Quantitative Economics 14 (2023), 117–159 1759-7331/20230117

Forecasting with a panel Tobit model

Laura Liu
Department of Economics, Indiana University

Hyungsik Roger Moon
Department of Economics, University of Southern California and Yonsei

Frank Schorfheide
Department of Economics, University of Pennsylvania, CEPR, NBER, and PIER

We use a dynamic panel Tobit model with heteroskedasticity to generate 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. In addition to density forecasts, 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 loan charge-off rates for small banks.

Keywords. Bayesian inference, density forecasts, loan charge-offs, panel data, set forecasts, Tobit model.

JEL classification. C11, C14, C23, C53, G21.

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. A charge-off occurs if a loan is deemed unlikely to be collected because the borrower has become delinquent. The prediction of charge-off rates is interesting to banks, regulators, and investors because they are losses on loan portfolios. If charge-off rates 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.

Laura Liu: lauraliu@iu.edu
Hyungsik Roger Moon: moonr@usc.edu
Frank Schorfheide: schorf@ssc.upenn.edu

We thank Mitchell Berlin, Siddhartha Chib, Tim Armstrong, and participants at various seminars and conferences for helpful comments and suggestions. Moon and Schorfheide gratefully acknowledge financial support from the National Science Foundation under Grants SES 1625586 and SES 1424843, respectively.

 https://doi.org/10.3982/QE1505

In a panel data setting, cross-sectional heterogeneity in the data is modeled through unit-specific parameters. The more precisely they are estimated, the more accurate the forecasts are. The challenge in forecasting panels with a short time dimension is that the data set does not contain a lot of information about the heterogeneous parameters. A natural way of adding information to the estimation of these parameters is the use of prior distributions. The key insight in panel data applications is that one can extract information from the cross-section and equate the prior distribution with the cross-sectional distribution of unit-specific coefficients. An empirical Bayes implementation of this idea creates a point estimate of the cross-sectional distribution of the heterogeneous coefficients and then conditions the subsequent posterior calculations on the estimated prior distribution. The classic James–Stein estimator for a vector of means can be interpreted as an empirical Bayes estimator.[1]

Rather than pursuing an empirical Bayes approach, we conduct a full Bayesian analysis by specifying a hyperprior for the distribution of heterogeneous coefficients and constructing a joint posterior for the coefficients of this hyperprior as well as the actual unit-specific coefficients. This approach can in principle handle quite general nonlinearities and generate predictions under a wide variety of loss functions. It is preferable for interval and density forecasts, because it captures all sources of uncertainty.

The contributions of our paper are threefold. First, we extend the full Bayesian estimation and prediction with a linear panel data model in Liu (2022) to a dynamic panel Tobit model with heteroskedastic innovations and correlated random effects. We hereby build on work on the Bayesian estimation of static, dynamic, and panel Tobit models by Chib (1992), Wei (1999), Baranchuk and Chib (2008), and Li and Zheng (2008).

Second, we construct interval forecasts that target average posterior coverage probability across all units in our panel instead of pointwise coverage probability for each unit. We show that it is optimal to generate these forecasts as the highest posterior density sets that use 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 take the form of the union of one or more intervals and the value zero, and thus we refer to them as set forecasts subsequently. We prove that the empirical coverage frequency converges to the average nominal coverage frequency of the sets as the cross-sectional dimension of the panel tends to infinity. This result is connected to similar findings in the literature on nonparametric function estimation and dates back to Wahba (1983) and Nychka (1988). The underlying insights also have been recently used in concurrent research by Armstrong, Kolesár, and Plagborg-Møller (2022) to construct empirical Bayes confidence intervals for vectors of means that are valid for multiple priors. In the Monte Carlo study and the empirical application, the proposed Bayesian set forecasts have good finite sample frequentist coverage properties in the cross-section.

Third, we present a novel application in which we forecast bank-level loan charge-off rates. Our empirical analysis is based on more than 100 short panel data sets with a

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

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. We also include local economic conditions as bank-specific regressors with homogeneous coefficients. For each data set, we document the density forecasting performance of several model specifications. We find that allowing for heteroskedasticity is important for good density and set forecasting performance. Overall, a specification with flexibly modeled correlated random effects and heteroskedasticity performs well in terms of density forecasting and is used in the subsequent analysis. In addition, we generate maps that compare the spatial distribution of predicted loan losses during and after the Great Recession and plot cross-sectional distribution of set forecasts. We document how set forecasts change as we move from targeting pointwise coverage probability to targeting average coverage probability. The latter approach smooths out differences among the lengths of the set forecasts and overall improves the forecasts with respect to both coverage probability and average length.

The heterogeneous intercepts in our model can be interpreted as estimates of the quality of the banks' loan portfolios. Loan quality is potentially determined by many factors: the risk taking behavior of the bank, the potential customer base, and its ability to efficiently screen borrowers. In regressing heterogeneous coefficient estimates on bank characteristics, we find that bank size as measured in total assets is positively related to inverse quality of the loan portfolio. A favorable interpretation of this finding is that larger banks are able to take higher risks on loans because they are better diversified or have a higher tolerance for risk. However, overall bank characteristics explain only a very small fraction of the estimated heterogeneity.

Because the Tobit model is nonlinear, the effect of a change in local economic conditions that enter the model with homogeneous coefficients depends on the heterogeneous intercept and is thereby bank specific. We are able to compute a posterior distribution of the "treatment" effect for each bank and decompose it into an extensive-margin effect (a bank switches from no charge-offs to positive charge-offs during an economic downturn) and an intensive-margin effect (a bank increases its positive charge-offs during a downturn). We find that the variation in charge-off rates generated by local economic conditions is very small compared to the variation due to the heterogeneous intercept estimates.

Our paper relates to several branches of the literature. We build on the Bayesian literature on the estimation of censored regression models.[2] The approach of using data augmentation for limited-dependent variable models that impute the latent uncensored variables dates back to Chib (1992) and Albert and Chib (1993). To sample the latent 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). Bayesian panel Tobit models have been estimated by Baranchuk and Chib (2008) and Li and Zheng (2008). Our flexible benchmark model is most closely related to the semiparametric model of Li and Zheng (2008), which we generalize by introducing heteroskedasticity through a latent unit-specific error variance and allowing for

[2]A general survey of the literature on Bayesian estimation of univariate and multivariate censored regression models can be found, for instance, in the handbook chapter by Li and Tobias (2011).

a more flexible form of correlated random effects. As mentioned previously, the former is very important for the density and set forecast performance.[3]

We model the unknown distribution of the heterogeneous coefficients (intercepts and innovation variances) as Dirichlet process mixtures (DPM) 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 DPM.[4] Examples of econometrics papers that use DPMs in the panel data context are Hirano (2002), Burda and Harding (2013), Rossi (2014), and Fisher and Jensen (2022). The implementation of our Gibbs sampler relies on Ishwaran and James (2001, 2002).

As an alternative to a full Bayesian analysis, recent papers by Gu and Koenker (2017a,b) and Liu, Moon, and Schorfheide (2020) have pursued an empirical Bayes strategy to generate predictions based on linear panel data models with heterogeneous coefficients. Forecasts from empirical Bayes and full Bayesian estimation approaches have desirable optimality properties as the cross-sectional dimension of the data set gets large. Liu, Moon, and Schorfheide (2020) generalize optimality results for the estimation of a vector of means in Brown and Greenshtein (2009) to a linear dynamic panel data forecasting setting. Liu (2022) shows that the predictive density obtained from the full Bayesian analysis of a linear panel data model converges to the predictive density derived from the true cross-sectional distribution of the heterogeneous coefficients as the cross-section gets large.

