Forecasting with a Panel Tobit Model

Laura Liu
Federal Reserve Board

Hyungsik Roger Moon
University of Southern California and Yonsei

Frank Schorfheide*
University of Pennsylvania CEPR, NBER, and PIER

Preliminary Version: September 13, 2018

Abstract

We use a dynamic panel Tobit model to generate point and density forecasts for a large cross-section of short time series of censored observations. Our fully Bayesian approach allows us to flexibly estimate the cross-sectional distribution of heterogeneous coefficients and then implicitly use this distribution to construct Bayes forecasts for the individual time series. We consider versions of the model with homoskedastic and heteroskedastic innovations. In Monte Carlo experiments and the empirical application to loan charge-off rates of small banks, we compare Bayesian point, interval, and density forecasts obtained from various versions of the panel Tobit model. In our empirical application the assumption of homoskedasticity yields more accurate point forecasts, whereas interval and density forecasts are more accurate if computed under the assumption of heteroskedastic innovations. Local house prices and unemployment rates do not appear to be significant determinants of charge-off rates.

JEL CLASSIFICATION: C11, C14, C23, C53, G21

KEY WORDS: Bayesian inference, density forecasts, interval forecasts, loan charge-offs, panel data, point forecasts, Tobit model.

*Correspondence: L. Liu: Board of Governors of the Federal Reserve System, 20th Street and Constitution Avenue N.W., Washington, D.C. 20551. Email: laura.liu@frb.gov. H.R. Moon: Department of Economics, University of Southern California, KAP 300, Los Angeles, CA 90089. E-mail: moonr@usc.edu. F. Schorfheide: Department of Economics, 3718 Locust Walk, University of Pennsylvania, Philadelphia, PA 19104-6297. Email: schorf@ssc.upenn.edu. Moon and Schorfheide gratefully acknowledge financial support from the National Science Foundation under Grants SES 1625586 and SES 1424843, respectively. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Board of Governors or the Federal Reserve System.

1 Introduction

This paper considers the problem of forecasting a large collection of short time series with censored observations. In particular, in our empirical application we forecast charge-off rates on loans for a panel of small banks. The prediction of charge-off rates is interesting from a regulator's perspective because charge-offs generate losses on loan portfolios. If these charge-offs are large, the bank may be entering a period of distress and require additional capital. Due to mergers and acquisitions, changing business models, and changes in regulatory environments the time series dimension that is useful for forecasting is often short. The general methods developed in this paper are not tied to the charge-off rate application and can be used in any setting in which a researcher would like to analyze a panel of censored data with a large cross-sectional and a short time series dimension.

The challenge in forecasting short time series is that they do not contain a lot of information about unit-specific parameters. An accurate estimation of the heterogeneous coefficients helps with an accurate forecast. A natural way of adding information is through the use of prior distributions. For each time series, the prior information can be combined with the unit-specific likelihood function to form a posterior. From a Bayesian perspective the posterior distribution then can be used to derive a forecast that minimizes posterior expected loss. From a frequentist perspective one obtains a forecast that will have some bias, but in a mean-squared-error sense, the introduction of bias might be dominated by a reduction in sampling variance. The key insight in panel data applications is that in the absence of any meaningful subjective prior information, one can extract information from the cross section and equate the prior distribution with the cross-sectional distribution of unit-specific coefficients.

There are several ways of implementing this basic idea. An empirical Bayes implementation would create a point estimate of the cross-sectional distribution of the heterogeneous coefficients and then condition the subsequent posterior calculations on the estimated prior distribution.[1] In fact, the classic James-Stein estimator for a vector of means can be interpreted as an estimator constructed as follows. In the first step a prior is generated by fitting a normal distribution to a cross-section of observations. In a second step, this prior is then combined with the unit-specific likelihood function to generate a posterior estimate of the unknown mean for that unit. In a panel setting the implementation is more involved but

[1]Empirical Bayes methods have a long history in the statistics literature going back to Robbins (1956); see Robert (1994) for a textbook treatment.

follows the same steps. If the model is linear in the coefficients and the forecast is evaluated under a quadratic forecast error loss function, then Tweedie's formula, which expresses the posterior mean of the heterogeneous coefficients as the maximum likelihood estimate corrected by a function of the cross-sectional density of sufficient statistic, can be used to construct a forecast without having to explicitly estimate a prior distribution for the heterogeneous coefficients. This insight has been recently used by Brown and Greenshtein (2009), Gu and Koenker (2016a,b), and Liu, Moon, and Schorfheide (2017).

Unfortunately, Tweedie's formula does not extend to nonlinear panel data models. Rather than pursuing an Empirical Bayes strategy, we will engage in a full Bayesian analysis by specifying a hyperprior for the distribution of heterogeneous coefficients and then constructing a joint posterior for the coefficients of this hyperprior as well as the actual unit-specific coefficients. While the computations are more involved, this approach can in principle handle quite general nonlinearities. Moreover, it is possible to generate point predictions under more general loss functions, as well as interval and density forecasts. For a linear panel data model, a full Bayesian analysis is implemented by Liu (2017). The contribution of our paper is to extend the implementation to the dynamic panel Tobit model and to apply it to the problem of forecasting loan charge-off rates. We compare Bayesian point, interval, and density forecasts obtained from various versions of the panel Tobit model. In our empirical application the assumption of homoskedasticity yields more accurate point forecasts, whereas interval and density forecasts are more accurate if computed under the assumption of heteroskedastic innovations. Local house prices and unemployment rates do not appear to be significant determinants of charge-off rates.

Our paper relates to several branches of the literature. The papers most closely related are Gu and Koenker (2016a,b), Liu (2017), and Liu, Moon, and Schorfheide (2017). All four of these papers focus on the estimation of the heterogeneous coefficients in linear panel data models and the subsequent use of the estimated coefficients for the purpose of prediction. Only the full Bayesian analysis in Liu (2017) has a natural extension to nonlinear models. Liu, Moon, and Schorfheide (2017), building on Brown and Greenshtein (2009), show that an empirical Bayes implementation based on Tweedie's formula can asymptotically (as the cross-sectional dimension tends to infinity) lead to forecasts that are as accurate as the so-called oracle forecasts. Here an oracle forecast is an infeasible benchmark that assumes that all homogeneous coefficients, as well as the distribution of the heterogeneous coefficients, are known to the forecaster. Liu (2017) shows that the predictive density obtained from the full Bayesian analysis converges strongly to the oracle's predictive density as the cross-section

gets large.

There is a Bayesian literature on the estimation of censored regression models. The idea of using data augmentation and Gibbs sampling to estimate a Tobit model dates back to Chib (1992). To sample the latent uncensored observations we rely on an algorithm tailored toward dynamic Tobit models by Wei (1999). Sampling from truncated normal distributions is implemented with a recent algorithm of Botev (2017). A broader survey of the literature on Bayesian estimation of univariate and multivariate censored regression models can be found in the handbook chapter by Li and Tobias (2011).

We consider two approaches of modeling the unknown distribution of the heterogeneous coefficients. First, we assume that this distribution belongs to the class of Normal distributions (or Normal and Inverse Gamma if we allow for heteroskedasticity). Second, we consider a more flexible setup in which the distribution is represented by Dirichlet process mixtures of Normals. Even though we do not emphasize the nonparametric aspect of this modeling approach (due to a truncation, our mixtures are strictly speaking finite and in that sense parametric), our paper is related to the literature on flexible / nonparametric density modeling using Dirichlet process mixtures (DPM). Examples of papers that use DPMs in the panel data context are Hirano (2002), Burda and Harding (2013), Rossi (2014), and Jensen, Fisher, and Tkac (2015). The implementation of our Gibbs sampler relies on Ishwaran and James (2001, 2002).

Our analysis raises the question to what extent the cross-sectional distribution of the heterogeneous coefficients is identifiable. Even though identification is not a necessary prerequisite for a Bayesian analysis as discussed, for instance in Moon and Schorfheide (2012) and references cited therein, it is important to understand in what directions of the parameter space one can expect to learn from the data. There exists a large literature on identification in nonlinear panel data models. Earlier censored panel regression analysis mostly focused on the established identification and estimating the homogeneous regression coefficients under fixed effects. Examples of these studies include Honoré (1992) (static model), Hu (2002) (dynamic model), and Abrevaya (2000) (generalized fixed effect model), Honore and Hu (2004) (endogenous censored panel regression), and Arellano and Honoré (2001) (survey on panel Tobit type regressions).

More recent papers investigated more general nonlinear and non-separable panel regression models. Here the conventional censored panel regression is a special case and the literature established identification of various parameters of interests such as regression co-

efficients, marginal effects, average structural function, and average treatment effects (e.g., Bester and Hansen (2009), Altonji and Matzkin (2005), Bonhomme (2012), and Hoderlein and White (2012). The paper most relevant to our analysis is that of Hu and Shiu (2018). They obtain identification results for censored dynamic panel data models with parametric specifications of the dynamics, censoring, and distribution of the shocks, but a nonparametric specification of the distribution of the unobserved individual effects. This covers the benchmark model considered in Section 2.

The remainder of this paper is organized as follows. The basic dynamic panel Tobit model with homoskedastic shocks is presented in Section 2. We discuss its specification and develop a sampler for posterior inference. This model is extended to allow for cross-sectional heteroskedasticity in Section 3 and further s are sketched out in Section 4. We conduct a Monte Carlo experiment in Section 5 to examine the performance of the proposed techniques in a controlled environment. An empirical application in which we forecast charge-off rates on various types of loans for a panel of banks is presented in Section 6. Finally, Section 7 concludes. A description of the data sets, additional empirical results, and some derivations are relegated to the Online Appendix.

2 The Basic Model

To fix ideas, we consider the following dynamic panel Tobit model with heterogeneous intercepts:

$$\begin{aligned} y_{it} &= y_{it^*}\mathbb{I}\{y_{it}^* \geq 0\}, \quad i = 1, \ldots, N, \quad t = 0, \ldots, T \\ y_{it}^* &= \lambda_i + \rho y_{it-1}^* + u_{it}, \quad u_{it} \overset{iid}{\sim} N(0, \sigma^2), \quad y_{i,0}^* \sim N(\mu_{i*}, \sigma_*^2), \end{aligned} \tag{1}$$

where $\mathbb{I}\{x \geq a\}$ is the indicator function that is equal to one if $x \geq a$ and equal to zero otherwise. Let $\theta = [\rho, \sigma^2]'$. The parameters that characterize the distribution of $y_{i,0}^*$ could either be a function of (θ, λ_i), e.g., $\mu_{i*} = \lambda_i/(1-\rho)$ and $\sigma_*^2 = \sigma^2/(1-\rho^2)$ or they could be fixed, e.g., $\mu_{i*} = 0$ and $\sigma_*^2 = 1$. We assume that conditional on the parameters the observations are cross-sectionally independent and that the heterogeneous intercepts λ_i are distributed according to an unknown random effects distribution

$$\lambda_i \overset{iid}{\sim} \pi(\lambda). \tag{2}$$

The density $\pi(\cdot)$ can be viewed as an unknown, possibly infinite-dimensional parameter. Our goal is to generate forecasts of y_{iT+h} conditional on the observations

$$Y_{1:N,0:T} = \big\{(y_{10}, \ldots, y_{N0}), \ldots, (y_{1T}, \ldots, y_{NT})\big\}.$$

A natural benchmark against which we can compare our forecasts is the so-called oracle forecast that is obtained under the assumption that θ and $\pi(\cdot)$ are known but λ_i is unknown. Due to the cross-sectional independence, the oracle forecast can be derived from the posterior predictive distribution of y_{iT+h} given $\big(Y_{i,0:T}, \theta, \pi(\cdot)\big)$. Define the posterior distribution of λ_i as

$$p(\lambda_i|Y_{i,0:T}, \theta, \pi) \propto p(Y_{i,1:T}|y_{i,0}, \theta, \lambda_i)\pi(\lambda_i), \tag{3}$$

where $p(Y_{i,1:T}|y_{i,0}, \theta, \lambda_i)$ is the conditional likelihood function associated with (1). Then, the posterior predictive distribution is given by

$$\begin{aligned} &p(y_{iT+h}|Y_{i,0:T}, \theta, \pi) \\ &= \int p(y_{iT+h}|y^*_{iT+h})p(y^*_{iT+h}|y^*_{iT}, \theta, \lambda_i)p(y^*_{iT}|Y_{i,0:T}, \theta, \lambda_i)p(\lambda_i|Y_{i,0:T}, \theta, \pi)dy^*_{iT}d\lambda_i. \end{aligned} \tag{4}$$

The distribution of $y_{iT+h}|y^*_{iT+h}$ is a unit point mass that is located at zero if $y^*_{iT+h} \leq 0$ or at y^*_{iT+h} if $y^*_{iT+h} > 0$. The density $p(y^*_{iT+h}|y^*_{iT}, \theta, \lambda_i)$ is generated by the forward simulation of the autoregressive law of motion for y^*_{it} in (1) and the remaining two densities are posteriors for y^*_{iT} and λ_i. In Section 2.1 we discuss the computation of the oracle predictor. Sections 2.2 and 2.3 focus on the estimation of θ and π, respectively. With the exception of the treatment of the latent variables $Y^*_{i,0:T}$, the computations for the Tobit model are very similar to the ones for the linear model studied in Liu (2017).

2.1 Oracle Forecast

In order to implement the oracle forecasts we will use a data augmentation algorithm (see Tanner and Wong (1987)) to generate draws from the posterior predictive distribution in (4). The implementation of this algorithm is based on Wei (1999) who studied the Bayesian estimation of univariate dynamic Tobit models. For each unit i, the data augmentation algorithm iterates over the following two conditional posteriors:

$$Y^*_{i,0:T}|(Y_{i,0:T}, \theta, \lambda_i) \quad \text{and} \quad \lambda_i|(Y_{i,0:T}, Y^*_{i,0:T}, \theta, \pi), \tag{5}$$

thereby generating draws from the joint distribution of $(Y^*_{i,0:T}, \lambda_i)|(Y_{i,0:T}, Y^*_{i,0:T}, \theta, \pi)$ that can be used to approximate the oracle predictor in equation (4).

Drawing from $Y^*_{i,0:T}|(Y_{i,0:T}, \theta, \lambda_i)$. To fix ideas, consider the following sequence of observations $y_{i0}, \ldots, y_{iT}$:

$$y^*_{i0},\ y^*_{i1},\ 0,\ 0,\ 0,\ y^*_{i5},\ y^*_{i6},\ 0,\ 0,\ 0,\ y^*_{i10}.$$

Our model implies that whenever $y_{it} > 0$ we can deduce that $y^*_{it} = y_{it}$. Thus, we can focus our attention on periods in which $y_{it} = 0$. In the hypothetical sample we observe two strings of censored observations: (y_{i2}, y_{i3}, y_{i4}) and (y_{i7}, y_{i8}, y_{i9}). We use t_1 for the start date of a string of censored observations and t_2 for the end date. In the example we have two such strings, we write $t^{(1)}_1 = 2$, $t^{(1)}_2 = 4$, $t^{(2)}_1 = 7$, $t^{(2)}_2 = 9$. The goal is to characterize $p(Y^*_{i,t^{(1)}_1:t^{(1)}_2}, Y^*_{i,t^{(2)}_1:t^{(2)}_2}|Y_{i,0:T}, \theta, \lambda_i)$. Because of the AR(1) structure, observations in periods $t < t_1 - 1$ and $t > t_2 + 1$ contain no additional information about $y^*_{it_1}, \ldots, y^*_{it_2}$. Thus, we obtain

$$\begin{aligned} & p(Y^*_{i,t^{(1)}_1:t^{(1)}_2}, Y^*_{i,t^{(2)}_1:t^{(2)}_2}|Y_{i,0:T}, \theta, \lambda_i) \\ &= \ p(Y^*_{i,t^{(1)}_1:t^{(1)}_2}|Y_{i,t^{(1)}_1-1:t^{(1)}_2+1}, \theta, \lambda_i) p(Y^*_{i,t^{(2)}_1:t^{(2)}_2}|Y_{i,t^{(2)}_1-1:t^{(2)}_2+1}, \theta, \lambda_i), \end{aligned}$$

which implies that we can sample each string of latent observations independently.

