x_i)p(\theta|\tilde{y}_{-b})\mathrm{d}\theta,$$

where $\tau_b$ contains the index set of the observations in the test data for the $b$th sample. $p(\tilde{y}_b|\tilde{y}_{-b}, x)$ is easily computed by averaging $\prod_{i\in\tau_b} p(y_i|\theta)$ over the posterior draws $p(\theta|\tilde{y}_{-b})$. This requires sampling from $B$ posterior distributions based on different training data but can be done independently for each data set so computer parallelism may be exploited.

As a very rough rule of thumb, a difference in LPS between 3 and 5 between two models is usually said to be strong evidence in favor of one model, and a difference of more than five LPS points is very strong evidence (Kass and Raftery, 1995).

## 6. Application: modeling Firm Bankruptcy risk

6.1. **Data.** Our data set contains yearly observations for Swedish firms in the time period 1991-2008 on bankruptcy status, firm-specific variables and two macro variables. This data set has been analyzed in Jacobson et al. (2011) and Giordani et al. (2013). Jacobson et al. (2011) uses a similar approach as Shumway (2001) with a multi-period logit model extended with macro economic variables. Giordani et al. (2013) extend by modeling the log odds of the firm failure probability as a non-linear function of covariates by introducing spline functions. They show substantial improvements in predictive power as a result of accounting for nonlinearities. The present paper consider the same predictors as in Giordani et al. (2013). These are three financial ratios, two firm-specific control variables and two macroeconomic variables. The financial ratios are: *EBIT/TA* - earnings before interest and taxes over total assets (earnings ratio); *TL/TA* - total liabilities over total assets (leverage ratio) and *CH/TL* - cash and liquid assets over total liabilities (cash ratio). The control variables are: *logTS* - logarithm of deflated total sales and *logAge* - logarithm of firm age in years since first registered as a corporate. Finally the macroeconomic variables included are: *GDPG* - yearly GDP-growth rate and *Repo* - the interest rate set by Sveriges Riksbank (the Central bank of Sweden). For a thorough description of the data set, definition of bankruptcy, and other details see Giordani et al. (2013).

6.2. **Models.** Although the spline model accounts for nonlinearities in a flexible way it has some drawbacks. First, the model assumes additivity, i.e. it rules out interactions between the covariates, and the extension to spline surface models with interactions is not computationally realistic for a data set of our size. Second, it can be hard to interpret spline models as the nonlinearities are not themselves explained by other covariates. Third, it cannot account for heterogeneity coming from missing explanatory variables. Fourth, it can

be computationally demanding for moderate to large data sets when doing Bayesian inference via MCMC. This is because the dimension of the covariate space can increase dramatically after expanding in basis functions. Variable selection can be used to keep the number of effective parameters at a minimum, but increases the computational burden.

We propose to analyze bankruptcy data for Swedish firms with a finite mixture of survival models . Such models can not only account for heterogeneity and nonlinearities, but also gives an interpretation of these features in terms of covariates. A mixture model can also be used for model based clustering which gives insights about firm dynamics. The use of covariates in the mixing function is extremely useful for understanding the role of the different mixture components. Many models in the bankruptcy literature are special cases of our model. For example the models in Shumway (2001) and Jacobson et al. (2011) are obtained with $K = 1$ and $h(x_{ij}) = \frac{\exp x_{ij}^T \beta}{1+\exp x_{ij}^T \beta}$. Likewise, the model in Giordani et al. (2013) has the same structure but in addition $x$ is expanded using spline functions. It is even possible to have $K > 1$ and use splines simultaneously as in Villani et al. (2009) for the case of heteroscedastic Gaussian regression. This paper omits splines to stress the fact that the finite mixture itself can capture the non-monotonic relationships. Adding spline terms in the mixture components would also increase the computing time dramatically.

We want each firm to have a sample space $t = \{1, 2, \dots\}$. This requires covariates for each observed time period, so we are restricted to consider firms with start-up year 1991 at the earliest. The analysis can be broaden to other type of firms but then one has to consider missing data issues so this is not pursued here. Thus the population studied in the present paper consist of Swedish firms that enter the sample in the period 1991-2008. The dataset is huge with a total of $228,589$ firms with $1,670,781$ firm-year observations, on average 7.3 time-periods per firm. To speed up computing times, we shall here analyze a randomly selected subset of $11,317$ firms with $82,831$ firm-year observations, on average 7.3 time-period per firm. We are currently working on an extension of the MCMC methods with the potential of handling essentially arbitrarily large data sets, but this will be reported elsewhere.

We estimate and compare both static and dynamic mixtures and also a one-component model with flexible baseline hazards. Two different distributions for the survival time are considered: exponential and Weibull as described in Section 4.1. The Weibull models are used with and without covariates in the shape parameter $\rho$. Weibull models with covariates in $\rho$ seems to be novel in the literature.

In all dynamic mixtures, exponential moving average covariates have been used to achieve persistence in component allocations over time, as described in Section 2. The choice $\alpha = 0.3$ was justified by computing for a range of values for $\alpha$ and then choose the one with highest in-sample LPS score. The choice of $\alpha$ does not affect the relative comparison between the dynamic and static models. It is also possible to estimate $\alpha$ from the data but this is not pursued here.

6.3. **Priors.** The prior for $\lambda$ is log-Normal with $E(\lambda) = 0.01405$ (the empirical hazard for another subset of the data) and $D(\lambda) = 0.05$ for both the exponential and Weibull model. The additional parameter $\rho$ in the Weibull model is also assigned a log-Normal prior with $E(\rho) = 1$ and $D(\rho) = 5$. Note that $\rho = 1$ gives the exponential model. Both priors are rather non-informative considering the scale and the log-link. The prior utilizing the Fisher information described in Section 5.1.1 is not needed in this particular example because of the enormous amounts of data, and we therefore assume prior independence between the regression coefficients for simplicity. For the mixing function the shrinkage factor $c_\gamma = 10$ gives a non-informative prior. The prior inclusion probability was set to 0.5 for each variable and in all parts of the model.

6.4. **Algorithmic considerations.** We use the Metropolis-within-Gibbs algorithm with tailored proposals and variable selection to sample from the posterior. The number of steps in the variable dimension Newton algorithm $R$ is set to 1 for the component model in all parameters and 3 for the mixing function. The degrees of freedom in the multivariate $t$ proposal is set to 10, for both the component and the mixing part of the model. Each variable selection indicator is proposed to change with probability 0.2 in each iteration of the algorithm.