There also exists a literature on estimating the determinants of loan losses. This literature often uses nonperforming loans (loans that have not been serviced for more than 90 days) and tends to ignore the censoring, which is reasonable if one uses an average across banks but can be problematic if one uses bank-level data. The two papers most closely related to our work are Ghosh (2015, 2017). We base our choice of bank-characteristic regressors on these papers.

The remainder of our 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 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. Detailed derivations and proofs, a description of the data sets, and additional simulation and empirical results are relegated to the Online Appendix (Liu, Moon, and Schorfheide (2023)).

[3] Baranchuk and Chib (2008) report some results on point forecasts of the probability of zeros versus nonzeros, whereas we focus on set and density forecasts.

[4] 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:

$$\begin{aligned} y_{it} &= y_{it}^* \mathbb{I}\{y_{it}^* \geq 0\}, \\ y_{it}^* | (Y_{1:N,0:t-1}^*, X_{1:N,-1:t-1}, \lambda_{1:N}, \sigma_{1:N}^2, \rho, \beta, \xi) &\overset{\text{indep}}{\sim} N(\lambda_i + \rho y_{it-1}^* + \beta' x_{it-1}, \sigma_i^2), \end{aligned} \tag{1}$$

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. Throughout the paper, we abbreviate sequences of the form $(a_1, \ldots, a_n)$ by $a_{1:n}$. For instance, $Y_{1:N,0:t-1}^* = \{(y_{10}^*, \ldots, y_{N0}^*), \ldots, (y_{1t-1}^*, \ldots, y_{Nt-1}^*)\}$, and $\lambda_{1:N} = (\lambda_1, \ldots, \lambda_N)$. The $n_x \times 1$ vector x_{it} comprises a set of sequentially exogenous regressors. ξ is a vector of hyperparameters defined in (3) below that does not affect the conditional distribution of y_{it}^*. It is assumed that conditional on the parameters and the regressors x_{it-1}, the observations y_{it} are cross-sectionally independent. The distributional assumption in (1) implies that we can write

$$p(y_{it}^* | Y_{1:N,0:t-1}^*, X_{1:N,-1:t-1}, \lambda_{1:N}, \sigma_{1:N}^2, \rho, \beta, \xi) = p(y_{it}^* | y_{it-1}^*, x_{it-1}, \lambda_i, \sigma_i^2, \rho, \beta), \tag{2}$$

which we will use subsequently to simplify formulas. 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

$$p(\lambda_i, y_{i0}^*, \sigma_i^2 | x_{i,-1}, \xi), \tag{3}$$

assuming cross-sectional independence of the heterogeneous coefficients.[5] Here, ξ is a hyperparameter vector that indexes a family of CRE distributions. For instance, the candidate distribution of $(\lambda_i, y_{i0}^*, \ln \sigma_i^2)$ could be jointly normal with a mean that is a linear function of $x_{i,-1}$. In this case, ξ would include the parameters of the conditional mean function and the nonredundant parameters of the covariance matrix. To achieve a flexible representation of the distribution of $(\lambda_i, y_{i0}^*, \sigma_i^2)$, we consider a family of mixtures of normal distributions in Section 3. We define the homogeneous parameter $\theta = [\rho, \beta']'$ and complete the model with the specification of a prior distribution for (θ, ξ).

Our model is closely related to the panel Tobit models of Baranchuk and Chib (2008) and Li and Zheng (2008), henceforth BC and LZ, respectively. However, the modeling approaches differ with respect to the treatment of coefficient heterogeneity and heteroskedasticity.[6] As in LZ, we restrict regression coefficient heterogeneity to the intercept. We also follow LZ in modeling the CRE distribution in (3) nonparametrically, albeit

[5] We consider period $t = -1$ for x in the conditioning set because of the timing assumption that charge-off rates can only respond with a one-period lag to changes in local economic conditions so as to accommodate possible sequentially exogenous regressors. See Section 2.1 for more details.

[6] As in the panel Probit model of Chib and Jeliazkov (2006), one could allow for additional lags of y_{it}^*.

the details are slightly different. Because the regressors x_{it} in our application are not assumed to be strictly exogenous, we condition the distribution of (λ_i, y_{i0}^*) only on the initial values $x_{i,-1}$ and not on other x_{it}s. The most important difference between our specification and that of LZ is that we allow for heterogeneous innovation variances σ_i^2, whereas LZ set $\sigma_i^2 = \sigma^2$ for all i. As documented in Section 5.2, σ_i^2 heterogeneity is very important for the construction of an accurate set and density forecasts in our empirical application.

BC restrict the distribution of the heterogeneous coefficients to be normal, but they do allow regression coefficients other than the intercept to be heterogeneous.[7] Rather than linking the heterogeneity to the regressors x_{it}, they let the mean of the distribution depend on additional unit-specific covariates. Instead of embedding additional covariates (such as bank characteristics) ex ante into (3), we run ex post regressions of estimates of the ratio $\widehat{\lambda_i/\sigma_i}$ on additional unit-specific covariates to explore potential relationships. The reasons for conducting an ex post analysis in our application are threefold: (i) it is not clear ex ante which bank characteristics are relevant, (ii) the relationship between bank characteristics and cross-sectional heterogeneity could be nonlinear, and (iii) bank characteristics may only explain a small fraction of the cross-sectional heterogeneity.

BC's interaction between regressors and the normal CRE distribution generates heteroskedasticity in what could be interpreted as composite error term that consists of a homoskedastic innovation in the regression equation for y_{it}^* and the randomness in the heterogeneous coefficients scaled by the regressors. In our model specification, the heteroskedasticity is unrelated to the regressors x_{it} because we are treating the σ_i^2 as random effects. A relationship to the regressors could be generated through a CRE specification for σ_i^2, but we did not pursue this extension because in our application the regressors, local unemployment, and house price growth, cannot explain the dispersion in σ_i^2.

In the remainder of this section, we discuss our assumptions about the simultaneous determination of outcomes y_{it} and regressors x_{it} in Section 2.1, the derivation of the posterior predictive density in Section 2.2, the density forecast evaluation criteria in Section 2.3, and the construction and evaluation of set forecasts in Section 2.4.

2.1 *Simultaneity and timing assumptions*

In our application, y_{it} corresponds to bank-level loan charge-off rates and the regressors x_{it} measure local economic conditions, such as unemployment and house prices, in the state in which the bank operates.[8] In this context, it is plausible to assume that there is feedback from the bank charge-offs, which affect profitability and overall health of the banking sector, to the local economic conditions.

The key assumption that we are making throughout the paper is that charge-off rates are only affected by lagged economic conditions and not by contemporaneous

[7] Our framework can be easily extended to accommodate heterogeneous slope coefficients (see Liu, Moon, and Schorfheide (2020) and Liu (2022)).

[8] We consider a sample of small banks that conduct most of their business locally.

economic conditions. For concreteness, suppose that x_{it} corresponds to economic conditions in the state in which bank i operates. We assume that the state-level conditions in period $t = 0, \ldots, T$ are described by the conditional density

$$\begin{aligned} &p\big(X_{1:N,t}|Y_{1:N,0:t}, Y^*_{1:N,0:t}, X_{1:N,-1:t-1}, \theta_x, \lambda_{1:N}, \sigma^2_{1:N}, \theta, \xi\big) \\ &\quad = p(X_{1:N,t}|Y_{1:N,0:t}, X_{1:N,-1:t-1}, \theta_x). \end{aligned} \tag{4}$$

Thus, we allow current charge-offs to affect current state-level conditions. However, we assume that $X_{1:N,t}$ does not separately depend on the latent variables $Y^*_{1:N,0:t}$ and the heterogeneous coefficients (λ_i, σ_i^2). In our application, only actual charge-off rates are assumed to matter for economic outcomes. θ_x is a vector of parameters determining the law of motion for the state-level conditions.