Let $s = t_2 - t_1 + 2$ be the length of the segment that includes the string of censored observations as well as the adjacent uncensored observations. Iterating the AR(1) law of motion for y_{it} forward from period $t = t_1 - 1$ we deduce that the vector of random variables $[Y^*_{i,t_1:t_2}, y_{it_2+1}]'$ conditional on y_{it_1-1} is multivariate normal with mean

$$M_{1:s|0} = [\mu_1, \ldots, \mu_s]', \quad \mu_1 = \lambda + \rho y_{it_1-1}, \quad \mu_\tau = \lambda + \rho \mu_{\tau-1} \text{ for } \tau = 2, \ldots, s. \tag{6}$$

The covariance matrix takes the form

$$\Sigma_{1:s|0} = \sigma^2 \begin{bmatrix} \rho_{1,1|0} & \cdots & \rho_{1,s|0} \\ \vdots & \ddots & \vdots \\ \rho_{s,1|0} & \cdots & \rho_{s,s|0} \end{bmatrix}, \quad \rho_{i,j|0} = \rho_{j,i|0} = \rho^{j-i} \sum_{l=0}^{i-1} \rho^{2l} \text{ for } j \geq i. \tag{7}$$

We can now use the formula for the conditional mean and variance of a multivariate normal

distribution

$$\begin{aligned} M_{1:s-1|0,s} &= M_{1:s-1|0} - \Sigma_{1:s-1,s|0}\Sigma_{ss|0}^{-1}(y_{it_2+1} - \mu_s) \\ \Sigma_{1:s-1,1:s-1|0,s} &= \Sigma_{1:s-1,1:s-1|0} - \Sigma_{1:s-1,s|0}\Sigma_{ss|0}^{-1}\Sigma_{s,1:s-1|0} \end{aligned} \tag{8}$$

to deduce that

$$Y_{i,t_1:t_2}^* \sim TN_-\big(M_{1:s-1|0,s}, \Sigma_{1:s-1,1:s-1|0,s}\big). \tag{9}$$

Here we use $TN_-(\mu, \Sigma)$ to denote a normal distribution that is truncated to satisfy $y \leq 0$. Draws from this truncated normal distribution can be efficiently generated using the algorithm recently proposed by Botev (2017).

There are two important special cases. First, suppose that $t_2 = T$, meaning that the last observation in the sample is censored. Then the mean vector and the covariance matrix of the truncated normal distribution are given by (6) and (7) with the understanding that $s = t_2 - t_1 + 1$. Second, suppose that $t_1 = 0$, meaning that the initial observation in the sample $y_{i0} = 0$. Because in this case the observation $y_{it_1-1} = y_{i,-1}$ is missing, we need to modify the expressions in (6) and (7). According to (1), $y_0^* \sim N(\mu_*, \sigma_*^2)$. This leads to the mean vector

$$M_{1:s} = [\mu_1, \ldots, \mu_s], \quad \mu_1 = \mu_*, \quad \mu_\tau = \lambda + \rho\mu_{\tau-1} \text{ for } \tau = 2, \ldots, s. \tag{10}$$

and the covariance matrix

$$\Sigma_{1:s} = \sigma^2 \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & \rho_{1,1} & \cdots & \rho_{1,s-1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \rho_{s-1,1} & \cdots & \rho_{s-1,s-1} \end{bmatrix} + \sigma_*^2 \begin{bmatrix} \rho^{0+0} & \cdots & \rho^{0+(s-1)} \\ \vdots & \ddots & \vdots \\ \rho^{(s-1)+0} & \cdots & \rho^{(s-1)+(s-1)} \end{bmatrix}, \tag{11}$$

where the definition of $\rho_{i,j}$ is identical to the definition of $\rho_{i,j|0}$ in (7). One can then use the formulas in (8) to obtain the mean and covariance parameters of the truncated normal distribution.

Drawing from $\lambda_i|(Y_{i,0:T}, Y_{i,0:T}^*, \theta, \pi)$. The posterior inference with respect to λ_i becomes "standard" once we condition on the latent variables $Y_{i,0:T}^*$. It is based on the Gaussian location-shift model

$$y_{it}^* - \rho y_{it-1}^* = \lambda_i + u_{it}, \quad u_{it} \overset{iid}{\sim} N(0, \sigma^2). \tag{12}$$

If the prior distribution $\pi(\lambda)$ is Gaussian, the posterior of λ will be Gaussian and direct sampling from $\lambda_i|(Y_{i,0:T}, Y^*_{i,0:T}, \theta, \pi)$ is straightforward. If $\pi(\lambda)$ is non-Gaussian, a draw could be generated with a Metropolis-Hastings (MH) step.

Drawing from the predictive density. Conditional on $(y^*_{iT}, \theta, \lambda_i)$ the predictive distribution of y_{iT+h} is given by a censored (at zero) normal distribution (CN_+):

$$y_{iT+h}|(y^*_{iT}, \theta, \lambda_i) \sim CN_+\left(\lambda_i \sum_{l=0}^{h-1} \rho^l + \rho^h y^*_{iT},\ \sigma^2 \sum_{l=0}^{h-1} \rho^{2l}\right). \tag{13}$$

Draws from this distribution can be generated by direct sampling and analytical formula for the mean and the variance are available (see Greene (2008) and the Online Appendix).

2.2 Estimating θ

We now consider the conditional posterior distribution $\theta|(Y_{i,0:T}, Y^*_{i,0:T}, \lambda_i, \pi)$. Note that conditional on the latent observations $Y^*_{i,0:T}$ the actual observations $Y_{i,0:T}$ contain no additional information. Moreover, conditional on λ_i the density $\pi(\lambda)$ is irrelevant for inference on θ. We can rewrite the law-of-motion for y^*_{it} as

$$y^*_{it} - \lambda_i = \rho y^*_{it-1} + u_{it}, \quad u_{it} \overset{iid}{\sim} N(0, \sigma^2), \quad i = 1, \ldots, N \text{ and } t = 1, \ldots, T. \tag{14}$$

Conditional on the λ_i's we have a Gaussian linear regression model with unknown slope coefficient ρ and unknown variance σ^2. The temporal and spatial independence of the u_{it}'s allows us to pool observations across i and t. If we use a conjugate Normal-Inverse-Gamma (NIG) prior, then the conditional posterior $\theta|(Y^*_{i,0:T}, \lambda_i)$ also belongs to the NIG family.

2.3 Estimating $\pi(\lambda)$

To model the unknown random effects distribution $\pi(\lambda)$ we consider a tightly parameterized approach under which $\pi(\lambda)$ is assumed to be Gaussian with unknown scale and location parameter and a more flexible approach in which we use a Dirichlet Process Mixture (DPM) to generate a prior distribution for $\pi(\lambda)$.

Parametric treatment of $\pi(\lambda)$. We assume that π belongs to the family of normal distributions and write $p_N(x|\mu, \sigma^2)$ to denote the density of a $N(\mu, \sigma^2)$. Using this notation,

let

$$p(\lambda|\pi) = p_N\big(\lambda|\mu_\lambda, \omega_\lambda^2\big). \tag{15}$$

Thus, here the distribution $\pi(\lambda)$ is represented by the two-dimensional parameter vector $[\mu_\lambda, \omega_\lambda^2]'$. To induce a prior distribution for the hyperparameters $(\mu_\lambda, \omega_\lambda^2)$ we assume that

$$p\big(\omega_\lambda^2\big) = p_{IG}\big(\omega_\lambda^2|2,2\big), \quad p\big(\mu_\lambda|\omega_\lambda^2\big) = p_N\big(\mu_\lambda|0,\omega_\lambda^2\big), \tag{16}$$

where $p_{IG}(x|a,b)$ is the density of an Inverse Gamma $IG(2,2)$ distribution.

Recall from (12) that the likelihood function for λ_i obtained from the latent variables $Y^*_{i,0:T}$ is Gaussian. Thus, we deduce that the conditional posterior distribution

$$p(\lambda_i|Y_{i,0:T}, Y^*_{i,0:T}, \theta, \pi) \propto p(Y_{i,0:T}|Y^*_{i,0:T})p(Y^*_{i,0:T}|\lambda_i, \theta)p(\lambda_i|\pi) \tag{17}$$

is also Normal. The conditional posterior distribution for the parameters of the random-effects distribution can be characterized as follows:

$$p(\mu_\lambda, \omega_\lambda^2|Y_{1:N,0:T}, Y^*_{1:N,0:T}, \lambda_{1:N}, \theta) \propto p(\mu_\lambda, \omega_\lambda^2)\prod_{i=1}^{N} p(\lambda_i|\mu_\lambda, \omega_\lambda^2). \tag{18}$$

Here the densities $p(Y^*_{i,0:T}|\lambda_i, \theta)$ are absorbed in the constant of proportionality because they do not depend on $(\mu_\lambda, \omega_\lambda^2)$.

Flexible treatment of $\pi(\lambda)$. For a nonparametric treatment of $\pi(\lambda)$ we are essentially replacing the Gaussian density in (15) with a mixture of normals. This implies that in addition to specifying priors for the means and variances of the normal distributions as in (16), we need to specify the number of mixture components and probabilities associated with the mixture components. This leads to a truncated stick breaking (TSB) representation of a DPM prior.

We fix the maximum number of mixture components of $\pi(\lambda)$ at a large number K.[2] Each component k, $k = 1, \ldots, K$ corresponds to a $N(\mu_k, \omega_k^2)$ distribution. The probability assigned to component k is denoted by π_k and normalized such that $\sum_{k=1}^K \pi_k = 1$. Then,

$$p(\lambda|\pi) = \sum_{k=1}^{K} \pi_k p_N\big(\lambda|\mu_k, \omega_k^2\big). \tag{19}$$

[2]In our simulations we choose $K = 20$. This leads to the following uniform bound on the approximation error (see Theorem 2 of Ishwaran and James (2001)): $\|f^{\lambda,K} - f^\lambda\| \sim 4N\exp[-(K-1)/\alpha] \leq 2.24 \times 10^{-5}$, at the prior mean of α ($\bar{\alpha} = 1$) and a cross-sectional sample size $N = 1000$.

To induce a prior distribution over the random effects distribution, we need priors for the means and variances of the mixture components, extending (16) above:

$$p\big(\omega_k^2\big) = p_{IG}\big(\omega_k^2\big|2,2\big), \quad p\big(\mu_k|\omega_k^2\big) = p_N\big(\mu_k|0,\omega_k^2\big), \quad k = 1,\ldots,K. \tag{20}$$

The prior for the mixture probabilities π_k is generated by a $TSB(1,\alpha,K)$ process of the form

$$\pi_k|(\alpha,K) \sim \begin{cases} \zeta_1, & k=1, \\ \prod_{j=1}^{k-1}(1-\zeta_j)\,\zeta_k, & k=2,\ldots,K-1,, \\ 1-\sum_{j=1}^{K-1} p_j, & k=K, \end{cases} \quad \zeta_k \sim B(1,\alpha) \tag{21}$$

where $B(\alpha,\beta)$ is the Beta distribution. Finally, we use an IG prior for the hyperparameter α of the TSB:

$$p(\alpha) = p_{IG}(\alpha|2,2). \tag{22}$$

In order to characterize the posterior distribution, we introduce some additional notation. Because $\pi(\lambda)$ is a mixture, *ex post*, each λ_i is associated with one of the K mixture components. Let $\gamma_i \in \{1,\ldots,K\}$ be the component affiliation of unit i. Moreover, let n_k be the number of units and J_k the set of units affiliated with component k. Using the component indicator γ_i we can rewrite (17) as

$$p(\lambda_i|Y_{i,0:T},Y^*_{i,0:T},\theta,\pi,\gamma_i) \propto p(Y_{i,0:T}|Y^*_{i,0:T})p(Y^*_{i,0:T}|\lambda_i,\theta)\left[\sum_{k=1}^{K}\mathbb{I}\{\gamma_i=k\}p_N(\lambda_i|\mu_k,\omega_k^2)\right], \tag{23}$$

where $\mathbb{I}\{\gamma_i = k\}$ is the indicator function that is one if $\gamma_i = k$ and zero otherwise. Thus, conditional on γ_i the posterior of λ_i remains Normal. The analogue of (18) becomes:

$$p(\mu_k,\omega_k^2|Y_{1:N,0:T},Y^*_{1:N,0:T},\lambda_{1:N},\theta) \propto p(\mu_k,\omega_k^2)\prod_{i\in J_k} p(\lambda_i|\mu_k,\omega_k^2). \tag{24}$$

The prior probability that unit i is affiliated with component k is given by π_k. Let $\bar{\pi}_{ik}$ denote the posterior probability conditional on the set of means $\mu_{1:k}$ and variances $\omega^2_{1:K}$ as well as λ_i. The $\bar{\pi}_{ik}$'s are given by

$$\bar{\pi}_{ik} = \frac{\pi_k p_N\big(\lambda_i|\mu_k,\omega_k^2\big)}{\sum_{k=1}^{K}\pi_k p_N\big(\lambda_i|\mu_k,\omega_k^2\big)}. \tag{25}$$

Thus,

$$\gamma_i|(\mu_{1:k}, \omega^2_{1:K}, \lambda_i) = k \text{ with prob. } \bar{\pi}_{ik}. \tag{26}$$

Based on $\gamma_{1:N}$ one can determine n_k and J_k for $k = 1, \ldots, N$. The conditional posterior of the component probabilities takes the form

$$\pi_k|(n_{1:K}, \alpha, K) \sim TSB\left(1 + n_k, \alpha + \sum_{j=k+1} n_j, K\right), \tag{27}$$

meaning that the ζ_k's in (21) have a $B\big(1 + n_k, \alpha + \sum_{j=k+1} n_j\big)$ distribution. Conditional on $\pi_{1:K}$ the hyperparameter α has a Gamma (G) posterior distribution of the form

$$\alpha|\pi_{1:K} \sim G(2 + K - 1, 2 - \ln \pi_K). \tag{28}$$

Gibbs Sampling. A Gibbs sampler can be obtained by sampling from the various conditional posterior distributions above. The sampler for the flexible mixture representation of $\pi(\lambda)$ is based on Ishwaran and James (2001, 2002). The Gibbs sampler iterates over the following posterior distributions: the latent variables $Y^*_{i,0:T}$ in (9), the homogeneous parameters θ in Section 2.2, the heterogeneous parameters λ_i in (23), the scale parameter α of the Dirichlet distribution in (28), the mixture weights π_k in (27), the mixture component parameters (μ_k, ω^2_k) in (24), and the mixture component memberships γ_i in (26). The Gibbs sampler for the model with Gaussian $\pi(\lambda)$ omits the steps to draw α, π_k, and γ_i. Moreover, it only has to generate a single pair (μ, ω^2).

3 A Model With Heteroskedasticity

In the benchmark model (1) we assumed that the innovations u_{it} are homoskedastic. We now generalize this assumption to allow for heteroskedasticity:

$$u_{it} \overset{iid}{\sim} N(0, \sigma^2_i). \tag{29}$$

Now σ^2_i is an additional heterogeneous parameter. We will specify a prior distribution according to which λ_i and σ^2_i are independent of each other, i.e., $\pi(\lambda, \sigma^2) = \pi(\lambda)\pi(\sigma^2)$. We consider a parametric treatment of $\pi(\sigma^2)$ that can be combined with the Gaussian prior for λ and a more flexible class of priors for σ^2 that we combine with the mixture prior for λ.

Parametric treatment of $\pi(\sigma^2)$. Here we assume that $\pi(\sigma^2)$ belongs to the Inverse Gamma (IG) family and is indexed by the hyperparameters a and b: $p_{IG}(\sigma^2|a,b)$. In specifying the prior for the hyperparameters, we follow Llera and Beckmann (2016) and let

$$b \sim G(\underline{\alpha}_b, \underline{\beta}_b), \quad p(a|b) \propto \frac{\underline{\alpha}_a^{-(1+a)} b^{a\underline{\gamma}_a}}{\Gamma(a)^{\underline{\beta}_a}}. \tag{30}$$

The parameters $(\underline{\alpha}_a, \underline{\beta}_a, \underline{\gamma}_a, \underline{\alpha}_b, \underline{\beta}_b)$ need to be chosen by the researcher. We use $\underline{\alpha}_a = 1$, $\underline{\beta}_a = \underline{\gamma}_a = \underline{\alpha}_b = \underline{\beta}_b = 0.01$, which specifies relatively uninformative priors for hyperparameters a and b. Suppose we have a sequence of observations σ_i^2, $i = 1, \ldots, N$. Then the joint posterior for (a, b) takes the form

$$\begin{aligned} p(a,b|\sigma_{1:N}^2) &\propto p_{IG}(\sigma^2|a,b) p_G(a_{b0}, b_{b0}) p(a|b) \\ &\propto \left(\frac{b^a}{\Gamma(a)}\right)^N \left(\prod_{i=1}^N \left(\sigma_i^2\right)^{-(a+1)}\right) \exp\left\{-b\sum_{i=1}^N 1/\sigma_i^2\right\} \\ &\quad \times b^{\underline{\alpha}_b - 1} \exp\left\{-\underline{\beta}_b b\right\} \frac{\underline{\alpha}_a^{-(1+a)} b^{a\underline{\gamma}_a}}{\Gamma(a)^{\underline{\beta}_a}} \end{aligned} \tag{31}$$

Thus, the conditional posterior of $b|a$ takes the form of a Gamma $G(\bar{\alpha}_b, \bar{\beta}_b)$ distribution with parameters $\bar{\alpha}_b = \underline{\alpha}_b + Na$ and $\bar{\beta}_b = \underline{\beta}_b + \sum_{i=1}^N 1/\sigma_i^2$. The conditional posterior of $a|b$ has the same kernel as the prior with parameters $\ln \bar{\alpha}_a = \ln \underline{\alpha}_a + \sum_{i=1}^N \ln \sigma_i^2$, $\bar{\beta}_a = \underline{\beta}_a + N$, and $\bar{\gamma}_a = \underline{\gamma}_a + N$. Because it is not possible to directly sample from this distribution, we use a Metropolis-within-Gibbs sampler with a random-walk proposal density to draw from $p(a,b|\sigma_{1:N}^2)$.[3] Conditional on the hyperparameters (a,b) it is straightforward to generate draws from the posterior of σ_i^2 by directly sampling from the appropriate Inverse Gamma distributions.