For all combinations of models in Section 6.2, $20,000$ iterations with the MCMC algorithm where performed and $5,000$ of them discarded as burn-in period, leaving $15,000$ draws from the posterior distribution. The efficiency of the sampler is measured by the inefficiency factor, which is defined as

$$IF \quad = \quad 1 + 2\sum_{l=1}^{L} \rho_l,$$

where $\rho_l$ is the autocorrelation at the $l$th lag in the MCMC chain and $L$ is an upper limit such that $\rho_l \approx 0$ when $l > L$. IF-values near 1 suggests a very efficient algorithm. We monitor convergence and measure performance using the cumulative means and IFs for the predictive mean $E(y|x)$ over a grid of $x$-values. The LPS was computed using $B = 4$ folds of the data.

6.5. **Results.** As a first attempt to investigate the fit of the models, Figure 6.1 compares the models' implied hazard function $h_t(x_t)$ as function of time to the empirical hazard rate. The models' hazard probabilities have been computed for each of the firms in the panel and then averaged across all firms. The posterior uncertainty regarding the hazard is illustrated with a box plot computed from the MCMC draws. In the case of the exponential model (left column), it is clear that the one-component model gives a very poor fit to the empirical hazard, but then quickly improves as more components are added to the model. A two-component exponential model gives a similar estimated hazard as a one-component model with flexible baseline hazards (top right). The one-component Weibull model without covariates in the shape parameter $\rho$ produces a similar hazard as the one-component exponential model, but by adding covariates in $\rho$ the Weibull model can capture the non-monotonic relationship of the empirical hazard fairly well.

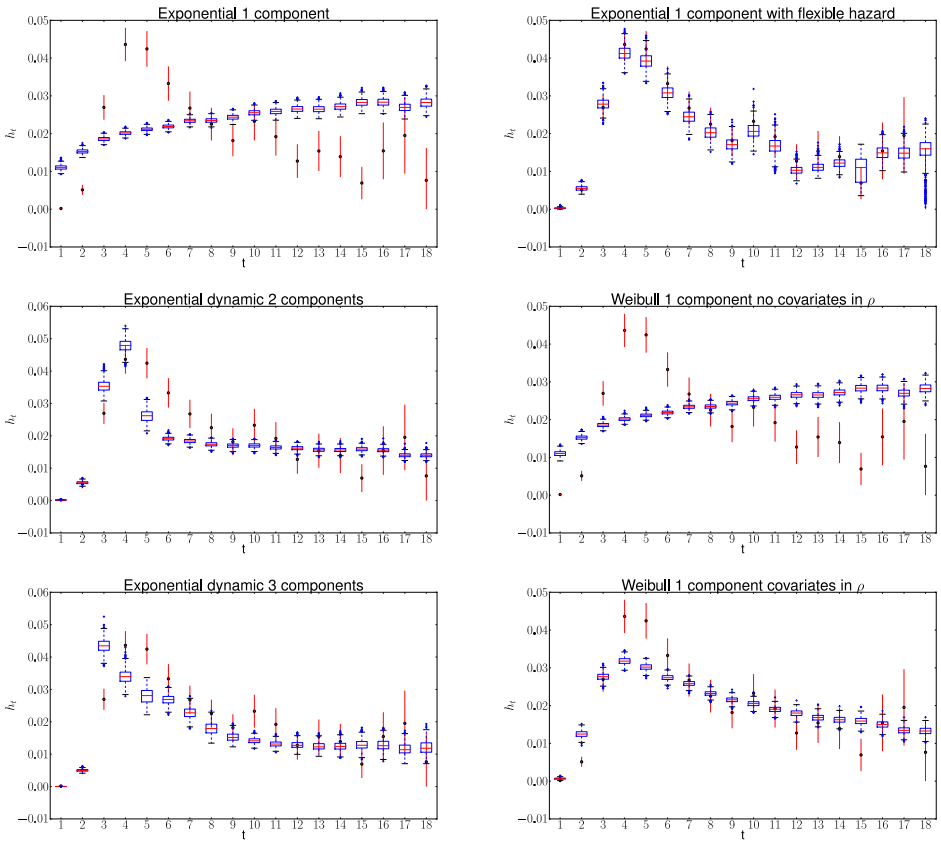


FIGURE 6.1. Hazard as a function of time for some models (box-plots) plotted against the empirical hazard (red vertical lines).

TABLE 1. Log Predictive Score (LPS) for the static and dynamic mixtures computed using 4-fold cross-validation. The best model for a given number of components are in bold typeface.

| **Static mixtures** | **Comp 1** | **Comp 2** | **Comp 3** |
|---|---|---|---|
| Exponential | -1784.83 | -1712.16 | -1687.54 |
| Weibull | -1785.17 | -1725.08 | -1683.60 |
| Weibull covariates in $\rho$ | -1696.96 | -1652.78 | -1648.73 |
| **Dynamic mixtures** | **Comp 1** | **Comp 2** | **Comp 3** |
| Exponential | -1784.83 | -1618.51 | -1570.20 |
| Weibull | -1785.17 | -1605.00 | -1561.97 |
| Weibull covariates in $\rho$ | -1696.96 | **-1585.07** | **-1553.43** |
| Exponential Flex Baseline | **-1686.41** | | |

The assessment of model fit in Figure 6.1 is visually appealing, but is very much a rather limited marginal view of the data. We will now turn to a comparison of the models' Log Predictive Scores (LPS), to compare the out-of-sample forecasting performance of the models. Table 1 reports the LPS for static and dynamic mixtures, using either exponential or Weibull components, with and without covariates in the Weibull shape parameter. The most striking result in Table 1 is the dramatically better out-of-sample predictive performance of the dynamic mixtures compared to their static counterparts. As an example, the three-components dynamic mixture of exponentials is 117.34 LPS units better than the three-components static mixture of exponentials. Table 1 also reports the LPS of the one-component exponential model with a free baseline hazard parameter estimated for each year. Using a flexible baseline hazard clearly improves the LPS, but also this model model is clearly outperformed by the dynamic mixtures; the LPS difference between the flexible baseline model and the three-component dynamic mixture of exponentials is 116.21. This suggests that these data are truly heterogeneous even after controlling for age and size effects and different baseline hazards.

Another interesting observation from Table 1 is that the LPS for the Weibull model improves considerably when allowing for covariates in both model parameters. This is true for models with multiple components as well. Covariates in the shape parameter of the Weibull is rare or perhaps even non-existent in practical work, but this is clearly an extension that should be considered.

In all models, the LPS improves for each added component but the rate of improvement decreases. It is worthwhile to mention that variable selection implies that adding components does not necessarily give a more complex model. See the Lidar example in Li et al. (2011) for a clear demonstration of how variable selection in mixture-of-experts models can be a very effective guard against overfitting.

To illustrate some of the interpretations of our models, Tables 2-4 presents parameter estimates for some selected one- and two-component models. Data have been standardized to have zero mean and unit variance for all covariates, hence all parameter estimates are on

the same scale. The posterior mean and standard deviation are computed conditional on the covariate belonging to the model.

