FORECASTING WITH A PANEL TOBIT MODEL

Laura Liu
Hyungsik Roger Moon
Frank Schorfheide

ABSTRACT

We use a dynamic panel Tobit model with heteroskedasticity to generate point, set, 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 as prior to construct Bayes forecasts for the
individual time series. We construct set forecasts that explicitly target the average coverage
probability for the cross-section. We present a novel application in which we forecast bank-level
charge-off rates for credit card and residential real estate loans, comparing various versions of the
panel Tobit model.

Laura Liu
Department of Economics
Indiana University
100 S. Woodlawn Avenue
Bloomington, IN 47405
liuyu1237@gmail.com

Hyungsik Roger Moon
University of Southern California
Department of Economics
KAP 300
University Park Campus
Los Angeles, CA 90089
hyungsikmoon@gmail.com

Frank Schorfheide
University of Pennsylvania
Department of Economics
The Ronald O. Perelman Center for
Political Science and Economics (PCPSE), Room 621
133 South 36th Street
Philadelphia, PA 19104-6297
and NBER
schorf@ssc.upenn.edu

1   Introduction

This paper considers the problem of forecasting a large collection of short time series with censored observations. In the 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.

In a panel data setting, we often capture cross-sectional heterogeneity in the data with unit-specific parameters. The more precisely these heterogeneous coefficients are estimated, the more accurate are the forecasts. The challenge in forecasting panels with a short time dimension is that they do not contain a lot of information about unit-specific parameters. A natural way of adding information to the estimation of the unit-specific parameters 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 resulting reduction in sampling variance might dominate the introduction of bias. 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 creates 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

---

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

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 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 (2017a,b), and Liu, Moon, and Schorfheide (2019).

Unfortunately, Tweedie's formula does not extend to nonlinear panel data models. Thus, 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 (2018).

The contributions of our paper are threefold. First, we extend the implementation of the full Bayesian estimation in Liu (2018) to the dynamic panel Tobit model with heteroskedastic innovations and correlated random effects. Second, in regard to interval forecasting, we construct forecasts that target average posterior coverage probability across all units in our panel instead of pointwise coverage probability for each unit. These forecasts are obtained from highest posterior density sets that are constructed using the same threshold for each unit instead of unit-specific thresholds. Because the predictive distributions associated with the Tobit models are mixtures of discrete and continuous distributions, interval forecasts may also take the form of the union of an interval and the value zero or simply the value zero (singleton). For this reason we will refer to these forecasts as set instead of interval forecasts throughout this paper. Both in the Monte Carlo study and the empirical application the proposed Bayesian set forecasts have good frequentist coverage properties in the cross-section. This basic insight has been used in the literature on nonparametric function estimation and dates back to Wahba (1983) and Nychka (1988).

Third, we present a novel application in which we forecast bank-level charge-off rates. Our empirical analysis is based on more than 100 short panel data sets with a time dimension of $T = 10$. These panel data sets include predominantly credit card (CC) and residential

real estate (RRE) loans and cover various (overlapping) time periods. For each data set, we consider several model specifications that differ with respect to assumptions about the correlation between heterogeneous coefficients and initial conditions (random versus correlated random effects), heterogeneity in the variances of the error terms (homoskedasticity versus heteroskedasticity) and the distributional assumptions about the distribution of the heterogeneous coefficients (Normal versus flexible). The model with flexibly modeled correlated random effects and heteroskedasticity serves as a benchmark. We conduct a variety of posterior predictive checks to document that this model is able to capture the salient features of our data.

We compare Bayesian density, set, and point forecasts for the various specifications. Forecasts from a Tobit model that allows for heterogeneous coefficients, clearly dominate forecasts from a model that imposes homogeneity. Modeling heteroskedasticity improves the density and set forecasts substantially. Set forecasts are evaluated based on average (in the cross section) length and coverage probability. We show how to construct Bayesian set forecasts that target average posterior coverage probability in the cross section. With respect to average length, these forecasts dominate traditional set forecasts that are constructed by targeting pointwise posterior coverage probability. Our empirical model includes local house prices and unemployment rates as additional predictors. The estimated effects of these predictors in the benchmark specification are qualitatively and quantitatively plausible.

Our paper relates to several branches of the literature. The papers most closely related are Gu and Koenker (2017a,b), Liu (2018), and Liu, Moon, and Schorfheide (2019). 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 (2018) has a natural extension to nonlinear models. Liu, Moon, and Schorfheide (2019), 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 the 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 (2018) 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). Our benchmark model is most closely related to the Bayesian semiparametric panel Tobit model of Li and Zheng (2008). They used their model to study female labor supply and estimated average partial effects and average transition probabilities. We generalize Li and Zheng's model by introducing heteroskedasticity through a latent unit-specific error variance and allowing for a more flexible form of correlated random effects. As mentioned previously, the former is very important for the density and set forecast performance. 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 model the unknown distribution of the heterogeneous coefficients (intercepts and innovation variances) as 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 nonparametric density modeling using Dirichlet process mixtures (DPM).[2] 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). If we restrict the number of mixture components to be equal to one, then we obtain a Normal correlated random effects model as a special case.

The remainder of this paper is organized as follows. Section 2 presents the specification of our dynamic panel Tobit model, a characterization of the posterior predictive distribution for future observations, and discusses the construction and evaluation of point, density, and set forecasts. Section 3 provides details on how we model the correlated random effects distribution and heteroskedasticity. It also presents the prior distributions for the parametric and flexible components of the model, and outlines a posterior sampler. We conduct a Monte Carlo experiment in Section 4 to examine the performance of the proposed techniques in a controlled environment. The empirical application in which we forecast charge-off rates on various types of loans for a panel of banks is presented in Section 5. Finally, Section 6 concludes. A description of the data sets, additional empirical results, and some derivations are relegated to the Online Appendix.

---

[2]Keane and Stavrunova (2011) introduce a smooth mixture of Tobits to model a cross-section of healthcare expenditures. Our model is related, but different in that we are using a DPM to average across different intercept values and innovation variances.

2 Model Specification and Forecast Evaluation

Throughout this paper we consider the following dynamic panel Tobit model with heterogeneous intercepts and innovation variances:

$$y_{it} = y_{it}^* \mathbb{I}\{y_{it}^* \geq 0\}, \quad y_{it}^* | (y_{it-1}^*, x_{it}, \lambda_i, \sigma_i^2, \rho, \beta) \sim N\big(\lambda_i + \rho y_{it-1}^* + \beta' x_{it}, \, \sigma_i^2\big), \qquad (1)$$

where $i = 1, \ldots, N$, $t = 1, \ldots, T$, and $\mathbb{I}\{y \geq a\}$ is the indicator function that is equal to one if $y \geq a$ and equal to zero otherwise. The $n_x \times 1$ vector x_{it} comprises a set of regressors and we define the homogeneous parameter $\theta = [\rho, \beta']'$. It is assumed that conditional on the heterogenous parameters and the regressors x_{it} the observations y_{it} are cross-sectionally independent. Our specification uses the lagged latent variable y_{it-1}^* on the right-hand side because it is more plausible for our empirical application. The Bayesian computations described in Section 3.2 below can be easily adapted to the alternative model, in which the lagged censored variable y_{it-1} appears on the right-hand side.

We model the heterogeneous parameters as correlated random effects (CRE) with density

$$p(\lambda_i, y_{i0}^*, \sigma_i^2 | x_{i0}, \xi),$$

assuming cross-sectional independence of the heterogeneous coefficients.[3] Here ξ is the parameter vector that indexes the CRE distribution, which we flexibly represent by a mixture of Normal distributions. The model is completed with the specification of a prior distribution for (θ, ξ). In the remainder of this section, we will discuss the posterior predictive density in Section 2.1, point and density forecasting in Section 2.2, and the construction and evaluation of set forecasts in Section 2.3.

2.1 Posterior Predictive Densities

Our goal is to generate forecasts of $Y_{1:N,T+h} = \{y_{1,T+h}, \ldots, y_{N,T+h}\}$ conditional on the observations

$$
\begin{aligned}
Y_{1:N,0:T} &= \big\{(y_{10}, \ldots, y_{N0}), \ldots, (y_{1T}, \ldots, y_{NT})\big\} \\
X_{1:N,0:T+h} &= \big\{(x_{10}, \ldots, x_{N0}), \ldots, (x_{1T+h}, \ldots, x_{NT+h})\big\}.
\end{aligned}
$$

---

[3]In principle the conditional distribution of $(\lambda_i, y_{i0}^*, \sigma_i^2)$ could also depend on x_{it} for $t > 0$, but in our applications we simply condition on the initial value of the regressors.

Because y^*_{it+h} depends on x_{it+h} we are implicitly assuming that the sequence $X_{1:N,T+1:T+h}$ is known once the forecast is made. In our application we focus on $h = 1$-step-ahead forecasts and define x_{it} as house price growth and unemployment growth in period $t - 1$.

Define $\lambda_{1:N} = (\lambda_1, \ldots, \lambda_N)$ and $\sigma^2_{1:N} = (\sigma^2_1, \ldots, \sigma^2_N)$. Assuming that $X_{1:N,T+1:T+h}$ contains no information about $(\lambda_{1:N}, \sigma^2_{1:N}, \theta, \xi)$ and is strictly exogenous, we can write the posterior distribution of the parameters and time-T latent variables as

$$p\big(Y^*_{1:N,T}, \lambda_{1:N}, \sigma^2_{1:N}, \theta, \xi | Y_{1:N,0:T}, X_{1:N,0:T+h}\big) \tag{2}$$

$$\propto \left[\prod_{i=1}^N \int p(Y_{i,1:T}|Y^*_{i,1:T}) p\big(Y^*_{i,1:T}|y^*_{i,0}, X_{i,1:T}, \lambda_i, \sigma^2_i, \theta\big) \right.$$

$$\left. \times p(y_{i0}|y^*_{i0}) p\big(\lambda_i, y^*_{i0}, \sigma^2_i | x_{i0}, \xi\big) \cdot dY^*_{i,0:T-1} \right] p(\theta) p(\xi),$$

where $\propto$ denotes proportionality. Here, $p(Y_{i,1:T}|Y^*_{i,1:T})$ and $p(y_{i0}|y^*_{i0})$ represent the censoring (in slight abuse of notation). The distribution of $y_{it}|y^*_{it}$ is a unit point mass that is located at 0 if $y^*_{it} \leq 0$ or at y^*_{it} if $y^*_{it} > 0$. The density $p(Y^*_{i,1:T}|\cdot)$ can be derived from (1), $p(\lambda_i, y^*_{i0}, \sigma^2_i|\cdot)$ is the CRE density, and $p(\theta)$ and $p(\xi)$ are priors for θ and ξ, respectively.

The posterior predictive distribution for unit i is given by

$$p(y_{iT+h}|Y_{1:N,0:T}, X_{1:N,0:T+h}) \tag{3}$$

$$= \int \int \int p(y_{iT+h}|y^*_{iT+h}) p\big(y^*_{iT+h}|y^*_{iT}, X_{i,T+1:T+h}, \lambda_i, \sigma^2_i, \theta\big)$$

$$\times p\big(y^*_{iT}, \lambda_i, \sigma^2_i, \theta, \xi | Y_{1:N,0:T}, X_{1:N,0:T+h}\big) \cdot dy^*_{iT} \cdot d(\lambda_i, \sigma^2_i) \cdot d(\theta, \xi).$$

Draws from the density $p(y^*_{iT+h}|y^*_{iT}, X_{i,T+1:T+h}, \lambda_i, \sigma^2_i, \theta)$ can be generated by forward simulation of the autoregressive law of motion for y^*_{it} in (1). Using decision-theoretic arguments, one can derive point and set forecasts from the predictive density $p(y_{iT+h}|Y_{i,0:T}, X_{i,0:T+h})$. To simplify the notation, we will drop $X_{1:N,0:T+h}$ from the conditioning set in the remainder of this section. We denote expectations and probabilities under the posterior predictive distribution by $\mathbb{E}^{y_{iT+h}}_{Y_{1:N,0:T}}[\cdot]$ and $\mathbb{P}^{y_{iT+h}}_{Y_{1:N,0:T}}\{\cdot\}$, respectively. More generally, we use subscripts to denote the conditioning set and superscripts to denote the random variables over which the operators integrate.

The predictive distribution is a mixture of a point mass at zero and a continuous distri-

bution for realizations of y_{iT+h} that are greater than zero:

$$p(y_{iT+h}|Y_{1:N,0:T}) = \mathbb{P}^{y_{iT+h}}_{Y_{1:N,0:T}}\{y_{iT+h} = 0\}\delta_0(y_{iT+h}) + p_c(y_{iT+h}|Y_{1:N,0:T})\mathbb{I}\{y_{iT+h} \geq 0\}. \quad (4)$$

Here $\delta_0(y)$ is the Dirac function with the property $\delta_0(y) = 0$ for $y \neq 0$ and $\int \delta_0(y)dy = 1$. The density $p_c(y_{iT+h}|Y_{1:N,0:T})$ represents the continuous part of the predictive distribution.

2.2 Point and Density Forecasts

Point forecasts are evaluated under the quadratic compound loss function

$$L_N(\widehat{Y}_{1:N,T+h|T}, Y_{1:N,T+h}) = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_{iT+h} - y_{iT+h})^2.$$

The posterior risk associated with a particular model specification M is given by

$$R_N(\widehat{Y}_{1:N,T+h|T}, Y_{1:N,T+h}) = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}^{y_{i,T+h}}_{Y_{1:N,0:T}}\left[(\hat{y}_{i,T+h|T} - y_{iT+h})^2|M\right]$$

and minimized by the posterior mean forecasts

$$\hat{y}^o_{i,T+h|T}(M) = \mathbb{E}^{y_{iT+h}}_{Y_{1:N,0:T}}[\hat{y}_{iT+h}|M], \quad i = 1,\ldots,N.$$

In the Monte Carlo study and the empirical application we report the root mean squared error (RMSE)

$$\text{RMSE}_h(M) = \sqrt{L_N\left(\widehat{Y}^o_{1:N,T+h|T}(M), Y_{1:N,T+h}\right)}. \quad (5)$$

To compare the density forecast performance of various model specifications M we report the average log predictive scores

$$LPS_h(M) = \frac{1}{N}\sum_{i=1}^{N}\ln\left(\mathbb{I}\{y_{iT+h} = 0\} \cdot \mathbb{P}^{y_{iT+h}}_{Y_{1:N,0:T}}\{y_{iT+h} = 0|M\}\right. \quad (6)$$
$$\left. +\mathbb{I}\{y_{iT+h} > 0\}p(y_{iT+h}|Y_{1:N,0:T})\right)$$

and contiuous ranked probability scores (CRPSs). The CRPS measures the L_2 distance between the cumulative density function $F^{y_{iT+h}}_{Y_{1:N,0:T}}(y|M)$ 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+h}. Then,

$$CRPS_h(M) = \frac{1}{N} \sum_{i=1}^{N} \int_0^\infty \left(F_{Y_{1:N,0:T}}^{y_{iT+h}}(y|M) - \mathbb{I}\{y_{iT+h} \le y\} \right)^2 dy. \tag{7}$$

Both LPS and CRPS are proper scoring rules, meaning that it is optimal for the forecaster to truthfully reveal her predictive density.

2.3 Constructing and Evaluating Set Forecasts

We construct set forecasts $C_{i,T+h|T}(Y_{1:N,0:T})$ from the posterior predictive distribution $p(y_{iT+h}|Y_{1:N,0:T})$ in (3). We consider the following types of set forecasts:

$$\{0\}, \quad [0, b_i], \quad \{0\} \cup [a_i, b_i], \quad \emptyset. \tag{8}$$

The $\{0\}$ values are generated by the discrete portion of the predictive density, whereas the interval components are obtained from the continuous portion of the predictive density; see the decomposition in (4). Throughout this paper, we restrict the interval portions to be connected. The empty set $\emptyset$ may arise for some units if we target an average coverage probability in the cross section. We measure the volume of the sets using the Euclidean length of the interval portions of the sets, which are 0, b_i, $b_i - a_i$, and 0, respectively.