Flexible treatment of $\pi(\sigma^2)$. We use the same approach as for $\pi(\lambda)$, described in Section 2.3. A DPM of normal distributions is used to generate a prior for $\ln \sigma^2$. The conditional posteriors of hyperparameters are essentially identical to the ones that arose in the context of $\pi(\lambda)$. Given mixture component k, the conditional posterior distribution of σ_i^2 is proportional to

$$p(\sigma_i^2|\cdot) \propto (\sigma_i^2)^{-1} p_N(\ln \sigma_i^2 | \mu_{\sigma^2,k}, \omega_{\sigma^2 k}) \prod_{t=1}^T p_N(y_{it}|\rho y_{it-1}^* + \lambda_i, \sigma_i^2). \tag{32}$$

We sample from this non-standard posterior via an adaptive random walk Metropolis-

[3] We use an adaptive procedure based on Atchadé and Rosenthal (2005) and Griffin (2016), which adaptively adjusts the random walk step size to keep acceptance rates around 30%.

Hastings (RWMH) step.

4 Generalizations

The basic dynamic panel Tobit model given by (1) and (29) can be generalized in several dimensions. First, one can include additional covariates with a homogeneous coefficient, which increases the dimension of the vector θ. This is done in the empirical analysis in Section 6; see (34). Second, it is fairly straightforward to allow for randomly missing observations by modifying the inference about the latent variables y_{it^*} in Section 2.1. Rather than drawing the latent variables in the data-augmentation algorithm from a truncated normal distribution, we need to draw them from a regular normal distribution. Third, we can allow for correlated random effects by replacing the unconditional distribution $\pi(\lambda)$ with a conditional distribution $\pi(\lambda)$. Liu (2017) discusses both the Normal as well as the flexible implementation of a correlated random effects specification. It requires an adjustment to the prior distribution of the λ_i's and an adjustment in the part of the Gibbs sampler that generates draws from the hyperparameters that index the random effects distribution.

Fourth, the panel setup can be extended to richer limited-dependent variable models. Let $Y_{it} = [y_{1,it}, \ldots, y_{K,it}]'$ and $Y_{it}^* = [y_{1,it}^*, \ldots, y_{K,it}^*]'$ and consider

$$y_{it} = m(Y_{it}^*), \quad Y_{it}^* | (Y_{it-1}^*, \lambda_i, \theta) \sim p(Y_{it}^* | Y_{it-1}^*, \lambda_i, \theta), \quad (\lambda_i, Y_{i0}^*) | \theta \sim \pi(\lambda_i, Y_{i0}^* | \theta), \tag{33}$$

where $m(\cdot)$ is a known function, $p(\cdot)$ is a known homogeneous transition density for Y_{it}^*, and $\pi(\cdot)$ is the correlated random-effects distribution. In the benchmark model (1) the dependent variable is a scalar, i.e., $K = 1$, the transformation of the latent variable is given by $m(y_{it}^*) = y_{it}^* \mathbb{I}\{y_{it}^* \geq 0\}$, and the transition density is $N(\lambda_i + \rho y_{it}^*, \sigma^2)$. In addition to this standard Tobit model, Amemiya (1985) defines four generalizations. For instance, in the so-called Type 2 Tobit model $K = 2$ and the $m(\cdot)$ function takes the form

$$m(\cdot): \quad y_{1,it} = \mathbb{I}\{y_{1,it}^* \geq 0\}, \quad y_{2,it} = y_{2,it}^* \mathbb{I}\{y_{1,it}^* \geq 0\},$$

in which the censoring of observation $y_{2,it}$ depends on the observed sign of the latent variable $y_{1,it}^*$. In order to implement richer Tobit models in our dynamic panel framework one has to modify the sampler for the conditional posterior distribution of the latent variables $Y_{i,0:T}^* | (Y_{i,0:T}, \lambda_i, \theta)$ described in Section 2.1. For instance, a posterior sampler for the (static)

Type 2 Tobit model is discussed in Li and Tobias (2011). The current version of this paper does not spell out any of these generalizations in detail, except that we allow for additional covariates in our empirical application.

5 Monte Carlo Experiment

The Monte Carlo experiment is based on the basic dynamic panel Tobit model in (1). The design of the experiment is summarized in Table 1. We set the autocorrelation parameter to $\rho = 0.8$ and normalize the innovation variance of the u_{it}'s to $\sigma^2 = 1$. We consider four distributions for the heterogeneous intercept λ_i: a normal distribution as well as three mixtures of normals that exhibit skewness, kurtosis, and bi-modality, respectively. The distributions are parameterized such that $\mathbb{E}[\lambda_i] = 1/2$ and $\mathbb{V}[\lambda_i] = 1$. The simulated panel data sets consist of $N = 1,000$ cross-sectional units and the number of time periods in the estimation sample is $T = 10$. We generate one-step-ahead forecasts for period $t = T + 1$.

Measures of forecast accuracy are computed as follows. For the evaluation of point forecasts (and estimators) based on forecast (estimation) error statistics (i.e, RMSE, bias, and standard deviation), we pool all forecast errors across $i = 1, \ldots, N = 1,000$ and across the $n_{sim} = 100$ Monte Carlo samples $Y_{1:N,1:T+1}$, which means we are averaging across 10^5 errors. For the evaluation of interval forecasts, density forecasts, and the probability of $Y_{iT+1} = 0$, we compute the performance statistic based on the cross-section $i = 1, \ldots, N = 1,000$ and then average the performance statistics over the n_{sim} Monte Carlo samples.[4]

We will compare the performance of six predictors described below: the oracle predictor, Bayes predictors derived from three versions of the dynamic panel Tobit model that differ in terms of the treatment of $\pi(\lambda)$, and predictors derived from pooled Tobit and pooled linear models. The prior distributions used for the estimation of the underlying models are summarized in Table 2.

Oracle Forecast. The oracle knows the parameters $\theta = (\rho, \sigma^2)$ as well as the random effects distribution $\pi(\lambda)$. However, the oracle does not know the specific λ_i values.

Posterior Predictive Distributions Based on Dynamic Panel Tobit Model. We construct posterior predictive distributions for the dynamic Tobit panel data model under the following assumptions on $\pi(\lambda)$: (i) $\pi(\lambda)$ is Gaussian; see (15). (ii) $\pi(\lambda)$ is a mixture of

[4]If the performance statistic is linear, e.g., the coverage probability or the average length of credible sets, then averaging the statistic is the same as pooling across i and across Monte Carlo samples.

Table 1: Monte Carlo Design 1

Law of Motion: $y_{it}^* = \lambda_i + \rho y_{it-1}^* + u_{it}$ where $u_{it} \overset{iid}{\sim} N(0, \sigma^2)$. $\rho = 0.8$, $\sigma^2 = 1$
Initial Observations: $y_{i0}^* \sim N(0,1)$
Random Effects Distributions:
(a) Normal: $\pi(\lambda_i|Y_{i0}) = p_N\left(\lambda_i|\frac{1}{2}, 1\right)$
(b) Skewed: $\pi(\lambda_i|Y_{i0}) = \frac{1}{9}p_N\left(\lambda_i|\frac{5}{2}, \frac{1}{2}\right) + \frac{8}{9}p_N\left(\lambda_i|\frac{1}{4}, \frac{1}{2}\right)$
(c) Fat Tailed: $\pi(\lambda_i|Y_{i0}) = \frac{4}{5}p_N\left(\lambda_i|\frac{1}{2}, \frac{1}{4}\right) + \frac{1}{5}p_N\left(\lambda_i|\frac{1}{2}, 4\right)$
(d) Bimodal: $\pi(\tilde{\lambda}_i|Y_{i0}) = 0.35 \cdot p_N\left(\tilde{\lambda}_i|0,1\right) + 0.65 \cdot p_N\left(\tilde{\lambda}_i|10,1\right)$, $\lambda_i = \frac{1}{2} + (\tilde{\lambda}_i - \mathbb{E}[\tilde{\lambda}_i])/\sqrt{\mathbb{V}[\tilde{\lambda}_i]}$
Sample Size: $N = 1,000$, $T = 10$
Number of Monte Carlo Repetitions: $N_{sim} = 100$
Fraction of zeros: (a) 31%, (b) 32%, (c) 21%, (d) 35%

normal distributions; see (19). (iii) The posterior for each unit i is computed based on the improper prior $\lambda \sim U[-9.5, 10.5]$. Thus, for this predictor, we do not use any cross-sectional information to make inference about $\pi(\lambda)$.

Pooled Tobit. This specification ignores the heterogeneity in λ_i, sets $\lambda_i = \lambda$ for all i, and treats λ as another homogeneous parameter, just like (ρ, σ^2). Posterior sampling is implemented by iteratively sampling from $Y_{i,0:T}^*|(Y_{i,0:T}, \lambda, \rho, \sigma^2)$ as described in Section 2.1 and from the distribution of $(\lambda, \rho, \sigma^2)|Y_{i,0:T}^*$ as discussed in Section 2.2. Point, interval, and density forecasts are obtained by generating draws y_{iT+h}^* from the posterior predictive distribution and then applying the censoring $y_{iT+h} = y_{iT+h}^* \mathbb{I}\{y_{iT+h}^* \geq 0\}$.

Pooled Linear. This specification imposes heterogeneity of the λ_i's, i.e., $\lambda_i = \lambda$ for all i, and, in addition, ignores the censoring of the observations, i.e., assumes $y_{it} = y_{it}^*$. Because in the absence of the censoring the model is linear, the posterior of $(\rho, \sigma^2, \lambda)$ belongs to the NIG family and sampling is straightforward. To generate point, interval, and density forecasts, we simulate trajectories from the posterior predictive distribution and apply the censoring $y_{iT+h} = y_{iT+h}^* \mathbb{I}\{y_{iT+h}^* \geq 0\}$ *ex post*.

Implementation Details. The estimators of the Tobit models are based on the initial distribution $y_{i0}^* \sim N(0, \sigma^2)$. The Gibbs sampler for the model specification with flexible $\pi(\lambda)$ is initialized as follows: $y_{1:N,0:T}^*$ with $y_{1:N,0:T}$; (ρ, σ^2) with GMM(AB); λ_i with $\frac{1}{T}\sum_{t=1}^{T}(y_{it}^* - \rho y_{it-1}^*)$; α with its prior mean; γ_i with k-means with 10 clusters; $(\mu_k, \omega_k^2, \pi_k)$ are drawn from the conditional posteriors described in Section 2.3. The Gibbs samplers for the other specifications are special cases in which some of the parameter blocks drop out. We generate a total of $M_0 + M = 10,000$ draws using the Gibbs sampler and discard the first $M_0 = 1,000$

Table 2: Summary of Prior Distributions

Specification	λ	$\pi(\lambda)$
Normal $\pi(\lambda)$	$\lambda \sim N(\mu_\lambda, \omega_\lambda^2)$	$\omega_\lambda^2 \sim IG(2,2)$, $\mu_\lambda \vert \omega_\lambda^2 \sim N(0, \omega_\lambda^2)$
Flexible $\pi(\lambda)$	$\lambda \sim \sum_{k=1}^K \pi_k N(\mu_k, \omega_k^2)$	$\omega_k^2 \sim IG(2,2)$, $\mu_k \vert \omega_k^2 \sim N(0, \omega_k^2)$ $\pi_k \sim TSB(1, \alpha, K)$, $\alpha \sim IG(2,2)$
Flat $\pi(\lambda)$	$\lambda \sim U[-9.5, 10.5]$	
Pooled Tobit	$\lambda \sim N(0, \sigma^2)$	
Pooled Linear	$\lambda \sim N(0, \sigma^2)$	
Prior for θ	$\sigma^2 \sim IG(2,2)$, $\rho \vert \sigma^2 \sim N(0, \sigma^2)$	
Heteroskedasticity	$\sigma_i^2 \sim IG(a,b)$, prior for hyperparameters is given by (30)	
(Parametric $\pi(\sigma^2)$)	with $\underline{\alpha}_a = 1$, $\underline{\beta}_a = \underline{\gamma}_a = \underline{\alpha}_b = \underline{\beta}_b = 0.01$.	
Heteroskedasticity	$\ln \sigma_i^2 \sim \sum_{k=1}^K \pi_k N(\mu_k, \omega_k^2)$, where	
(Flexible $\pi(\sigma^2)$)	$\omega_k^2 \sim IG(2,2)$, $\mu_k \vert \omega_k^2 \sim N(0, \omega_k^2)$, $\pi_k \sim TSB(1, \alpha, K)$, $\alpha \sim IG(2,2)$.	

draws.

Point Forecasts and Parameter Estimates. In Table 3 we report root-mean-squared errors (RMSEs) for posterior mean point forecasts. We decompose the RMSEs into a bias and standard deviation component. We also report the average point estimates of ρ and σ^2 across the repetitions of the Monte Carlo experiment. The results for the oracle serve as a benchmark. For all four DGPs the predictors obtained under the flat prior for λ are dominated by the predictors that estimate $\pi(\lambda)$ from the cross-sectional distribution. Modeling $\pi(\lambda)$ as Gaussian distribution works well under Gaussian, skewed, and fat-tailed random effects. For the bi-modal random effects DGP, the predictor obtained by approximating $\pi(\lambda)$ with a mixture of normals dominates the Gaussian predictor. The pooled linear specification performs significantly worse than all competitors because it ignores the censoring which leads to a large bias. For all but the pooled linear and Tobit estimators and, in some instances the predictor obtained under a flat prior for λ, the bias component of the RMSE is negligible.

Each model delivers a forecast of the probability that $y_{iT+1} = 0$. We use this probability as a point forecast of $\mathbb{I}\{y_{iT+1} = 0\}$ and compute the corresponding RMSEs, which are also reported in Table 3. The RMSE differential between the first five model specifications is negligible. The predictor derived from the pooled linear model performs significantly worse than the other predictors in all four Monte Carlo designs. We also computed RMSEs for a perfect predictor that is constructed based on knowing each (λ_i, y_{iT}^*) in addition to θ. The

Table 3: Monte Carlo Experiment: Point Forecast Performance

	Forecast Error Stats				θ Estimates	
	RMSE	Bias	StdD	RMSE $\mathbb{I}\{y_{iT+1}=0\}$	Bias($\hat{\rho}$)	Bias($\hat{\sigma}^2$)
(a) Gaussian Random Effects						
Oracle	0.85	.004	0.85	0.22	N/A	N/A
Normal $\pi(\lambda)$, estimated θ	0.85	.004	0.85	0.22	0	0
Flexible $\pi(\lambda)$, estimated θ	0.85	.003	0.85	0.22	0	0
Flat $\pi(\lambda)$, estimated θ	0.85	0.07	0.85	0.22	-0.08	0
Pooled Tobit	0.88	-0.18	0.86	0.22	0.22	0.37
Pooled Linear	0.93	-0.32	0.87	0.36	0.21	-0.15
(b) Skewed Random Effects						
Oracle	0.83	.005	0.83	0.25	N/A	N/A
Normal $\pi(\lambda)$, estimated θ	0.83	-0.01	0.83	0.24	0	0.01
Flexible $\pi(\lambda)$, estimated θ	0.83	-0.01	0.83	0.24	0.01	0.01
Flat $\pi(\lambda)$, estimated θ	0.84	0.05	0.84	0.25	-0.08	0.01
Pooled Tobit	0.88	-0.19	0.85	0.25	0.23	0.42
Pooled Linear	0.92	-0.34	0.85	0.37	0.22	-0.13
(c) Fat-tailed Random Effects						
Oracle	0.89	0.02	0.89	0.24	N/A	N/A
Normal $\pi(\lambda)$, estimated θ	0.89	-0.01	0.89	0.24	0	0.01
Flexible $\pi(\lambda)$, estimated θ	0.89	-0.02	0.89	0.24	0.01	0.02
Flat $\pi(\lambda)$, estimated θ	0.89	0.07	0.89	0.24	-0.09	0
Pooled Tobit	0.92	-0.19	0.90	0.24	0.21	0.34
Pooled Linear	0.95	-0.30	0.90	0.33	0.21	-0.09
(d) Bi-modal Random Effects						
Oracle	0.82	.004	0.82	0.08	N/A	N/A
Normal $\pi(\lambda)$, estimated θ	0.84	0.03	0.84	0.09	0	-0.02
Flexible $\pi(\lambda)$, estimated θ	0.82	0.02	0.82	0.08	-0.01	-0.01
Flat $\pi(\lambda)$, estimated θ	0.85	0.09	0.85	0.09	-0.07	-0.01
Pooled Tobit	0.88	-0.15	0.86	0.10	0.22	0.35
Pooled Linear	0.92	-0.29	0.88	0.38	0.18	0.17

Notes: The design of the experiment is summarized in 1. The true values for ρ and σ^2 are 0.8 and 1.0, respectively.

values are 0.20, 0.23, 0.23, and 0.08, respectively, and thus slightly smaller than the oracle which only knows $\pi(\lambda)$ but not (λ_i, y_{iT}^*) for each unit.[5]

Finally, the last two columns of Table 3 show the means of the Bayes estimators of ρ

[5]If $Y \in \{0,1\}$ and the probability of $Y=0$ is p, then the RMSE associated with the optimal forecast $\widehat{\mathbb{I}\{Y=0\}} = 1-p$ is $\sqrt{p(1-p)} \leq 0.5$.

and σ^2. Recall that the true values are 0.8 and 1.0, respectively. Because the oracle is assumed to know the two parameters, we report N/A. The Bayes estimators from the panel Tobit model under the Normal and the flexible prior are unbiased. Under the flat prior, the estimate of the autocorrelation parameter ρ is downward biased. Its mean across the Monte Carlo repetitions is 0.71 instead of 0.80. The estimators of ρ obtained from the pooled Tobit and pooled linear specifications are around 1.0 and thus strongly upward biased.