TABLE 2. Estimation results for exponential model with one component. IF: min = 0.55, median = 1, max = 1.70.

| **Component 1** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Intercept | Earnings | Leverage | Cash | Size | Age | GDPG | Repo |
| **Post Mean** | -4.397 | -0.258 | 0.25 | -1.174 | -0.02 | 0.399 | 0.057 | 0.1 |
| **Post Std** | 0.045 | 0.019 | 0.018 | 0.11 | 0.023 | 0.033 | 0.033 | 0.029 |
| **Post Incl Prob** | - | 1 | 1 | 1 | 0.012 | 1 | 0.051 | 0.837 |
| **Mean Acc Prob** | 0.404 | | | | | | | |

Starting with the results for the one component exponential model in Table 2, we see that the most significant variables are cash, age, earnings, and leverage, all with a posterior inclusion probability of unity. The variable selection effectively removes size, GDPG, and to some extent Repo. In this model, a positive sign corresponds to increased hazard probability as a variable increases, and vice versa. For the Weibull model with covariates in both parameters in Table 3, the important covariates in $\rho$ are cash and age, both with a posterior inclusion probability 1 and for the rest of the covariates the inclusion probability is near zero.

TABLE 3. Estimation results for Weibull model with one component and covariates in both parameters. IF: min = 0.92, median = 7.01, max = 14.5.

| **Parameter $\lambda$** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Intercept | Earnings | Leverage | Cash | Size | Age | GDPG | Repo |
| **Post Mean** | -4.985 | -0.265 | 0.213 | 0.81 | -0.04 | 1.971 | -0.02 | 0.057 |
| **Post Std** | 0.244 | 0.022 | 0.02 | 0.073 | 0.026 | 0.118 | 0.026 | 0.028 |
| **Post Incl Prob** | - | 1 | 1 | 1 | 0.029 | 1 | 0.014 | 0.063 |
| **Mean Acc Prob** | 0.712 | | | | | | | |
| **Parameter $\rho$** | | | | | | | | |
| | Intercept | Earnings | Leverage | Cash | Size | Age | GDPG | Repo |
| **Post Mean** | 0.231 | 0.021 | -0.013 | -1.07 | -0.008 | -0.782 | -0.003 | 0.015 |
| **Post Std** | 0.093 | 0.015 | 0.014 | 0.071 | 0.008 | 0.04 | 0.009 | 0.009 |
| **Post Incl Prob** | - | 0.007 | 0.01 | 1 | 0.002 | 1 | 0.004 | 0.016 |
| **Mean Acc Prob** | 0.788 | | | | | | | |

Moving to the dynamic mixture of two exponential components in Table 4 it is evident that the most significant covariates in the mixing function are age and cash. There is also a posterior inclusion probability of 1 for GDPG and Repo, but the magnitude of their effects are smaller. This means that the separation of the data into the two different classes is mostly determined by age and cash. Our parametrization is such that when age increases it is more likely to belong to the first component and the same holds for cash. To illustrate the interpretation of the mixture models, let us consider a newly founded firm. Since a newly founded firm is by definition of low age, such a firm tends to belongs to the second component with a large probability, everything else equal. Since age has a large positive coefficient in the second component, this young firm will initially experience a rapidly increasing hazard as it grows older. If the firm manages to survive the early years, it will eventually move over to the first mixture component where age is no longer a significant determinant of the hazard. The firm has managed to survive the first risky years and can now grow older without accelerating risk on account of its age. Figure 6.2 shows the posterior allocation of firms over their time: firms that have survived for a long time are classified to component 1 in their later time periods, while firms in early time periods are classified to the second component. Cash has a similar interpretation as age; with a large probability, a firm with low cash belongs to the second component where the coefficient on cash is strongly negative. This means that a low cash firm can drastically reduce the bankruptcy probability by increasing its holdings of cash. As the firm continues to improve its liquidity, it will eventually reach a point where it switches over to the first component. In this component, cash remains a positive factor for decreasing bankruptcy risk, but its effect is much smaller.

TABLE 4. Estimation results for a dynamic exponential model with two components. Covariates in the mixing function are exponentially moving averages. Parameters in the mixing function corresponds to $P(s_t = 2|z_t)$. IF: min = 0.84, median = 1.23, max = 110.84.

| **Component 1** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Intercept | Earnings | Leverage | Cash | Size | Age | GDPG | Repo |
| **Post Mean** | -4.311 | -0.251 | 0.339 | -0.559 | 0.033 | 0.096 | -0.024 | 0.039 |
| **Post Std** | 0.058 | 0.026 | 0.023 | 0.108 | 0.035 | 0.086 | 0.05 | 0.067 |
| **Post Incl Prob** | - | 1 | 1 | 1 | 0.016 | 0.055 | 0.016 | 0.026 |
| **Mean Acc Prob** | 0.751 | | | | | | | |
| **Component 2** | | | | | | | | |
| **Post Mean** | -2.522 | -0.367 | 0.042 | -2.873 | -0.004 | 4.544 | -0.066 | 0.044 |
| **Post Std** | 0.238 | 0.036 | 0.053 | 0.501 | 0.052 | 0.237 | 0.045 | 0.043 |
| **Post Incl Prob** | - | 1 | 0.027 | 1 | 0.013 | 1 | 0.04 | 0.019 |
| **Mean Acc Prob** | 0.782 | | | | | | | |
| **Mixing** | | | | | | | | |
| **Post Mean** | -4.777 | -0.113 | 0.031 | -1.698 | 0.039 | -8.296 | 0.788 | 0.735 |
| **Post Std** | 0.496 | 0.088 | 0.094 | 0.39 | 0.089 | 0.815 | 0.184 | 0.196 |
| **Post Incl Prob** | - | 0.092 | 0.067 | 1 | 0.066 | 1 | 1 | 1 |
| **Mean Acc Prob** | 0.835 | | | | | | | |

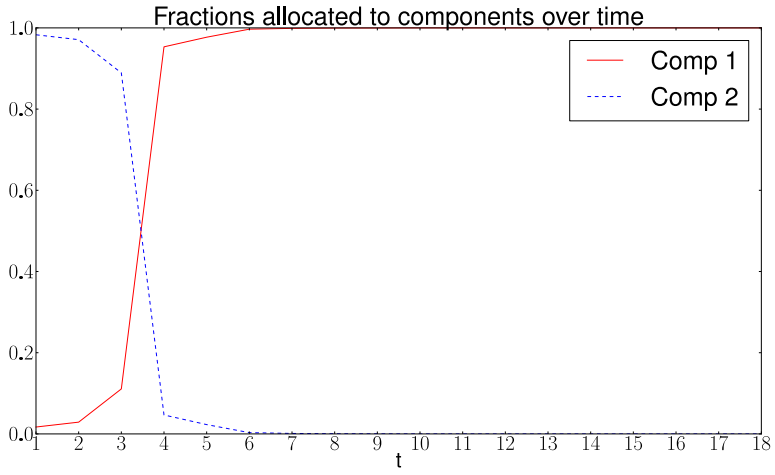


FIGURE 6.2. Fraction allocated to respective component over time for the dynamic exponential mixture.