The assessment of the set forecasts in our simulation study and the empirical application is based on the cross-sectional coverage frequency

$$\frac{1}{N} \sum_{i=1}^{N} \mathbb{I}\{y_{iT+h} \in C_{i,T+h|T}(Y_{1:N,0:T})\} \tag{9}$$

and the average length of the sets $C_{i,T+h|T}(Y_{1:N,0:T})$. Rather than trading off average length against deviations of average coverage frequency from the nominal coverage probability in a single criterion, we will simply report both.[4]

To generate the set forecasts, we adopt a Bayesian approach and require that the probability of $\{y_{iT+h} \in C_{i,T+h|T}(Y_{1:N,0:T})\}$ conditional on having observed $Y_{1:N,0:T}$ is at least $1 - \alpha$. Given that the estimation of the Tobit model is executed with Bayesian techniques, the use of posterior predictive credible sets is natural. Moreover, it is known from the literature on

---

[4]Various approaches to rank interval forecasts are discussed in Askanazi, Diebold, Schorfheide, and Shin (2018).

nonparametric function estimation, that Bayesian credible sets can have good frequentist average coverage probability; see Wahba (1983) and Nychka (1988). We will distinguish between forecasts that are constructed to satisfy the coverage probability constraint pointwise, that is,

$$\mathbb{P}^{y_{iT+h}}_{Y_{1:N,0:T}}\left\{y_{iT+h} \in C_{i,T+h|T}(Y_{1:N,0:T})\right\} \geq 1 - \alpha \quad \text{for all } i,$$

and sets that are constructed to satisfy the constraint on average:

$$\frac{1}{N}\sum_{i=1}^{N}\mathbb{P}^{y_{iT+h}}_{Y_{1:N,0:T}}\left\{y_{iT+h} \in C_{i,T+h|T}(Y_{1:N,0:T})\right\} \geq 1 - \alpha.$$

The latter approach allows the sets $C_{i,T+h|T}(Y_{1:N,0:T})$ for some units i to be "shortened" in the sense that their posterior credible level drops below $1 - \alpha$, whereas sets for other units are "lengthened."

In the remainder of this subsection we describe how the endpoints a_i and b_i of the interval portion of the set forecasts are determined. Let $\alpha_{i0} = \mathbb{P}^{y_{iT+h}}_{Y_{1:N,0:T}}\{y_{iT+h} = 0\}$ and define the interval length $l_i = b_i - a_i$. The problem of minimizing the average length subject to the constraint on coverage probability is given by

$$\min_{a_i,l_i} \ l_i \quad \text{s.t.} \quad \int_{a_i}^{a_i+l_i} p_c(y_{iT+h}|Y_{1:N,0:T})dy_{iT+h} = 1 - \alpha - \alpha_{i0}, \qquad (10)$$
$$a_i \geq 0.$$

Because $p_c(y_{iT+h}|Y_{1:N,0:T})$ represents the continuous part of the predictive density, we deduce that $l_i > 0$ provided that $1 - \alpha - \alpha_{i0} > 0$. The first-order conditions are

$$(l_i) \ : \ 1 = \mu_\alpha p_c(a_i + l_i|Y_{1:N,0:T})$$
$$(a_i) \ : \ 0 = \mu_\alpha\left[p_c(a_i + l_i|Y_{1:N,0:T}) - p_c(a_i|Y_{1:N,0:T})\right] + \mu_{i,a}, \quad \mu_{i,a}\cdot a_i = 0, \quad \mu_{i,a} \geq 0,$$

where μ_α and $\mu_{i,a}$ are the Lagrange multipliers for the coverage probability constraint and the non-negativity constraint for a_i, respectively. If $\mu_{i,a} > 0$ and $a_i = 0$, then the set forecast is of the form $[0, b_i]$; otherwise it takes the form $\{0\} \cup [a_i, b_i]$, where $p_c(b_i|Y_{1:N,0:T}) = p_c(a_i|Y_{1:N,0:T})$. If, in addition, the continuous portion of the predictive density is unimodal, then the interval part of the set forecast is the HPD interval $\{y_{iT+h} \mid p_c(y_{iT+h}|Y_{1:N,0:T}) \geq C\}$, where $C = p_c(a_i + l_i|Y_{1:N,0:T})$.

Targeting average coverage probability changes the derivation as follows. Define $\bar{\alpha}_0 =$

$\frac{1}{N}\sum_{i=1}^{N}\mathbb{P}_{Y_{1:N,0:T}}^{y_{iT+h}}\{y_{iT+h}=0\}$. If $\bar{\alpha}_0 > 1 - \alpha$ then we can set the forecasts for some units to $C_{i,T+h|T}(Y_{1:N,0:T}) = \emptyset$ and for the other units to $C_{i,T+h|T}(Y_{1:N,0:T}) = \{0\}$ such that the constraint on the coverage probability is satisfied. Otherwise, we need to solve

$$\min_{\{a_i, l_i\}_{i=1}^{N}} \frac{1}{N}\sum_{i=1}^{N} l_i \quad \text{s.t.} \quad \frac{1}{N}\sum_{i=1}^{N}\int_{a_i}^{a_i+l_i} p_c(y_{iT+h}|Y_{1:N,0:T})dy_{iT+h} = 1 - \alpha - \bar{\alpha}_0, \qquad (11)$$
$$a_i \geq 0 \text{ and } l_i \geq 0 \; \forall i.$$

Because it is conceivable that the solution implies that for some i the forecast is given by $\{0\}$, we now also account for the constraint $l_i \geq 0$ with Lagrange multiplier μ_{il}. The first-order conditions take the form

$$(l_i) \;\; : \;\; \frac{1}{N} = \frac{\mu_\alpha}{N}p(a_i + l_i|Y_{1:N,0:T}) + \mu_{i,l}, \quad \mu_{i,l} \cdot l_i = 0, \quad \mu_{i,l} \geq 0$$
$$(a_i) \;\; : \;\; 0 = \frac{\mu_\alpha}{N}\big[p(a_i + l_i|Y_{1:N,0:T}) - p(a_i|Y_{1:N,0:T})\big] + \mu_{i,a}, \quad \mu_{i,a} \cdot a_i = 0, \quad \mu_{i,a} \geq 0.$$

The first-order condition for l_i implies that for all units with $l_i > 0$ the densities at the endpoint $b_i = a_i + l_i$ are identical. If the continuous parts of the predictive densities are unimodal for each i, then the first-order conditions imply that there is a common HPD threshold C for all i such that the interval components of the set forecasts are of the form $\big\{y_{iT+h} \,|\, p_c(y_{iT+h}|Y_{1:N,0:T}) \geq C\big\}$.

3 Correlated Random Effects, Priors, and Posteriors

We provide a characterization of the CRE distribution $p(\lambda_i, y_{i0}^*, \sigma_i^2|x_{i0}, \xi)$ and a specification of the prior distribution for (θ, ξ) in Section 3.1. Section 3.2 provides a description of the posterior sampler and Section 3.3 discusses potential generalizations of the dynamic panel Tobit model.

3.1 Correlated Random Effects and Prior Distributions

We will start out with a CRE specification in which we allow for heteroskedasticity, the heterogeneous parameters λ_i and σ_i^2 are independent, and $p(\lambda_i, y_{i0}^*, \sigma_i^2|x_{i0}, \xi)$ belongs to a flexible family of mixtures. Subsequently, we will consider various special cases of this

model, including a more tightly parameterized $p(\lambda_i, y_{i0}^*, \sigma_i^2 | x_{i0}, \xi)$, random effects (RE), and homoskedasticity. At last, we will consider a prior under which λ_i and σ_i^2 are correlated.

Before we estimate the model specifications, we standardize the regressors x_{it} to have zero mean and unit variance. This will facilitate the scaling of the prior distributions. Our prior depends on various hyperparameters that can be chosen by the user. The prior variances for the conditional mean parameters are scaled by τ. We set $\tau = 5$ in the simulation study in Section 4 and the empirical analysis in Section 5, which makes these priors fairly diffuse. In order to scale the prior distribution for σ_i^2, we define the cross-sectional average of the time-series variances of y_{it}: let $V^* = \frac{1}{N} \sum_{i=1}^{N} \mathbb{V}_i(y_{it})$. For all other hyperparameters we report a default choice based on the use of similar priors in other studies. These default choices seem to work well in the application in that they provide a reasonable trade-off between flexibility and model complexity.

The description of the CRE and prior distribution involves various parametric probability distributions in addition to the Normal distribution that appeared in (1). We use $B(a, b)$, $G(a, b)$, and $IG(a, b)$ to denote the Beta, Gamma, and Inverse Gamma distributions, respectively. The pair (θ, σ^2) has a Normal-Inverse-Gamma distribution $NIG(m, v, a, b)$ if $\sigma^2 \sim IG(a, b)$ and $\theta | \sigma^2 \sim N(m, \sigma^2 v)$. Finally, the pair (Φ, Σ) has a matricvariate Normal-Inverse-Wishart distribution $MNIW(M, V, S, \nu)$ if $\Sigma \sim IW(S, \nu)$ has an inverse Wishart distribution and $\text{vec}(\Phi)|\Sigma \sim N(\text{vec}(M), \Sigma \otimes V)$.

Flexible CRE with heteroskedasticity. We assume that the baseline CRE distribution factorizes as follows:

$$p(\lambda_i, y_{i0}^*, \sigma_i^2 | x_{i0}, \xi) = p(\lambda_i, y_{i0}^* | x_{i0}, \xi) p(\sigma_i^2 | \xi).$$

The distribution of (λ_i, σ_i^2) is a mixture of Normal distributions:

$$\begin{bmatrix} \lambda_i \\ y_{i0}^* \end{bmatrix} \bigg| x_{i0} \overset{iid}{\sim} N\big([1, \, x_{i0}']\Phi_k, \Sigma_k\big) \text{ with prob. } \pi_{\lambda,k}, \quad k = 1, \ldots, K, \tag{12}$$

where Φ_k is an $(n_x + 1) \times 2$ matrix and Σ_k is a 2×2 matrix.[5] Similarly, we model $\ln \sigma_i^2$ as

---

[5]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 $N = 1000$.

a mixture of Normals:

$$\ln \sigma_i^2 \sim N\left(\psi_k, \omega_k^2\right) \text{ with prob. } \pi_{\sigma,k}, \quad k = 1, \ldots, K. \tag{13}$$

Our prior for (θ, ξ) takes the following form: the homogeneous regression parameters are distributed according to

$$\theta \overset{iid}{\sim} N(0, \tau I_{n_x+1}), \quad \tau = 5. \tag{14}$$

To generate a prior for the CRE distribution we use Dirichlet Process Mixtures (DPM). We fix the maximum number of mixture components at a large number K which we set equal to 20 in the simulation exercise and the empirical analysis. The prior for the mixture probabilities $\pi_{\lambda,1:K}$ is generated by a truncated stick breaking process $TSB(1, \alpha_\lambda, K)$ of the form

$$\pi_{\lambda,1:K}|(\alpha_\lambda, K) \sim \begin{cases} \zeta_1, & k = 1, \\ \prod_{j=1}^{k-1} (1 - \zeta_j) \zeta_k, & k = 2, \ldots, K-1,, \quad \zeta_k \sim B(1, \alpha_\lambda). \\ 1 - \sum_{j=1}^{K-1} p_j, & k = K, \end{cases} \tag{15}$$

We use $G(2,2)$ priors for the hyperparameter α_λ of the $TSB(1, \alpha_\lambda, K)$ process. The prior for the mixture probabilities $\pi_{\sigma,1:K}$ takes the same form, with α_λ replaced by α_σ.

The coefficient matrices Φ_k and Σ_k for $p(\lambda_i, y_{i0}^*|x_{i0}, \xi)$ are assumed to follow a MNIW distribution:

$$(\Phi_k, \Sigma_k) \overset{iid}{\sim} MNIW\left(0, \tau I_3, 4I_2, 7\right), \quad k = 1, \ldots, K. \tag{16}$$

The coefficients ψ_k and ω_k have NIG priors:

$$(\psi_k, \omega_k) \overset{iid}{\sim} NIG\left(\ln V^* - \ln(2)/2, 1, 3, (3-1)\ln 2\right), \quad k = 1, \ldots, K. \tag{17}$$

In the specification of these two distributions we already fixed some of the parameters of the MNIW and NIG distributions in light of the subsequent application. This parameterization of the heteroskedasticity prior in (17) is chosen so that the implied prior mean $\mathbb{E}[\sigma_i^2]$ and prior variance $\mathbb{V}[\sigma_i^2]$ matches the one implied by the prior used in the homoskedastic versions of the Tobit model (see (18) below).

Normal CRE. We replace the mixture of Normal distributions by a single Normal distribution $(K = 1)$.

Homoskedasticity imposes the restriction that $\sigma_i^2 = \sigma^2$. To simplify the posterior compu-

tations for the homoskedastic specification we use a conjugate prior for σ^2 of the form

$$\sigma^2 \sim IG\big(3, (3-1)V^*\big). \tag{18}$$

Random Effects (RE). The CRE specification can be simplified to an RE specification as follows. We factorize

$$p(\lambda_i, y_{i0}^*, \sigma_i^2 | x_{i0}, \xi) = p(\lambda_i | \xi) p(y_{i0}^* | \xi) p(\sigma_i^2 | \xi).$$

While $p(\sigma_i^2 | \xi)$ remains the same as in (13), we set

$$\begin{aligned}
\lambda_i &\overset{iid}{\sim} N(\phi_{\lambda,k}, \Sigma_{\lambda,k}) \text{ with prob. } \pi_{\lambda,k}, \quad k = 1, \ldots, K \\
y_{i0}^* &\overset{iid}{\sim} N(\phi_y, \Sigma_y).
\end{aligned} \tag{19}$$

The prior for the mixture probabilities is the same as in (15) and the priors for the coefficients of the normal distributions are

$$(\phi_{\lambda,k}, \Sigma_{\lambda,k}) \overset{iid}{\sim} NIG(0, \tau, 3, 2), \quad (\phi_y, \Sigma_y) \sim NIG(0, \tau, 3, 2), \quad \tau = 5. \tag{20}$$

By selecting between CRE versus RE, heteroskedastic versus homoskedastic, and flexible $(k > 1)$ versus Normal distributions $(k = 1)$ we can generate eight specifications.

3.2   Posterior Sampling

Draws from the posterior distribution can be obtained with a Gibbs sampling algorithm. We will subsequently describe the conditional distributions over which the Gibbs sampler iterates. We will focus on the flexible CRE specification with heteroskedasticity, which is the most complicated specification. A key feature of the Gibbs sampler is that it uses data augmentation by sampling the sequences of latent variables $Y_{i,0:T}^*$, $i = 1, \ldots, N$; see Tanner and Wong (1987) and for the Tobit model Chib (1992) and Wei (1999). The sampler for the flexible mixture representation of the CRE distribution is based on Ishwaran and James (2001, 2002). 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 (2018).

In order to characterize the conditional posterior distributions for the Gibbs sampler, we introduce some additional notation. Because $p(\lambda_i, y_{i0}^*|x_{i0}, \xi)$ and $p(\sigma_i^2|\xi)$ are mixture distributions, *ex post* each (λ_i, y_{i0}^*) and σ_i^2 is associated with one of the K mixture components, respectively. We denote the component affiliations by $\gamma_{i,\lambda}$ and $\gamma_{i,\sigma} \in \{1, \ldots, K\}$, respectively.