Timing restrictions such as the one above have traditionally been widely used in the macroeconometric literature on structural vector autoregressions; see, for instance, the survey by Ramey (2016). Here, we are assuming that a deterioration of macroeconomic conditions affects banks' decisions to write off loans with a one-period delay, where the length of a period is a quarter in our application.[9] Combining (1), (2), and (4), we can write

$$\begin{aligned} &p\big(Y_{1:N,1:T}, Y^*_{1:N,1:T}, X_{1:N,0:T}|Y_{1:N,0}, Y^*_{1:N,0}, X_{1:N,-1}, \lambda_{1:N}, \sigma^2_{1:N}, \theta, \xi, \theta_x\big) \\ &\quad = \prod_{t=1}^{T}\left\{ p(X_{1:N,t}|Y_{1:N,0:t}, X_{1:N,-1:t-1}, \theta_x) \times \left[\prod_{i=1}^{N} p\big(y_{it}|y^*_{it}\big)\, p\big(y^*_{it}|y^*_{it-1}, x_{it-1}, \lambda_i, \sigma_i^2, \theta\big)\right]\right\} \\ &\qquad \times p(X_{1:N,0}|Y_{1:N,0}, X_{1:N,-1}, \theta_x) \\ &\quad = \left\{\prod_{i=1}^{N}\left[\prod_{t=1}^{T} p\big(y_{it}|y^*_{it}\big)\, p\big(y^*_{it}|y^*_{it-1}, x_{it-1}, \lambda_i, \sigma_i^2, \theta\big)\right]\right\} \\ &\qquad \times \prod_{t=0}^{T} p(X_{1:N,t}|Y_{1:N,0:t}, X_{1:N,-1:t-1}, \theta_x). \end{aligned} \tag{5}$$

In slight abuse of notation, $p(y_{i0}|y^*_{i0})$ represents the censoring. 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 $y^*_{it} > 0$. Because the system is triangular, the panel Tobit component in (1) can be estimated independently of (4) and without the use of instrumental variables.

2.2 *Posterior predictive densities*

Our goal is to generate forecasts of $Y_{1:N,T+h}$ conditional on the observations $(Y_{1:N,0:T}, X_{1:N,-1:T})$. In the empirical analysis in Section 5, we focus on $h = 1$-step-ahead forecasts, which require the predictor x_{iT}, which is known at the forecast origin $t = T$. The extension to multistep forecasts is discussed in Section 3.3. Because in a

[9] Relaxing this assumption is beyond the scope of this paper.

Bayesian framework uncertainty with respect to parameters, latent variables, and future shocks is treated identically through the use of random variables, it is conceptually straightforward to construct a predictive distribution of $Y_{1:N,T+1}$ conditional on $(Y_{1:N,0:T}, X_{1:N,-1:T})$ by integrating out all sources of uncertainty. The general approach is summarized, for instance, in Geweke and Whiteman (2006). We subsequently describe the integration steps required for our panel Tobit model.

According to (3), the distribution of $(Y_{1:N,0}, Y^*_{1:N,0})$ conditional on $X_{1:N,-1}$ does not depend on θ_x. Using the factorization in (5), the CRE density (3), and the prior $p(\theta, \xi) = p(\theta)p(\xi)$, we can write the posterior distribution of the parameters and time-T latent variables as

$$\begin{aligned} &p\big(Y^*_{1:N,T}, \lambda_{1:N}, \sigma^2_{1:N}, \theta, \xi | Y_{1:N,0:T}, X_{1:N,-1:T}\big) \\ &\quad \propto \Bigg[\prod_{i=1}^{N} \int \Bigg(\prod_{t=1}^{T} p\big(y_{it}|y^*_{it}\big) p\big(y^*_{it}|y^*_{it-1}, x_{it-1}, \lambda_i, \sigma^2_i, \theta\big)\Bigg) \\ &\qquad \times p\big(y_{i0}|y^*_{i0}\big) p\big(\lambda_i, y^*_{i0}, \sigma^2_i | x_{i,-1}, \xi\big)\, dY^*_{i,0:T-1}\Bigg] p(\theta)p(\xi), \end{aligned} \tag{6}$$

where $\propto$ denotes proportionality. The posterior predictive distribution for units $i = 1, \ldots, N$ is given by

$$\begin{aligned} &p(Y_{1:N,T+1}|Y_{1:N,0:T}, X_{1:N,-1:T}) \\ &\quad = \int \prod_{i=1}^{N} \Bigg[\int \int p\big(y_{iT+1}|y^*_{iT+1}\big) p\big(y^*_{iT+1}|y^*_{iT}, x_{iT}, \lambda_i, \sigma^2_i, \theta\big) \\ &\qquad \times p\big(y^*_{iT}, \lambda_i, \sigma^2_i | \theta, \xi, Y_{1:N,0:T}, X_{1:N,-1:T}\big)\, dy^*_{iT}\, d\big(\lambda_i, \sigma^2_i\big)\Bigg] \\ &\qquad \times p(\theta, \xi | Y_{1:N,0:T}, X_{1:N,-1:T})\, d(\theta, \xi). \end{aligned} \tag{7}$$

Draws from $p(Y_{1:N,T+1}|Y_{1:N,0:T}, X_{1:N,-1:T})$ can be generated by sampling $(Y^*_{1:N,T}, \lambda_{1:N}, \sigma^2_{1:N}, \theta, \xi)$ from the posterior (6) and then evaluating the autoregressive law of motion for y^*_{it} in (1) for $t = T + 1$.

To simplify the notation, we drop $X_{1:N,-1:T}$ from the conditioning set in the remainder of this section. Moreover, we denote the forecast horizon by h again with the understanding that the discussion of multistep forecasts is deferred to Section 3.3. 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 indicate 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 distribution for realizations of y_{iT+h} that are greater than zero:

$$p(y_{iT+h}|Y_{1:N,0:T}) = \mathbb{P}^{y_{iT+h}}_{Y_{1:N,0:T}}\{y_{iT+h} = 0\}\delta_0(y_{iT+h}) + p_c(y_{iT+h}|Y_{1:N,0:T})\mathbb{I}\{y_{iT+h} \geq 0\}. \tag{8}$$

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.

2.3 *Evaluating density forecasts*

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

$$\begin{aligned}\mathrm{LPS}_h(M) = \frac{1}{N}\sum_{i=1}^{N} \ln\big(&\mathbb{I}\{y_{iT+h} = 0\} \cdot \mathbb{P}^{y_{iT+h}}_{Y_{1:N,0:T}}\{y_{iT+h} = 0|M\} \\ &+ \mathbb{I}\{y_{iT+h} > 0\} p(y_{iT+h}|Y_{1:N,0:T})\big)\end{aligned} \tag{9}$$

and continuous ranked probability scores (CRPSs). The CRPS measures the L_2 distance between the cumulative distribution function $F^{y_{iT+h}}_{Y_{1:N,0:T}}(y|M)$ associated with $p(y_{iT+1}|Y_{1:N,0:T})$ and a "perfect" density forecast, which assigns probability one to the realized y_{iT+h}. Then

$$\mathrm{CRPS}_h(M) = \frac{1}{N}\sum_{i=1}^{N} \int_0^{\infty} \big(F^{y_{iT+h}}_{Y_{1:N,0:T}}(y|M) - \mathbb{I}\{y_{iT+h} \leq y\}\big)^2\, dy. \tag{10}$$

Both LPS and CRPS are proper scoring rules, meaning that it is optimal for the forecaster to truthfully reveal her predictive density (Gneiting and Raftery (2007)).