To further explore the difference between the static and dynamic mixtures we plot the overall predictive hazard $h_t(x_t)$ in Figures 6.3 and 6.4 for a firm that is born in the beginning

of the sample period, i.e. 1991. Each subgraph shows the predictive hazard $h_t(x_t)$ as a function of the covariate cash for a given year. The analysis in Figures 6.3 and 6.4 is conditioned on fixed paths for the other covariates. We have chosen to set the covariate paths for Repo, GDPG and age as the realized values at each time point but with a one year lag for repo and GDPG; when predicting bankruptcy at period $t$, macro variables from $t-1$ are used. For the financial ratios and the size variable, the average covariate value in the sample for each respective year is used as conditioning paths. The covariate paths together with their moving averages are presented in Figure 6.5. Figure 6.3 plots the predictive hazard $h_t(x_t)$ for the early years $1991-1994$ corresponding to a survival time $t = 1, 2, 3, 4$, while Figure 6.4 covers the years $2004-2007$ corresponding to $t = 14, 15, 16, 17$. This example clearly illustrates the main difference in these models; the dynamically evolving proportions in the dynamic mixture (left panel) gives a much more flexible hazard than the static mixture (right panel) where the mixture weights are constant thoughout time and the overall shape of the hazard is forced to remain unchanged.

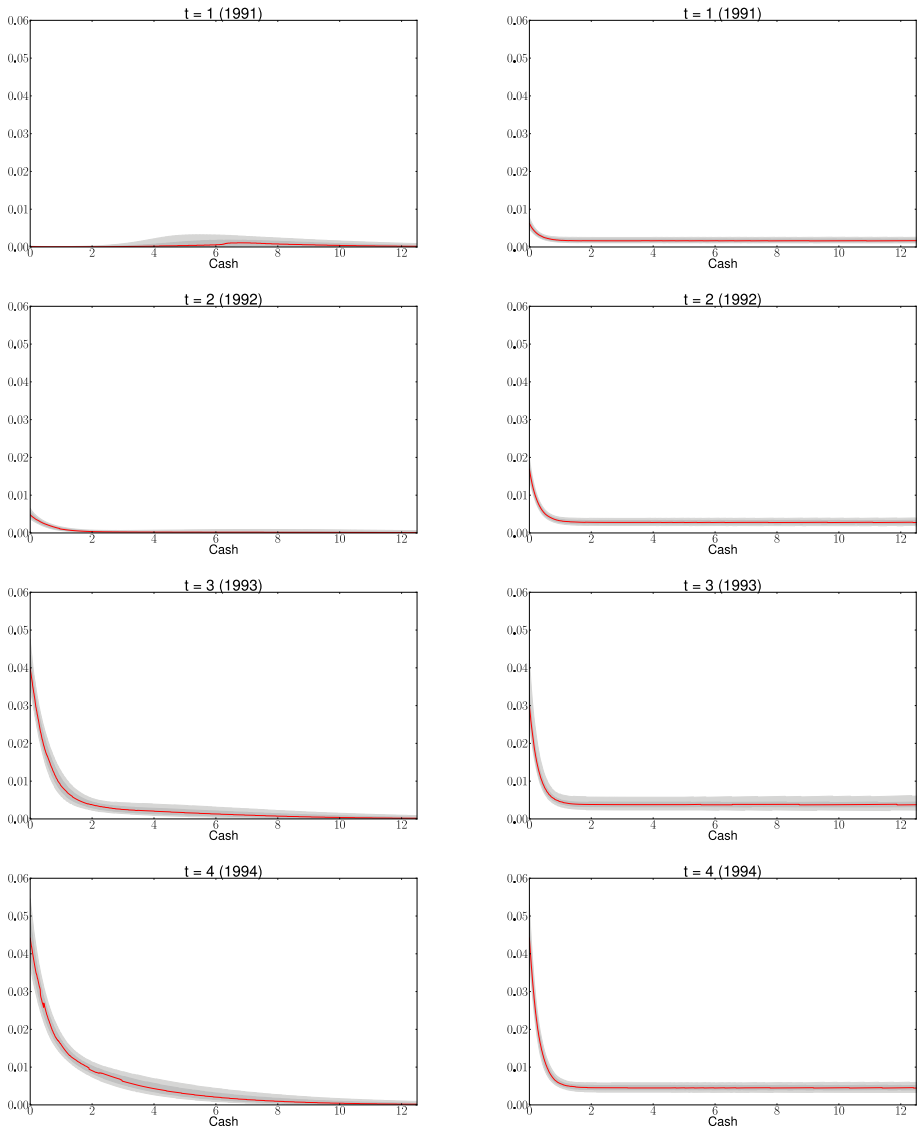


FIGURE 6.3. Posterior distribution of the hazard probability of the representative firm as a function of cash for a dynamic (left panel) and a static (right panel) exponential mixture with two components for $t = 1, \ldots, 4$. The dark shaded area corresponds to 68% Highest Posterior Density (HPD) regions and lighter shaded area are the 95% HPD regions. The red solid line is the posterior mode.