Step 1: Drawing from $Y_{i,0:T}^*|(Y_{i,0:T}, \lambda_i, \sigma_i^2, \gamma_{i,y}, \gamma_{i,\sigma}, \theta, \xi)$. 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 $t_1^{(1)} = 2$, $t_2^{(1)} = 4$, $t_1^{(2)} = 7$, $t_2^{(2)} = 9$. The goal is to characterize $p(Y_{i,t_1^{(1)}:t_2^{(1)}}^*, Y_{i,t_1^{(2)}:t_2^{(2)}}^*|Y_{i,0:T}, \ldots)$. 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

$$
\begin{aligned}
&p(Y_{i,t_1^{(1)}:t_2^{(1)}}^*, Y_{i,t_1^{(2)}:t_2^{(2)}}^*|Y_{i,0:T}, \ldots) \\
&= \; p(Y_{i,t_1^{(1)}:t_2^{(1)}}^*|Y_{i,t_1^{(1)}-1:t_2^{(1)}+1}, \ldots)p(Y_{i,t_1^{(2)}:t_2^{(2)}}^*|Y_{i,t_1^{(2)}-1:t_2^{(2)}+1}, \ldots),
\end{aligned}
$$

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

$$M_{1:s|0} = [\mu_1, \ldots, \mu_s]', \quad \mu_1 = \lambda_i + \rho y_{it_1-1} + \beta'x_{it}, \quad \mu_\tau = \lambda_i + \rho\mu_{\tau-1} + \beta'x_{it} \text{ for } \tau = 2, \ldots, s. \tag{21}$$

The covariance matrix takes the form

$$\Sigma_{1:s|0} = \sigma_i^2 \begin{bmatrix} \rho_{1,1|0} & \cdots & \rho_{1,s|0} \\ \vdots & \ddots & \vdots \\ \rho_{s,1|0} & \cdots & \rho_{s,s|0} \end{bmatrix}, \quad \rho_{i,j|0} = \rho_{j,i|0} = \rho^{j-i}\sum_{l=0}^{i-1}\rho^{2l} \text{ for } j \geq i. \tag{22}$$

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

distribution

$$M_{1:s-1|0,s} = M_{1:s-1|0} - \Sigma_{1:s-1,s|0}\Sigma_{ss|0}^{-1}(y_{it_2+1} - \mu_s) \tag{23}$$

$$\Sigma_{1:s-1,1:s-1|0,s} = \Sigma_{1:s-1,1:s-1|0} - \Sigma_{1:s-1,s|0}\Sigma_{ss|0}^{-1}\Sigma_{s,1:s-1|0}$$

to deduce that

$$Y_{i,t_1:t_2}^* \sim TN_-\big(M_{1:s-1|0,s}, \Sigma_{1:s-1,1:s-1|0,s}\big). \tag{24}$$

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 (21) and (22) 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 (21) and (22). According to (12), the joint distribution of (λ_i, y_{i0}^*) is a mixture of normals. Using the mixture component affiliation $\gamma_{i,\lambda}$, we can express $y_{i0}^*|(\lambda_i x_{i0}) \sim N(\mu_*(\lambda_i, x_{i0}), \sigma_*^2)$. This leads to the mean vector

$$M_{1:s} = [\mu_1, \ldots, \mu_s], \quad \mu_1 = \mu_*(\lambda_i, x_{i0}), \quad \mu_\tau = \lambda_i + \rho\mu_{\tau-1} + \beta' x_{it} \text{ for } \tau = 2, \ldots, s \tag{25}$$

and the covariance matrix

$$\Sigma_{1:s} = \sigma_i^2 \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & \rho_{1,1} & \cdots & \rho_{1,s-1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \rho_{s-1,1} & \cdots & \rho_{s-1,s-1} \end{bmatrix} + \sigma_*^2 \begin{bmatrix} \rho^{0+0} & \cdots & \rho^{0+(s-1)} \\ \vdots & \ddots & \vdots \\ \rho^{(s-1)+0} & \cdots & \rho^{(s-1)+(s-1)} \end{bmatrix}, \tag{26}$$

where the definition of $\rho_{i,j}$ is identical to the definition of $\rho_{i,j|0}$ in (22). One can then use the formulas in (23) to obtain the mean and covariance parameters of the truncated normal distribution.

Step 2: Drawing from $\lambda_i|(Y_{i,0:T}, Y_{i,0:T}^*, \sigma_i^2, \gamma_{i,y}, \gamma_{i,\sigma}, \theta, \xi)$. Posterior inference with respect to λ_i becomes "standard" once we condition on the latent variables $Y_{i,0:T}^*$ and the component

affiliation $\gamma_{i,\lambda}$. It is based on the Normal location-shift model

$$y_{it}^* - \rho y_{it-1}^* - \beta' x_{it} = \lambda_i + u_{it}, \quad u_{it} \stackrel{iid}{\sim} N(0, \sigma_i^2), \quad t = 1, \ldots, T. \tag{27}$$

Because the conditional prior distribution $\lambda_i | (y_{i0}^*, x_{i0}, \gamma_{i,\lambda})$ is Normal, the posterior of λ_i is also Normal and direct sampling is possible.

Step 3: Drawing from $\sigma_i^2 | (Y_{i,0:T}, Y_{i,0:T}^*, \lambda_i, \gamma_{i,y}, \gamma_{i,\sigma}, \theta, \xi)$. Posterior inference with respect to σ_i^2 is based on the Normal scale model

$$y_{it}^* - \rho y_{it-1}^* - \beta' x_{it} - \lambda_i = u_{it}, \quad u_{it} \stackrel{iid}{\sim} N(0, \sigma_i^2), \quad t = 1, \ldots, T. \tag{28}$$

However, even conditional on the mixture component affiliation $\gamma_{i,\sigma}$, the prior for σ_i^2 in (13) is not conjugate and direct sampling is not possible. Instead, we sample from this non-standard posterior via an adaptive random walk Metropolis-Hastings (RWMH) step.[6]

Step 4: Drawing from $\theta | (Y_{1:N,0:T}, Y_{1:N,0:T}^*, \lambda_{1:N}, \sigma_{1:N}^2, \gamma_{1:N,\lambda}, \gamma_{1:N,\sigma}, \xi)$. Conditional on the latent variables $Y_{i,0:T}^*$ and the heterogeneous coefficients λ_i, σ_i^2, we can express our model as

$$y_{it}^* - \lambda_i = \rho y_{it-1}^* + \beta' x_{it} + u_{it}, \quad u_{it} \stackrel{iid}{\sim} N(0, \sigma_i^2), \quad i = 1, \ldots, N, \quad t = 1, \ldots, T. \tag{29}$$

The temporal and spatial independence of the u_{it}'s allows us to pool observations across i and t. Under the Normal prior in (14), the posterior distribution of $\theta = [\rho, \beta']'$ is also Normal and we can obtain draws by direct sampling.

Step 5: Drawing from $(\gamma_{i,\lambda}, \gamma_{i,\sigma}) | (Y_{i,0:T}, Y_{i,0:T}^*, \lambda_i, \sigma_i^2, \theta, \xi)$. We describe how to draw the component affiliation $\gamma_{i,\lambda}$. Straightforward modifications lead to a sampler for $\gamma_{i,\sigma}$. Note that ξ contains the elements of $\Phi_{1:K}$, $\Sigma_{1:K}$, and $\pi_{\lambda,1:K}$. The prior probability that unit i is affiliated with component k is given by $\pi_{\lambda,k}$. Let $\bar{\pi}_{i,\lambda,k}$ denote the posterior probability of unit i belonging to component k conditional on the set of means $\Phi_{1:K}$ and variances $\Sigma_{1:K}$ as well as λ_i. The $\bar{\pi}_{i,\lambda,k}$'s are given by

$$\bar{\pi}_{i,\lambda,k} = \frac{\pi_{\lambda,k} p_N\left(\lambda_i | y_{i0}^*, x_{i0}, \Phi_k, \Sigma_k\right)}{\sum_{k=1}^K \pi_{\lambda,k} p_N\left(\lambda_i | y_{i0}^*, x_{i0}, \Phi_k, \Sigma_k\right)}. \tag{30}$$

Note that the conditional distribution $\lambda_i | (y_{i0}^*, x_{i0}, \Phi_k, \Sigma_k)$ is Normal, indicated by the notation $p_N(\cdot)$, and can be derived from the joint normal distributions of the mixture components

---

[6]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%.

in (12). Thus,

$$\gamma_{i,\lambda}|(\Phi_{1:k}, \Sigma_{1:K}, \lambda_i) = k \text{ with prob. } \bar{\pi}_{i,\lambda,k}. \tag{31}$$

Step 6: Drawing from $\xi|(Y_{1:N,0:T}, Y_{1:N,0:T}^*, \lambda_{1:N}, \sigma_{1:N}^2, \gamma_{1:N,\lambda}, \gamma_{1:N,\sigma}, \theta)$**.** Sampling from the conditional posterior of $\Phi_{1:K}$, $\Sigma_{1:K}$, and $\pi_{\lambda,1:K}$ can be implemented as follows. Let $n_{\lambda,k}$ be the number of units and $J_{\lambda,k}$ the set of units affiliated with component k. Both $n_{\lambda,k}$ and $J_{\lambda,k}$ can be determined based on $\gamma_{1:N,\lambda}$. The conditional posterior of the component probabilities takes the form of a generalized truncated stick breaking process

$$\pi_{\lambda,1:K}|(n_{\lambda,1:K}, \alpha, K) \sim TSB\left(\{1 + n_{\lambda,k}\}_{k=1}^K, \left\{\alpha_\lambda + \sum_{j=k+1}^K n_{\lambda,j}\right\}_{k=1}^K, K\right), \tag{32}$$

meaning that the ζ_k's in (15) have a $B\left(1 + n_{\lambda,k}, \alpha_\lambda + \sum_{j=k+1}^K n_{\lambda,j}\right)$ distribution. Conditional on $\pi_{\lambda,1:K}$ the hyperparameter α_λ has a Gamma posterior distribution of the form

$$\alpha_\lambda|\pi_{\lambda,1:K} \sim G(2 + K - 1, 2 - \ln \pi_{\lambda,K}). \tag{33}$$

The conditional posterior for (Φ_k, Σ_k) takes the form

$$p(\Phi_k, \Sigma_k|Y_{1:N,0:T}, Y_{1:N,0:T}^*, \lambda_{1:N}, \sigma_{1:N}^2, \gamma_{1:N,\lambda}, \gamma_{1:N,\sigma}, \theta) \propto p(\Phi_k, \Sigma_k) \prod_{i \in J_{\lambda,k}} p(\lambda_i, y_{i0}^*|\Phi_k, \Sigma_k) \tag{34}$$

Because here the prior $p(\Phi_k, \Sigma_k)$ is MNIW and the likelihood $\prod_{i \in J_{\lambda,k}} p(\lambda_i, y_{i0}^*|\Phi_k, \Sigma_k)$ is derived from a multivariate Normal linear regression model, the conditional posterior of (Φ_k, Σ_k) is also MNIW. All three conditional posteriors allow direct sampling. The derivations can be modified to obtain the conditional posterior of $\psi_{1:K}$, $\omega_{1:K}$, and $\pi_{\sigma,1:K}$.

Step 7: Drawing from the predictive density. Conditional on $(y_{iT}^*, \lambda_i, \sigma_i^2, \theta)$ and $x_{i,T+1:T+H}$, paths from the predictive distribution for $y_{i,T+1:T+h}$ can be easily generated by simulating (1) forward.

Modifications for the simplified model specifications. If the CRE distribution is modeled parametrically instead of flexibly, then the drawing of the component affiliations $(\gamma_{i,\lambda}, \gamma_{i,\sigma})$ in Step 5 and the drawing of $\pi_{.,1:K}$ and α in Step 6 are unnecessary. One only has to draw from the MNIW posterior of (Φ_1, Σ_1) and the NIG posterior of (ψ_1, ω_1). Under homoskedasticity, i.e., $\sigma_i^2 = \sigma^2$ for all i, we can pool (28) in Step 3 across t and i. In combination with the prior in (18) this leads to an IG posterior for σ^2 from which one can sample directly. The RE specification requires modifications to Step 1, because the distribution

of y_{i0} is now simplified to $y_{i0}^* \sim N(\phi_y, \Sigma_y)$, to Step 2 because the prior distribution of λ_i is different, and to Step 6 because the pairs of VAR coefficients (Φ_k, Σ_k) are replaced by $(\phi_{\lambda,k}, \Sigma_{\lambda,k})$ and (ϕ_y, Σ_y), which leads to NIG posteriors.

3.3 Generalizations

The basic dynamic panel Tobit model in (1) can be generalized in several dimensions. First, it is fairly straightforward to allow for randomly missing observations by modifying the inference about the latent variables y_{it^*} in Step 1 of Section 3.2. Rather than drawing the latent variables in Step 1 of the posterior sampler from a truncated normal distribution, we need to draw them from a regular normal distribution.

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

$$
\begin{aligned}
Y_{it} &= f(Y_{it}^*), \quad Y_{it}^*|(Y_{it-1}^*, x_{it}, \lambda_i, \sigma_i^2, \theta) \sim p(Y_{it}^*|Y_{it-1}^*, x_{it}, \lambda_i, \sigma_i^2, \theta), \qquad (35) \\
&\quad (\lambda_i, Y_{i0}^*, \sigma_i^2)|\theta \sim p(\lambda_i, Y_{i0}^*, \sigma_i^2|\theta, \xi),
\end{aligned}
$$

where $f(\cdot)$ is a known function, $p(Y_{it}^*|\cdot)$ is a known homogeneous transition density for Y_{it}^*, and $p(\lambda_i, Y_{i0}^*, \sigma_i^2|\cdot)$ is the correlated random-effects distribution. In the benchmark model (1) the dependent variable is a scalar, i.e., $M = 1$, the transformation of the latent variable is given by $f(y_{it}^*) = y_{it}^* \mathbb{I}\{y_{it}^* \geq 0\}$, and the transition density is $N(\lambda_i + \rho y_{it}^* + \beta' x_{it}, \sigma_i^2)$. In addition to this standard Tobit model, Amemiya (1985) defines four generalizations. For instance, in the so-called Type 2 Tobit model $M = 2$ and the $f(\cdot)$ function takes the form

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

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 $Y_{i,0:T}^*|(Y_{i,0:T}, \lambda_i, \sigma_i^2, \gamma_{i,\lambda}, \gamma_{i,\sigma}, \theta, \pi)$ in Step 1 above. For instance, a posterior sampler for the (static) Type 2 Tobit model is discussed in Li and Tobias (2011). These extensions may be of interest for other panel data forecasting applications with limited dependent variables.

Table 1: Monte Carlo Design

Law of Motion: $y_{it}^* = \lambda_i + \rho y_{it-1}^* + u_{it}$ where $u_{it} \sim N(0, \sigma_i^2)$ and $\rho = 0.8$
Initial Observations: $y_{i0}^* \sim N(0,1)$
Skewed Random Effects Distributions:
$\quad p(\lambda_i|y_{i0}^*) = \frac{1}{9} p_N\left(\lambda_i|\frac{5}{2}, \frac{1}{2}\right) + \frac{8}{9} p_N\left(\lambda_i|\frac{1}{4}, \frac{1}{2}\right)$
$\quad p(\ln \sigma_i^2|y_{i0}^*) = \frac{1}{9} p_N\left(\ln \sigma_i^2 - c|\frac{5}{2}, \frac{1}{2}\right) + \frac{8}{9} p_N\left(\ln \sigma_i^2 - c|\frac{1}{4}, \frac{1}{2}\right)$, c is chosen such that $\mathbb{E}[\sigma_i^2] = 1$
Sample Size: $N = 1,000$, $T = 10$
Number of Monte Carlo Repetitions: $N_{sim} = 100$
Fraction of zeros: 45%, Fraction of all-zeros: 15%

4 Monte Carlo Experiment

The Monte Carlo experiment is based on the dynamic panel Tobit model in (1), which we simplify by omitting the additional predictors x_{it} and using an RE version. We will endow the forecaster with knowledge of the true $p(y_{i0}^*)$ and factorize $p(\lambda_i, y_{i0}^*, \ln \sigma_i^2|\xi)$ as $p(\lambda_i|\xi)p(y_{i0}^*)p(\ln \sigma_i|\xi)$. The data generating process (DGP) is summarized in Table 1. We set the autocorrelation parameter to $\rho = 0.8$ and consider skewed random effects distributions for λ_i and $\ln \sigma_i^2$ that are generated as mixtures of Normals.

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 $t = T + 1$. The fraction of zeros across all samples is 45% and for roughly 15% of the cross-sectional units the sample consists of $T = 10$ zeros ("all zeros"). The measures of forecast accuracy discussed in Sections 2.2 and 2.3 are first computed for the cross section $i = 1, \ldots, N = 1,000$ and then average the performance statistics over the $n_{sim} = 100$ Monte Carlo repetitions.[7]

Model Specifications and Predictors. We will compare the performance of six predictors described below: four Bayes predictors derived from different versions of the dynamic panel Tobit model, a predictor derived from a Tobit model with homogeneous coefficients, and a predictor from a linear model with homogeneous coefficients that ignores the censoring. The prior distributions used for the estimation of the various models were described in Section 3.1 and are summarized in Table 2. Further implementation details are provided in the Online Appendix.