The panels of Figure 1 show true RE densities $\pi(\lambda)$, estimated densities $\hat{\pi}(\lambda)$, and histograms of the point estimates $\mathbb{E}[\lambda_i|Y_{1:N,0:T}]$. The panels of the left are obtained under the assumption that $\pi(\lambda)$ is Normal and the panels on the right correspond to specification with a flexible $\pi(\lambda)$. The four rows of the figure correspond to the four Monte Carlo designs. Under the Normal RE distribution the estimates $\hat{\pi}(\lambda)$ under the Normal and the flexible prior are essentially identical and match the true $\pi(\lambda)$. Under the other three experimental designs the two density estimates $\hat{\pi}(\lambda)$ are different and this difference is most pronounced in the bi-modal RE setting. In all of these cases the flexible estimate of $\hat{\pi}(\lambda)$ continues to deliver a good approximation of the λ_i distribution.

To interpret the histograms of $\mathbb{E}[\lambda_i|Y_{1:N,0:T}]$ in view of the plotted $\pi(\lambda)$, we will consider two stylized examples that capture important aspects of our setup. First, suppose that the model is linear, i.e., $y_{it} = \lambda_i + u_{it}$ and $\lambda_i \sim N(\mu_\lambda, \omega^2)$, and the hyperparameters are known. Therefore, $\hat{\lambda}_i \sim N\big(\mu_\lambda, \omega^2 + \sigma^2/T\big)$ and the posterior means have the cross-sectional distribution

$$\mathbb{E}[\lambda_i|Y_{1:N,1:T}] = \frac{T/\sigma^2}{T/\sigma^2 + 1/\omega^2}\hat{\lambda}_i + \frac{1/\omega^2}{T/\sigma^2 + 1/\omega^2}\mu_\lambda \sim N\left(\mu_\lambda, \omega^2\frac{\omega^2}{\omega^2 + \sigma^2/T}\right).$$

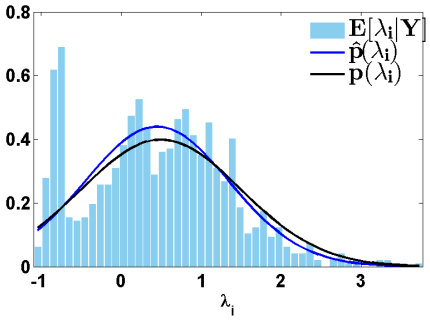
In this example, the distribution of the posterior mean estimates is less dispersed than the distribution of the λ_i's, but centered at the same mean. In our design (a) we have $\omega^2 = \sigma^2 = 1$ and $T = 10$. This implies that the standard deviation of the posterior mean estimates should be 5% smaller than the standard deviation of the λ_i's, which is visually consistent with Figure 1.

Second, to understand the effect of censoring, suppose that $y^*_{it} = \lambda_i + u_{it}$ and we observe a sequence of zeros. The likelihood associated with this sequence of zeros is given by $\Phi_N^T(-\lambda_i/\sigma)$. The posterior mean for a sequence of zeros is then given by

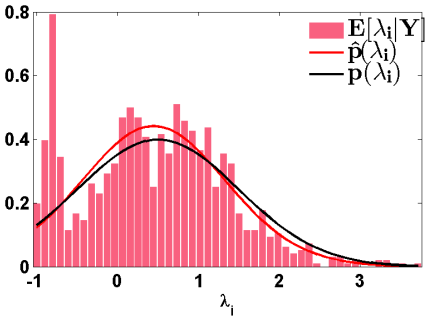
$$\mathbb{E}[\lambda_i|Y_{1:N,1:T} = 0] = \frac{\int \lambda \Phi_N^T(-\lambda/\sigma)\pi(\lambda)d\lambda}{\int \Phi_N^T(-\lambda/\sigma)\pi(\lambda)d\lambda}$$

Figure 1: Estimates $\mathbb{E}[\lambda_i|Y_{1:N,0:T}]$ and Estimated Random Effects Distributions

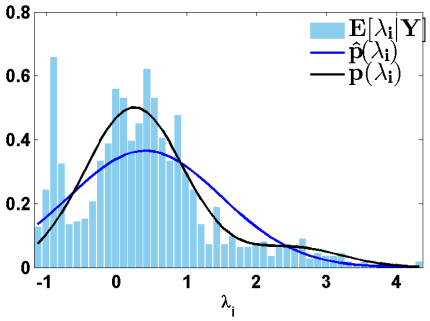
Normal RE: Normal $\pi(\lambda)$



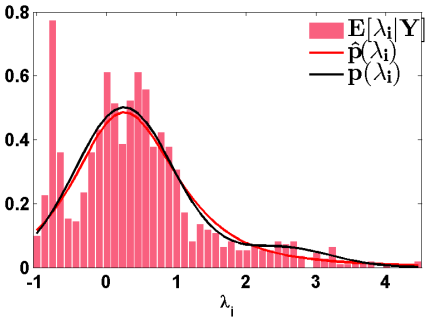
Normal RE: Flexible $\pi(\lambda)$



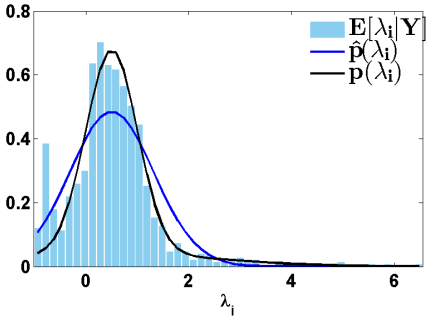
Skewed RE: Normal $\pi(\lambda)$



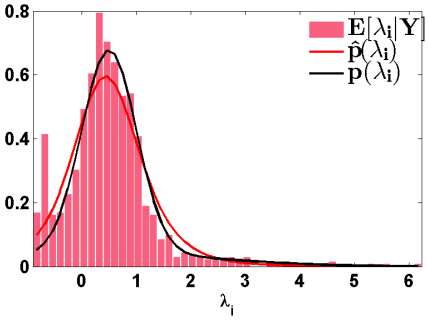
Skewed RE: Flexible $\pi(\lambda)$



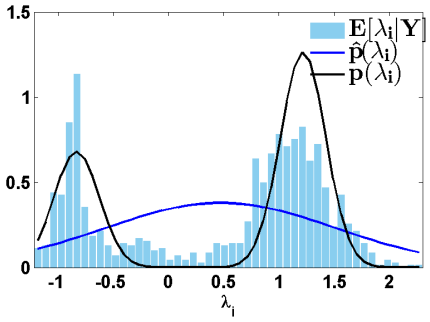
Fat-tail RE: Normal $\pi(\lambda)$



Fat-tail RE: Flexible $\pi(\lambda)$



Bi-modal RE: Normal $\pi(\lambda)$



Bi-modal RE: Flexible $\pi(\lambda)$

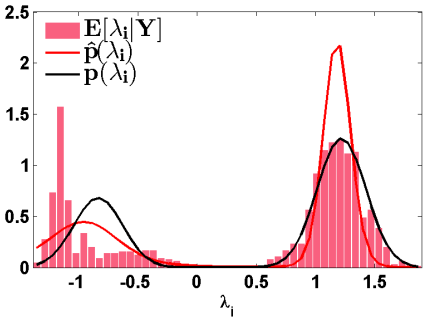


Table 4: Monte Carlo Experiment: Interval Forecast and Density Forecast Performance

	Interval Forecast		Density Forecast	
	Cover Freq	CI Len	LPS	CRPS
(a) Normal Random Effects				
Oracle	0.91	2.23	-1.08	0.40
Normal $\pi(\lambda)$, estimated θ	0.91	2.22	-1.08	0.40
Flexible $\pi(\lambda)$, estimated θ	0.91	2.22	-1.08	0.40
Flat $\pi(\lambda)$, estimated θ	0.91	2.23	-1.10	0.41
Pooled Tobit	0.93	2.51	-1.12	0.42
Pooled Linear	0.93	2.74	-1.31	0.47
(b) Skewed Random Effects				
Oracle	0.91	2.14	-1.08	0.40
Normal $\pi(\lambda)$, estimated θ	0.91	2.15	-1.08	0.40
Flexible $\pi(\lambda)$, estimated θ	0.91	2.15	-1.08	0.40
Flat $\pi(\lambda)$, estimated θ	0.91	2.15	-1.10	0.40
Pooled Tobit	0.93	2.45	-1.12	0.42
Pooled Linear	0.93	2.67	-1.30	0.46
(c) Fat-tailed Random Effects				
Oracle	0.90	2.44	-1.20	0.45
Normal $\pi(\lambda)$, estimated θ	0.90	2.44	-1.20	0.45
Flexible $\pi(\lambda)$, estimated θ	0.90	2.45	-1.20	0.45
Flat $\pi(\lambda)$, estimated θ	0.90	2.43	-1.22	0.45
Pooled Tobit	0.92	2.72	-1.23	0.46
Pooled Linear	0.92	2.90	-1.35	0.50
(d) Bi-modal Random Effects				
Oracle	0.93	2.17	-0.96	0.38
Normal $\pi(\lambda)$, estimated θ	0.93	2.22	-0.99	0.39
Flexible $\pi(\lambda)$, estimated θ	0.93	2.19	-0.96	0.38
Flat $\pi(\lambda)$, estimated θ	0.93	2.24	-0.99	0.39
Pooled Tobit	0.95	2.50	-1.03	0.40
Pooled Linear	0.93	2.79	-1.31	0.47

and provides a lower bound for the estimator $\hat{\lambda}_i$. If the λ_i's are sampled from the prior, we should observe this posterior mean with probability $\int \Phi_N^T\big(-\lambda/\sigma\big)\pi(\lambda)d\lambda$. Thus, according to this example, there should be a spike in the left tail of the distributions of $\mathbb{E}[\lambda_i|Y_{1:N,1:T}]$. Such spikes are clearly visible in the panels of Figure 1.

Interval and Density Forecasts. The dynamic panel Tobit model generates a posterior predictive density for y_{iT+1} from which one can derive interval and density forecasts.

Both types of forecasts reflect parameter uncertainty, potential uncertainty about y_{iT}^*, and uncertainty about future shocks. To assess the interval forecasts we compute the coverage frequency and the average length of 90% predictive intervals. As a measure of density forecast accuracy we evaluate compute log predictive scores $p(y_{iT+1}|Y_{1:N,1:T})$ as well as the continuous ranked probability score (CRPS). The CRPS measures the L_2 distance between the cumulative density function (cdf) associated with $p(y_{iT+1}|Y_{1:N,1:T})$ and a "perfect" density forecasts which assigns probability one to the realized y_{iT+1}.

The results are summarized in Table 4. Under the Normal RE, the skewed RE, and the fat-tailed RE designs the coverage frequencies for the panel Tobit models are very close to the nominal coverage probability of 90%. Under the bi-modal RE design, they are slightly larger (93%). The coverage frequencies of the interval predictions derived from the pooled Tobit and pooled linear specifications generally exceed the nominal coverage frequency. For the first three simulation designs the average lengths of the panel Tobit intervals are generally very close to the length of the oracle forecasts, indicating that parameter uncertainty does not play a significant role. Except for the last design, in which the Normal RE distribution is grossly misspecified, the performance of the predictors under Normal and flexible $\pi(\lambda)$ are generally similar. The interval predictions obtained from the pooled Tobit and pooled linear specifications are clearly inferior because the average lengths of the intervals are much larger than for the specifications with heterogeneous λ_i.

The density forecast rankings mirror the interval forecast rankings. Note that we prefer forecasts with a high LPS and a low CRPS. The forecasts derived from the estimated dynamic panel Tobit models are (almost) as good as the oracle density forecasts. The pooled Tobit and pooled linear forecasts, on the other hand, tend to be substantially worse than the oracle forecasts.

6 Empirical Analysis

We will now use the dynamic panel Tobit model to forecast loan charge-off rates for a panel of "small" banks, which we define to be banks with total assets of less than one billion dollars. For these banks it is reasonable to assume that they operate in local markets. Thus, we will use local economic indicators as additional predictors. A detailed description of the data set is provided in the Online Appendix.

6.1 Model Specification

Starting point is a homoskedastic random effects model:

$$\begin{aligned} y_{it} &= y_{it}^* \mathbb{I}\{y_{it}^* \geq 0\}, \quad i = 1, \ldots, N, \quad t = 0, \ldots, T \\ y_{it}^* &= \lambda_i + \rho y_{it-1}^* + \beta_1 \ln \mathrm{HPI}_{it} + \beta_2 \mathrm{UR}_{it} + u_{it}, \quad u_{it} \overset{iid}{\sim} N(0, \sigma^2), \quad y_{i0}^* \sim N(\mu_{i*}, \sigma_*^2), \end{aligned} \tag{34}$$

where y_{it} are charge-off rates, HPI_{it} is a house price index, and UR_{it} is the unemployment rate. The subsequent results are based on $\mu_{i*} = 0$ and $\sigma_*^2 = \sigma^2$.

The time period is one quarter. In addition, we also consider a heteroskedastic law of motion for y_{it}^* with

$$u_{it} \overset{iid}{\sim} N(0, \sigma_i^2), \quad y_{i0}^* \sim N(\mu_{i*}, \sigma_{i*}^2). \tag{35}$$

The charge-off rates are measured in annualized percentages. The distribution of charge-off rates is highly skewed. In any given quarter at least 30% of the banks report zero charge-offs. The cross-sectional means of the charge-off rates range from 0.2% to 3.4%. The 75th quantiles range from 0% to 4.8%. The maximum, in many quarters, reaches 400% (we eliminated observations for which the charge-off rates were larger than the annualized 400%). A with summary statistics for the charge-off data is provided in the Online Appendix.

6.2 Estimation Results

We begin by inspecting some of the parameter estimates. We focus on a sample of credit card charge-off rates that starts in 2001Q2 and ends in 2003Q4. The first period, 2001Q2 is used to initialize the lag in the dynamic specification. Thus, $T = 10$. We use the estimates to construct one-step-ahead forecasts for 2004Q1. We consider the same estimators as in the Monte Carlo experiments of Section 5 and, with three qualifications, use the prior distributions listed in Table 2.