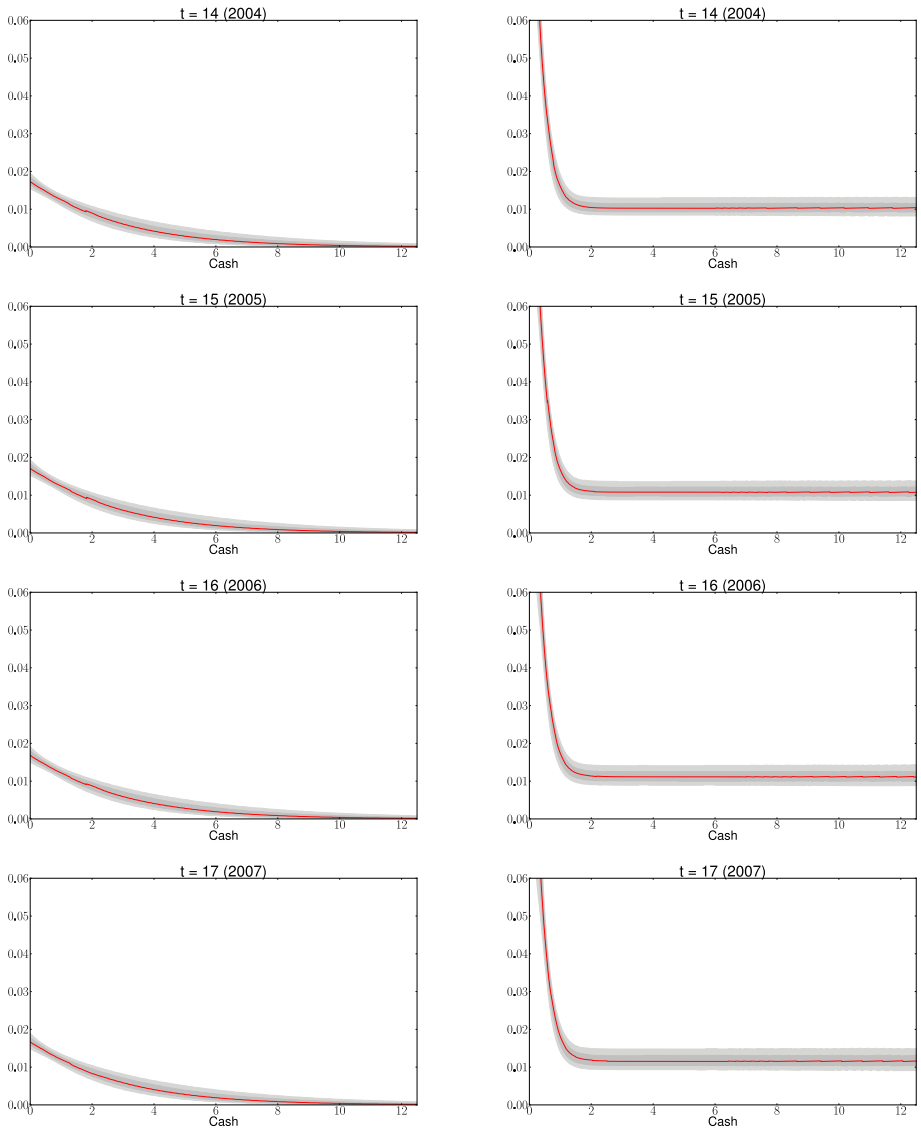


FIGURE 6.4. Posterior distribution of the hazard probability of the representative firm as a function of cash for a dynamic (left panel) and a static (right panel) exponential mixture with two components for $t = 14, \ldots, 17$. The dark shaded area corresponds to 68% Highest Posterior Density (HPD) regions and lighter shaded area are the 95% HPD regions. The red solid line is the posterior mode.

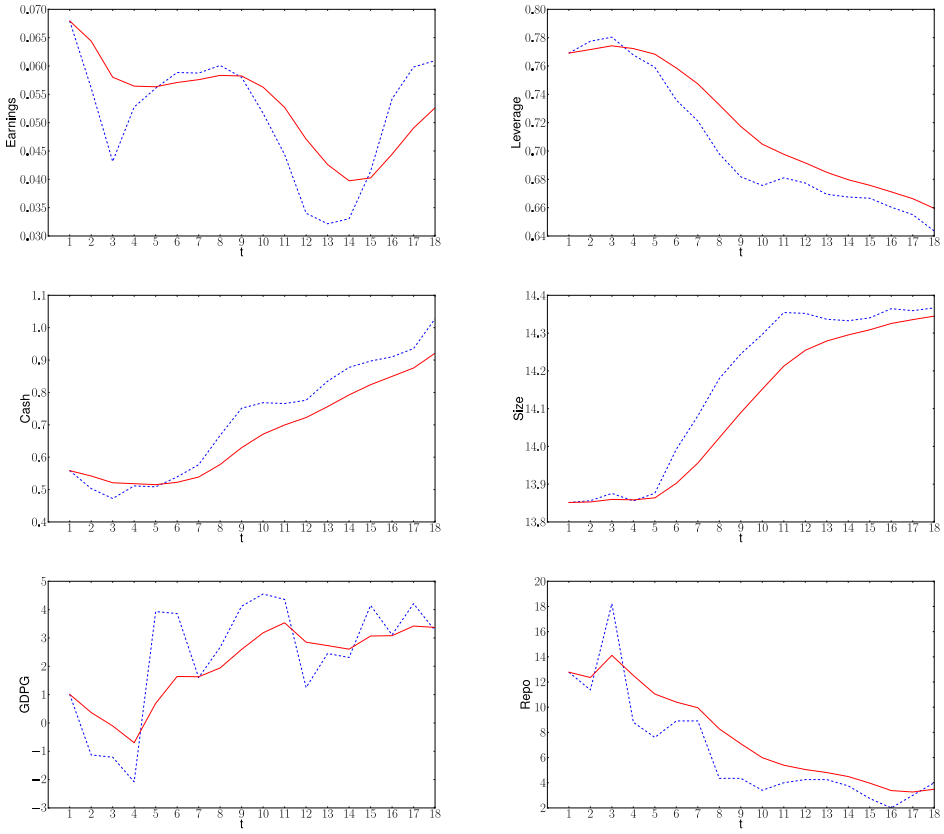


FIGURE 6.5. Covariate paths for a representative firm. The dashed blue line corresponds to realized covariates and the solid red line are the exponentially moving averages with $\alpha = 0.3$.

## 7. CONCLUSIONS

We propose flexible smooth mixture models for longitudinal data, with special emphasis on models for survival data in discrete time. We discuss how the longitudinal dimension opens up for two different types of mixture models, the static and dynamic mixture. In the static mixture, subjects have to remain in the same component in all time periods, whereas in the dynamic mixture they can move between mixture components over time. We argue that the obvious Markov transition model would be prohibitively time-consuming for datasets with a large number of subjects, and we propose an alternative approach where the within-subject dynamics is determined by subject-specific time-varying covariates.

We compare the static and dynamic mixtures in bankruptcy modeling for a large panel of Swedish firms over the time period 1991-2008. The main result is that the dynamic

mixture formulation dramatically outperforms the static mixture, a result that holds both when exponential or Weibull mixture components are used. We also show that the MCMC algorithm with variable selection in Villani et al. (2012) can be straightforwardly extended to the longitudinal case and we document a high MCMC efficiency in our application to firm bankruptcy.

It is also shown that the firm bankruptcy data are heterogeneous even after the standard firm specific variables in the literature are included in the model and when a flexible baseline hazard is used. This result suggests that there are different classes of firms and the effect of the covariates on the hazard probability is different in each class. Furthermore, it is also shown that model with multiple classes is able to generate a non-monotonic hazard function which agrees with the empirical hazard and also with models that uses a flexible baseline hazard with a separate parameter for each time period.

Although our way of modeling within-subject dynamics by mixture-of-experts with time-varying mixing covariates is computationally attractive in comparison to other standard approaches, data sets with millions of observations remains a challenge. We are currently working on extensions of the MCMC algorithm presented here that may reduce computing times substantially for large data sets. In terms of model extensions it would be interesting to explore the role of a continuous frailty in the components. The hierarchical structure of such a model requires two extra steps in the MCMC scheme; sampling the frailty and the parameters in its distribution. This is in principle straightforward, but will add to the computing time, which again requires innovations in the MCMC methodology.