---

[7]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 2: Summary of Prior Distributions

Specification	λ	$p(\lambda\vert\xi)$	σ^2	$p(\sigma^2\vert\xi)$
Flexible RE & Heterosk.	$\lambda \sim N(\phi_{\lambda,k}, \Sigma_{\lambda,k})$ w.p. $\pi_{\lambda,k}$	$(\phi_{\lambda,k}, \Sigma_{\lambda,k}) \sim NIG(0,5,3,2)$ $\pi_{\lambda,k} \sim TSB(1, \alpha_\lambda, K)$ $\alpha_\lambda \sim G(2,2)$	$\ln \sigma^2 \sim N(\psi_k, \omega_k)$ w.p. $\pi_{\sigma,k}$	$(\psi_k, \omega_k) \sim NIG(\ln V^* - \ln(2)/2,$ $1,3,2\ln 2)$ $\pi_{l,k} \sim TSB(1, \alpha_l, K)$ $\alpha_l \sim G(2,2)$
Normal RE & Heterosk.	$\lambda \sim N(\phi_\lambda, \Sigma_\lambda)$	$(\phi_\lambda, \Sigma_\lambda) \sim NIG(0,5,3,2)$	$\ln \sigma^2 \sim N(\psi, \omega)$	$(\psi, \omega) \sim NIG(\ln V^* - \ln(2)/2,$ $1,3,2\ln 2)$
Flexible RE & Homosk.	$\lambda \sim N(\phi_{\lambda,k}, \Sigma_{\lambda,k})$ w.p. $\pi_{\lambda,k}$	$(\phi_{\lambda,k}, \Sigma_{\lambda,k}) \sim NIG(0,5,3,2)$ $\pi_{\lambda,k} \sim TSB(1, \alpha_\lambda, K)$ $\alpha_\lambda \sim IG(2,2)$	$\sigma^2 \sim IG(3, 2V^*)$	N/A
Normal RE & Homosk.	$\lambda \sim N(\phi_\lambda, \Sigma_\lambda)$	$(\phi_\lambda, \Sigma_\lambda) \sim NIG(0,5,3,2)$	$\sigma^2 \sim IG(3, 2V^*)$	N/A
Pooled Tobit / Linear	$\lambda \sim N(0,5)$	N/A	$\sigma^2 \sim G(3, 2V^*)$	N/A
Prior for ρ	$\rho \sim N(0,5)$			
Prior for y_{i0}^*	$y_{i0}^* \sim N(\phi_y, \Sigma_y)$, $(\phi_y, \Sigma_y) \sim NIG(0,5,3,2)$			

Notes: We set $V^* = \frac{1}{N}\sum_{i=1}^{N} \widehat{\mathbb{V}}_i(Y_{it})$.

Table 3: Monte Carlo Experiment: Parameter Estimates and Point Forecast Performance

| | Forecast Error Stats | | | | Parameter Estimates | |
| | RMSE | Bias | StdD | RMSE | Bias($\hat{\rho}$) | StdD($\hat{\rho}$) |
	y_{iT+1}			$\mathbb{I}\{y_{iT+1} = 0\}$		
Flexible & Heterosk.	0.740	-0.007	0.740	0.225	-0.001	0.006
Normal & Heterosk.	0.742	-0.008	0.742	0.225	-0.005	0.005
Flexible & Homosk.	0.746	-0.013	0.745	0.228	0.008	0.009
Normal & Homosk.	0.747	-0.014	0.747	0.228	0.003	0.009
Pooled Tobit	0.789	-0.142	0.776	0.228	0.252	0.004
Pooled Linear	0.822	-0.302	0.764	0.406	0.230	0.004

Notes: The design of the experiment is summarized in Table 1. The true values for ρ is 0.8.

We consider four versions of the dynamic panel Tobit model with random effects (see Section 3.1 for details): (i) flexible RE and heteroskedasticity; (ii) Normal RE and heteroskedasticity; (iii) flexible RE and homoskedasticity; and (iii) Normal RE and homoskedasticity. Versions (b)-(d) are misspecified in light of the DGP. The pooled Tobit specification ignores the heterogeneity in λ_i, setting $\lambda_i = \lambda$ for all i, and imposes homoskedasticity. Finally, the pooled linear specification imposes $\lambda_i = \lambda$, $\sigma_i = \sigma^2$ for all i, and, in addition, ignores the censoring of the observations during the estimation stage.

To generate point, set, and density forecasts, we first generate draws from the posterior distribution of the model parameters and the latent variable y_{iT}^*, and then, conditional on each of these draws, simulate a trajectory $\{y_{iT+s}^*, y_{iT+s}\}_{s=1}^{h}$ from the predictive distribution. While we ignore the censoring in the estimation of the pooled linear specification, we do account for it when we generate forecasts from the linear model.

Point Forecasts and Parameter Estimates. In Table 3 we report RMSEs for posterior mean point forecasts. We decompose the RMSEs into a bias and standard deviation component. We also report the bias and standard deviation of the posterior mean estimate of ρ across Monte Carlo repetitions. The correctly specified dynamic Tobit model that approximates the RE distribution flexibly and allows for heteroskedasticity attains the smallest RMSE. Forcing the RE distributions to be Normal leads to a slight deterioration in RMSE. Incorrectly imposing homoskedasticity increases the RMSE more substantially because of a loss in efficiency. The bias of all four panel Tobit predictors is essentially negligible and the RMSEs are determined by the standard deviations of the forecast errors. The behavior of forecast errors across model versions mirrors the behavior of the ρ estimates: they are un-

Figure 1: Posterior Means and Estimated RE Distributions for λ_i

Flexible & Heteroskedastic	Normal & Heteroskedastic
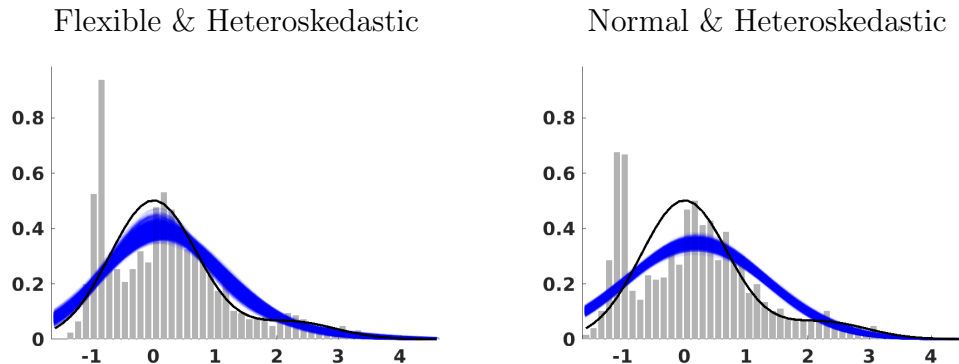
Notes: The histograms depict $\mathbb{E}[\lambda_i|Y_{1:N,0:T}]$, $i = 1, \ldots, N$, for two different model specifications. The shaded areas are hairlines obtained by generating draws from the posterior distribution of ξ and plotting the corresponding random effects densities $p(\lambda|\xi)$. The black lines represent the true $p(\lambda)$.

biased and the standard deviation increases if the heteroskedasticity is ignored. The pooled Tobit and pooled linear predictors perform significantly worse as their estimates of ρ and the resulting forecast errors are severely biased.

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 two heteroskedastic specifications dominate all other predictors. Conditional on using a homoskedastic specification, shutting down heterogeneity (pooled Tobit) does not lead to a further increase in RMSE. Ignoring the censoring when estimating the parameters of the homogeneous specification (pooled linear), raises the RMSE for predicting zero versus non-zero from 0.23 to 0.41.[8]

The panels of Figure 1 show the true RE density $p(\lambda)$, hairlines that represent $p(\lambda|\xi)$ generated from posterior draws of ξ, and histograms of the point estimates $\mathbb{E}[\lambda_i|Y_{1:N,0:T}]$. The left panel corresponds to the flexible specification, whereas the panel on the right displays results for the Normal specification. In both cases we allow for heteroskedasticity. The posterior distribution of $p(\lambda|\xi)$ under the flexible specification concentrates near the true density, whereas, not surprisingly, the parametric specification yields to larger discrepancies between the true RE density and the draws from the posterior distribution.

To interpret the histograms of $\mathbb{E}[\lambda_i|Y_{1:N,0:T}]$ in view of the plotted $p(\lambda)$'s, we consider two stylized examples that capture important aspects of our setup. First, suppose that the model

---

[8]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 $\sqrt{p(1-p)} \leq 0.5$.

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

	Density Forecast		Set Forecast "Average"		Set Forecast "Pointwise"	
	LPS	CRPS	Coverage	Avg. Len.	Coverage	Avg. Len.
Flexible & Heterosk.	-0.761	0.277	0.910	1.260	0.932	1.506
Normal & Heterosk.	-0.762	0.278	0.908	1.249	0.931	1.502
Flexible & Homosk.	-0.899	0.294	0.928	1.505	0.942	1.698
Normal & Homosk.	-0.901	0.294	0.927	1.499	0.941	1.699
Pooled Tobit	-0.934	0.314	0.935	1.704	0.947	1.912
Pooled Linear	-1.236	0.358	0.921	1.921	0.931	1.952

is static, linear, and homoskedastic, i.e., $y_{it} = \lambda_i + u_{it}$, $u_{it} \sim N(0, \sigma^2)$ and $\lambda_i \sim N(\phi_\lambda, 1)$, and ϕ_λ is known (which implies $p(\lambda)$ is known). Therefore, the maximum likelihood estimator (MLE) $\hat{\lambda}_i = \lambda_i + \frac{1}{T}\sum_{t=1}^{T} u_{it}$ has the cross-sectional distribution $\hat{\lambda}_i \sim N(\phi_\lambda, 1 + \sigma^2/T)$ and the posterior means have the distribution

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

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, which is qualitatively 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

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

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 $\int \Phi_N^T(-\lambda/\sigma)p(\lambda)d\lambda$. Thus, according to this example, there should be a spike in the left tail of the distributions of $\mathbb{E}[\lambda_i|Y_{1:N,1:T}]$. This spike is clearly visible in the two panels of Figure 1.

Density and Set Forecasts. The dynamic panel Tobit model generates a posterior predictive density for y_{iT+1} from which one can derive density and set forecasts. Both types of forecasts reflect parameter uncertainty, potential uncertainty about y_{iT}^*, and uncertainty about future shocks. Accuracy statistics are reported in Table 4. To assess the density fore-

casts we compute LPS and CRPS; see Section 2.2. The larger LPS and the smaller CRPS the better the forecast. As expected, the flexible specification with heteroskedasticity that nests the DGP delivers the most accurate density forecasts. While replacing the flexible representations of the RE distributions with Normal distributions only leads to a marginal deterioration of forecast performance, imposing homoskedasticity generates a substantial drop in accuracy.

As discussed in Section 2.3, we consider two types of set forecasts. The first type targets the average coverage probability in the cross-section ("average"), whereas the other type targets the correct coverage probability for each unit i ("pointwise"). To assess the set forecasts we compute the coverage frequency and the average length of 90% predictive sets. The "average" sets constructed from the heteroskedastic specification have good frequentist coverage properties. They attain coverage frequencies of 91.0% and 90.8%, respectively.

The relationship between the nominal credible level of the set forecasts and the empirical coverage frequency is delicate. Sampling in a Bayesian framework involves drawing parameters from the appropriate distribution and generating data conditional on these parameters. Define $U_{i,T+1:T+h} = \{u_{iT+1}, \ldots, u_{iT+h}\}$. In our model $(\lambda_i, \sigma_i^2, U_{i,T+1:T+h})$ are cross-sectionally independent conditional on (ρ, ξ), but not unconditionally. Let $(\tilde{\rho}_N, \tilde{\xi}_N)$ be a draw from the posterior $p(\rho, \xi | Y_{1:N,0:T})$. Then, $(\lambda_i, \sigma_i^2, U_{i,T+1:T+h})$ and hence $Y_{i,T+1:T+h}$ are cross-sectionally independent conditional on $(\tilde{\rho}_N, \tilde{\xi}_N, Y_{1:N,0:T})$ and we can apply a law of large numbers for independently and identically distributed random variables to deduce that (9) approximates the random variable

$$\frac{1}{N} \sum_{i=1}^{N} \mathbb{P}_{Y_{1:N,0:T},\tilde{\rho}_N,\tilde{\xi}_N}^{y_{iT+h}} \left\{ y_{iT+h} \in C_{i,T+h|T}(Y_{1:N,0:T}) \right\}. \tag{36}$$

Recall that by construction of the set forecasts we have

$$\frac{1}{N} \sum_{i=1}^{N} \mathbb{P}_{Y_{1:N,0:T}}^{y_{iT+h}} \left\{ y_{iT+h} \in C_{i,T+h|T}(Y_{1:N,0:T}) \right\} = 1 - \alpha. \tag{37}$$

If the posterior distribution of (ρ, ξ) concentrates around a limit point, then under suitable regularity conditions the discrepancy between (36) and (37) will asymptotically vanish.

A comparison between the "average" and the "pointwise" set forecasts from the heteroskedastic models highlights that the average length of the "average" sets is indeed smaller. Moreover, the coverage frequency of the "pointwise" sets exceeds the nominal coverage level

of 90% by a larger amount. We observe a similar pattern also for the set forecasts from the homoskedastic model specifications. Overall, the homoskedastic specifications generate worse set forecasts, in terms of coverage frequency *and* average length, than the heteroskedastic specifications.

5   Empirical Analysis

We will now use different versions of the dynamic panel Tobit model to forecast loan charge-off rates (charge-offs divided by the stock of loans in the previous period, multiplied by 400) 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 include local changes in house prices and the unemployment rate as additional predictors in the empirical model. As mentioned in the introduction, 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.

5.1   Data

The raw data are obtained from "call reports" (FFIEC 031 and 041) that the banks have to file with their regulator and are available through the website of the Federal Reserve Bank of Chicago. Due to missing observations and outliers we restrict our attention to four loan categories: credit card loans (CC), other consumer credit (CON), construction and land development (CLD), and residential real estate (RRE). We construct rolling panel data sets for each loan category that have a time dimension of twelve quarterly observations: one observation y_0 to initialize the estimation, $T = 10$ observations for estimation, and one observation to evaluate the one-step-ahead forecast. The number of banks N in the cross section varies depending on market size and date availability. The earliest sample considered in the estimation starts ($t = 0$) in 2001Q2 and the most recent sample starts in 2016Q1. A detailed description of the construction of the data set is provided in the Online Appendix.

In the remainder of this section, we will present two types of results: (i) forecast evaluation statistics and parameter estimates for RRE and CC charge-off rates based on samples that cover the Great Recession and range from 2007Q2 ($t = 0$) to 2009Q4 ($t = T + 1$); (ii)

Table 5: Summary Statistics for Baseline Samples

Loan Category	N	Zeros [%]	All Zeros [%]	Mean	75th	Max
Residential Real Estate (RRE)	2,576	76	61	0.25	0.00	33.1
Credit Card (CC)	561	43	22	3.27	4.07	260

Notes: The estimation sample ranges from 2007Q2 ($t = 0$) to 2009Q4 ($t = T = 10$). We forecast 2010Q1 observations. "Zeros" refers to the fraction of zeros in the overall sample of observations (all i and all t), "All Zeros" is the fraction of banks for which charge-off rates are zero in all periods. Mean, 75th percentile, and maximum are computed based on the overall sample.

scatter plots summarizing forecast evaluation statistics for the 111 rolling samples that we constructed (based on data availability) for the above-mentioned four loan categories.

Table 5 contains some summary statistics for the two baseline samples. For the small banks in our sample, residential real estate loans are an important part of their loan portfolio. For approximately 45% of the banks these loans account for 20% to 50% of their loan portfolio. For 25% of the banks in the sample, RREs sum up to more than 50% of their loan portfolio. Credit card loans, on the other hand, make up less than 2% of the loans held by the banks in our sample. Both baseline samples contain a substantial fraction of zero charge-off observations: 76% for RREs and 43% for CC which makes it more challenging to estimate the coefficients of our panel data models. Moreover, 61% of the banks in the RRE sample never write off any loans between 2007 and 2009. The distribution of charge-off rates, across banks and time, is severely skewed. For RREs the 75th percentile is 0 and the maximum is 33.1% annualized. For CCs the corresponding figures are 4.07% and 260%, respectively. A table with summary statistics for the remaining samples is provided in the Online Appendix.

5.2 Charge-Off Rates and the Tobit Model

The forecasts are generated from model (1) with $x_{it} = [\Delta \ln \text{HPI}_{it-1}, \Delta \text{UR}_{it-1}]'$, where y_{it} are charge-off rates, Δ is the temporal difference operator, HPI_{it} is a house price index, and UR_{it} is the unemployment rate. Define $\theta = [\rho, \beta]'$. The model is completed by the specification of a CRE distribution and a prior:

$$p\big(\lambda_i, y_{i0}^*, \sigma_i | x_{i0}, \xi\big), \quad p\big(\theta, \xi\big).$$

In the empirical analysis below we consider various choices of $p(\lambda_i, y_{i0}^*, \sigma_i | x_{i0}, \xi)$ that were described in Section 3.1.