2.4 *Constructing and evaluating set forecasts*

We construct set forecasts from the posterior predictive distribution $p(y_{iT+h}|Y_{1:N,0:T})$ in (7) of the form:

$$C_{iT+h|T}(Y_{1:N,0:T}) = \{0\} \cup \left(\bigcup_{k=1}^{K_i} [a_{ik}, b_{ik}]\right) \tag{11}$$

with the understanding that (i) $C_i = \{0\}$ if $K_i = 0$, (ii) a_{i1} may be equal to zero, and (iii)

$$a_{i1} < b_{i1} < a_{i2} < b_{i2} < \cdots < a_{iK_i} < b_{iK_i}.$$

The $\{0\}$ value arises from 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 (8).[10] The disjoint interval segments may arise if the continuous part of the predictive density is multimodal. If we target an average coverage probability in the cross-section, then for some units i we might obtain the empty set, that is, $C_{iT+h|T}(Y_{1:N,0:T}) = \emptyset$.

[10]Because in our model the support of the posterior predictive distribution of y^*_{iT+h} includes $y < 0$, the probability of censoring is strictly positive and the set that includes $\{0\}$ is strictly shorter than the one without zero.

Constructing set forecasts To generate the set forecasts, we adopt a Bayesian approach and require that the probability of $\{y_{iT+h} \in C_{iT+h|T}(Y_{1:N,0:T})\}$ conditional on having observed $Y_{1:N,0:T}$ reaches a prespecified level. Given that the estimation of the Tobit model is executed with Bayesian techniques, the use of posterior predictive credible sets is natural. We distinguish between forecasts that are constructed to satisfy the coverage probability constraint pointwise, that is,

$$\mathbb{P}^{y_{iT+h}}_{Y_{1:N,0:T}}\big\{y_{iT+h} \in C_{iT+h|T}(Y_{1:N,0:T})\big\} \geq 1-\alpha \quad \text{for all } i, \tag{12}$$

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

$$\frac{1}{N}\sum_{i=1}^{N} \mathbb{P}^{y_{iT+h}}_{Y_{1:N,0:T}}\big\{y_{iT+h} \in C_{iT+h|T}(Y_{1:N,0:T})\big\} \geq 1-\alpha. \tag{13}$$

The latter approach allows the sets $C_{iT+h|T}(Y_{1:N,0:T})$ 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."

It is well known that the shortest credible sets take the form of highest posterior density (HPD) sets. Suppose that we require to satisfy the coverage constraint for each i individually. If $\mathbb{P}^{y_{iT+h}}_{Y_{1:N,0:T}}\{y_{iT+h}=0\} \geq 1-\alpha$, then $C_{iT+h|T}(Y_{1:N,0:T}) = \{0\}$. Otherwise, the set takes the form

$$C_{iT+h|T}(Y_{1:N,0:T}) = \{0\} \cup \big\{y_{iT+h} \mid p_c(y_{iT+h}|Y_{1:N,0:T})\mathbb{I}\{y_{iT+h} \geq 0\} \geq \kappa_i\big\}, \tag{14}$$

where the threshold κ_i is chosen such that

$$\int_{y_{iT+h} \in C_{iT+h|T}(Y_{1:N,0:T})} p_c(y_{iT+h}|Y_{1:N,0:T})\mathbb{I}\{y_{iT+h} \geq 0\}\, dy_{iT+h} = 1-\alpha - \mathbb{P}^{y_{iT+h}}_{Y_{1:N,0:T}}\{y_{iT+h}=0\}.$$

Because $p_c(y|\cdot)$ is a continuous density, the HPD set can be represented as a collection of disjoint intervals as in (11).

If the objective is to minimize average length across i conditional on the constraint on coverage probability holding only on average, then the unit-specific thresholds κ_i in (14) are replaced by a common threshold κ that applies to all units i. One can establish the optimality of the common threshold as follows. Suppose that one lowers the threshold for unit i ($\kappa_i < \kappa$) and raises it for unit j ($\kappa_j > \kappa$). This lengthens the set for unit i by $\delta_i > 0$ and shortens the set for unit j by $\delta_j < 0$. The increase in coverage probability for unit i, $\Delta\pi_i > 0$, is less than $\delta_i\kappa$, whereas the decrease in coverage probability for unit j, $\Delta\pi_j < 0$, is less than $\delta_j\kappa$. Because we are holding the overall coverage probability constant, we obtain

$$\delta_i\kappa > \Delta\pi_i = -\Delta\pi_j > -\delta_j\kappa.$$

Thus, $\delta_i > -\delta_j$, which means that the overall average length increases and the uniform threshold of κ dominates.

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

$$\frac{1}{N}\sum_{i=1}^{N}\mathbb{I}\big\{y_{iT+h}\in C_{iT+h|T}(Y_{1:N,0:T})\big\} \tag{15}$$

and the average length of the sets $C_{iT+h|T}(Y_{1:N,0:T})$,

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K_i}(b_{ik}-a_{ik}). \tag{16}$$

Rather than trading off average length against deviations of average coverage frequency from the nominal coverage probability in a single criterion, we simply report both.[11]

The relationship between the nominal credible level of the set forecasts and the empirical coverage frequency is delicate. In Theorem 2.1 below, we provide high-level regularity conditions under which

$$\frac{1}{N}\sum_{i=1}^{N}\mathbb{I}\big\{y_{iT+h}\in C_{iT+h|T}(Y_{1:N,0:T})\big\}\xrightarrow{p}1-\alpha \tag{17}$$

in $\mathbb{P}^{Y_{1:N,0:T},Y_{1:N,T+h}}$ probability as $N\longrightarrow\infty$. Underlying these results is the well-known insight (see, for instance, the textbook by Robert (1994)) that, for a generic parameter ς and data set Y, the following relationship between credible sets and confidence sets holds:

$$\mathbb{P}^{\varsigma,Y}\big\{\varsigma\in C(Y)\big\}=\int_{Y}\mathbb{P}_{Y}^{\varsigma}\big\{\varsigma\in C(Y)\big\}\,d\mathbb{P}^{Y}=\int_{\varsigma}\mathbb{P}_{\varsigma}^{Y}\big\{\varsigma\in C(Y)\big\}\,d\mathbb{P}^{\varsigma}.$$

Thus, $1-\alpha$ Bayesian credible sets have on average $1-\alpha$ frequentist coverage probability, but not pointwise for each ς. In our framework, the cross-sectional averaging across i approximates the integration under the prior distribution. The basic insight has previously been used in the literature on nonparametric function estimation, dating back to Wahba (1983) and Nychka (1988), to obtain results that link average coverage probabilities to Bayesian credible levels. More recently, Armstrong, Kolesár, and Plagborg-Møller (2022) constructed empirical Bayes confidence intervals for vectors of means that are valid for multiple priors.

Let $\vartheta=(\theta,\xi)$. To state the theorem, we define the following probability associated with the interval $[a_{ik,N},b_{ik,N}]$ conditional on $(Y_{i,0:T},\vartheta)$:

$$F_{ik,N}(\vartheta)=\int_{a_{ik,N}}^{b_{ik,N}}p\big(y_{iT+h}^{*}|Y_{i,0:T},\vartheta\big)\,dy_{iT+h}^{*}. \tag{18}$$

THEOREM 2.1. *Suppose the following assumptions hold*:

(i) *The future observations are sampled from the predictive density* $p(y_{1:N,T+h}|Y_{1:N,0:T})$.