First, while the Monte Carlo experiment assumed homoskedasticity, in the empirical analysis we also allow for heteroskedasticity. The last row of Table 2 describes the distribution of the σ_i^2 and its hyperparameters. Second, the homogeneous parameter vector θ now comprises $(\rho, \beta_1, \beta_2, \sigma^2)$. The prior for $(\rho, \beta_1, \beta_2)|\sigma^2)$ is $N(0, \sigma^2 I)$ for the homoskedastic model and $N(0, I)$ for the heteroskedastic model. Third, some extra care is necessary for the construction of a flat prior $\pi(\lambda)$. If the domain of this prior is too large, then for some units the posterior sampler becomes unreliable. If the domain is too small, then

the prior adds information to the estimation of the heterogeneous coefficient. We construct bounds for the uniform distribution as follows: (i) we replace y_{it}^* by y_{it} and use a fixed-effect GMM estimator to get approximate estimates of the common coefficients $\hat{\beta}$. (ii) We then average the residuals to obtain estimates of the heterogeneous intercepts: $\hat{\lambda}_i = \frac{1}{T}\sum_{t=1}^{T}(y_{it} - (\hat{\rho}y_{it-1} + \hat{\beta}_1 \ln \mathrm{HPI}_{it} + \hat{\beta}_2 \mathrm{UR}_{it}))$. (iii) Finally, we set the bounds for the uniform distribution to $\mathrm{mean}(\hat{\lambda}_i) \pm 10\mathrm{StdD}(\hat{\lambda}_i)$. For instance, in the 2001Q2 to 2003Q4 credit-card charge-off rate sample these bounds are -42.3 and 36.4, respectively.

Parameter estimates (posterior means and credible sets) are reported in Table 5. We estimate some mild serial correlation. The estimates vary across specifications. The largest estimates of ρ, 0.53 and 0.67, are obtained from the pooled estimators. The credit card charge-off rates essentially do not respond to house prices. A house price increase of 1% corresponds to $\Delta \ln \mathrm{HPI}_{it} = 0.01$, which under the homoskedastic model with flexible $\pi(\lambda)$ prior translates into a $0.01 \cdot 0.15$ percentage increase in annualized charge-off rates. Moreover, for most of the specifications, the credible intervals of the house-price coefficient cover zero. The response to the unemployment rate is also generally small. A coefficient of one implies that a one percent increase in the unemployment rate ($\Delta UR_{it} = 0.01$ because we divided the unemployment rate by 100) raises the charge-off rate by 0.01%. The sign of the β_2 point estimates does not conform with economic intuition. An increase in unemployment seems to lower rather than raise the charge-off rates. However, all but one interval cover zero. The estimated shock variances in the homoskedastic specification are large because they have to capture occasionally large charge-off rates.

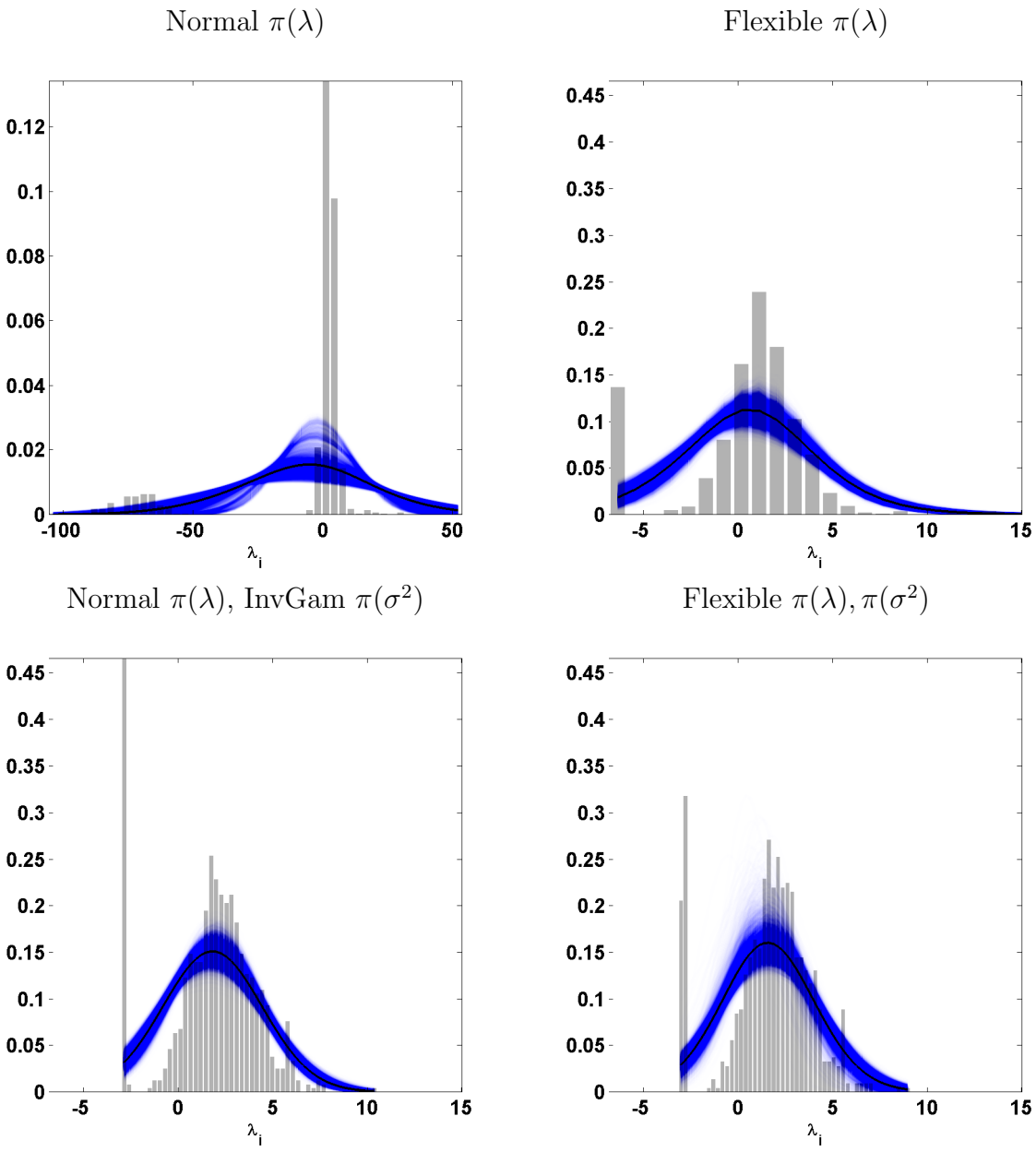
Each panel of Figure 2 depicts a histogram for $\mathbb{E}[\lambda_i|Y_{1:N,0:T}]$ and hairlines that represent draws from the posterior distribution of $\pi(\lambda)|Y_{1:N,0:T}$. The graphs differ from Figure 1 in two dimensions. First, the population object $\pi(\lambda)$ is unknown and therefore cannot be displayed. Second, the point estimator $\hat{\pi}(\lambda)$ is replaced by draws from the posterior distribution of the RE distribution to illustrate estimation uncertainty. The results in the top right and the two bottom panels are very similar. The posterior mean estimates from λ_i range from about -5 to 10. The histograms peak around 2 and have a spike in the left tail at around -3. The spikes are a consequence of the censoring in our model. We also observe that the distribution of the posterior mean estimates is more concentrated than the estimated RE distributions. These features are consistent with the simplified examples discussed in Section 5. The results from the homoskedastic model with Gaussian RE distribution shown in the top left panel are different. Here there is a mass for values of $\mathbb{E}[\lambda_i|Y_{1:N,0:T}]$ between -80 and -60 and the estimates of the $\pi(\lambda)$ distributions are much more diffuse. Even though we are not

Table 5: Parameter Estimates

Estimator	$\mathbb{E}[\hat{\rho}]$ (y^*_{it-1})		$\mathbb{E}[\hat{\beta}_1]$ $(\ln \mathrm{HPI}_{it})$		$\mathbb{E}[\hat{\beta}_2]$ (UR_{it})		$\mathbb{E}[\hat{\sigma}^2]$	
Credit Card Charge-Off Rates – Homoskedastic Models								
Normal $\pi(\lambda)$	0.17	[0.09,0.24]	0.05	[-0.04,0.15]	-1.90	[-8.69,4.90]	19.73	[18.86,20.67]
Flexible $\pi(\lambda)$	0.27	[0.24,0.31]	0.15	[0.06,0.24]	-2.99	[-9.93,3.85]	20.59	[19.67,21.56]
Flat $\pi(\lambda)$	0.16	[0.13,0.20]	0.07	[-0.01,0.15]	-4.83	[-11.68,2.00]	20.15	[19.48,20.87]
Pooled Tobit	0.67	[0.64,0.70]	0.19	[-0.26,0.64]	-3.67	[-10.37,3.03]	28.81	[27.29,30.43]
Pooled Linear	0.53	[0.52,0.54]	0.15	[-0.21,0.51]	-7.55	[-12.95,-2.16]	19.14	[18.67,19.62]
Credit Card Charge-Off Rates – Heteroskedastic Models								
Normal $\pi(\lambda)$, InvGam $\pi(\sigma^2)$	0.12	[0.09,0.15]	-.002	[-0.03,0.02]	-0.03	[-1.64,1.59]	N/A	
Flexible $\pi(\lambda), \pi(\sigma^2)$	0.17	[0.13,0.23]	.004	[-0.02,0.03]	0.08	[-1.62,1.77]	N/A	

Notes: The estimation sample ranges from 2001Q2 to 2003Q4 ($T = 10$). The table contains posterior mean parameter estimates as well as 90% credible intervals in brackets.

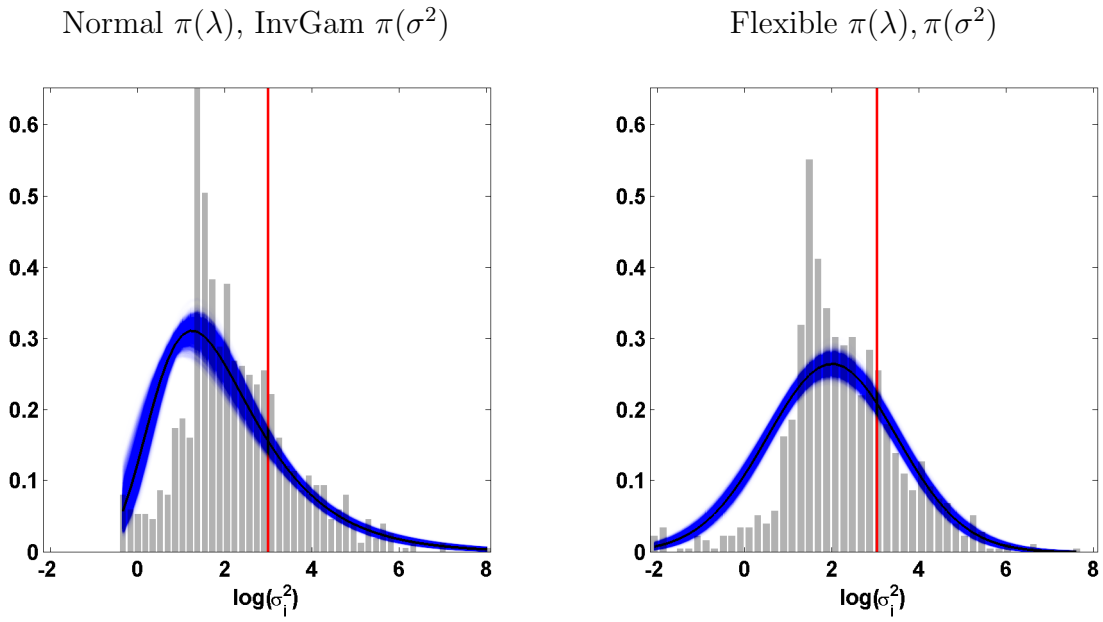
Figure 2: Posterior Means vs. Estimated Random-Effects Distributions



Notes: The figure depicts histograms for $\mathbb{E}[\lambda_i|Y_{1:N,0:T}]$, $i = 1, \ldots, N$ for four different model specifications. The shaded areas are obtained by generating draws from the posterior distribution of the random effects density: $\pi(\lambda)|Y_{1:N,0:T}$. The estimation sample for the dynamic Tobit models ranges from 2001Q2 to 2003Q4 ($T = 10$).

conducting a formal specification test at this stage, we view this as evidence that the fit of the homoskedastic model with Normal $\pi(\lambda)$ is not as good as the fit of the other three specifications.

Figure 3: Posterior Means vs. Estimated Random-Effects Distributions



Notes: The panels depict histograms for $\ln \mathbb{E}[\sigma_i^2|Y_{1:N,0:T}]$, $i = 1, \ldots, N$. The shaded areas are obtained by generating draws from the posterior distribution of the random effects density: $\pi(\sigma_i)|Y_{1:N,0:T}$. The estimation sample for the dynamic Tobit models ranges from 2001Q2 to 2003Q4 ($T = 10$). The point estimates under the homoskedastic specifications reported in Table correspond to 2.98 (Normal $\pi(\lambda)$) and 3.02 (Flexible $\pi(\lambda)$), respectively, and are indicated using red vertical lines.

In Figure 3 we overlay the posterior means $\mathbb{E}[\sigma_i^2|Y_{1:N,0:T}]$ as well as draws from the posterior distribution $\pi(\sigma^2)|Y_{1:N,0:T}$ for model specifications with heteroskedasticity. The shape of the histogram mimics the shape of the densities, except for the pronounced spike of the histogram near $\ln \sigma_i^2 = 2$. We generated bivariate scatter plots of $\hat{\lambda}_i$ and $\ln \hat{\sigma}_i^2$ (shown in the Online Appendix) that illustrate that the spikes in the histograms depicted in Figures 2 and 3 correspond to the same units. The point estimates of σ^2 for the homoskedastic model specifications, translated into the log scale of Figure 3, are 2.98 (Normal $\pi(\lambda)$) and 3.02 (Flexible $\pi(\lambda)$), respectively. The density plots indicate that the assumption of homoskedasticity is quite restrictive: for some banks the estimated innovation variances are greater than four, whereas for others they are less than one.

6.3 Forecasting Performance for Credit Card Charge-Off Rates

Point Forecast Evaluation. Point forecast error statistics are reported in Table 6. The homoskedastic versions of the Normal and the flexible random effects model perform well.

Table 6: Point Forecast Performance

Estimator	RMSE	Bias	StdD	RMSE($y_{iT+1} = 0$)
Credit Card Charge-Off Rates – Homoskedastic Models				
Normal $\pi(\lambda)$	4.24	-0.55	4.20	0.41
Flexible $\pi(\lambda)$	4.23	-0.48	4.20	0.41
Flat $\pi(\lambda)$	4.37	-0.56	4.33	0.41
Pooled Tobit	4.54	-0.64	4.49	0.44
Pooled Linear	4.72	-1.20	4.57	0.48
Credit Card Charge-Off Rates – Heterokedastic Models				
Normal $\pi(\lambda)$, InvGam $\pi(\sigma^2)$	4.76	-0.03	4.76	0.39
Flexible $\pi(\lambda), \pi(\sigma^2)$	4.76	-0.02	4.76	0.39

Notes: The estimation sample ranges from 2001Q2 to 2003Q4 ($T = 10$). We forecast the observation in 2004Q1. The last column reports the RMSE for the probability forecast of a zero charge-off.

Under the flat prior, the RMSE increases. The pooled Tobit and pooled linear specifications perform noticeably worse than the homoskedastic random effects specifications. The heteroskedastic specifications are associated with the largest RMSEs. The last column of Table 6 reports the RMSE for the probability forecast of $\mathbb{I}\{y_{iT+1} = 1\}$. The models with Normal, flexible, and flat priors perform similarly and attain RMSEs for the event forecasts of approximately 0.41. The pooled Tobit model and the pooled linear model perform worse.[6] The best predictions of a zero charge-off are generated by the two heteroskedastic specifications.

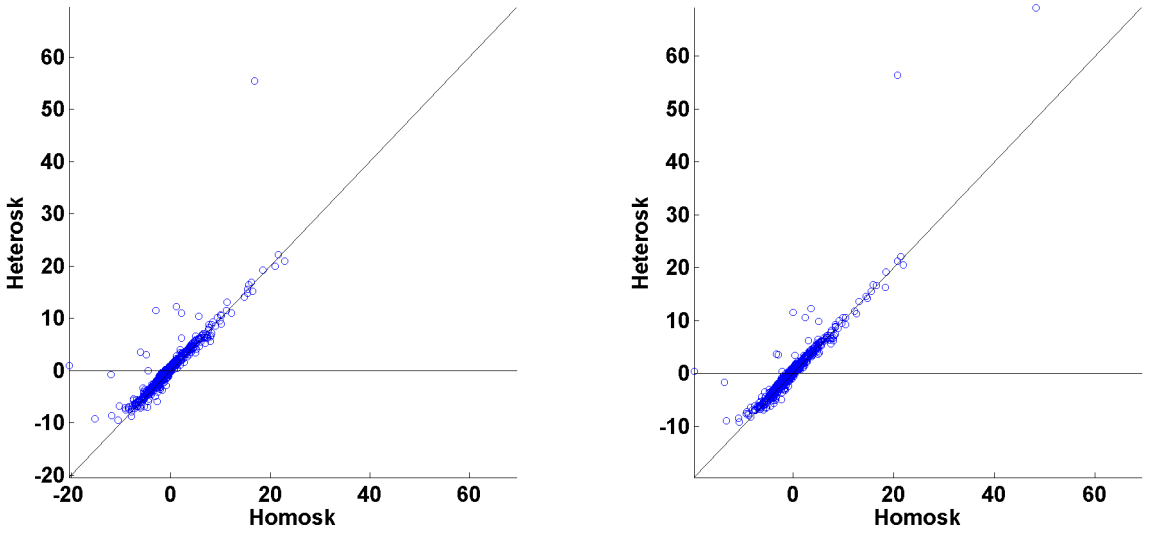
To gain a deeper understanding of the RMSE statistics, we plot bank-level forecast errors in Figure 4. The two panel of the figures show scatter plots of the forecast errors under the homoskedastic (x-axis) and the heteroskedastic (y-axis) model specifications. Due to the censoring, the distribution of forecast errors is skewed to the right. Note that for a few institutions, the forecast errors are very large (and positive). For those banks the heteroskedastic predictor tends to perform significantly worse.