The Tobit model is consistent with the following stylized model. Suppose that bank i has issued a continuum of loans indexed by $j \in [0, 1]$ in the constant amount of ℓ and let $\lambda(j)$ be the probability of repayment of loan j. We assume that the bank writes off the loan j if the repayment probability falls below a threshold κ. Then, we can write the charge-off as

$$x(j) = \ell \mathbb{I}\{\lambda(j) \leq \kappa\}. \tag{38}$$

Let $F_{it}(\kappa) = F(\kappa - \mu_{it})$ be the cumulative distribution function of repayment probabilities and assume that it belongs to a location family with a location parameter μ_{it} that varies across banks and time. Then the charge-off rate is given by

$$y_{it} = F(\kappa - \mu_{it}) = F(y_{it}^*), \quad \text{where} \quad y_{it}^* = \kappa - \mu_{it}. \tag{39}$$

Under the assumption that $F(\cdot)$ is the cdf of a random variable that is uniformly distributed on the unit interval (or has a cdf that is locally linear near zero) we obtain for values of $y_{it} \leq 1$:[9]

$$y_{it} = y_{it}^* \mathbb{I}\{y_{it}^* \geq 0\}, \tag{40}$$

which is the censoring used in (1). In our specification for y_{it}^* the heterogenous intercept λ_i can be interpreted as a bank-specific measure of the quality of the loan portfolio, the autoregressive term captures the persistence of the composition of the loan portfolio over time, and the covariates shift the density of repayment probabilities.

5.3   Density Forecasts

We begin the empirical analysis by comparing the density forecast performance of several variants of (1) for the two baseline samples. This comparison includes forecasts from a Tobit model and a linear model with homogeneous intercepts and homoskedastic innovation variances. Table 6 reports LPS (the larger the better) and CRPS (the smaller the better). Several observations stand out. First, allowing for heteroskedasticity improves the density forecasts unambiguously. For RRE the CRE specification leads to more accurate density forecasts than the RE specification. However, the gain from modeling the distribution of

---

[9]In the empirical analysis we use a scaling for y_{it}^* which implies that it is between 0 and 400 instead of 0 and 1.

Table 6: Density Forecast Performance

Specification	RRE Charge Offs		CC Charge Offs	
	LPS	CRPS	LPS	CRPS
Heteroskedastic Models				
Flexible CRE	-0.52	0.24	-1.92	1.98
Normal CRE	-0.52	0.24	-1.89	1.89
Flexible RE	-0.75	0.26	-1.92	1.98
Normal RE	-0.53	0.24	-1.90	1.93
Homoskedastic Models				
Flexible CRE	-0.75	0.27	-2.41	2.45
Normal CRE	-0.75	0.27	-2.46	2.32
Flexible RE	-0.52	0.24	-2.64	2.59
Normal RE	-0.75	0.26	-2.53	2.36
Pooled Tobit	-0.82	0.30	-2.64	2.60
Pooled Linear	-1.58	0.37	-3.00	2.78

Notes: The estimation sample ranges from 2007Q2 ($t = 0$) to 2009Q4 ($t = T = 10$). We forecast 2010Q1 observations.

the heterogeneous coefficients in a flexible manner is small. For CC charge-off rates the predictive scores for the flexible CRE and RE specifications are essentially the same and slightly dominate the scores of the Normal specifications.

Figure 2 summarizes the LPS comparisons for all 111 samples. The left panel compares predictive scores from the heteroskedastic specifications versus homoskedastic specifications using flexibly modeled correlated random effects. The center panel contrasts LPS obtained from flexible and Normal CRE distributions. Finally, the right panel assesses the importance of the correlation between the heterogeneous intercept and the initial observations and the regressors. The solid lines are 45-degree lines and the blue and red circles correspond to the scores associated with the baseline RRE and CC samples reported in Table 6. The figure shows that the results for the baseline samples are qualitatively representative: incorporating heteroskedasticity is important for density forecasting whereas accuracy differentials between Normal and flexible CREs on the one hand, and CREs versus REs are small. In view of these results, we will subsequently focus on the flexible CRE specification with heteroskedasticity.

Figure 2: Log Predictive Density Scores – All Samples
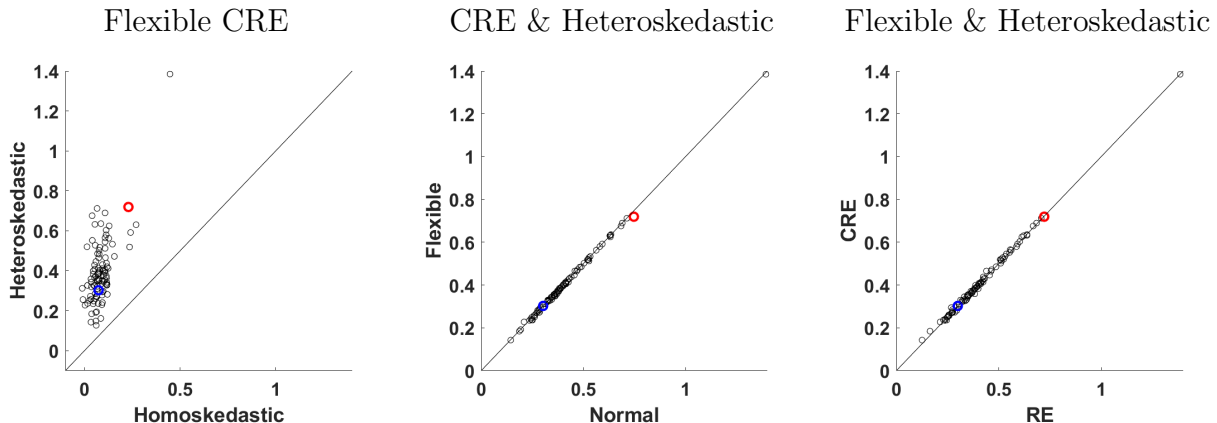
Notes: The panels provide pairwise comparisons of log predictive scores. We also show the 45-degree line. Log probability scores are depicted as differentials relative to pooled Tobit. The blue (red) circle corresponds to RRE (CC).

Table 7: Estimates of Common Parameters

	y_{it-1}^*		$\Delta \ln \mathrm{HPI}_{it-1}$		$\Delta \mathrm{UR}_{it-1}$	
	Mean	CI	Mean	CI	Mean	CI
RRE	0.22	[0.19, 0.26]	-3.63	[-4.57, -2.34]	11.61	[9.33, 13.47]
CC	0.46	[0.42, 0.50]	-4.72	[-7.75, -1.71]	2.97	[-0.63, 6.55]

Notes: Heteroskedastic flexible CRE specification. The estimation sample ranges from 2007Q2 ($t = 0$) to 2009Q4 ($t = T = 10$). The table contains posterior means and 90% credible intervals in brackets.

5.4   Parameter Estimates and Predictive Checks

Parameter estimates of the common coefficients for the flexible CRE specification with heteroskedasticity are reported in Table 7. We report posterior means and 90% credible intervals. Both samples exhibit mild autocorrelation. The point estimate of ρ is 0.22 for RRE and 0.46 for CC. A reduction in house prices leads to an increase in the charge-off rates for RRE and CC which is consistent with the narrative of the Great Recession. Likewise, a rise in the unemployment rate is associated with higher charge-off rates. The effect of unemployment changes is more pronounced for RRE loans than for CC loans. In fact, the credible interval for the CC sample also contains some negative values.

Because we are using standardized regressors in our estimation, the numerical values in Table 7 are not directly interpretable. The mean of the pooled log HPI changes in our sample

Figure 3: Heterogeneous Coefficient Estimates

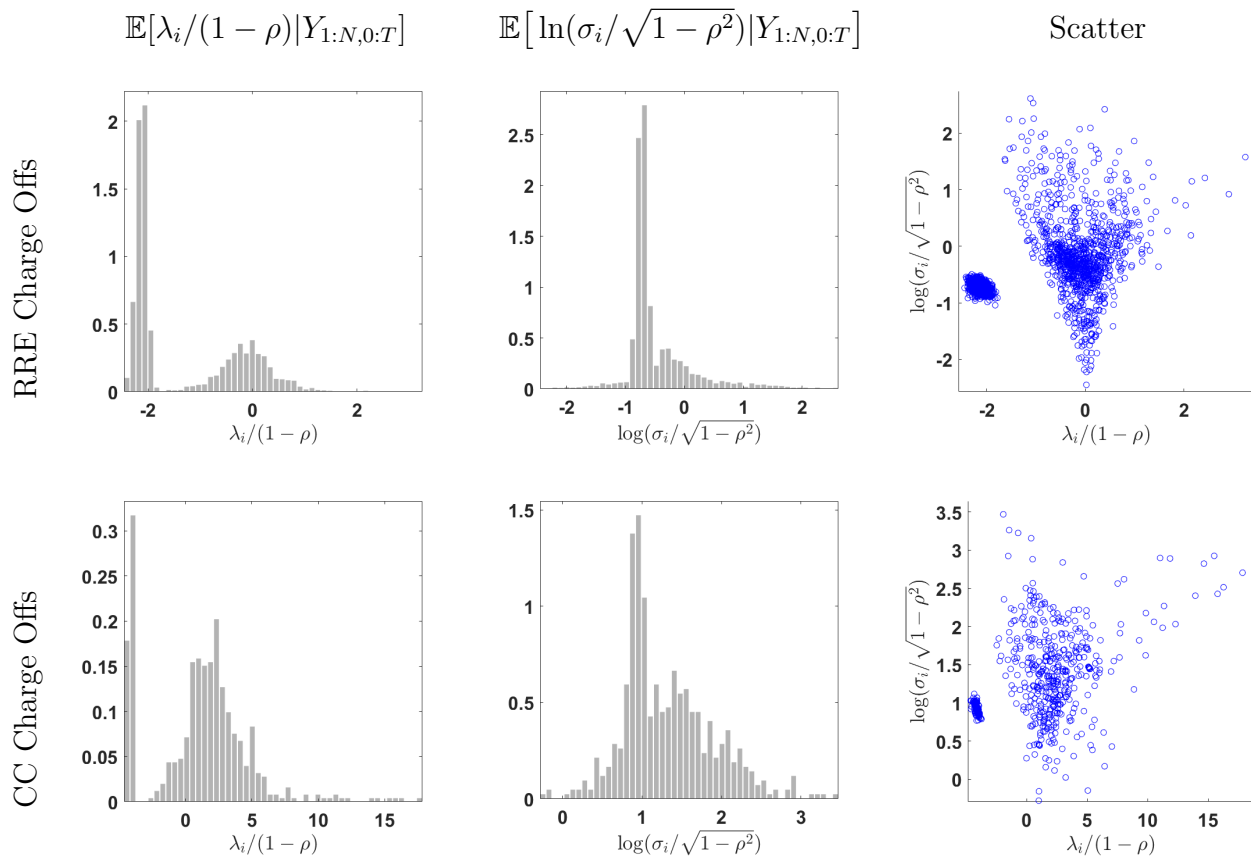$$\mathbb{E}[\lambda_i/(1-\rho)|Y_{1:N,0:T}] \qquad \mathbb{E}\big[\ln(\sigma_i/\sqrt{1-\rho^2})|Y_{1:N,0:T}\big] \qquad \text{Scatter}$$

Notes: Heteroskedastic flexible CRE specification. The estimation sample ranges from 2007Q2 ($t = 0$) to 2009Q4 ($t = T = 10$). A few extreme observations are not visible in the plots.

is -0.012 and corresponds to a 4.8% drop at an annualized rate. The standard deviation is 0.014. Suppose that house prices drop by an additional one percent within a quarter. For $\beta_1 = 4$, this leads to an increase of the charge-off rate by $4 * (-0.01)/0.014 = 2.86$ percentage points relative to the baseline value. The average change in the unemployment rate is 0.47 percentage points per quarter and the standard deviation is 0.46. Suppose that $\beta_2 = 5$ and the increase in the unemployment rate within a quarter is 0.5 percentages higher than its baseline value. In response, the charge-off rate rises by $5 * 0.5/0.46 = 5.43$ percentage points.

The distributions of posterior mean estimates of the heterogeneous coefficients are depicted in Figure 3. We use the AR coefficient ρ to rescale λ_i and σ_i. The panels on the left and in the center of the figure show histograms for the posterior means of λ_i and σ_i, respectively, whereas the panels on the right contain scatter plots that illustrate the correlation between the posterior means of intercepts and shock standard deviations. A notable

feature of the histograms for the posterior means of $\lambda_i/(1-\rho)$ are the spikes in the left tail of the distribution.[10] The spikes correspond to banks with predominantly zero charge-off rates. For these banks, the sample contains very little information about λ_i other than that it has to be sufficiently small to explain the zero charge-off rates. In turn, the posterior mean estimate is predominantly driven by the prior. Similar spikes are visible in the histograms for the posterior means of the re-scaled log standard deviations and the right panels show that the σ_i spikes and the λ_i spikes are associated with the same banks.

Recall from Table 5 that the RRE sample contains more zero charge-off observations than the CC sample. Accordingly, the spikes in the RRE histograms are associated with larger mass. The unconditional distribution of charge-off rates in the CC sample has a non-zero mean and a longer right tail. This is captured in the histograms by larger posterior mean estimates of λ_i and a more pronounced right tail. The estimated re-scaled standard deviations are also larger for the CC sample than for the RRE sample. The large dispersion of σ_i estimates in both samples is consistent with the substantially better density forecast performance of the heteroskedastic models. For the RRE loan sample small estimates of σ_i are associated with near zero estimates of λ_i, whereas large estimates of σ_i are associated with a broad range of λ_i estimates. This pattern is less pronounced for the CC sample.

In order to assess the fit of the estimated panel Tobit model, we report posterior predictive checks in Figure 4. A posterior predictive check examines the extent to which the estimated model can generate artificial data with sample characteristics that are similar to the characteristics of the actual data that have been used for estimation.[11] Consider the top left panel of the figure. Here, the particular characteristic, or sample statistic, under consideration is the cross-sectional density of y_{iT+1} conditional on $y_{iT+1} > 0$. The black line is computed from the actual RRE loan sample. Each blue hairline is generated as follows: (i) take a draw of (ρ, β, ξ) from the posterior distribution; (ii) conditional on these draws generate $\lambda_{1:N}$, $Y^*_{1:N,0}$, and $\sigma^2_{1:N}$; (iii) simulate a panel of observations $\tilde{Y}_{1:N,0+T+1}$; (iv) compute a kernel density estimate based on $\tilde{Y}_{1:N,T+1}$. The swarm of hairlines visualizes the posterior predictive distribution. A model passes a posterior predictive check if the observed value of the sample statistic does not fall too far into the tails of the posterior predictive distribution. Rather than formally computing p-values, we will focus on a qualitative assessment of the model fit.

---

[10]Recall that these spikes were also present in the Monte Carlo simulation; see Figure 1.

[11]Textbook treatments of posterior predictive checks can be found, for instance, in Lancaster (2004) and Geweke (2005).

Figure 4: Posterior Predictive Checks: Cross-sectional Distribution of Sample Statistics
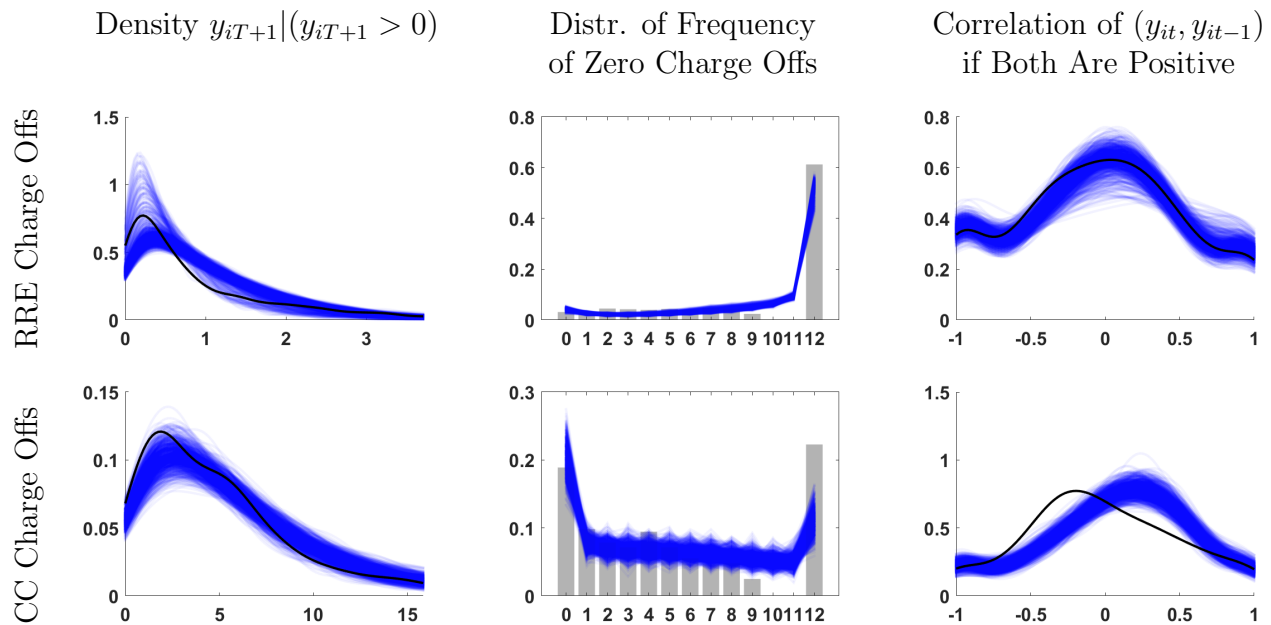
Notes: Heteroskedastic flexible CRE specification. The estimation sample ranges from 2007Q2 ($t = 0$) to 2009Q4 ($t = T = 10$). The black lines (left and right panels) and the histogram (center panels) are computed from the actual data. Each hairline corresponds to a simulation of a sample $\tilde{Y}_{1:N,0:T+1}$ of the panel Tobit model based on a parameter draw from the posterior distribution.