[11] For various approaches to rank set forecasts, see Askanazi, Diebold, Schorfheide, and Shin (2018).

(ii) *The posterior distribution* $p(\vartheta|Y_{1:N,0:T})$ *has the unique mode* $\bar{\vartheta}_N$. *There exists a sequence of shrinking neighborhoods* $\mathcal{N}_N(\bar{\vartheta}_N)$ *with complements* $\mathcal{N}_N^c(\bar{\vartheta}_N)$ *and a sequence* δ_N, *such that* $\|\vartheta - \bar{\vartheta}_N\| \leq \delta_N$ *for all* $\vartheta \in \mathcal{N}_N(\bar{\vartheta}_N)$ *and*

$$\mathbb{P}^{\vartheta}_{Y_{1:N,0:T}}\{\vartheta \in \mathcal{N}_N^c(\bar{\vartheta}_N)\} \xrightarrow{p} 0, \qquad \delta_N \xrightarrow{p} 0$$

in $\mathbb{P}^{Y_{1:N,0:T}}$ *probability as* $N \longrightarrow \infty$.

(iii) *The functions* $F_{ik,N}(\vartheta)$ *defined in* (18) *are locally Lipschitz in any compact neighborhood* $\mathcal{N}_N(\vartheta)$ *with Lipschitz constants* $M_{ik,N}(\mathcal{N}_N(\vartheta))$.

(iv) *For some* $M < \infty$ *independent of* N, *the Lipschitz constants satisfy*

$$\mathbb{P}^{Y_{1:N,0:T}}\left\{\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K_i} M_{ik,N}\big(\mathcal{N}_N(\bar{\vartheta}_N)\big) > M\right\} \longrightarrow 0.$$

(v) *The Bayesian coverage probability constraint* (*see* (12) *or* (13)) *holds with equality.*

Then the empirical coverage frequency converges to the Bayesian credible level in the sense of (17).

A proof of this theorem is provided in the Online Appendix. Assumption (i) states that the future observations are generated from the "true" predictive density $p(Y_{1:N,T+h}|Y_{1:N,0:T})$. In Assumption (ii), we require the posterior distribution of ϑ to concentrate. Throughout the paper, we represent the CRE distribution through finite-dimensional mixtures; see Section 3.1. Thus, ϑ is finite-dimensional and the concentration results can be obtained from the literature on the consistency and asymptotic normality of posterior distributions; see Hartigan (1983), van der Vaart (1998), Ghosh and Ramamoorthi (2003), or Ghosal and van der Vaart (2017) for textbook treatments. The only difference to many of the results stated in the literature is that we assume that the convergence in probability to occur under the marginal distribution of $Y_{1:N,0:T}$ rather than its distribution conditional on a "true" parameter, which imposes some restrictions on the prior for ϑ. Assumptions (iii) and (iv) require the probabilities $F_{ik,N}$ to be smooth functions of ϑ. In our model, the probabilities are computed from finite-dimensional mixtures of normal distributions, which are smooth functions of the underlying parameters. However, the Lipschitz constants are generally sample dependent and one needs to require that their average across i and k is stochastically bounded. In the Online Appendix, we verify the conditions for a simple model without censoring.

3. Correlated random effects, priors, and posteriors

We provide a characterization of the CRE distribution $p(\lambda_i, y_{i0}^*, \sigma_i^2|x_{i,-1}, \xi)$ and a specification of the prior distribution for (θ, ξ) in Section 3.1. Section 3.2 contains a description of the posterior sampler, and Section 3.3 outlines multistep forecasting approaches.

3.1 *(Correlated) random effects and prior distributions*

We now describe the prior distribution for θ, the parametrization of the distribution of (λ_i, y_{i0}^*), and the prior distribution for the hyperparameter vector ξ. We begin with a homoskedastic random effects (RE) setup in which λ_i and y_{i0}^* are independent of each other and of $x_{i,-1}$. We then introduce heteroskedasticity and finally extend the model specification to CRE. The prior distribution involves a small number of tuning constants, denoted by τ, that allow the researcher to scale the prior in various dimensions.

The subsequent exposition involves various parametric probability distributions in addition to the normal distribution that appeared in (1). We use $B(a, b)$, $G(a, b)$, and $\mathrm{IG}(a, b)$ to denote the beta, gamma, and inverse gamma distributions, respectively. The pair (θ, σ^2) has a normal-inverse-gamma distribution $\mathrm{NIG}(m, v, a, b)$ if $\sigma^2 \sim \mathrm{IG}(a, b)$ and $\theta|\sigma^2 \sim N(m, \sigma^2 v)$. Finally, the pair (Φ, Σ) has a matrix variate normal-inverse-Wishart distribution $\mathrm{MNIW}(M, V, \nu, S)$ if $\Sigma \sim \mathrm{IW}(\nu, S)$ has an inverse Wishart distribution and $\mathrm{vec}(\Phi)|\Sigma \sim N(\mathrm{vec}(M), \Sigma \otimes V)$.

Prior for θ We standardize the regressors x_{it} to have zero mean and unit variance and use the following normal prior for the regression coefficients θ:

$$\theta \sim N(0, \tau_\theta I_{n_x+1}), \tag{19}$$

where τ_θ is a tuning constant that controls the prior variance.

Flexible RE with homoskedasticity Under RE, the distribution of λ_i and y_{i0}^* does not depend on $x_{i,-1}$. Moreover, we assume that λ_i and y_{i0}^* are independent. Thus,

$$p\big(\lambda_i, y_{i0}^* | x_{i,-1}, \xi\big) = p(\lambda_i|\xi)\, p\big(y_{i0}^*|\xi\big).$$

We consider a mixture representation for $p(\lambda_i|\xi)$ while assuming that the initial values y_{i0}^* are normally distributed:

$$\begin{aligned} &\lambda_i|\xi \overset{\text{i.i.d.}}{\sim} N(\phi_{\lambda,k}, \Sigma_{\lambda,k}) \text{ with prob. } \pi_{\lambda,k}, \quad k = 1, \ldots, K, \\ &y_{i0}^*|\xi \overset{\text{i.i.d.}}{\sim} N(\phi_y, \Sigma_y). \end{aligned} \tag{20}$$

The maximum number of mixture components K is assumed to be prespecified.[12]

A prior over the RE distributions is induced through a prior $p(\xi)$ for the hyperparameter vector

$$\xi = [\phi_{\lambda,1}, \Sigma_{\lambda,1}, \pi_{\lambda,1}, \ldots, \phi_{\lambda,K}, \Sigma_{\lambda,K}, \pi_{\lambda,K}, \phi_y, \Sigma_y]'.$$

During the Bayesian inference stage, the prior is updated in view of the data and we obtain a posterior distribution for ξ, and hence a posterior distribution for the RE distribution. The priors for the coefficients of the normal distributions are

$$(\phi_{\lambda,k}, \Sigma_{\lambda,k}) \overset{\text{i.i.d.}}{\sim} \mathrm{NIG}\big(0, \tau_\phi, 3, 2\tau_\sigma^\lambda\big), \qquad (\phi_y, \Sigma_y) \sim \mathrm{NIG}\big(0, \tau_\phi^y, 3, 2\tau_\sigma^y\big). \tag{21}$$

[12]We use $K = 20$ in the simulation exercise and the empirical analysis. 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 (= 1) and a cross-sectional sample size $N = 1000$.