Abstracting from the effect of censoring, the posterior mean of λ_i is a linear combination of the MLE and the prior mean. The relative weights depend on the relative precision of likelihood and prior. At least for the flexible specification, see Figure 2 the estimated $\pi(\lambda)$ are quite similar under homoskedasticity and heteroskedasticity. Thus, the key difference in the predictors is due to the precision of the likelihood. Under the heteroskedastic specifications,

[6]Recall that we censor the draws from the posterior predictive distribution of the pooled linear model, which leads to a non-trivial probability of a zero charge-off.

Figure 4: Scatter Plot of Forecast Errors

Normal $\pi(\lambda)$ vs Normal $\pi(\lambda)$, InvGam $\pi(\sigma^2)$ Flexible $\pi(\lambda)$ vs Flexible $\pi(\lambda), \pi(\sigma^2)$



Notes: The panels depict scatter plots of bank-level forecast errors of the homoskedastic (x-axis) versus heteroskedastic (y-axis) specification under a parametric and a flexible prior distribution, respectively. We overlay 45-degree lines in each panel.

there are several banks for which $\hat{\sigma}_i^2$ is large, which means that the precision of the likelihood is low and the forecast is essentially generated from the prior mean of λ. This seems to generate poor forecast performance, compared to the homoskedastic predictor which imposes more shrinkage on these extreme observations. In the Online Appendix, we provide scatter plots of forecast error differentials versus $\ln \hat{\sigma}_i^2$ that confirm this conjecture.

Interval and Density Forecast Evaluation. Interval and density forecast results are presented in Table 7. The coverage frequency is close to the nominal coverage probability of 90% for all forecasts, except the ones from the pooled linear specification. In terms of striking a good balance between coverage probability and length of the confidence interval, the heteroskedastic specifications clearly come out ahead. The heteroskedastic specifications also come out ahead in terms of log probability score and continuous ranked probability score. The differences between the parametric and the flexible specification are very small across all criteria.

Histograms of probability integral transforms (PIT) are plotted in Figure 5. If the density forecasts are perfectly calibrated, then the PITs should be uniform. Because we are using 20

Table 7: Interval Forecast and Density Forecast Performance

Estimator	Interval Forecast Cover Freq	CI Len	Density Forecast LPS	CRPS
Credit Card Charge-Off Rates – Homoskedastic Models				
Normal $\pi(\lambda)$	0.92	7.90	-2.19	1.83
Flexible $\pi(\lambda)$	0.92	7.89	-2.20	1.82
Flat $\pi(\lambda)$	0.92	7.97	-2.22	1.86
Pooled Tobit	0.92	8.84	-2.28	1.98
Pooled Linear	0.93	9.96	-2.39	2.11
Credit Card Charge-Off Rates – Heteroskedastic Models				
Normal $\pi(\lambda)$, InvGam $\pi(\sigma^2)$	0.89	6.88	-1.94	1.71
Flexible $\pi(\lambda), \pi(\sigma^2)$	0.88	6.84	-1.96	1.71

Notes: The estimation sample ranges from 2001Q2 to 2003Q4 ($T = 10$). We forecast the observation in 2004Q1.

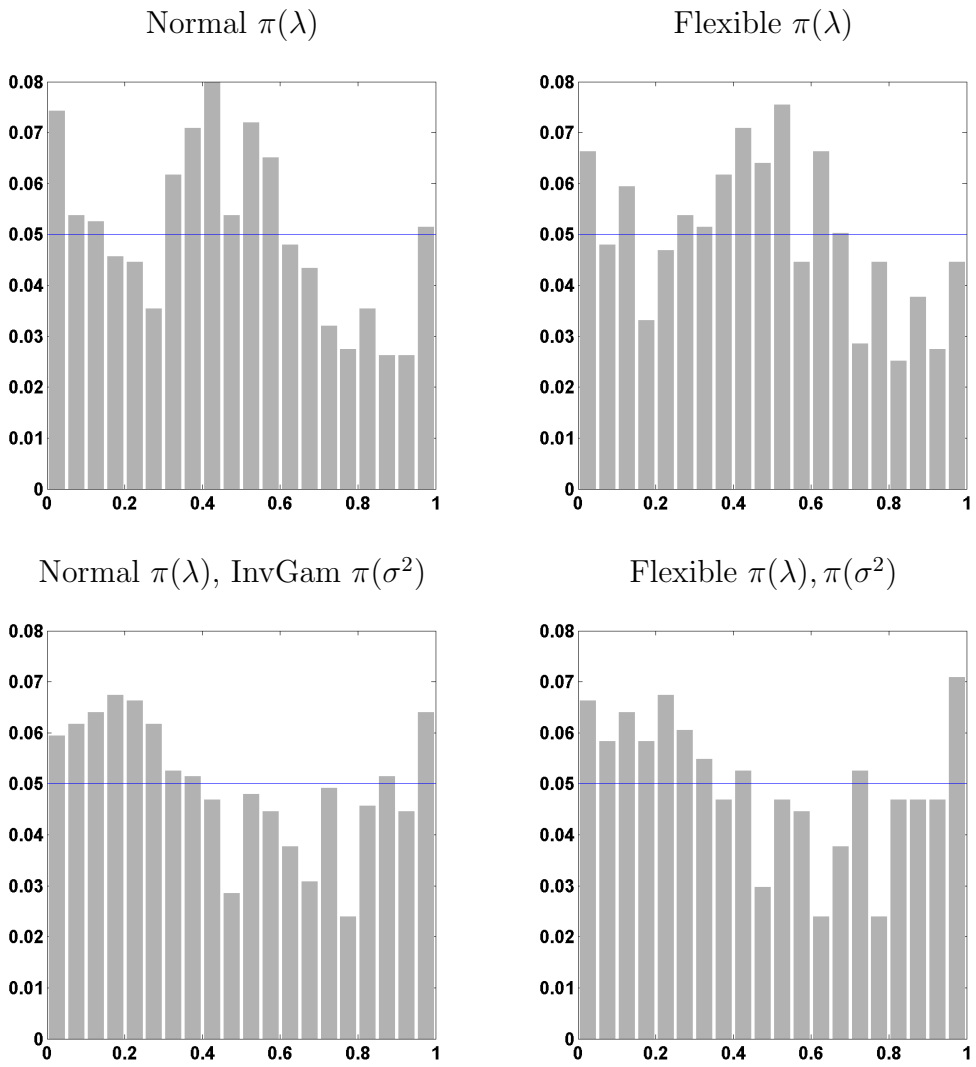
bins for the histogram, uniformity would imply a mass of 5% (indicated by the horizontal line) for each bin. For all model specifications, it appears that the distribution of the PITs differs from the uniform distribution. The heteroskedastic model with a flexible $\pi(\lambda)\pi(\sigma^2)$ seems to come closes to the uniform benchmark. In general, the frequency of PITs between 0 and 0.3 in the bottom panels is higher than the target value. This means we are overpredicting the frequency of zero and low charge-off rates. Note that the probability of a zero is relatively strongly influenced by prior.

6.4 Forecasting Performance for Other Samples

We now generate forecast error statistics for other charge-off rates and other sample periods. The total number of samples is 54. Figure 6 consists of two scatter plots that compare RMSEs for predictors derived from homoskedastic and heteroskedastic specifications. The plot on the left refers to the specification in which $\pi(\lambda)$ and $\pi(\sigma^2)$ are represented by a Normal and an Inverse Gamma distribution, respectively. The plot on the right corresponds to specifications in which the RE distributions are flexible. In order to make the RMSEs comparable across different samples, we normalize them by the RMSE of the (homoskedastic) pooled linear specification. For instance, a value of 0.9 means that the RMSE obtained from the panel Tobit model is 10% smaller than the RMSE from the pooled linear specification.

According to Figure 6, the homoskedastic panel Tobit predictors tend to dominate the

Figure 5: Probability Integral Transforms

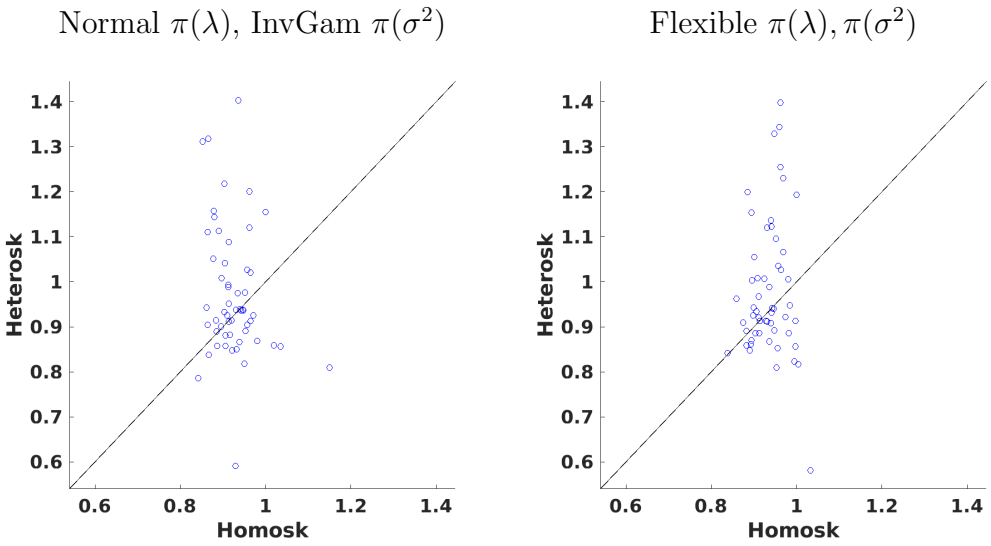


Notes: The panels depict histograms of probability integral transforms.

predictions from the pooled linear specification in all but four samples. In about 20 of the samples, the pooled linear predictors seem to perform better than the heteroskedastic panel Tobit predictors. The majority of the points in the scatter plot lie above the 45 degree line, which means that imposing homoskedasticity tends, on average, generate better forecasts. We saw this already for the credit card charge-off rates in Section 6.3.

Figure 7 shows scatter plots for log probability score and CRPS differentials relative to forecasts from the pooled Tobit specification. Recall that positive LPS and negative CRPS favor the heterogeneous coefficient models. In terms of density forecasts the heteroskedastic

Figure 6: RMSE Ratios



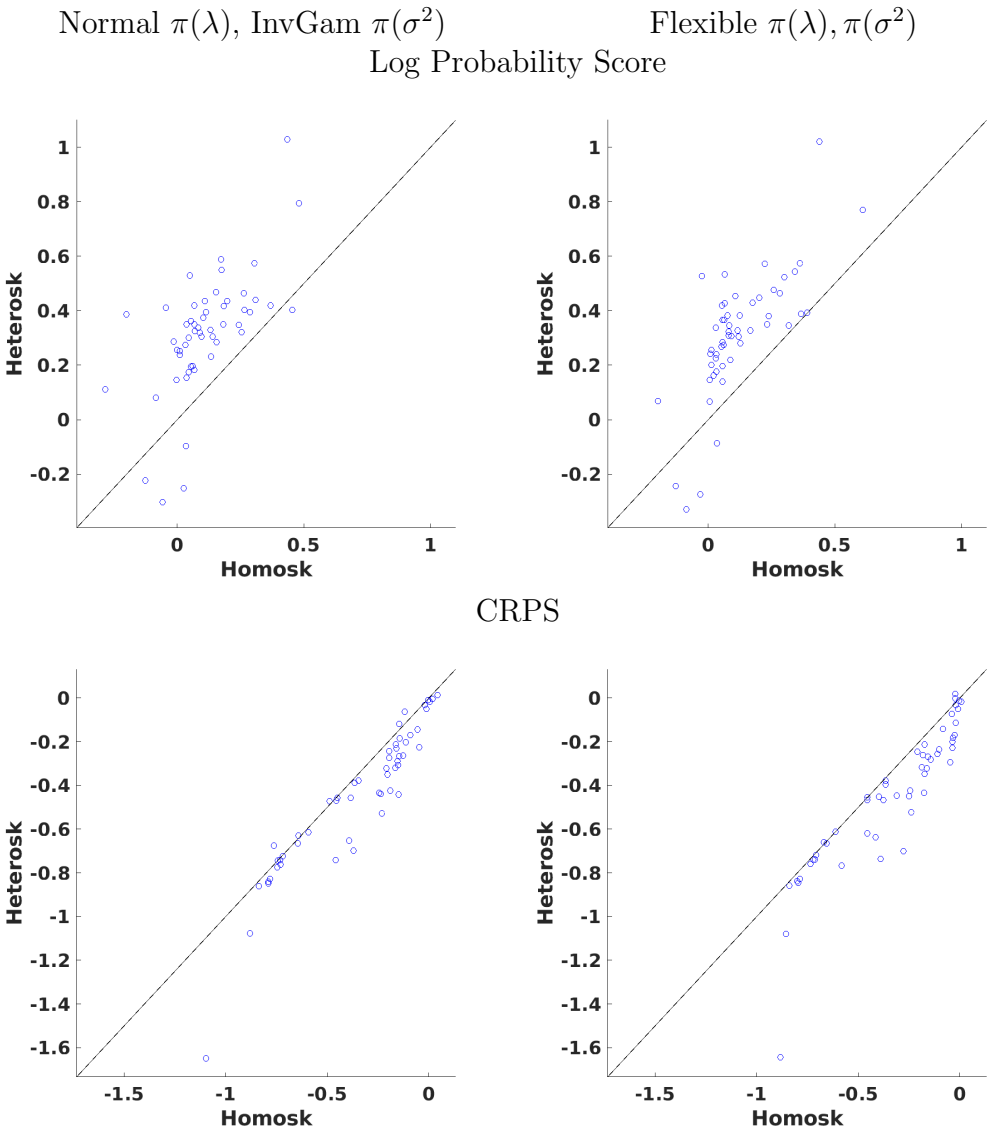
Notes: The panels compare the RMSEs of posterior mean predictors derived under the parametric prior and the flexible prior for the heterogeneous coefficients for homoskedastic and heteroskedastic specifications. We also show the 45-degree line. RMSEs are depicted as ratios relative to the RMSE obtained by the pooled linear specification.

specifications perform consistently better than the homoskedastic specifications across the 54 samples. Moreover, because the LPS values are almost all positive and the CRPS values are almost all negative, we deduce that the pooled Tobit predictor is clearly inferior.

7 Conclusion

The limited dependent variable panel with unobserved individual effects is a common data structure but not extensively studied in the forecasting literature. This paper constructs forecasts based on a flexible dynamic panel Tobit model to forecast individual future outcomes based on a panel of censored data with large N and small T dimensions. Our empirical application to loan charge-off rates of small banks shows that the flexible predictors provide more accurate point, interval, and density forecasts than naive competitors, which constitutes a helpful tool for bank supervision. Our framework can be extended to dynamic panel versions of more general multivariate censored regression models. We can also allow for missing observations in our panel data set. Finally, even though we focused on the analysis of charge-off data, there are many other potential applications for our methods.

Figure 7: Density Forecast Evaluations



Notes: The panels compare the log probability scores (top) and CRPS (bottom) of density forecasts derived under the parametric prior and the flexible prior for the heterogeneous coefficients for homoskedastic and heteroskedastic specifications. We also show the 45-degree line. Log probability scores and CRPS are depicted as differentials relative to pooled Tobit.

References

Abrevaya, J. (2000): "Rank estimation of a generalized fixed-effects regression model," *Journal of Econometrics*, 95(1), 1–23.

Altonji, J. G., and R. L. Matzkin (2005): "Cross section and panel data estimators for nonseparable models with endogenous regressors," *Econometrica*, 73(4), 1053–1102.