By and large, the estimated models for RRE and CC charge-off rates do a fairly good job in reproducing the cross-sectional densities of y_{iT+1} in that some of the hairlines generated from the posterior cover the observed densities. The only discrepancies arise for charge-off values close to zero. With high probability, the densities computed from simulated data have less mass than the observed RRE and CC densities. Moreover, the modes of the simulated densities are slightly to the right and lower than the mode in the two actual densities. The hairlines depict the densities conditional on $y_{iT+1} > 0$. In the RRE sample the fraction of $y_{iT+1} = 0$ is 0.71. The 90% posterior predictive interval is [0.73, 0.79]. For CC charge-off rates, the value in the data is 0.43 and the posterior predictive interval is [0.37, 0.47].

The center panels of Figure 4 focus on the estimated models' ability to reproduce the number of zero charge-off observations. For each unit i we compute the number of periods in which $y_{it} = 0$. Because $T = 10$ the maximum number of zeros between $t = 0$ and $t = T + 1$ is 12. The histogram is generated from the actual data, whereas the hairlines are computed from the simulated data. For instance, 61% of the banks do not write off any RRE loans in the twelve quarters of the sample and roughly 5% of the banks write off RRE loans in every

period. Overall, the estimated models do remarkably well in reproducing the patterns in the data. For RRE loans, the model captures the large number of all-zero samples and the fairly uniform distribution of the number of samples with zero to nine instances of $y_{it} = 0$. The only deficiency is that the model cannot explain the absence of samples with ten or eleven instances of zero charge-off rates. In the case of CC loans, the estimated model underpredicts the number of all-zero samples but generally is able to match the rest of the distribution.

The last column of Figure 4 provides information about the models' ability to capture some of the dynamics of the charge-off data. Here the test statistic is the first-order sample autocorrelation of the $y_{i0:T+1}$ sequence, conditional on both y_{it} and y_{it-1} being greater than zero. The panels in the figure depict the cross-sectional density of these sample autocorrelations. For the RRE loans the density computed from the actual data is covered by the hairlines generated from the posterior predictive distribution. For the CC loans the estimated model generates somewhat higher sample autocorrelations than what is present in the data.

In the Online Appendix (see Figure A-1) we consider three additional predictive checks based on (i) the time series mean of y_{it} after observing a zero (and, if applicable, before observing the next zero), (ii) the time series mean of y_{it} before observing a zero (and, if applicable, after observing the previous zero), (iii) a robust estimate of the first-order autocorrelation of $y_{i,0:T+1}$ provided there are sufficiently many non-zero observations. With the exception of the autocorrelations in the CC charge-off sample, the two estimated models are able to reproduce the cross-sectional densities of the sample statistics.

5.5   Set Forecasts

Results on the accuracy of set forecasts for the 2007Q2 samples of RRE and CC charge-off rates are presented in Table 8. The nominal credible level is 90%. We distinguish forecasts that for each bank i are constructed from the posterior predictive distribution as 90% credible sets, denoted by "Pointwise", from forecasts that target conditional on the observations $Y_{1:N,0:T}$ an average coverage probability across banks of 90%, labeled as "Average." In addition to the flexible CRE specification with heteroskedastic innovations, we also consider a homoskedastic version. It turns out that the set forecasts generated by the homoskedastic specifications are substantially larger than the sets obtained from the models with heteroskedasticity, without improving the coverage probability. This finding is consistent with the density forecast results in Table 6.

Table 8: Set Forecast Performance

		Coverage	Ave. Length	Fraction of CIs of the Form		
				$\{0\}$	$[0,\,b]$	$\{0\} \cup [a,\,b]$
RRE Charge-Off Rates						
Heterosk.	Average	0.88	0.30	0.68	0.28	0.04
	Pointwise	0.94	0.74	0.61	0.21	0.18
Homosk.	Average	0.92	0.64	0.63	0.32	0.04
	Pointwise	0.96	1.03	0.61	0.18	0.20
CC Charge-Off Rates						
Heterosk.	Average	0.90	6.35	0.02	0.82	0.15
	Pointwise	0.91	7.71	0.22	0.25	0.53
Homosk.	Average	0.95	10.10	0.21	0.71	0.07
	Pointwise	0.95	10.19	0.00	0.34	0.65

Notes: Flexible CRE specification. The estimation sample ranges from 2007Q2 ($t = 0$) to 2009Q4 ($t = T = 10$). We forecast 2010Q1 observations. The nominal coverage probability is 90%.

The set forecasts that are constructed by targeting the average coverage probability have a cross-sectional coverage frequency that is closer to the nominal coverage probability of 90% and they tend to be shorter. The reduction in the average length is particularly pronounced for the heteroskedastic specifications, because due to the cross-sectional variation in the variance of the posterior predictive distributions, there is more scope for tightening wide forecast sets and enlarging narrow forecast sets while maintaining the average coverage probability.

We also report the frequency of the three types of set forecasts. Due to the large number of zero observations in the RRE sample, there is a large fraction of banks, between 60% and 68%, for which the posterior predictive probability of observing $y_{iT+1} = 0$ exceeds 90%. This leads to a forecast of $\{0\}$. For the CC sample the fraction of $\{0\}$ forecasts is considerably smaller. As one switches from targeting pointwise coverage probability to average coverage probability the composition of the set types changes. Recall that the set forecasts correspond to HPD sets constructed from the posterior predictive distributions. If the forecaster targets average instead of pointwise coverage probability, then, roughly speaking, she should widen the "narrow" sets (small σ_i) by lowering their HPD threshold, and tighten the wide sets (large σ_i) by raising their HPD threshold. For RRE we observe that the fraction of disconnected set forecasts decreases when one switches to targeting average coverage probability, which means that the HPD threshold for these sets has to fall. Thus, we can infer that the short

Figure 5: Sets Forecasts – All Samples
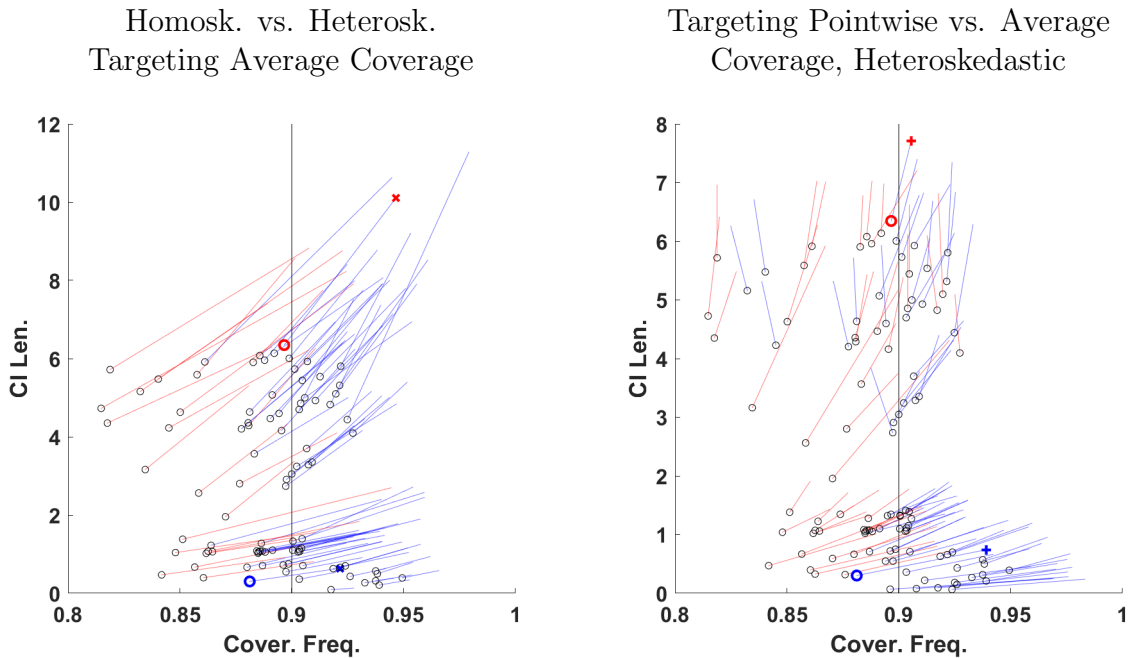
Notes: Flexible CRE specification. The blue (red) symbols corresponds to RRE (CC). The two endpoints of each hairline indicate the coverage probability and length for a particular combination of loan category and estimation sample. Left panel: circle corresponds heteroskedastic specification. Right panel: circles correspond to targeting average coverage probability. Blue hairlines indicate samples for which the coverage frequency gets closer to the nominal coverage probability of 90%. Red hairlines indicate samples for which the coverage frequency deteriorates.

pointwise sets for low σ_i tend to be disconnected.

In Figure 5 we provide information about the coverage frequency and average length size of the set forecast for all samples. The left panel compares the quality of the set forecasts for the homoskedastic specification to that of the heteroskedastic specification. Each hairline corresponds to one of the 111 different samples and the endpoints of the hairlines indicate coverage and length. The circled endpoint corresponds to the heteroskedastic specification whereas the unmarked endpoint indicates the performance of the homoskedastic specification. In all samples allowing for heteroskedastic errors shortens the the average length of the set forecasts. In addition, for 76% of the samples (blue hairlines) the discrepancy between coverage frequency and nominal coverage probability is also reduced.

The right panel of Figure 5 compares set forecasts that target pointwise (unmarked endpoint) and average (circled endpoint) coverage probability. Targeting the latter unambiguously reduces the average length. For 59% of the samples it also improves the empirical

Figure 6: Point Forecast Accuracy: All Samples
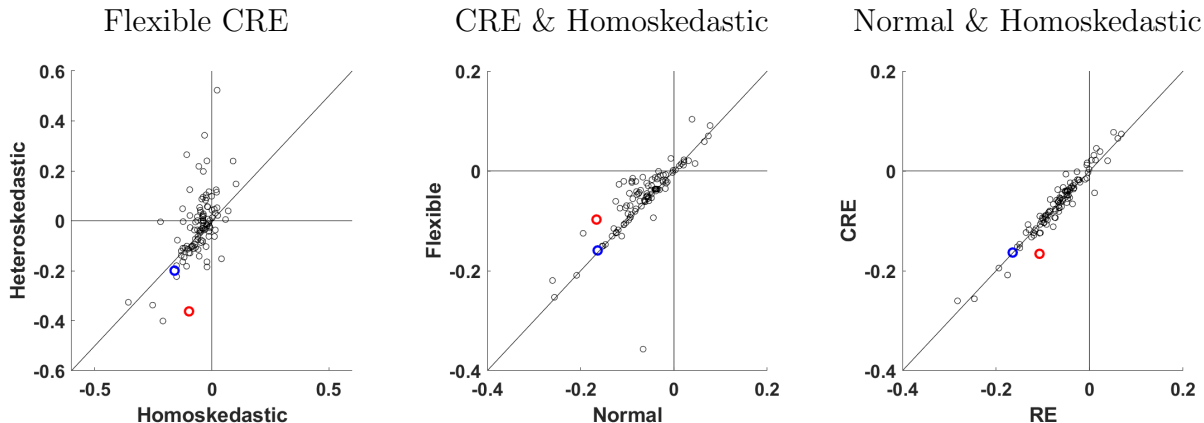
Notes: The panels provide pairwise comparisons of log RMSE differentials with respect to Pooled Tobit under the benchmark prior. The red (blue) circle corresponds to CC (RRE).

coverage frequency (blue hairlines). For many of the remaining 41% of the samples the deterioration of the coverage frequency is relatively small. In all samples, the coverage frequency falls. We conclude that, by and large, directly targeting the average posterior coverage probability improves the empirical coverage frequency in the cross section and produces shorter set forecasts.

5.6 Point Forecasts

Finally, we provide an assessment of the performance of point forecasts, measured through RMSEs. We consider the same model specifications as in Section 5.3. In Table 9 we report results for forecasts of y_{iT+1} and the event that charge-offs are zero, $\mathbb{I}\{y_{iT+1} = 0\}$. For the two samples considered in the table, the heteroskedastic specifications deliver more accurate forecasts than the homoskedastic representations, just as for density and set forecasts. The RE specifications perform slightly better than the CRE specifications and in the CC charge-off sample the Normal specifications deliver more precise point forecasts than the flexible specifications. Thus, imposing parsimony on the Tobit model improves the point forecasts in these two samples, at least marginally. We also consider a pooled Tobit model and a pooled linear model (with homoskedasticity), but they perform generally worse than the Tobit models that allow for intercept heterogeneity.[12]

---

[12]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.

Table 9: Point Forecast Accuracy: RMSEs

	RRE Charge Offs		CC Charge Offs	
Specification	y_{iT+1}	$\mathbb{I}\{y_{iT+1}=0\}$	y_{iT+1}	$\mathbb{I}\{y_{iT+1}=0\}$
Heteroskedastic Models				
Flexible CRE	1.10	0.28	6.42	0.39
Normal CRE	1.10	0.28	5.82	0.38
Flexible RE	1.08	0.28	6.28	0.38
Normal RE	1.09	0.29	6.01	0.38
Homoskedastic Models				
Flexible CRE	1.15	0.30	8.37	0.41
Normal CRE	1.14	0.30	7.82	0.40
Flexible RE	1.14	0.30	9.24	0.41
Normal RE	1.14	0.30	8.29	0.40
Flat	1.14	0.30	8.05	0.39
Pooled Tobit	1.35	0.31	9.23	0.41
Pooled Linear	1.20	0.52	8.91	0.47

Notes: The estimation sample ranges from 2007Q2 ($t=0$) to 2009Q4 ($t=T=10$). We forecast 2010Q1 observations.

Figure 6 compares the accuracy of point forecasts for the full set of our 111 samples. The left panel compares log RMSE differentials – using the RMSE from the pooled Tobit models as baseline – from the heteroskedastic specifications versus homoskedastic specifications using flexibly modeled correlated random effects. A value below zero corresponds to a sample in which the heterogeneous coefficient Tobit model leads to a more precise forecast than its homogeneous counterpart. A point below the 45-degree line corresponds to a sample in which the heteroskedastic specification delivers a lower RMSE than the homoskedastic specification. Unlike in the case of density forecasts (see Figure 2), the heteroskedastic specification does not dominate the homoskedastic specification. In fact, for about 60% of the samples, the ranking is reversed. Because 86% of the dots lie to the left of the vertical zero line, it remains the case that the model with heterogeneous λ_i's dominates the pooled Tobit model that imposes homogeneity.

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. For our benchmark prior the variance of λ_i is not scaled by the innovation variance σ_i^2. Thus, while for the homoskedastic model the relative weight of MLE and prior mean is similar for all banks in the panel, under the heteroskedastic specification

the MLE receives little weight for banks that are associated with a large σ_i^2. We investigated whether this feature contributed to the poor point forecast performance by considering an alternative prior in which we scale the prior variance of λ_i by σ_i^2:

$$\begin{bmatrix} \lambda_i \\ y_{i0}^* \end{bmatrix} \Bigg| (x_{i0}, \sigma_i^2) \overset{iid}{\sim} N\big([1, \, x_{i0}']\Phi_k, \sigma_i^2\Sigma_k\big) \text{ with prob. } \pi_{\lambda,k}, \quad k = 1, \dots, K. \quad (41)$$

The multiplication of the prior variance by σ_i^2 disassociates the degree of shrinkage toward the prior mean from the innovation variance σ_i^2. While changing the prior leads to small improvements in the point forecast performance in some samples, it created large distortions in many other samples and therefore does not provide a robust remedy against a weak point forecast performance of the heteroskedastic models in many of the samples that we considered. A figure documenting this finding is provided in the Online Appendix.