We parameterized the IG distribution such that the variances $\Sigma_{\lambda,k}$ and Σ_y have a prior distribution with mean τ_σ and variance τ_σ^2 (omitting the superscripts).[13] Conditional on Σ, the mean parameter ϕ has a $N(0, \tau_\phi\Sigma)$ distribution (omitting the subscripts). The marginal distribution of y_{i0}^* implied by (20) and (21) is a Student-t distribution, whereas the distribution of λ_i is a mixture of Student-t distributions. The tuning constants can be used to control the spread of the means of the mixture components as well as the magnitude and variation of the variances of the mixture components.

The prior for the probabilities $\pi_{\lambda,1:K}$ is generated by a mixture of truncated stick breaking processes TSB$(1, \alpha_\lambda, K)$ of the form

$$\pi_{\lambda,k}|(\alpha_\lambda, K) \sim \begin{cases} \zeta_1, & k = 1, \\ \prod_{j=1}^{k-1}(1-\zeta_j)\zeta_k, & k = 2, \ldots, K-1, \\ 1 - \sum_{j=1}^{K-1}\pi_{\lambda,j}, & k = K, \end{cases} \tag{22}$$

$$\zeta_k \sim B(1, \alpha_\lambda), \qquad \alpha_\lambda \sim G(2, 2).$$

Note that the $B(1, \alpha_\lambda)$ prior has a density $p(\zeta_k) \propto (1-\zeta_k)^{(\alpha_\lambda - 1)}$. If α_λ is close to zero, then a lot of the mass of the distribution is concentrated near $\zeta_k = 1$. This means that the first mixture component has a probability that is close to one, whereas the remaining mixture components have very small probabilities. If α_λ is close to two, then most of the mass of the distribution of ζ_k is concentrated on values of ζ_k that are close to zero. In turn, a larger number of mixture components receive nontrivial probabilities. The $G(2, 2)$ distribution is recommended by Ishwaran and James (2002). It has a mean of one and draws fall with 95% probability into the interval [0.12, 2.75], which means that the prior covers both mixtures dominated by few components and mixtures with many nontrivial components.

In the homoskedastic specification, we use the conjugate prior for σ^2 that arises in the context of a linear regression model:

$$\sigma^2 \sim \text{IG}\big(3, 2\tau_v V^*\big). \tag{23}$$

The IG distribution is parameterized in a similar way as the IG distributions in (21). $V^* = \frac{1}{N}\sum_{i=1}^N \widehat{\mathbb{V}}_i(y_{it})$ is the cross-sectional average of the time-series variances of y_{it} and the tuning constant τ_v provides additional flexibility to scale the prior for σ^2.

Heteroskedasticity To generate heteroskedasticity, one could simply replace (23) by $\sigma_i^2 \sim \text{IG}(3, 2\tau_v V^*)$. However, to make the distribution a bit more flexible, we augment the hyperparameter vector ξ and also represent the distribution of $\ln \sigma_i^2$ as a mixture of

[13] Under our parametrization of the $X \sim \text{IG}(a, b)$ distribution, $\mathbb{E}[X] = b/(a-1)$ for $a > 1$, and $\mathbb{V}[X] = (\mathbb{E}[X])^2/(a-2)$ for $a > 2$.

normals:[14]

$$\ln \sigma_i^2 | \xi \sim N\big(\psi_k, \omega_k^2\big) \text{ with prob. } \pi_{\sigma,k}, \quad k = 1, \ldots, K. \tag{24}$$

A straightforward change-of-variables yields the distribution $p(\sigma_i^2|\xi)$. As for the RE distribution, the coefficients ψ_k and ω_k have NIG priors:

$$\big(\psi_k, \omega_k^2\big) \overset{\text{i.i.d.}}{\sim} \text{NIG}\big(\ln\big(\tau_v V^*\big) - \ln(2)/2, 1, 3, 2\ln 2\big), \quad k = 1, \ldots, K. \tag{25}$$

The parametrization is chosen so that the implied prior mean $\mathbb{E}[\sigma_i^2]$ and prior variance $\mathbb{V}[\sigma_i^2]$ for each mixture component k matches the one implied by the prior used in the homoskedastic version of the Tobit model; see (23).[15] Moreover, we verified by simulation that the marginal density of σ_i^2 under this prior is very similar to the IG$(3, (3-1)\tau_v V^*)$ distribution used for the homoskedastic specification. It does, however, have fatter tails as it is a mixture of log t distributions.

Flexible CRE with heteroskedasticity We extend the RE specification in two directions: first, we allow for correlation of λ_i and y_{i0}^* with $x_{i,-1}$, and second, we allow λ_i and y_{i0}^* to be correlated with each other conditional on $x_{i,-1}$. The CRE distribution is given by the following location and scale mixture of normal distributions:

$$\big[\lambda_i, y_{i0}^*\big] \,|\, (x_{i,-1}, \xi) \overset{\text{i.i.d.}}{\sim} N\big(\big[1, x_{i,-1}'\big]\Phi_k, \Sigma_k\big) \text{ with prob. } \pi_{\lambda,k}, \quad k = 1, \ldots, K, \tag{26}$$

where Φ_k is an $(n_x + 1) \times 2$ matrix and Σ_k is a 2×2 matrix. The hyperparameter vector ξ is now defined to include the nonredundant elements of $(\Phi_k, \Sigma_k, \pi_{\lambda,k})$.

For the mixture probabilities $\pi_{\lambda,1:K}$, we use the same prior distribution as in (22). The prior distribution for the coefficient matrices Φ_k and Σ_k is a multivariate generalization of the RE distribution. We assume

$$(\Phi_k, \Sigma_k) \overset{\text{i.i.d.}}{\sim} \text{MNIW}\big(0, \tau_\phi I_{n_x+1}, 7, 4D(\tau_\sigma)\big), \quad k = 1, \ldots, K, \quad D(\tau_\sigma) = \begin{bmatrix} \tau_\sigma^\lambda & 0 \\ 0 & \tau_\sigma^y \end{bmatrix}. \tag{27}$$

Under this parametrization, the marginal IW distribution of the 2×2 matrix Σ_k has mean $D(\tau_\sigma)$. The conditional distribution of $\Phi_k|\Sigma_k$ is $MN(0, \tau_\phi \Sigma_k \otimes I_{n_z+1})$, where τ_ϕ scales the variance of the normal distribution. The dimension of Σ_k is 2×2, and hence, the marginal distribution of λ_i is identical to the RE case.[16]

[14] In an earlier version of the paper, we used a mixture of IG distributions. We switched to a mixture of normals for $\ln \sigma_i^2$ for a more symmetric treatment of λ_i and σ_i^2. Alternatively, Chib and Hamilton (2002) used Dirichlet process prior with an IG base measure to generate scale mixtures of normals.

[15] The marginal IG distribution implies $\mathbb{E}[\omega_k^2] = \ln 2$. Conditional on $\omega_k^2 = \ln 2$, the transformed parameter $\exp(\psi_k)$ has a log-normal distribution with mean $\tau_v V_*$ and variance $(\tau_v V_*)^2$.

[16] The marginal distribution of the (1,1) element of the IW$(7, 4D(\tau_\Sigma))$ distribution is IW$(6, 4D_{11}(\tau_\Sigma))$. Converted into the parametrization of the Gamma distribution, this corresponds to an IG$(3, 2D_{11}(\tau_\Sigma)) =$ IG$(3, 2\tau_\sigma^\lambda)$ distribution.