Amemiya, T. (1985): *Advanced Econometrics*. Harvard University Press, Cambridge.

Arellano, M., and B. Honoré (2001): "Panel data models: some recent developments," *Handbook of econometrics*, 5, 3229–3296.

Atchadé, Y. F., and J. S. Rosenthal (2005): "On adaptive Markov chain Monte Carlo algorithms," *Bernoulli*, 11(5), 815–828.

Bester, C. A., and C. Hansen (2009): "Identification of marginal effects in a nonparametric correlated random effects model," *Journal of Business & Economic Statistics*, 27(2), 235–250.

Bonhomme, S. (2012): "Functional differencing," *Econometrica*, 80(4), 1337–1385.

Botev, Z. I. (2017): "The Normal Law under Linear Restrictions: Simulation and Estimation via Minimax Tilting," *Journal of the Royal Statistical Society B*, 79(1), 125–148.

Brown, L. D., and E. Greenshtein (2009): "Nonparametric empirical Bayes and compound decision approaches to estimation of a high-dimensional vector of normal means," *The Annals of Statistics*, pp. 1685–1704.

Burda, M., and M. Harding (2013): "Panel probit with flexible correlated effects: quantifying technology spillovers in the presence of latent heterogeneity," *Journal of Applied Econometrics*, 28(6), 956–981.

Chib, S. (1992): "Bayes Inference in the Tobit Censored Regression Model," *Journal of Econometrics*, 51, 79–99.

Covas, F. B., B. Rump, and E. Zakrajsek (2014): "Stress-Testing U.S. Bank Holding Companies: A Dynamic Panel Quantile Regression Approach," *International Journal of Forecasting*, 30(3), 691–713.

Gneiting, T., and A. E. Raftery (2007): "Strictly Proper Scoring Rules, Prediction, and Estimation," *Journal of the American Statistical Association*, 102(477), 359–378.

Greene, W. H. (2008): *Econometric Analysis*. Pearson Prentice Hall.

Griffin, J. E. (2016): "An adaptive truncation method for inference in Bayesian nonparametric models," *Statistics and Computing*, 26(1), 423–441.

Gu, J., and R. Koenker (2016a): "Empirical Bayesball Remixed: Empirical Bayes Methods for Longitudinal Data," *Journal of Applied Economics (Forthcoming)*.

——— (2016b): "Unobserved Heterogeneity in Income Dynamics: An Empirical Bayes Perspective," *Journal of Business & Economic Statistics (Forthcoming)*.

Hirano, K. (2002): "Semiparametric Bayesian inference in autoregressive panel data models," *Econometrica*, 70(2), 781–799.

Hoderlein, S., and H. White (2012): "Nonparametric identification in nonseparable panel data models with generalized fixed effects," *Journal of Econometrics*, 168(2), 300–314.

Honoré, B. E. (1992): "Trimmed LAD and least squares estimation of truncated and censored regression models with fixed effects," *Econometrica: journal of the Econometric Society*, pp. 533–565.

Honore, B. E., and L. Hu (2004): "Estimation of cross sectional and panel data censored regression models with endogeneity," *Journal of Econometrics*, 122(2), 293–316.

Hu, L. (2002): "Estimation of a censored dynamic panel data model," *Econometrica*, 70(6), 2499–2517.

Hu, Y., and J.-L. Shiu (2018): "Identification and estimation of semi-parametric censored dynamic panel data models of short time periods," *The Econometrics Journal*, pp. 55–85.

Ishwaran, H., and L. F. James (2001): "Gibbs Sampling Methods for Stick-Breaking Priors," *Journal of the American Statistical Association*, 96(453), 161–173.

——— (2002): "Approximate Drichlet Process Computing in Finite Normal Mixtures," *Journal of Computational and Graphical Statistics*, 11(3), 508–532.

Jensen, M. J., M. Fisher, and P. Tkac (2015): "Mutual Fund Performance When Learning the Distribution of Stock-Picking Skill," .

Khan, S., M. Ponomareva, and E. Tamer (2016): "Identification of panel data models with endogenous censoring," *Journal of Econometrics*, 194(1), 57–75.

Li, M., and J. Tobias (2011): "Bayesian Methods in Microeconometrics," in *Oxford Handbook of Bayesian Econometrics*, ed. by J. Geweke, G. Koop, and H. van Dijk, no. 221-292. Oxford University Press, Oxford.

Liu, L. (2017): "Density Forecasts in Panel Data Models: A Semiparametric Bayesian Perspective," *Manuscript, University of Pennsylvania*.

Liu, L., H. R. Moon, and F. Schorfheide (2017): "Forecasting with Dynamic Panel Data Models," .

Llera, A., and C. Beckmann (2016): "Estimating an Inverse Gamma distribution," *arXiv preprint arXiv:1605.01019*.

Moon, H. R., and F. Schorfheide (2012): "Bayesian and frequentist inference in partially identified models," *Econometrica*, 80(755–782).

Robbins, H. (1956): "An Empirical Bayes Approach to Statistics," in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press, Berkeley and Los Angeles.

Robert, C. (1994): *The Bayesian Choice*. Springer Verlag, New York.

Rossi, P. E. (2014): *Bayesian Non- and Semi-parametric Methods and Applications*. Princeton University Press.

Shiu, J.-L., and Y. Hu (2013): "Identification and estimation of nonlinear dynamic panel data models with unobserved covariates," *Journal of Econometrics*, 175(2), 116 – 131.

Tanner, M. A., and W. H. Wong (1987): "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association*, 82(398), 528–540.

Wei, S. X. (1999): "A Bayesian Approach to Dynamic Tobit Models," *Econometric Reviews*, 18(4), 417–439.

Supplemental Appendix to "Forecasting with a Panel Tobit Model"

Laura Liu, Hyungsik Roger Moon, and Frank Schorfheide

A Data Set

Charge-off rates. The raw data are obtained from the website of the *Federal Reserve Bank of Chicago*:
https://www.chicagofed.org/banking/financial-institution-reports/commercial-bank-data.
The raw data are available at a quarterly frequency. The charge-off rates are defined as charge-offs divided by the stock of loans and constructed in a similar manner as in Tables A-1 and A-2 of Covas, Rump, and Zakrajsek (2014). However, the construction differs in the following dimensions: (i) We focus on charge-off rates instead of net charge-off rates. (ii) We standardize the charge-offs by the lagged stock of loans instead of the current stock of loans to reduce the timing issue.[7] (iii) For banks with domestic offices only (Form FFIEC 041), RIAD4645 (CI numerator) is not reported, so we switch to the corresponding variable RIAD4638.

The charge-offs are reported as year-to-date values. Thus, in order to obtain quarterly data, we take differences: $Q1 \mapsto Q1$, $(Q2 - Q1) \mapsto Q2$, $(Q3 - Q2) \mapsto Q3$, and $(Q4 - Q3) \mapsto Q4$. The loans are stock variables and no further transformation is needed. We multiply the charge-off rates by 400 to convert them into annualized percentages. We construct charge-off rates for the following types of loans:

- CI = commercial & industrial;
- CLD = construction & land development;
- MF = multifamily real estate;
- CRE = (nonfarm) nonresidential commercial real estate;
- HLC = home equity lines of credit (HELOCs);
- RRE = residential real estate, excluding HELOCs;

[7] According to bank report forms (e.g. FFIEC 041), the stocks of loans are given by quarterly averages. "For all items, banks have the option of reporting either (1) an average of DAILY figures for the quarter, or (2) an average of WEEKLY figures (i.e., the Wednesday of each week of the quarter)."

- CC = credit card;
- CON = consumer, excluding credit card loans.

Because in our econometric model we are relating the charge-off rates to local economic conditions, we restrict our analysis to "small" banks that operate in well-defined local markets. We include a bank in the sample if its assets are below one billion dollars. The raw data set contains missing observations and outliers that we are unable to explain with our econometric model. Thus, we proceed as follows to select a subset of observations from the raw data:

1. Eliminate banks for which domestic total assets are missing for all time periods in the sample.
2. Compute average non-missing domestic total assets and eliminate banks with average assets above 1 billion dollars.
3. For each loan category, eliminate banks for which the target charge-off rate is missing for at least one period of the sample.
4. For each loan category, eliminate banks for which the target charge-off rate is negative or greater than 400% for at least one period of the sample.
5. For loan category proceed as follows: First, for each bank, drop the two largest observations $y_{it},\ t = 0, \cdots, T+1$, and calculate the standard deviation (stdd) of the remaining observations. Then, eliminate a bank if any successive change $|y_{it} - y_{it-1}| + |y_{it+1} - y_{it}| >$ 10stdd. For $t = 0$ and $t = T + 1$, we only have one of the two terms and we set the other term in this selection criterion to zero.

The remaining sample sizes after each of these steps for the credit card loan charge-off rates as well as some summary statistics are reported in Table A-1.

Local Market. We use the annual *Summary of Deposits* data from the *Federal Deposit Insurance Corporation* to determine the local market for each bank. This data set contains information about the locations (at ZIP code level) in which deposits were made. Based on this information, for each bank in the charge-off data set we compute the amount of deposits received by state. We then associate each bank with the state from which it received the largest amount of deposits.

Unemployment Rate ($\mathbf{UR}_{it}$). Obtained from the *Bureau of Labor Statistics*. We use seasonally adjusted monthly data, time-aggregated to quarterly frequency by simple averaging. We divide the unemployment rate by 100 so that its variation is in the same order of magnitude as the house-price variation.

Housing Price Index ($\mathbf{HPI}_{it}$). Obtained from the *Federal Housing Finance Agency* on all transactions, not seasonally adjusted. The index is available at a quarterly frequency.

B Computational Details

B.1 Forecasts and Forecast Error Statistics

For each unit i, the posterior sampler generates draws $(y_{iT}^{*j}, \theta^j, \lambda_i^j)$, $j = 1, \ldots, M$, from the posterior distribution $p(y_{iT}^*, \theta, \lambda_i | Y_{1:N,0:T})$. Conditional on the parameter draws, the predictive density is a censored normal distribution. Define

$$\mu_{iT+h|T}^j = \lambda_i^j \sum_{l=0}^{h-1} \big(\rho^j\big)^l + \big(\rho^j\big)^h y_{iT}^{*j}, \quad \sigma_{iT+h|T}^{2,j} = \sigma^{2,j} \sum_{l=0}^{h-1} \big(\rho^j\big)^{2l}. \tag{A.1}$$

Then, a draw from the predictive distribution can be generated according to

$$y_{iT+h}^{*j} | (y_{iT}^{*j}, \theta^j, \lambda_i^j) \sim N\big(\mu_{iT+h|T}^j, \sigma_{iT+h|T}^{2,j}\big), \quad y_{iT+h}^j = y_{iT+h}^{*j} \mathbb{I}\{y_{iT+h}^{*j} \geq 0\}. \tag{A.2}$$

The expected value of the censored random variable y_{iT}^j is given by

$$\begin{aligned} &\mathbb{E}\big[y_{iT+h}^j | (y_{iT}^{*j}, \theta^j, \lambda_i^j)\big] \\ &= \left[\mu_{iT+h|T}^j + \sigma_{iT+h|T}^j \frac{\phi_N\big(\mu_{iT+h|T}^j / \sigma_{iT+h|T}^j\big)}{\Phi_N\big(\mu_{iT+h|T}^j / \sigma_{iT+h|T}^j\big)}\right] \Phi_N\big(\mu_{iT+h|T}^j / \sigma_{iT+h|T}^j\big) \\ &= \mu_{iT+h|T}^j \Phi_N\big(\mu_{iT+h|T}^j / \sigma_{iT+h|T}^j\big) + \sigma_{iT+h|T}^j \phi_N\big(\mu_{iT+h|T}^j / \sigma_{iT+h|T}^j\big), \end{aligned} \tag{A.3}$$

where $\phi_N(\cdot)$ and $\Phi_N(\cdot)$ are the pdf and the cdf of a standard $N(0,1)$. The probability of observing a zero is

$$\mathbb{P}\big[y_{iT+h}^j = 0 | (y_{iT}^{*j}, \theta^j, \lambda_i^j)\big] = \Phi_N\big(-\mu_{iT+h|T}^j / \sigma_{iT+h|T}^j\big). \tag{A.4}$$

Mean forecasts and forecasts of observing a zero can be approximated by Monte Carlo averaging:

$$\mathbb{E}\big[y_{iT+h}|Y_{1:N,0:T}\big] \approx \frac{1}{M}\sum_{j=1}^{M}\mathbb{E}\big[y_{iT+h}^{j}|(y_{iT}^{*j},\theta^{j},\lambda_{i}^{j})\big] \tag{A.5}$$

$$\mathbb{P}\big[y_{iT+h}=0|Y_{1:N,0:T}\big] \approx \frac{1}{M}\sum_{j=1}^{M}\mathbb{P}\big[y_{iT+h}^{j}=0|(y_{iT}^{*j},\theta^{j},\lambda_{i}^{j})\big]. \tag{A.6}$$

The log-predictive density can be approximated by

$$\ln p\big(y_{iT+h}|Y_{1:N,0:T}\big) \approx \begin{cases} \ln\mathbb{P}\big[y_{iT+h}=0|Y_{1:N,0:T}\big] & \text{if } y_{iT+h}=0 \\ \ln\Big(\frac{1}{M}\sum_{j=1}^{M}p_N\big(y_{iT+h}|\mu_{iT+h|T}^{j},\sigma_{iT+h|T}^{2,j}\big)\Big) & \text{otherwise} \end{cases}. \tag{A.7}$$

$1-\alpha$ Credible intervals can be obtained as follows: (i) sort $\{y_{iT+h}^{j}\}_{j=1}^{M}$ in increasing order and denote the elements of the sorted sequence by $y_{iT+h}^{(j)}$. (ii) For $j=1$ to $\lfloor\alpha M\rfloor$, compute the length of the j'th connected $1-\alpha$ interval $\Delta^{(j)}=y_{iT+h}^{(j+\lfloor(1-\alpha)M\rfloor)}-y_{iT+h}^{(j)}$. Then choose the shortest interval that contains $\lfloor(1-\alpha)M\rfloor$ draws.

Define the empirical distribution function based on the draws from the posterior predictive distribution as

$$\hat{F}(y_{iT+h})=\frac{1}{M}\sum_{j=1}^{M}\mathbb{I}\{y_{iT+h}^{(j)}\leq y_{iT+h}\}. \tag{A.8}$$

Then the probability integral transform associated with the density forecast of y_{iT+h} can be approximated as

$$PIT(y_{iT+h})\approx\hat{F}(y_{iT+h}). \tag{A.9}$$

The continuous ranked probability score associated with the density can be approximated as

$$CRPS(\hat{F},y_{iT+h})=\int_{0}^{\infty}\big(\hat{F}(x)-\mathbb{I}\{y_{iT+h}\leq x\}\big)^{2}dx. \tag{A.10}$$

Because the density $\hat{F}(y_{iT+h})$ is a step function, we can express the integral as a Riemann sum. To simplify the notation we drop the $iT+h$ subscripts and add an o superscript for the observed value at which the score is evaluated. Drawing a figure will help with the subsequent formulas. Define

$$M_{*}=\sum_{j=1}^{M}\mathbb{I}\{y^{(j)}\leq y^{o}\}.$$

Case 1: $M_* = M$. Then,

$$CRPS(\hat{F}, y^o) = \sum_{j=2}^{M} \big[\hat{F}(y^{(j-1)}) - 0\big]^2 (y^{(j)} - y^{(j-1)}) + \big[1 - 0\big]^2 (y^o - y^{(M)}). \tag{A.11}$$

Case 2: $M_* = 0$. Then,

$$CRPS(\hat{F}, y^o) = \big[0 - 1\big]^2 (y^{(1)} - y^o) + \sum_{j=2}^{M} \big[\hat{F}(y^{(j-1)}) - 1\big]^2 (y^{(j)} - y^{(j-1)}). \tag{A.12}$$

Case 3: $1 \leq M_* \leq M - 1$. Then,

$$\begin{aligned} & CRPS(\hat{F}, y^o) \qquad\qquad (A.13) \\ = & \sum_{j=2}^{M_*} \big[\hat{F}(y^{(j-1)}) - 0\big]^2 (y^{(j)} - y^{(j-1)}) + \big[\hat{F}(y^{(M_*)}) - 0\big]^2 (y^o - y^{(M_*)}) \\ & + \big[\hat{F}(y^{(M_*)}) - 1\big]^2 (y^{(M_*+1)} - y^o) + \sum_{j=M_*+2}^{M} \big[\hat{F}(y^{(j-1)}) - 1\big]^2 (y^{(j)} - y^{(j-1)}). \end{aligned}$$