The center panel and the right panel of Figure 6 focus on the homoskedastic specifications. The center panel indicates that representing the CRE distribution by a multivariate Normal distribution works as well and in many instances better than modeling it more flexibly as a mixture of Normals (65% of the points are above the 45-degree line, and most of the remaining points are very close to the line). The right panel focuses on a comparison between CRE and RE. Here the ranking is mixed. In about half of the samples modeling the correlation between intercept and initial observation improves the forecast accuracy. In the other have of the samples the RE specification that assumes independence between intercept and initial conditions leads to lower RMSEs. All three panels of Figure 6 that modeling the heterogeneity in the intercept is important for point forecasting: once heteroskedasticity is imposed, in the vast majority of samples (about 82%) the forecast from the dynamic panel Tobit model with intercept heterogeneity is more accurate than the pooled Tobit forecast.

6 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 estimation of heterogeneous intercepts and conditional variances improves density and set forecasting performance

in the more than 100 samples that are considering. Posterior predictive checks conducted for two particular samples indicate that the Tobit model is able to capture salient features of the charge-off panel data sets. 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.

References

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

ARELLANO, M., AND O. BOVER (1995): "Another look at the instrumental variable estimation of error-components models," *Journal of econometrics*, 68(1), 29–51.

ASKANAZI, R., F. X. DIEBOLD, F. SCHORFHEIDE, AND M. SHIN (2018): "On the Comparison of Interval Forecasts," *Journal of Time Series Analysis*, 39(6), 953–965.

ATCHADÉ, Y. F., AND J. S. ROSENTHAL (2005): "On Adaptive Markov Chain Monte Carlo Algorithms," *Bernoulli*, 11(5), 815–828.

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.

CHANG, C. C., AND D. N. POLITIS (2016): "Robust Autocorrelation Estimation," *Journal of Computational and Graphical Statistics*, 25(1), 144–166.

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.

GEWEKE, J. (2005): *Contemporary Bayesian Econometrics and Statistics*. John Wiley & Sons, Inc.

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

Griffin, J. E. (2016): "An Adaptive Truncation Method for Inference in Bayesian Non-parametric Models," *Statistics and Computing*, 26(1), 423–441.

Gu, J., and R. Koenker (2017a): "Empirical Bayesball Remixed: Empirical Bayes Methods for Longitudinal Data," *Journal of Applied Economics*, 32(3), 575–599.

——— (2017b): "Unobserved Heterogeneity in Income Dynamics: An Empirical Bayes Perspective," *Journal of Business & Economic Statistics*, 35(1), 1–16.

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

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," .

Keane, M., and O. Stavrunova (2011): "A Smooth Mixture of Tobits Model for Healthcare Expenditure," *Health Economics*, 20, 1126–1153.

Lancaster, T. (2004): *An Introduction to Modern Bayesian Econometrics*. Blackwell Publishing.

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.

Li, T., and X. Zheng (2008): "Semiparametric Bayesian Inference for Dynamic Tobit Panel Data Models with Unobserved Heterogeneity," *Journal of Applied Econometrics*, 23(6), 699–728.

Liu, L. (2018): "Density Forecasts in Panel Data Models: A Semiparametric Bayesian Perspective," *arXiv preprint arXiv:1805.04178*.

Liu, L., H. R. Moon, and F. Schorfheide (2018): "Forecasting with Dynamic Panel Data Models," *NBER Working Paper*, 25102.

——— (2019): "Forecasting with Dynamic Panel Data Models," *Econometrica*, Forthcoming.

Nychka (1988): "Douglas," *Journal of the American Statistical Association*, 83(404), 1134–1143.

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.

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.

WAHBA, G. (1983): "Bayesian "Confidence Intervals" for the Cross-Validated Smoothing Spline," *Journal of the Royal Statistical Society, Series B*, 45(1), 133–150.

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

Supplemental Online Appendix to "Forecasting with a Panel Tobit Model"

Laura Liu, Hyungsik Roger Moon, and Frank Schorfheide

This Online Appendix consists of the following sections:

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.[13] (iii) For banks with domestic offices only (Form FFIEC 041), RIAD4645 (numerator for commercial and industrial loans) 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;

---

[13]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)."

- 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;

- 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.

Table A-1: Sample Sizes After Selection Steps and Summary Statistics for 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	3,146	1,304	77	1.5	0.0	106.8
CLD	2007Q4	7,835	7,835	7,219	3,244	3,088	1,264	74	1.9	0.1	106.8
CLD	2008Q1	7,692	7,692	7,084	3,204	3,032	1,257	71	2.2	0.5	180.2
RRE	2007Q1	7,991	7,991	7,393	6,260	5,993	2,654	77	0.2	0.0	33.1
RRE	2007Q2	7,993	7,993	7,383	6,152	5,894	2,576	76	0.3	0.0	33.1
RRE	2007Q3	7,903	7,903	7,299	6,193	5,920	2,606	73	0.3	0.0	35.9
RRE	2007Q4	7,835	7,835	7,219	6,146	5,859	2,581	70	0.4	0.1	69.2
RRE	2008Q1	7,692	7,692	7,084	6,106	5,792	2,561	68	0.4	0.2	45.6
RRE	2008Q2	7,701	7,701	7,080	6,029	5,721	2,492	67	0.4	0.2	63.6
RRE	2008Q3	7,631	7,631	7,008	6,052	5,743	2,577	65	0.5	0.3	39.2
RRE	2008Q4	7,559	7,559	6,938	6,005	5,679	2,600	63	0.5	0.3	45.6
RRE	2009Q1	7,480	7,480	6,849	5,971	5,634	2,588	62	0.5	0.3	45.0
RRE	2009Q2	8,103	8,103	7,381	5,895	5,564	2,536	62	0.5	0.3	45.0
RRE	2009Q3	8,016	8,016	7,302	5,899	5,568	2,563	61	0.5	0.4	47.6
RRE	2009Q4	7,940	7,940	7,229	5,846	5,508	2,553	60	0.5	0.4	45.0
RRE	2010Q1	7,770	7,770	7,077	5,765	5,426	2,494	61	0.5	0.4	45.0
RRE	2010Q2	7,770	7,770	7,072	5,635	5,308	2,420	61	0.5	0.4	45.0
RRE	2010Q3	7,707	7,707	7,013	5,632	5,298	2,441	61	0.5	0.4	45.6
RRE	2010Q4	7,608	7,608	6,910	5,583	5,255	2,443	61	0.5	0.3	38.2
RRE	2011Q1	7,469	7,469	6,784	5,520	5,220	2,437	62	0.4	0.3	38.2
RRE	2011Q2	7,472	7,472	6,783	5,398	5,110	2,385	62	0.4	0.3	38.2
RRE	2011Q3	7,402	7,402	6,716	5,395	5,110	2,397	64	0.4	0.2	38.2
RRE	2011Q4	7,334	7,334	6,649	5,341	5,059	2,395	65	0.3	0.2	38.2
RRE	2012Q1	7,236	7,236	6,546	5,284	5,008	2,349	67	0.3	0.2	38.2
RRE	2012Q2	7,234	7,234	6,534	5,584	5,283	2,430	66	0.3	0.2	38.2
RRE	2012Q3	7,170	7,170	6,465	5,576	5,267	2,416	67	0.2	0.1	28.4
RRE	2012Q4	7,073	7,073	6,358	5,495	5,197	2,362	69	0.2	0.1	22.2
RRE	2013Q1	6,931	6,849	6,212	5,420	5,121	2,341	71	0.2	0.1	28.7
RRE	2013Q2	6,932	6,857	6,200	5,296	5,008	2,298	71	0.2	0.1	28.7
RRE	2013Q3	6,882	6,807	6,144	5,291	4,999	2,307	72	0.2	0.0	28.7
RRE	2013Q4	6,801	6,726	6,061	5,212	4,932	2,271	74	0.1	0.0	28.7
RRE	2014Q1	6,648	6,576	5,913	5,144	4,870	2,258	75	0.1	0.0	27.2
RRE	2014Q2	6,648	6,580	5,897	5,020	4,754	2,195	76	0.1	0.0	16.9
RRE	2014Q3	6,580	6,512	5,821	5,012	4,750	2,183	77	0.1	0.0	22.2
RRE	2014Q4	6,500	6,433	5,729	4,953	4,699	2,215	78	0.1	0.0	16.9
RRE	2015Q1	6,340	6,273	5,564	4,882	4,619	2,205	79	0.1	0.0	11.1
RRE	2015Q2	6,346	6,280	5,560	4,759	4,508	2,139	79	0.1	0.0	11.1
CC	2001Q2	9,031	9,031	8,532	1,691	1,540	875	33	3.4	4.7	162.5
CC	2001Q3	8,995	8,995	8,491	1,666	1,515	844	33	3.4	4.8	88.9
CC	2001Q4	8,887	8,887	8,382	1,636	1,489	836	34	3.3	4.6	88.9

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 time periods. 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 Charge-Off Rates (cont.)

		Sample Sizes					Cross-sectional Statistics			
Loan	t_0	Initial	Step1	Step2	Step3	Step5	% 0s	Mean	75%	Max
CC	2002Q1	8,723	8,723	8,228	1,612	1,466	814	35	3.3	4.4	400.0
CC	2002Q2	8,823	8,823	8,312	1,670	1,519	817	38	3.2	4.3	88.9
CC	2002Q3	8,805	8,805	8,286	1,631	1,488	821	38	3.2	4.3	88.9
CC	2002Q4	8,728	8,728	8,199	1,606	1,468	813	39	3.1	4.1	88.9
CC	2003Q1	8,611	8,611	8,077	1,573	1,445	811	40	3.0	4.0	128.5
CC	2003Q2	8,754	8,754	8,203	1,544	1,422	787	40	3.0	3.9	136.1
CC	2003Q3	8,755	8,755	8,198	1,513	1,395	754	41	2.9	3.8	136.1
CC	2003Q4	8,671	8,671	8,120	1,500	1,387	724	42	2.8	3.6	136.1
CC	2004Q1	8,526	8,526	7,989	1,468	1,355	707	43	2.7	3.6	136.1
CC	2004Q2	8,662	8,662	8,108	1,440	1,331	677	42	2.8	3.6	136.1
CC	2004Q3	8,626	8,626	8,067	1,411	1,308	664	43	2.7	3.5	136.1
CC	2004Q4	8,552	8,552	7,989	1,391	1,284	657	44	2.6	3.3	140.9
CC	2005Q1	8,384	8,384	7,829	1,369	1,271	639	44	2.5	3.2	151.3
CC	2005Q2	8,507	8,507	7,938	1,332	1,236	611	44	2.6	3.2	175.0
CC	2005Q3	8,482	8,482	7,897	1,315	1,218	596	45	2.6	3.2	175.0
CC	2005Q4	8,404	8,404	7,816	1,290	1,203	604	46	2.6	3.2	210.5
CC	2006Q1	8,263	8,263	7,674	1,275	1,188	614	47	2.6	3.1	175.0
CC	2006Q2	8,307	8,307	7,708	1,247	1,164	594	47	2.7	3.2	269.2
CC	2006Q3	8,240	8,240	7,639	1,231	1,156	594	46	2.8	3.4	269.2
CC	2006Q4	8,137	8,137	7,537	1,211	1,139	595	45	3.0	3.6	269.2
CC	2007Q1	7,991	7,991	7,393	1,197	1,129	574	44	3.2	3.9	269.2
CC	2007Q2	7,993	7,993	7,383	1,173	1,107	561	43	3.3	4.1	269.2
CC	2007Q3	7,903	7,903	7,299	1,159	1,091	544	44	3.2	4.2	175.0
CC	2007Q4	7,835	7,835	7,219	1,133	1,066	534	43	3.3	4.2	175.0
CC	2008Q1	7,692	7,692	7,084	1,123	1,056	527	44	3.3	4.2	175.0
CC	2008Q2	7,701	7,701	7,080	1,101	1,035	512	45	3.2	4.1	158.3
CC	2008Q3	7,631	7,631	7,008	1,096	1,036	509	44	3.1	4.0	158.3
CC	2008Q4	7,559	7,559	6,938	1,082	1,020	506	45	3.1	3.9	149.4
CC	2009Q1	7,480	7,480	6,849	1,059	999	498	46	3.0	3.7	147.3
CC	2009Q2	8,103	8,103	7,381	1,045	989	492	45	2.8	3.7	78.5
CC	2009Q3	8,016	8,016	7,302	1,042	988	492	47	2.7	3.5	77.6
CC	2009Q4	7,940	7,940	7,229	1,032	978	479	49	2.7	3.3	400.0
CC	2010Q1	7,770	7,770	7,077	1,020	963	459	49	2.5	3.2	100.0
CC	2010Q2	7,770	7,770	7,072	997	940	454	50	2.3	3.0	62.0
CC	2010Q3	7,707	7,707	7,013	994	940	450	50	2.2	2.8	62.0
CC	2010Q4	7,608	7,608	6,910	976	920	454	51	2.1	2.6	56.3
CC	2011Q1	7,469	7,469	6,784	961	906	451	52	2.0	2.5	68.6
CC	2011Q2	7,472	7,472	6,783	941	889	450	53	1.9	2.4	67.9
CC	2011Q3	7,402	7,402	6,716	933	879	443	54	1.9	2.3	67.9
CC	2011Q4	7,334	7,334	6,649	920	869	430	55	1.8	2.2	67.9

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 time periods. 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 Charge-Off Rates (cont.)

Loan	t_0	Initial	Step1	Step2	Step3	Step5	% 0s	Mean	75%	Max	
CC	2012Q1	7,236	7,236	6,546	913	862	438	56	1.7	2.1	67.9
CC	2012Q2	7,234	7,234	6,534	916	862	430	54	1.8	2.2	67.9
CC	2012Q3	7,170	7,170	6,465	907	853	409	55	1.7	2.1	67.9
CON	2009Q2	8,103	8,103	7,381	5,837	5,698	2,600	77	0.4	0.0	77.4
CON	2009Q3	8,016	8,016	7,302	5,872	5,693	2,672	71	0.5	0.2	202.2
CON	2009Q4	7,940	7,940	7,229	5,814	5,584	2,723	65	0.5	0.5	202.2
CON	2010Q1	7,770	7,770	7,077	5,735	5,461	2,680	58	0.7	0.7	202.2
CON	2010Q2	7,770	7,770	7,072	5,602	5,339	2,600	53	0.7	0.8	202.2
CON	2010Q3	7,707	7,707	7,013	5,596	5,311	2,555	47	0.8	0.9	202.2
CON	2010Q4	7,608	7,608	6,910	5,545	5,227	2,473	42	0.9	1.0	202.2
CON	2011Q1	7,469	7,469	6,784	5,482	5,133	2,427	36	1.0	1.1	202.2
CON	2011Q2	7,472	7,472	6,783	5,361	5,026	2,328	37	1.0	1.1	202.2
CON	2011Q3	7,402	7,402	6,716	5,377	5,028	2,333	38	1.0	1.1	202.2
CON	2011Q4	7,334	7,334	6,649	5,324	4,979	2,377	38	0.9	1.0	202.2
CON	2012Q1	7,236	7,236	6,546	5,266	4,932	2,403	39	0.9	1.0	202.2
CON	2012Q2	7,234	7,234	6,534	5,544	5,195	2,530	42	0.8	1.0	76.0
CON	2012Q3	7,170	7,170	6,465	5,536	5,184	2,541	43	0.8	0.9	76.0
CON	2012Q4	7,073	7,073	6,358	5,457	5,117	2,526	43	0.8	0.9	44.7
CON	2013Q1	6,931	6,849	6,212	5,379	5,042	2,548	44	0.8	0.9	100.0
CON	2013Q2	6,932	6,857	6,200	5,254	4,932	2,465	43	0.8	0.9	100.0
CON	2013Q3	6,882	6,807	6,144	5,246	4,917	2,512	44	0.7	0.9	76.0
CON	2013Q4	6,801	6,726	6,061	5,165	4,843	2,470	44	0.7	0.9	76.0
CON	2014Q1	6,648	6,576	5,913	5,094	4,767	2,415	44	0.7	0.9	76.0
CON	2014Q2	6,648	6,580	5,897	4,976	4,658	2,332	44	0.7	0.9	35.7
CON	2014Q3	6,580	6,512	5,821	4,961	4,645	2,303	44	0.7	0.9	35.7
CON	2014Q4	6,500	6,433	5,729	4,901	4,592	2,330	43	0.7	0.9	76.0
CON	2015Q1	6,340	6,273	5,564	4,834	4,522	2,304	43	0.7	0.9	35.7
CON	2015Q2	6,346	6,280	5,560	4,711	4,413	2,219	43	0.7	0.9	52.9
CON	2015Q3	6,269	6,206	5,479	4,696	4,408	2,214	43	0.7	0.9	52.9
CON	2015Q4	6,181	6,120	5,395	4,632	4,343	2,227	43	0.8	0.9	113.8
CON	2016Q1	6,057	5,996	5,256	4,545	4,258	2,185	43	0.7	0.9	52.9

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 time periods. The descriptive statistics are computed across N and T dimension of each panel.