Tuning of the prior The scale of the prior distribution is controlled by a vector of tuning constants:

$$\tau = \left[\tau_\theta, \tau_\phi, \tau_\sigma^\lambda, \tau_\sigma^y, \tau_v\right]'.$$

While these tuning constants could in principle be determined in a data-driven way, using a marginal data density criterion (see the approach used in the Bayesian vector autoregression (VAR) literature, for instance, Del Negro, Schorfheide, Smets, and Wouters (2007) and Giannone, Lenza, and Primiceri (2015)), we do not pursue that route in this paper. Instead, we choose τ ex ante in an informal calibration step. While τ_θ has a straightforward interpretation after the regressors have been normalized, the implications of the remaining constants are less transparent because they control priors that are specified over a set of distributions. We recommend the researcher makes an initial choice and then samples from the prior. We found it useful to examine plots of moments or number of modes associated with the distributions. Similar plots can be generated based on the posterior. If a researcher finds that the posterior is located in an area that has essentially no prior mass, then the scaling of the prior can be adjusted to examine whether the initial prior unduly biases the posterior estimates. An example in the context of our empirical application is provided in the Online Appendix.

3.2 *Posterior sampling*

Draws from the posterior distribution can be obtained with a Gibbs sampling algorithm. We subsequently describe the conditional distributions over which the Gibbs sampler iterates. We 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 $Y^*_{i,0:T}$, $i = 1, \ldots, N$. In this regard, we are building on Tanner and Wong (1987) (data augmentation for a general state-space model), Chib (1992) (static Tobit model), Albert and Chib (1993) (Probit model), Carter and Kohn (1994) (linear state space model), and Wei (1999) (dynamic Probit model). The general blocking of parameters in the Gibbs sampler is related to Baranchuk and Chib (2008) and Li and Zheng (2008). The sampler for the flexible mixture representation of the CRE distribution is based on Ishwaran and James (2001, 2002). In terms of the actual implementation, the computations for the Tobit model are very similar to the ones for the linear model studied in Liu (2022). The only exception is the treatment of the latent variables $Y^*_{i,0:T}$, which closely follows Wei (1999).

In order to characterize the conditional posterior distributions for the Gibbs sampler, we introduce some additional notation. Because $p(\lambda_i, y^*_{i0}|x_{i,-1}, \xi)$ 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 membership indicators by $\gamma_{i,\lambda}$ and $\gamma_{i,\sigma} \in \{1, \ldots, K\}$, respectively.

Step 1: Drawing from $Y^*_{i,0:T}|(Y_{1:N,0:T}, X_{1:N,-1:T}, \lambda_{1:N}, \sigma^2_{1:N}, \gamma_{1:N,y}, \gamma_{1:N,\sigma}, \theta, \xi)$ To fix ideas, consider the following sequence of observations $y_{i0}, \ldots, y_{iT}$:

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

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

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

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 $t_1 - 1$, 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

$$\begin{aligned} M_{1:s|0} = [\mu_1, \ldots, \mu_s]', \qquad \mu_1 = \lambda_i + \rho y_{it_1-1} + \beta' x_{it_1-1}, \\ \mu_\tau = \lambda_i + \rho \mu_{\tau-1} + \beta' x_{i\tau-1} \quad \text{for } \tau = 2, \ldots, s. \end{aligned} \tag{28}$$

The covariance matrix takes the form

$$\Sigma_{1:s|0} = \sigma_i^2 \begin{bmatrix} \rho_{1,1|0} & \cdots & \rho_{1,s|0} \\ \vdots & \ddots & \vdots \\ \rho_{s,1|0} & \cdots & \rho_{s,s|0} \end{bmatrix}, \qquad \rho_{i,j|0} = \rho_{j,i|0} = \rho^{j-i} \sum_{l=0}^{i-1} \rho^{2l} \quad \text{for } j \geq i. \tag{29}$$

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

$$\begin{aligned} M_{1:s-1|0,s} &= M_{1:s-1|0} - \Sigma_{1:s-1,s|0} \Sigma^{-1}_{ss|0} (y_{it_2+1} - \mu_s), \\ \Sigma_{1:s-1,1:s-1|0,s} &= \Sigma_{1:s-1,1:s-1|0} - \Sigma_{1:s-1,s|0} \Sigma^{-1}_{ss|0} \Sigma_{s,1:s-1|0} \end{aligned} \tag{30}$$

to deduce that

$$Y^*_{i,t_1:t_2} \sim TN_-(M_{1:s-1|0,s}, \Sigma_{1:s-1,1:s-1|0,s}). \tag{31}$$

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 (28) and (29) 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 (28) and (29). According to (26), the joint distribution of (λ_i, y_{i0}^*) is a mixture of normals. Using the mixture component membership indicator $\gamma_{i,\lambda}$ and deriving the conditional normal distribution $y_{i0}^*|(\lambda_i, x_{i,-1})$ from the joint normal distribution $(\lambda_i, y_{i0}^*)|x_{i,-1}$, we can express $y_{i0}^*|(\lambda_i, x_{i,-1}) \sim N(\mu_*(\lambda_i, x_{i,-1}), \sigma_*^2)$. This leads to the mean vector

$$\begin{aligned} M_{1:s} &= [\mu_1, \ldots, \mu_s]', \qquad \mu_1 = \mu_*(\lambda_i, x_{i,-1}), \\ \mu_\tau &= \lambda_i + \rho\mu_{\tau-1} + \beta' x_{i\tau-1} \quad \text{for } \tau = 2, \ldots, s \end{aligned} \tag{32}$$

and the covariance matrix

$$\Sigma_{1:s} = \sigma_i^2 \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & \rho_{1,1} & \cdots & \rho_{1,s-1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \rho_{s-1,1} & \cdots & \rho_{s-1,s-1} \end{bmatrix} + \sigma_*^2 \begin{bmatrix} \rho^{0+0} & \cdots & \rho^{0+(s-1)} \\ \vdots & \ddots & \vdots \\ \rho^{(s-1)+0} & \cdots & \rho^{(s-1)+(s-1)} \end{bmatrix}, \tag{33}$$

where the definition of $\rho_{i,j}$ is identical to the definition of $\rho_{i,j|0}$ in (29). One can then use the formulas in (30) to obtain the mean and covariance parameters of the truncated normal distribution.

Step 2: Drawing from $\lambda_i|(Y_{1:N,0:T}, Y_{1:N,0:T}^*, X_{1:N,-1:T}, \sigma_{1:N}^2, \gamma_{1:N,y}, \gamma_{1:N,\sigma}, \theta, \xi)$ Posterior inference with respect to λ_i becomes "standard" once we condition on the latent variables $Y_{i,0:T}^*$ and the component membership $\gamma_{i,\lambda}$. It is based on the normal location-shift model

$$y_{it}^* - \rho y_{it-1}^* - \beta' x_{it-1} = \lambda_i + u_{it}, \qquad u_{it} \overset{\text{i.i.d.}}{\sim} N\big(0, \sigma_i^2\big), \quad t = 1, \ldots, T. \tag{34}$$

Because the conditional prior distribution $\lambda_i|(y_{i0}^*, x_{i,-1}, \gamma_{i,\lambda})$ is normal, the posterior of λ_i is also normal and direct sampling is possible.