Equivalently, based on Gneiting and Raftery (2007) Equation (21), we have

$$CRPS(\hat{F}, y^o) = \frac{1}{M} \sum_{j=1}^{M} |y^{(j)} - y^o| - \frac{1}{M^2} \sum_{1 \leq i < j \leq M} (y^{(j)} - y^{(i)}). \tag{A.14}$$

To see their equivalence, note that (A.14) can be re-written as follows:

$$\begin{aligned} & \frac{1}{M} \sum_{j=1}^{M} |y^{(j)} - y^o| - \frac{1}{M^2} \sum_{1 \leq i < j \leq M} (y^{(j)} - y^{(i)}) \qquad\qquad (A.15) \\ = & \frac{1}{M} \left[\sum_{j > M_*} y^{(j)} - \sum_{j \leq M_*} y^{(j)} + \big(M_* - (M - M_*)\big) y^o \right] - \frac{1}{M^2} \sum_{j=1}^{M} (2j - M - 1) y^{(j)}. \\ = & \frac{1}{M^2} \left[-\sum_{j=1}^{M_*} (2j - 1) y^{(j)} + \sum_{j=M_*+1}^{M} (2M - 2j + 1) y^{(j)} \right] + \frac{2M_* - M}{M} y^o. \end{aligned}$$

Considering that $\hat{F}(y^{(j)})$ is the empirical distribution, we have

$$\hat{F}(y^{(j)}) = \frac{j}{M}.$$

First, let us look at the more general Case 3. After replacing $\hat{F}(y^{(j)})$, the RHS of (A.13) becomes

$$\begin{aligned}
&\sum_{j=2}^{M_*}\big[\hat{F}(y^{(j-1)})-0\big]^2(y^{(j)}-y^{(j-1)})+\big[\hat{F}(y^{(M_*)})-0\big]^2(y^o-y^{(M_*)})\\
&+\big[\hat{F}(y^{(M_*)})-1\big]^2(y^{(M_*+1)}-y^o)+\sum_{j=M_*+2}^{M}\big[\hat{F}(y^{(j-1)})-1\big]^2(y^{(j)}-y^{(j-1)})\\
&= \sum_{j=2}^{M_*}\frac{(j-1)^2}{M^2}(y^{(j)}-y^{(j-1)})+\frac{M_*^2}{M^2}(y^o-y^{(M_*)})\\
&\quad+\frac{(M-M_*)^2}{M^2}(y^{(M_*+1)}-y^o)+\sum_{j=M_*+2}^{M}\frac{(M-(j-1))^2}{M^2}(y^{(j)}-y^{(j-1)})\\
&= \frac{1}{M^2}\Big[-y^{(1)}+\sum_{j=2}^{M_*}\big((j-1)^2-j^2\big)y^{(j)}+\sum_{j=M_*+1}^{M-1}\big((M-(j-1))^2-(M-j)^2\big)y^{(j)}\\
&\quad+y^{(M)}+\big(M_*^2-(M-M_*)^2\big)y^o\Big]\\
&= \frac{1}{M^2}\Big[-\sum_{j=1}^{M_*}(2j-1)y^{(j)}+\sum_{j=M_*+1}^{M}(2M-2j+1)y^{(j)}\Big]+\frac{2M_*-M}{M}y^o,
\end{aligned}$$

which is the same as (A.15). Similarly, for Case 1, after substituting $\hat{F}$, the RHS of (A.11) becomes

$$\begin{aligned}
&\sum_{j=2}^{M}\big[\hat{F}(y^{(j-1)})-0\big]^2(y^{(j)}-y^{(j-1)})+\big[1-0\big]^2(y^o-y^{(M)})\\
&= \sum_{j=2}^{M}\frac{(j-1)^2}{M^2}(y^{(j)}-y^{(j-1)})+(y^o-y^{(M)})\\
&= \frac{1}{M^2}\big[-y^{(1)}+\sum_{j=2}^{M}\big((j-1)^2-j^2\big)y^{(j)}\big]+y^o\\
&= -\frac{1}{M^2}\sum_{j=1}^{M_*}(2j-1)y^{(j)}+y^o,
\end{aligned}$$

which is equal to (A.15) when $M_*=M$. And for Case 2, after substituting $\hat{F}$, the RHS of

(A.12) becomes

$$\begin{aligned}
&\big[0-1\big]^2(y^{(1)}-y^o)+\sum_{j=2}^{M}\big[\hat{F}(y^{(j-1)})-1\big]^2(y^{(j)}-y^{(j-1)}) \\
&= (y^{(1)}-y^o)+\sum_{j=2}^{M}\frac{(M-(j-1))^2}{M^2}(y^{(j)}-y^{(j-1)}) \\
&= \frac{1}{M^2}\big[\sum_{j=1}^{M-1}\big((M-(j-1))^2-(M-j)^2\big)y^{(j)}+y^{(M)}\big]-y^o \\
&= \frac{1}{M^2}\sum_{j=1}^{M}(2M-2j+1)y^{(j)}-y^o,
\end{aligned}$$

which is equal to (A.15) when $M_*=0$.

Table A-1: Sample Sizes After Selection Steps and Summary Statistics for Loan Charge-Off Rates

		Sample Sizes					Cross-sectional Statistics			
Loan	t_0	Initial	Step1	Step2	Step3	Step5	% 0s	Mean	75%	Max
CLD	2007Q3	7,903	7,903	7,299	3,290	1,304	77	1.5	0.0	106.8
CLD	2007Q4	7,835	7,835	7,219	3,244	1,264	74	1.9	0.1	106.8
CLD	2008Q1	7,692	7,692	7,084	3,204	1,257	71	2.2	0.5	180.2
RRE	2007Q1	7,991	7,991	7,393	6,260	2,654	77	0.2	0.0	33.1
RRE	2007Q2	7,993	7,993	7,383	6,152	2,576	76	0.3	0.0	33.1
RRE	2007Q3	7,903	7,903	7,299	6,193	2,606	73	0.3	0.0	35.9
RRE	2007Q4	7,835	7,835	7,219	6,146	2,581	70	0.4	0.1	69.2
RRE	2008Q1	7,692	7,692	7,084	6,106	2,561	68	0.4	0.2	45.6
RRE	2008Q2	7,701	7,701	7,080	6,029	2,492	67	0.4	0.2	63.6
RRE	2008Q3	7,631	7,631	7,008	6,052	2,577	65	0.5	0.3	39.2
RRE	2008Q4	7,559	7,559	6,938	6,005	2,600	63	0.5	0.3	45.6
RRE	2009Q1	7,480	7,480	6,849	5,971	2,588	62	0.5	0.3	45.0
RRE	2009Q2	8,103	8,103	7,381	5,895	2,536	62	0.5	0.3	45.0
RRE	2009Q3	8,016	8,016	7,302	5,899	2,563	61	0.5	0.4	47.6
RRE	2009Q4	7,940	7,940	7,229	5,846	2,553	60	0.5	0.4	45.0
RRE	2010Q1	7,770	7,770	7,077	5,765	2,494	61	0.5	0.4	45.0
CC	2001Q2	9,031	9,031	8,532	1,691	875	33	3.4	4.7	162.5
CC	2001Q3	8,995	8,995	8,491	1,666	844	33	3.4	4.8	88.9
CC	2001Q4	8,887	8,887	8,382	1,636	836	34	3.3	4.6	88.9
CC	2002Q1	8,723	8,723	8,228	1,612	814	35	3.3	4.4	400.0
CC	2002Q2	8,823	8,823	8,312	1,670	817	38	3.2	4.3	88.9
CC	2002Q3	8,805	8,805	8,286	1,631	821	38	3.2	4.3	88.9
CC	2002Q4	8,728	8,728	8,199	1,606	813	39	3.1	4.1	88.9
CC	2003Q1	8,611	8,611	8,077	1,573	811	40	3.0	4.0	128.5
CC	2003Q2	8,754	8,754	8,203	1,544	787	40	3.0	3.9	136.1
CC	2003Q3	8,755	8,755	8,198	1,513	754	41	2.9	3.8	136.1
CC	2003Q4	8,671	8,671	8,120	1,500	724	42	2.8	3.6	136.1
CC	2004Q1	8,526	8,526	7,989	1,468	707	43	2.7	3.6	136.1
CC	2004Q2	8,662	8,662	8,108	1,440	677	42	2.8	3.6	136.1
CC	2004Q3	8,626	8,626	8,067	1,411	664	43	2.7	3.5	136.1
CC	2004Q4	8,552	8,552	7,989	1,391	657	44	2.6	3.3	140.9
CC	2005Q1	8,384	8,384	7,829	1,369	639	44	2.5	3.2	151.3
CC	2005Q2	8,507	8,507	7,938	1,332	611	44	2.6	3.2	175.0
CC	2005Q3	8,482	8,482	7,897	1,315	596	45	2.6	3.2	175.0
CC	2005Q4	8,404	8,404	7,816	1,290	604	46	2.6	3.2	210.5
CC	2006Q1	8,263	8,263	7,674	1,275	614	47	2.6	3.1	175.0
CC	2006Q2	8,307	8,307	7,708	1,247	594	47	2.7	3.2	269.2
CC	2006Q3	8,240	8,240	7,639	1,231	594	46	2.8	3.4	269.2
CC	2006Q4	8,137	8,137	7,537	1,211	595	45	3.0	3.6	269.2

Notes: This table provides summary statistics for samples with cross-sectional dimension $N > 400$ and percentage of zeros less than 80%. The date assigned to each panel refers to $t = t_0$, which is the conditioning information used to initialize the lag in the dynamic Tobit. We assume that $T = 10$, which means that each sample has 12 observations. The descriptive statistics are computed across N and T dimension of each panel.

Table A-1: Sample Sizes After Selection Steps and Summary Statistics for Loan Charge-Off Rates (cont.)

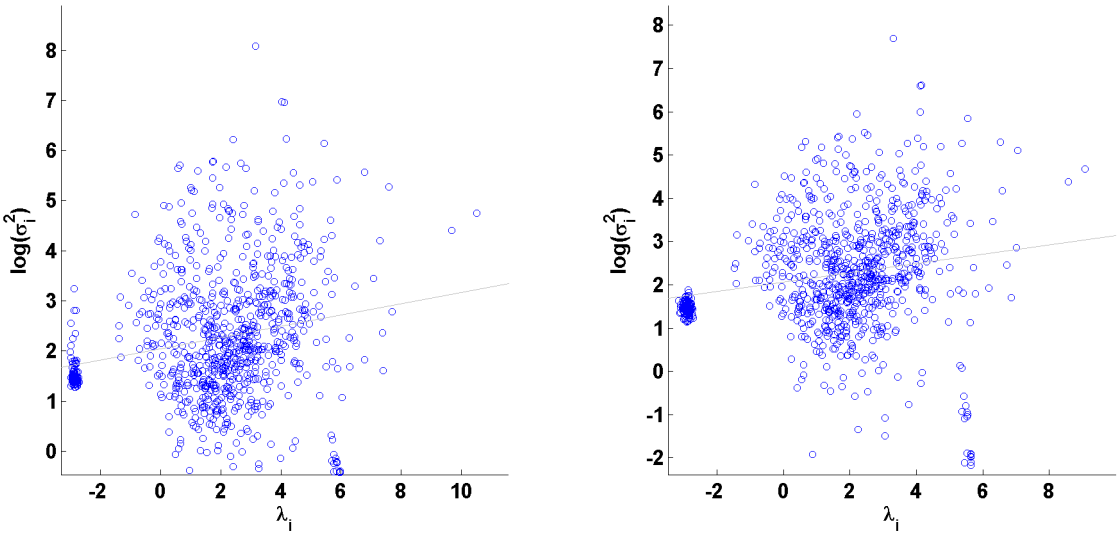
		Sample Sizes					Cross-sectional Statistics			
Loan	t_0	Initial	Step1	Step2	Step3	Step5	% 0s	Mean	75%	Max
CC	2007Q1	7,991	7,991	7,393	1,197	574	44	3.2	3.9	269.2
CC	2007Q2	7,993	7,993	7,383	1,173	561	43	3.3	4.1	269.2
CC	2007Q3	7,903	7,903	7,299	1,159	544	44	3.2	4.2	175.0
CC	2007Q4	7,835	7,835	7,219	1,133	534	43	3.3	4.2	175.0
CC	2008Q1	7,692	7,692	7,084	1,123	527	44	3.3	4.2	175.0
CC	2008Q2	7,701	7,701	7,080	1,101	512	45	3.2	4.1	158.3
CC	2008Q3	7,631	7,631	7,008	1,096	509	44	3.1	4.0	158.3
CC	2008Q4	7,559	7,559	6,938	1,082	506	45	3.1	3.9	149.4
CC	2009Q1	7,480	7,480	6,849	1,059	498	46	3.0	3.7	147.3
CC	2009Q3	8,016	8,016	7,302	1,042	492	47	2.7	3.5	77.6
CC	2009Q4	7,940	7,940	7,229	1,032	479	49	2.7	3.3	400.0
CC	2010Q1	7,770	7,770	7,077	1,020	459	49	2.5	3.2	100.0
CON	2009Q2	8,103	8,103	7,381	5,837	2,600	77	0.4	0.0	77.4
CON	2009Q3	8,016	8,016	7,302	5,872	2,672	71	0.5	0.2	202.2
CON	2009Q4	7,940	7,940	7,229	5,814	2,723	65	0.5	0.5	202.2
CON	2010Q1	7,770	7,770	7,077	5,735	2,680	58	0.7	0.7	202.2

Notes: This table provides summary statistics for samples with cross-sectional dimension $N > 400$ and percentage of zeros less than 80%. The date assigned to each panel refers to $t = t_0$, which is the conditioning information used to initialize the lag in the dynamic Tobit. We assume that $T = 10$, which means that each sample has 12 observations. The descriptive statistics are computed across N and T dimension of each panel.

C Additional Empirical Results

Figure A-1: Posterior Means of λ_i vs. σ_i^2

Normal $\pi(\lambda)$, InvGam $\pi(\sigma^2)$ Flexible $\pi(\lambda), \pi(\sigma^2)$

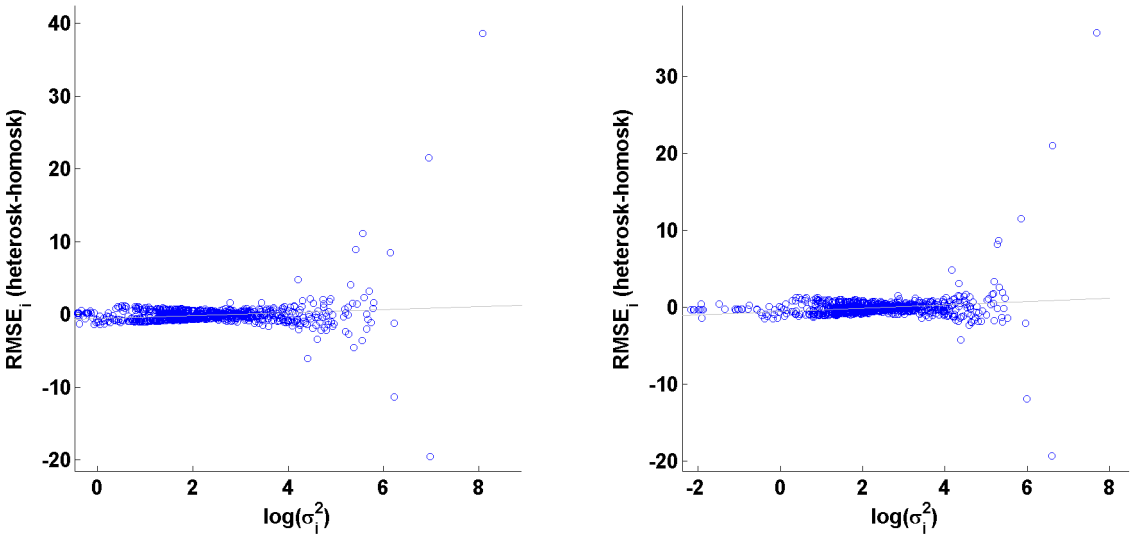


Notes: Credit card charge-off rates, heteroskedastic specification. The estimation sample for the dynamic Tobit models ranges from 2001Q2 to 2003Q4 ($T = 10$).

Figure A-1 showes that here is a cluster of banks for which $\hat{\lambda}_i$ is approximately -2.4 and $\ln \hat{\sigma}_i^2$ is approximately 1.5.

Figure A-2: Posterior Means of λ_i vs. σ_i^2

Normal $\pi(\lambda)$, InvGam $\pi(\sigma^2)$ Flexible $\pi(\lambda), \pi(\sigma^2)$



Notes: Credit card charge-off rates. The estimation sample for the dynamic Tobit models ranges from 2001Q2 to 2003Q4 ($T = 10$).