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 (UR_{it}). Obtained from the *Bureau of Labor Statistics*. We use seasonally adjusted monthly data, time-aggregated to quarterly frequency by simple averaging.

Housing Price Index (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   Gibbs Sampling

The Gibbs sampler for the flexible RE/CRE specification with heteroskedasticity is initialized as follows:

- $y^*_{1:N,0:T}$ with $y_{1:N,0:T}$;

- ρ with a generalized method of moments (GMM) estimator $\hat{\rho}$, such as the orthogonal differencing in Arellano and Bover (1995) (implementation details can be found in the working paper version of Liu, Moon, and Schorfheide (2018));

- λ_i with $\hat{\lambda}_i = \frac{1}{T}\sum_{t=1}^{T}(y^*_{it} - \hat{\rho}y^*_{it-1})$;

- σ_i^2 with the variance of the GMM orthogonal differencing residues for each individual i, i.e., let $y^{\perp}_{it}, t = 1, \cdots, T-1$, denote the data after orthogonal differencing transformation, then $\hat{\sigma}_i^2 = \mathbb{V}_i(y^{\perp}_{it} - \hat{\rho}y^{\perp}_{it-1})$;

- for $z = \lambda, \sigma$, α_z with its prior mean; $\gamma_{z,i}$ with k-means with 10 clusters; $\{\Phi_k, \Sigma_k, \pi_{\lambda,k}\}_{k=1}^{K}$ and $\{\psi_k, \omega_k, \pi_{\sigma,k}\}_{k=1}^{K}$ are drawn from the conditional posteriors described in Section 3.2.

The Gibbs samplers for the other dynamic panel Tobit specifications are special cases in which some of the parameter blocks drop out. The Gibbs sampler for the pooled Tobit and linear specifications are initialized via pooled OLS, which ignores the censoring. We generate

a total of $M_0 + M = 10,000$ draws using the Gibbs sampler and discard the first $M_0 = 1,000$ draws.

B.2   Point Forecasts

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

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

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

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

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:

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

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

B.3   Set Forecasts

Algorithm for $1 - \alpha$ Credible Set Targeting Pointwise Coverage Probability:[14]

1. Sort $\{y^j_{iT+h}\}^M_{j=1}$ in increasing order and denote the elements of the sorted sequence by $y^{(j)}_{iT+h}$.

2. Now, separate the draws that are zero from the draws that are non-zero. Define j_0, such that $y^{(j)}_{iT+h} > 0$ for all $j > j_0$. Define $p_0 = j_0/M$. Let $p_1 = \max\{(1-\alpha) - p_0, 0\}$.

3. If $p_1 = 0$, the credible set is $\{0\}$ and we can terminate.

4. If $p_1 > 0$, then

   (a) Define $\alpha_1 = 1 - p_1$.

   (b) For $j = j_0 + 1$ to $\lfloor \alpha_1 M \rfloor$, compute the length of the j'th connected $1 - \alpha_1$ interval
   $$\Delta^{(j)} = y^{(j-1+\lceil (1-\alpha_1)M \rceil)}_{iT+h} - y^{(j)}_{iT+h}.$$

   (c) Choose the shortest interval that contains $\lceil (1-\alpha_1)M \rceil$ draws.

   (d) If $p_0 > 0$, then take the union of the shortest connected interval and $\{0\}$. Note that if the shortest interval constructed from the non-zero draws starts with $y^{(j_0+1)}_{iT+h}$, then we interpreted the set as connected, ranging from 0 to some non-zero upper bound.

Algorithm for $1 - \alpha$ Credible Set Targeting Average Coverage Probability:

1. For $i = 1$ to N:

   (a) Generate draws Y^j_{iT+1}, $j = 1, \ldots, M$, from the posterior predictive distribution.

   (b) If $Y^j_{iT+1} > 0$ then set $\pi^j_i = p(Y^j_{iT+1}|Y_i)$. If $Y^j_{iT+1} = 0$ then set $\pi^j_i = \infty$ with the understanding that ∞ is replaced by a large number in the code.

   Note: maybe use a subsample of MCMC draws to make computation less costly.

2. Let $\Pi = \{\pi^j_i \mid i = 1, \ldots, N \text{ and } j = 1, \ldots, M\}$. Sort the elements of Π in descending order. Denote the sorted elements of Π by $\pi^{(s)}$.

3. Define s_0, such that $\pi^{(s)} < \infty$ for all $s > s_0$. Define $p_0 = s_0/(NM)$. Let $p_1 = \max\{(1-\alpha) - p_0, 0\}$.

---

[14]This algorithm assumes that the continuous part of the predictive density is unimodal.

4. If $p_1 = 0$ then proceed as follows:

   (a) For each i, compute the fraction of zero-draws.

   (b) Based on the fraction of zero-draws, sort the units i in descending order.

   (c) Assign the set $\{0\}$ to the units with the highest fraction of zeros until the desired coverage is reached. All other units i are assigned $\emptyset$.

5. If $p_1 > 0$, then

   (a) Let $\bar{\Pi}^0$ be the set $Y^j_{iT+1} = 0$, i.e., $\pi^j_i = \infty$, so we have

   $$\bar{\Pi}^0 = \{\pi^{(s)} : s = 1, \cdots, s_0\}.$$

   Define the set
   $$\mathcal{J}^0_i = \{j \mid \pi^j_i \in \bar{\Pi}^0\}, \quad i = 1, \ldots, N.$$

   (b) Define $\alpha_1 = 1 - p_1$.

   (c) Let $\bar{\Pi}^1$ be the set of $1 - \alpha_1$ largest density values, excluding ∞, i.e.,

   $$\bar{\Pi}^1 = \{\pi^{(s)} : s = s_0 + 1, \cdots, \lceil (1 - \alpha)NM \rceil\}.$$

   Define the set
   $$\mathcal{J}^1_i = \{j \mid \pi^j_i \in \bar{\Pi}^1\}, \quad i = 1, \ldots, N.$$

   (d) If $\mathcal{J}^0_i \bigcup \mathcal{J}^1_i = \emptyset$, then $C_i = \emptyset$.

   (e) If $\mathcal{J}^0_i \neq \emptyset$ and $\mathcal{J}^1_i = \emptyset$, then $C_i = \{0\}$.

   (f) If $\mathcal{J}^1_i \neq \emptyset$, then for each i compute

   $$
   \begin{aligned}
   \underline{Y}_{iT+1} &= \operatorname{argmin}_{Y^j_{iT+1}} Y^j_{iT+1} \text{ s.t. } j \in \mathcal{J}^1_i \\
   \overline{Y}_{iT+1} &= \operatorname{argmax}_{Y^j_{iT+1}} Y^j_{iT+1} \text{ s.t. } j \in \mathcal{J}^1_i
   \end{aligned}
   $$

   If $\mathcal{J}^0_i \neq \emptyset$, then $C_i = \{0\} \cup [\underline{Y}_{iT+1}, \overline{Y}_{iT+1}]$, else $C_i = [\underline{Y}_{iT+1}, \overline{Y}_{iT+1}]$. Check whether $\{0\}$ and $[\underline{Y}_{iT+1}, \overline{Y}_{iT+1}]$ are connected.

B.4 Density Forecasts

The log-predictive density can be approximated by

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

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

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

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

$$
CRPS(\hat{F}, y_{iT+h}) = \int_0^\infty \big(\hat{F}(x) - \mathbb{I}\{y_{iT+h} \le x\}\big)^2 dx. \tag{A.10}
$$

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

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

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,

$$
CRPS(\hat{F}, y^o) \tag{A.13}
$$
$$
= \sum_{j=2}^{M_*} \left[\hat{F}(y^{(j-1)}) - 0\right]^2 (y^{(j)} - y^{(j-1)}) + \left[\hat{F}(y^{(M_*)}) - 0\right]^2 (y^o - y^{(M_*)})
$$
$$
+ \left[\hat{F}(y^{(M_*)}) - 1\right]^2 (y^{(M_*+1)} - y^o) + \sum_{j=M_*+2}^{M} \left[\hat{F}(y^{(j-1)}) - 1\right]^2 (y^{(j)} - y^{(j-1)}).
$$

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:

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

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

$$\sum_{j=2}^{M_*} \left[\hat{F}(y^{(j-1)}) - 0\right]^2 (y^{(j)} - y^{(j-1)}) + \left[\hat{F}(y^{(M_*)}) - 0\right]^2 (y^o - y^{(M_*)})$$

$$+ \left[\hat{F}(y^{(M_*)}) - 1\right]^2 (y^{(M_*+1)} - y^o) + \sum_{j=M_*+2}^{M} \left[\hat{F}(y^{(j-1)}) - 1\right]^2 (y^{(j)} - y^{(j-1)})$$

$$= \sum_{j=2}^{M_*} \frac{(j-1)^2}{M^2} (y^{(j)} - y^{(j-1)}) + \frac{M_*^2}{M^2} (y^o - y^{(M_*)})$$

$$+ \frac{(M - M_*)^2}{M^2} (y^{(M_*+1)} - y^o) + \sum_{j=M_*+2}^{M} \frac{(M - (j-1))^2}{M^2} (y^{(j)} - y^{(j-1)})$$

$$= \frac{1}{M^2} \left[-y^{(1)} + \sum_{j=2}^{M_*} \left((j-1)^2 - j^2\right) y^{(j)} + \sum_{j=M_*+1}^{M-1} \left((M - (j-1))^2 - (M-j)^2\right) y^{(j)} \right.$$

$$\left. + y^{(M)} + \left(M_*^2 - (M - M_*)^2\right) y^o \right]$$

$$= \frac{1}{M^2} \left[-\sum_{j=1}^{M_*} (2j-1) y^{(j)} + \sum_{j=M_*+1}^{M} (2M - 2j + 1) y^{(j)} \right] + \frac{2M_* - M}{M} y^o,$$

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

$$\sum_{j=2}^{M} \left[\hat{F}(y^{(j-1)}) - 0\right]^2 (y^{(j)} - y^{(j-1)}) + \left[1 - 0\right]^2 (y^o - y^{(M)})$$

$$= \sum_{j=2}^{M} \frac{(j-1)^2}{M^2} (y^{(j)} - y^{(j-1)}) + (y^o - y^{(M)})$$

$$= \frac{1}{M^2} \left[-y^{(1)} + \sum_{j=2}^{M} \left((j-1)^2 - j^2\right) y^{(j)} \right] + y^o$$

$$= -\frac{1}{M^2} \sum_{j=1}^{M_*} (2j-1) y^{(j)} + y^o,$$

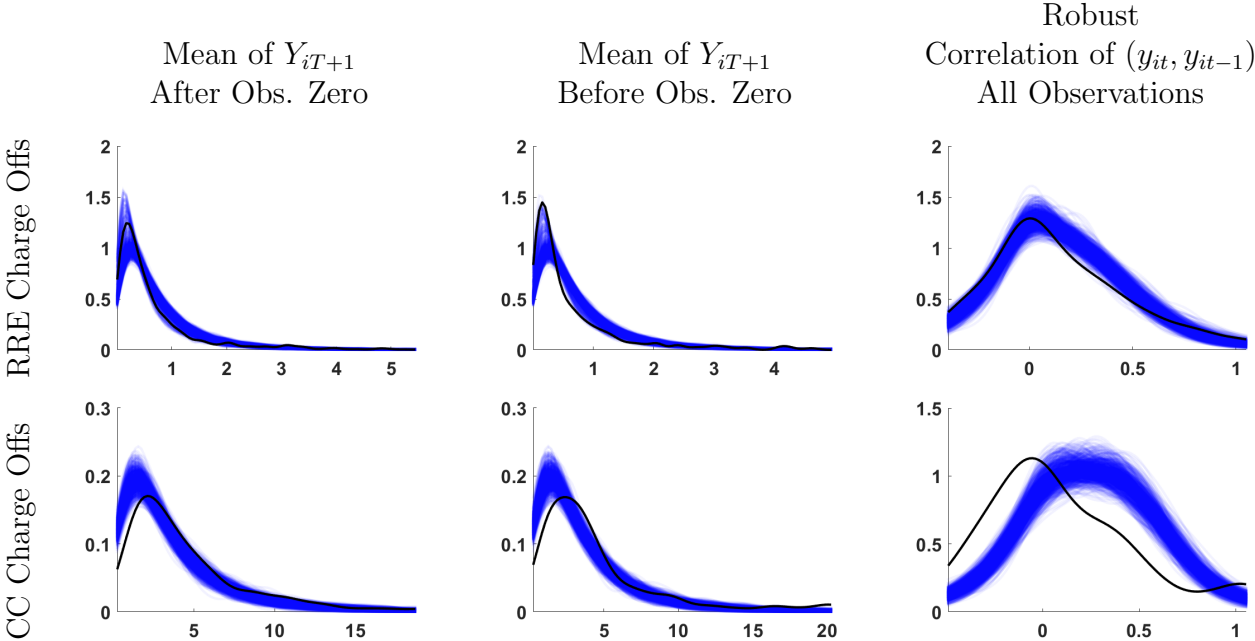which is equal to (A.15) when $M_* = M$. And for Case 2, after substituting $\hat{F}$, the RHS of

(A.12) becomes

$$
\left[0-1\right]^2(y^{(1)} - y^o) + \sum_{j=2}^{M} \left[\hat{F}(y^{(j-1)}) - 1\right]^2 (y^{(j)} - y^{(j-1)})
$$

$$
= (y^{(1)} - y^o) + \sum_{j=2}^{M} \frac{(M - (j-1))^2}{M^2}(y^{(j)} - y^{(j-1)})
$$

$$
= \frac{1}{M^2}\left[\sum_{j=1}^{M-1} \left((M - (j-1))^2 - (M-j)^2\right)y^{(j)} + y^{(M)}\right] - y^o
$$

$$
= \frac{1}{M^2}\sum_{j=1}^{M}(2M - 2j + 1)y^{(j)} - y^o,
$$

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

C   Additional Empirical Results

Figure A-1: Additional Posterior Predictive Checks: Cross-sectional Distribution of Sample Statistics

Notes: Heteroskedastic flexible CRE specification. The estimation sample ranges from 2007Q2 ($t = 0$) to 2009Q4 ($t = T = 10$). The black lines are computed from the actual data. Each hairline corresponds to a simulation of a sample $\tilde{Y}_{1:N,0:T+1}$ of the panel Tobit model based on a parameter draw from the posterior distribution. Robust autocorrelations are computed using the MM estimator in Chang and Politis (2016).

Figure A-2: Point Forecast Accuracy: All Samples

Benchmark vs. Alt. Prior
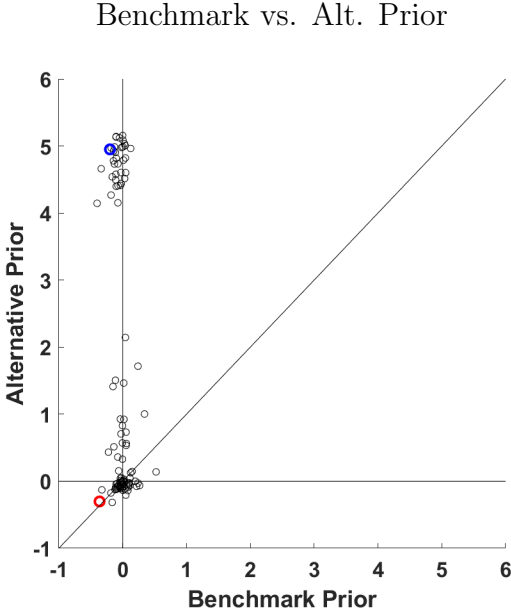
Notes: Flexible CRE specification with heteroskedasticity. The panels provide pairwise comparisons of log RMSE differentials with respect to Pooled Tobit under the benchmark prior. The red (blue) circle corresponds to CC (RRE).