## References

Allison, P. (1982). Discrete-time methods for the analysis of event histories. *Sociological methodology*, 13(1):61–98.

Baum, L. and Petrie, T. (1966). Statistical inference for probabilistic functions of finite state markov chains. *The Annals of Mathematical Statistics*, 37(6):1554–1563.

Carling, K., Edin, P.-A., Harkman, A., and Holmlund, B. (1996). Unemployment duration, unemployment benefits, and labor market programs in sweden. *Journal of Public Economics*, 59(3):313–334.

Cox, D. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2):187–220.

Doucet, A., Godsill, S., and Andrieu, C. (2000). On sequential monte carlo sampling methods for bayesian filtering. *Statistics and computing*, 10(3):197–208.

Franzén, J. (2008). *Bayesian Cluster Analysis: Some Extensions to Non-standard Situations*. PhD thesis, Stockholm University, Department of Statistics.

Frühwirth-Schnatter, S. (2006). *Finite mixture and Markov switching models*. Springer-Verlag.

Frühwirth-Schnatter, S. (2011). Panel data analysis: a survey on model-based clustering of time series. *Advances in Data Analysis and Classification*, 5(4):251–280.

Gamerman, D. (1997). Sampling from the posterior distribution in generalized linear mixed models. *Statistics and Computing*, 7(1):57–68.

Geweke, J. (2007). Interpretation and inference in mixture models: Simple mcmc works. *Computational Statistics & Data Analysis*, 51(7):3529–3550.

Geweke, J. and Keane, M. (2007). Smoothly mixing regressions. *Journal of Econometrics*, 138(1):252–290.

Giordani, P., Jacobson, T., Von Schedvin, E., and Villani, M. (2013). Taking the twists into account: Predicting firm bankruptcy risk with splines of financial ratios. *Journal of Financial and Quantitative Analysis*, forthcoming.

Huynh, K. and Voia, M. (2009). Mixed proportional hazard models with finite mixture unobserved heterogeneity: An application to nascent firm survival. Manuscript.

Ibrahim, J., Chen, M., and Sinha, D. (2005). *Bayesian survival analysis*. Wiley Online Library.

Jacobs, R., Jordan, M., Nowlan, S., and Hinton, G. (1991). Adaptive mixtures of local experts. *Neural computation*, 3(1):79–87.

Jacobson, T., Lindé, J., and Roszbach, K. (2011). Firm default and aggregate fluctuations. *Journal of European Economic Association*, forthcoming:to appear.

Jasra, A., Holmes, C., and Stephens, D. (2005). Markov chain monte carlo methods and the label switching problem in bayesian mixture modeling. *Statistical Science*, 20(1):50–67.

Jordan, M. and Jacobs, R. (1994). Hierarchical mixtures of experts and the em algorithm. *Neural computation*, 6(2):181–214.

Kass, R. (1993). Bayes factors in practice. *The Statistician*, 42(5):551–560.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the american statistical association*, 90(430):773–795.

Kim, C.-J. and Nelson, C. R. (2003). State-space models with regime switching: classical and gibbs-sampling approaches with applications. *MIT Press Books*, 1.

Kohn, R., Smith, M., and Chan, D. (2001). Nonparametric regression using linear combinations of basis functions. *Statistics and Computing*, 11(4):313–322.

Lancaster, T. (1979). Econometric methods for the duration of unemployment. *Econometrica: Journal of the Econometric Society*, 47(4):939–956.

Li, F., Villani, M., and Kohn, R. (2011). Modeling conditional densities using finite smooth mixtures. In Mengersen, K., Robert, C., and Titterington, M., editors, *Mixtures: estimation and applications*, pages 123–144. John Wiley & Sons.

McLachlan, G., McGiffin, D., et al. (1994). On the role of finite mixture models in survival analysis. *Statistical methods in medical research*, 3(3):211.

Miller, R., Gong, G., and Muñoz, A. (1981). *Survival analysis*. Wiley New York.

Mosler, K. (2003). Mixture models in econometric duration analysis. *Applied Stochastic Models in Business and Industry*, 19(2):91–104.

Muthén, B. and Masyn, K. (2005). Discrete-time survival mixture analysis. *Journal of Educational and Behavioral Statistics*, 30(1):27–58.

Nott, D. and Leonte, D. (2004). Sampling schemes for bayesian variable selection in generalized linear models. *Journal of Computational and Graphical Statistics*, 13(2):362–382.

Ntzoufras, I., Dellaportas, P., and Forster, J. (2003). Bayesian variable and link determination for generalised linear models. *Journal of statistical planning and inference*, 111(1):165–180.

Qi, Y. and Minka, T. (2002). Hessian-based markov chain monte-carlo algorithms. Manuscript.

Richardson, S. and Green, P. (2002). On bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society: series B (statistical methodology)*, 59(4):731–792.

Shumway, T. (2001). Forecasting bankruptcy more accurately: A simple hazard model*. *The Journal of Business*, 74(1):101–124.

Singer, J. and Willett, J. (1993). It's about time: Using discrete-time survival analysis to study duration and the timing of events. *Journal of Educational and Behavioral Statistics*, 18(2):155–195.

Vaupel, J., Manton, K., and Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16(3):439–454.

Villani, M., Kohn, R., and Giordani, P. (2009). Regression density estimation using smooth adaptive gaussian mixtures. *Journal of Econometrics*, 153(2):155–173.

Villani, M., Kohn, R., and Nott, D. (2012). Generalized smooth finite mixtures. *Journal of Econometrics*, 171(2):121–133.

Xue, X. and Brookmeyer, R. (1997). Regression analysis of discrete time survival data under heterogeneity. *Statistics in medicine*, 16(17):1983–1993.

# Earlier Working Papers:

For a complete list of Working Papers published by Sveriges Riksbank, see www.riksbank.se

| | |
|---|---|
| Estimation of an Adaptive Stock Market Model with Heterogeneous Agents<br>*by Henrik Amilon* | 2005:177 |
| Some Further Evidence on Interest-Rate Smoothing: The Role of Measurement Errors in the Output Gap<br>*by Mikael Apel and Per Jansson* | 2005:178 |
| Bayesian Estimation of an Open Economy DSGE Model with Incomplete Pass-Through<br>by Malin Adolfson, Stefan Laséen, Jesper Lindé and Mattias Villani | 2005:179 |
| Are Constant Interest Rate Forecasts Modest Interventions? Evidence from an Estimated Open Economy DSGE Model of the Euro Area<br>*by Malin Adolfson, Stefan Laséen, Jesper Lindé and Mattias Villani* | 2005:180 |
| Inference in Vector Autoregressive Models with an Informative Prior on the Steady State<br>*by Mattias Villani* | 2005:181 |
| Bank Mergers, Competition and Liquidity<br>*by Elena Carletti, Philipp Hartmann and Giancarlo Spagnolo* | 2005:182 |
| Testing Near-Rationality using Detailed Survey Data<br>*by Michael F. Bryan and Stefan Palmqvist* | 2005:183 |
| Exploring Interactions between Real Activity and the Financial Stance<br>*by Tor Jacobson, Jesper Lindé and Kasper Roszbach* | 2005:184 |
| Two-Sided Network Effects, Bank Interchange Fees, and the Allocation of Fixed Costs<br>*by Mats A. Bergman* | 2005:185 |
| Trade Deficits in the Baltic States: How Long Will the Party Last?<br>*by Rudolfs Bems and Kristian Jönsson* | 2005:186 |
| Real Exchange Rate and Consumption Fluctuations follwing Trade Liberalization<br>*by Kristian Jönsson* | 2005:187 |
| Modern Forecasting Models in Action: Improving Macroeconomic Analyses at Central Banks<br>*by Malin Adolfson, Michael K. Andersson, Jesper Lindé, Mattias Villani and Anders Vredin* | 2005:188 |
| Bayesian Inference of General Linear Restrictions on the Cointegration Space<br>*by Mattias Villani* | 2005:189 |
| Forecasting Performance of an Open Economy Dynamic Stochastic General Equilibrium Model<br>*by Malin Adolfson, Stefan Laséen, Jesper Lindé and Mattias Villani* | 2005:190 |
| Forecast Combination and Model Averaging using Predictive Measures<br>*by Jana Eklund and Sune Karlsson* | 2005:191 |
| Swedish Intervention and the Krona Float, 1993-2002<br>*by Owen F. Humpage and Javiera Ragnartz* | 2006:192 |
| A Simultaneous Model of the Swedish Krona, the US Dollar and the Euro<br>*by Hans Lindblad and Peter Sellin* | 2006:193 |
| Testing Theories of Job Creation: Does Supply Create Its Own Demand?<br>*by Mikael Carlsson, Stefan Eriksson and Nils Gottfries* | 2006:194 |
| Down or Out: Assessing The Welfare Costs of Household Investment Mistakes<br>*by Laurent E. Calvet, John Y. Campbell and Paolo Sodini* | 2006:195 |
| Efficient Bayesian Inference for Multiple Change-Point and Mixture Innovation Models<br>*by Paolo Giordani and Robert Kohn* | 2006:196 |
| Derivation and Estimation of a New Keynesian Phillips Curve in a Small Open Economy<br>*by Karolina Holmberg* | 2006:197 |
| Technology Shocks and the Labour-Input Response: Evidence from Firm-Level Data<br>*by Mikael Carlsson and Jon Smedsaas* | 2006:198 |
| Monetary Policy and Staggered Wage Bargaining when Prices are Sticky<br>*by Mikael Carlsson and Andreas Westermark* | 2006:199 |
| The Swedish External Position and the Krona<br>*by Philip R. Lane* | 2006:200 |

| | |
|---|---|
| Price Setting Transactions and the Role of Denominating Currency in FX Markets<br>*by Richard Friberg and Fredrik Wilander* | 2007:201 |
| The geography of asset holdings: Evidence from Sweden<br>*by Nicolas Coeurdacier and Philippe Martin* | 2007:202 |
| Evaluating An Estimated New Keynesian Small Open Economy Model<br>*by Malin Adolfson, Stefan Laséen, Jesper Lindé and Mattias Villani* | 2007:203 |
| The Use of Cash and the Size of the Shadow Economy in Sweden<br>*by Gabriela Guibourg and Björn Segendorf* | 2007:204 |
| Bank supervision Russian style: Evidence of conflicts between micro- and macro-prudential concerns<br>*by Sophie Claeys and Koen Schoors* | 2007:205 |
| Optimal Monetary Policy under Downward Nominal Wage Rigidity<br>*by Mikael Carlsson and Andreas Westermark* | 2007:206 |
| Financial Structure, Managerial Compensation and Monitoring<br>*by Vittoria Cerasi and Sonja Daltung* | 2007:207 |
| Financial Frictions, Investment and Tobin's q<br>*by Guido Lorenzoni and Karl Walentin* | 2007:208 |
| Sticky Information vs Sticky Prices: A Horse Race in a DSGE Framework<br>*by Mathias Trabandt* | 2007:209 |
| Acquisition versus greenfield: The impact of the mode of foreign bank entry on information and bank lending rates<br>*by Sophie Claeys and Christa Hainz* | 2007:210 |
| Nonparametric Regression Density Estimation Using Smoothly Varying Normal Mixtures<br>*by Mattias Villani, Robert Kohn and Paolo Giordani* | 2007:211 |
| The Costs of Paying – Private and Social Costs of Cash and Card<br>*by Mats Bergman, Gabriella Guibourg and Björn Segendorf* | 2007:212 |
| Using a New Open Economy Macroeconomics model to make real nominal exchange rate forecasts<br>*by Peter Sellin* | 2007:213 |
| Introducing Financial Frictions and Unemployment into a Small Open Economy Model<br>*by Lawrence J. Christiano, Mathias Trabandt and Karl Walentin* | 2007:214 |
| Earnings Inequality and the Equity Premium<br>*by Karl Walentin* | 2007:215 |
| Bayesian forecast combination for VAR models<br>*by Michael K. Andersson and Sune Karlsson* | 2007:216 |
| Do Central Banks React to House Prices?<br>*by Daria Finocchiaro and Virginia Queijo von Heideken* | 2007:217 |
| The Riksbank's Forecasting Performance<br>*by Michael K. Andersson, Gustav Karlsson and Josef Svensson* | 2007:218 |
| Macroeconomic Impact on Expected Default Freqency<br>*by Per Åsberg and Hovick Shahnazarian* | 2008:219 |
| Monetary Policy Regimes and the Volatility of Long-Term Interest Rates<br>*by Virginia Queijo von Heideken* | 2008:220 |
| Governing the Governors: A Clinical Study of Central Banks<br>*by Lars Frisell, Kasper Roszbach and Giancarlo Spagnolo* | 2008:221 |
| The Monetary Policy Decision-Making Process and the Term Structure of Interest Rates<br>*by Hans Dillén* | 2008:222 |
| How Important are Financial Frictions in the U S and the Euro Area<br>*by Virginia Queijo von Heideken* | 2008:223 |
| Block Kalman filtering for large-scale DSGE models<br>*by Ingvar Strid and Karl Walentin* | 2008:224 |
| Optimal Monetary Policy in an Operational Medium-Sized DSGE Model<br>*by Malin Adolfson, Stefan Laséen, Jesper Lindé and Lars E. O. Svensson* | 2008:225 |
| Firm Default and Aggregate Fluctuations<br>*by Tor Jacobson, Rikard Kindell, Jesper Lindé and Kasper Roszbach* | 2008:226 |

| | |
|---|---|
| Re-Evaluating Swedish Membership in EMU: Evidence from an Estimated Model<br>*by Ulf Söderström* | 2008:227 |
| The Effect of Cash Flow on Investment: An Empirical Test of the Balance Sheet Channel<br>*by Ola Melander* | 2009:228 |
| Expectation Driven Business Cycles with Limited Enforcement<br>*by Karl Walentin* | 2009:229 |
| Effects of Organizational Change on Firm Productivity<br>*by Christina Håkanson* | 2009:230 |
| Evaluating Microfoundations for Aggregate Price Rigidities: Evidence from Matched Firm-Level Data on Product Prices and Unit Labor Cost<br>*by Mikael Carlsson and Oskar Nordström Skans* | 2009:231 |
| Monetary Policy Trade-Offs in an Estimated Open-Economy DSGE Model<br>*by Malin Adolfson, Stefan Laséen, Jesper Lindé and Lars E. O. Svensson* | 2009:232 |
| Flexible Modeling of Conditional Distributions Using Smooth Mixtures of Asymmetric Student T Densities<br>*by Feng Li, Mattias Villani and Robert Kohn* | 2009:233 |
| Forecasting Macroeconomic Time Series with Locally Adaptive Signal Extraction<br>*by Paolo Giordani and Mattias Villani* | 2009:234 |
| Evaluating Monetary Policy<br>*by Lars E. O. Svensson* | 2009:235 |
| Risk Premiums and Macroeconomic Dynamics in a Heterogeneous Agent Model<br>*by Ferre De Graeve, Maarten Dossche, Marina Emiris, Henri Sneessens and Raf Wouters* | 2010:236 |
| Picking the Brains of MPC Members<br>*by Mikael Apel, Carl Andreas Claussen and Petra Lennartsdotter* | 2010:237 |
| Involuntary Unemployment and the Business Cycle<br>*by Lawrence J. Christiano, Mathias Trabandt and Karl Walentin* | 2010:238 |
| Housing collateral and the monetary transmission mechanism<br>*by Karl Walentin and Peter Sellin* | 2010:239 |
| The Discursive Dilemma in Monetary Policy<br>*by Carl Andreas Claussen and Øistein Røisland* | 2010:240 |
| Monetary Regime Change and Business Cycles<br>*by Vasco Cúrdia and Daria Finocchiaro* | 2010:241 |
| Bayesian Inference in Structural Second-Price common Value Auctions<br>*by Bertil Wegmann and Mattias Villani* | 2010:242 |
| Equilibrium asset prices and the wealth distribution with inattentive consumers<br>*by Daria Finocchiaro* | 2010:243 |
| Identifying VARs through Heterogeneity: An Application to Bank Runs<br>*by Ferre De Graeve and Alexei Karas* | 2010:244 |
| Modeling Conditional Densities Using Finite Smooth Mixtures<br>*by Feng Li, Mattias Villani and Robert Kohn* | 2010:245 |
| The Output Gap, the Labor Wedge, and the Dynamic Behavior of Hours<br>*by Luca Sala, Ulf Söderström and Antonella Trigari* | 2010:246 |
| Density-Conditional Forecasts in Dynamic Multivariate Models<br>*by Michael K. Andersson, Stefan Palmqvist and Daniel F. Waggoner* | 2010:247 |
| Anticipated Alternative Policy-Rate Paths in Policy Simulations<br>*by Stefan Laséen and Lars E. O. Svensson* | 2010:248 |
| MOSES: Model of Swedish Economic Studies<br>*by Gunnar Bårdsen, Ard den Reijer, Patrik Jonasson and Ragnar Nymoen* | 2011:249 |
| The Effects of Endogenuos Firm Exit on Business Cycle Dynamics and Optimal Fiscal Policy<br>*by Lauri Vilmi* | 2011:250 |
| Parameter Identification in a Estimated New Keynesian Open Economy Model<br>*by Malin Adolfson and Jesper Lindé* | 2011:251 |
| Up for count? Central bank words and financial stress<br>*by Marianna Blix Grimaldi* | 2011:252 |

| | |
|---|---|
| Wage Adjustment and Productivity Shocks<br>*by Mikael Carlsson, Julián Messina and Oskar Nordström Skans* | 2011:253 |
| Stylized (Arte) Facts on Sectoral Inflation<br>*by Ferre De Graeve and Karl Walentin* | 2011:254 |
| Hedging Labor Income Risk<br>*by Sebastien Betermier, Thomas Jansson, Christine A. Parlour and Johan Walden* | 2011:255 |
| Taking the Twists into Account: Predicting Firm Bankruptcy Risk with Splines of Financial Ratios<br>*by Paolo Giordani, Tor Jacobson, Erik von Schedvin and Mattias Villani* | 2011:256 |
| Collateralization, Bank Loan Rates and Monitoring: Evidence from a Natural Experiment<br>*by Geraldo Cerqueiro, Steven Ongena and Kasper Roszbach* | 2012:257 |
| On the Non-Exclusivity of Loan Contracts: An Empirical Investigation<br>*by Hans Degryse, Vasso Ioannidou and Erik von Schedvin* | 2012:258 |
| Labor-Market Frictions and Optimal Inflation<br>*by Mikael Carlsson and Andreas Westermark* | 2012:259 |
| Output Gaps and Robust Monetary Policy Rules<br>*by Roberto M. Billi* | 2012:260 |
| The Information Content of Central Bank Minutes<br>*by Mikael Apel and Marianna Blix Grimaldi* | 2012:261 |
| The Cost of Consumer Payments in Sweden<br>*by Björn Segendorf and Thomas Jansson* | 2012:262 |
| Trade Credit and the Propagation of Corporate Failure: An Empirical Analysis<br>*by Tor Jacobson and Erik von Schedvin* | 2012:263 |
| Structural and Cyclical Forces in the Labor Market During the Great Recession: Cross-Country Evidence<br>*by Luca Sala, Ulf Söderström and AntonellaTrigari* | 2012:264 |
| Pension Wealth and Household Savings in Europe: Evidence from SHARELIFE<br>*by Rob Alessie, Viola Angelini and Peter van Santen* | 2013:265 |
| Long-Term Relationship Bargaining<br>*by Andreas Westermark* | 2013:266 |
| Using Financial Markets To Estimate the Macro Effects of Monetary Policy: An Impact-Identified FAVAR*<br>*by Stefan Pitschner* | 2013:267 |

Sveriges Riksbank
Visiting address: Brunkebergs torg 11
Mail address: se-103 37 Stockholm

Website: www.riksbank.se
Telephone: +46 8 787 00 00, Fax: +46 8 21 05 31
E-mail: registratorn@riksbank.se