Step 3: Drawing from $\sigma_i^2|(Y_{1:N,0:T}, Y_{1:N,0:T}^*, X_{1:N,-1:T}, \lambda_{1:N}, \gamma_{1:N,y}, \gamma_{1:N,\sigma}, \theta, \xi)$ Posterior inference with respect to σ_i^2 is based on the normal scale model

$$y_{it}^* - \rho y_{it-1}^* - \beta' x_{it-1} - \lambda_i = u_{it}, \qquad u_{it} \overset{\text{i.i.d.}}{\sim} N\big(0, \sigma_i^2\big), \quad t = 1, \ldots, T. \tag{35}$$

However, even conditional on the mixture component membership indicator $\gamma_{i,\sigma}$, the prior for σ_i^2 in (24) is not conjugate and direct sampling is not possible. Instead, we sample from this nonstandard posterior via an adaptive random walk Metropolis–Hastings (RWMH) step.[17]

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

$$y_{it}^* - \lambda_i = \rho y_{it-1}^* + \beta' x_{it-1} + u_{it}, \qquad u_{it} \overset{\text{i.i.d.}}{\sim} N\big(0, \sigma_i^2\big), \quad i = 1, \ldots, N, \, t = 1, \ldots, T. \tag{36}$$

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

The temporal and spatial independence of the u_{it}'s allows us to pool observations across i and t. Under the normal prior in (19), the posterior distribution of $\theta = [\rho, \beta']'$ is also normal and we can obtain draws by direct sampling.

Step 5: Drawing from $(\gamma_{i,\lambda}, \gamma_{i,\sigma})|(Y_{1:N,0:T}, Y^*_{1:N,0:T}, X_{1:N,-1:T}, \lambda_{1:N}, \sigma^2_{1:N}, \theta, \xi)$ We describe how to draw the component membership indicator $\gamma_{i,\lambda}$. 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 a member of 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

$$\bar{\pi}_{i,\lambda,k} = \frac{\pi_{\lambda,k}\, p_N\big(\lambda_i|y^*_{i0}, x_{i,-1}, \Phi_k, \Sigma_k\big)}{\sum_{k=1}^{K} \pi_{\lambda,k}\, p_N\big(\lambda_i|y^*_{i0}, x_{i,-1}, \Phi_k, \Sigma_k\big)}. \tag{37}$$

Note that the conditional distribution $\lambda_i|(y^*_{i0}, x_{i,-1}, \Phi_k, \Sigma_k)$ is normal, indicated by the notation $p_N(\cdot)$, and can be derived from the joint normal distributions of the mixture components in (26). Thus,

$$\gamma_{i,\lambda}|(\Phi_{1:k}, \Sigma_{1:K}, \lambda_i) = k \text{ with prob. } \bar{\pi}_{i,\lambda,k}. \tag{38}$$

Step 6: Drawing from $\xi|(Y_{1:N,0:T}, Y^*_{1:N,0:T}, X_{1:N,-1:T}, \lambda_{1:N}, \sigma^2_{1:N}, \gamma_{1:N,\lambda}, \gamma_{1:N,\sigma}, \theta)$ 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 that are members of 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

$$\pi_{\lambda,1:K}|(n_{\lambda,1:K}, \alpha, K) \sim \text{TSB}\left(\{1 + n_{\lambda,k}\}_{k=1}^{K}, \left\{\alpha_\lambda + \sum_{j=k+1}^{K} n_{\lambda,j}\right\}_{k=1}^{K}, K\right), \tag{39}$$

meaning that the ζ_k's in (22) have a $B(1 + n_{\lambda,k}, \alpha_\lambda + \sum_{j=k+1}^{K} n_{\lambda,j})$ distribution. Conditional on $\pi_{\lambda,1:K}$, the hyperparameter α_λ has a Gamma posterior distribution of the form

$$\alpha_\lambda|\pi_{\lambda,1:K} \sim G(2 + K - 1, 2 - \ln \pi_{\lambda,K}). \tag{40}$$

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

$$\begin{aligned} &p\big(\Phi_k, \Sigma_k|Y_{1:N,0:T}, Y^*_{1:N,0:T}, \lambda_{1:N}, \sigma^2_{1:N}, \gamma_{1:N,\lambda}, \gamma_{1:N,\sigma}, \theta\big) \\ &\quad \propto p(\Phi_k, \Sigma_k) \prod_{i \in J_{\lambda,k}} p\big(\lambda_i, y^*_{i0}|x_{i,-1}, \Phi_k, \Sigma_k\big). \end{aligned} \tag{41}$$

Because here the prior $p(\Phi_k, \Sigma_k)$ is MNIW and the likelihood $\prod_{i \in J_{\lambda,k}} p(\lambda_i, y^*_{i0}|x_{i,-1}, \Phi_k, \Sigma_k)$ 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 $x_{i,T:T+h-1}$, paths from the predictive distribution for $y_{i,T+1:T+h}$ can be easily generated by simulating (1) forward; see Section 3.3 for further details.

Modifications for the simplified model specifications If the CRE distribution is modeled parametrically instead of flexibly, then the drawing of the component membership indicators $(\gamma_{i,\lambda}, \gamma_{i,\sigma})$ in Step 5 and the drawing of $\pi_{\cdot,1:K}$ and α in Step 6 are unnecessary. One only has to sample from the MNIW posterior of (Φ_1, Σ_1) and the NIG posterior of (ψ_1, ω_1). Under homoskedasticity, that is, $\sigma_i^2 = \sigma^2$ for all i, we can pool (35) in Step 3 across t and i. In combination with the prior in (23), 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 *Multistep forecasting*

In general, there are two ways of extending one-step-ahead to multistep-ahead forecasting: an iterated approach and a direct approach.

First, iterating the law of motion of y_{it}^* in (1) forward by h periods, starting from period $t = T$, yields

$$y_{iT+h}^* = \lambda_i\left(\sum_{s=0}^{h-1}\rho^s\right) + \rho^h y_{iT}^* + \beta'\left(\sum_{s=0}^{h-1}\rho^s x_{iT+h-1-s}\right) + \sum_{s=0}^{h-1}\rho^s u_{iT+h-s}. \tag{42}$$

Thus, forecasting y_{iT+h}^* iteratively requires the path $x_{i,T:T+h-1}$. We can distinguish the following scenarios: (i) the path is given at time T. For instance, in a stress-testing application of our framework the path of the exogenous variables would be specified by the regulator as part of the stressed macroeconomic scenario. (ii) x_{it} is strictly exogenous. In this case, the user has to specify a separate model for x_{it} to simulate future trajectories along which (42) is evaluated. Because of the exogeneity, this simulation can be conducted independently of the simulation of (42). Suppose one has draws $(\lambda_i^{(j)}, \rho^{(j)}, \beta^{(j)}, \sigma_i^{2(j)})$ and draws $x_{i,T:T+h-1}^{(j)}$ from the posterior predictive distribution of the exogenous regressors, then one can define

$$\mu_{iT+h|T}^{(j)} = \lambda_i^{(j)}\left(\sum_{s=0}^{h-1}\big(\rho^{(j)}\big)^s\right) + \big(\rho^{(j)}\big)^h y_{iT}^{*(j)} + \beta^{(j)}\left(\sum_{s=0}^{h-1}\big(\rho^{(j)}\big)^s x_{iT+h-1-s}^{(j)}\right),$$

$$\sigma_{iT+h|T}^{2(j)} = \left(\sum_{s=0}^{h-1}\big(\rho^{(j)}\big)^{2s}\right)\sigma_i^{2(j)}.$$

One can sample $y_{iT+h}^{*(j)}$ from a $N(\mu_{iT+h|T}^{(j)}, \sigma_{iT+h|T}^{2(j)})$ and apply the censoring to obtain a draw $y_{iT+h}^{(j)}$. (iii) The x_{it}s are endogenous and interact with the y_{it}s, which is the case in

our application. To capture the feedback from the dependent variables to the regressors, one has to simulate $(Y_{1:N,T+1:T+h}, Y^*_{1:N,T+1:T+h}, X_{1:N,T+1:T+h-1})$ jointly; see (5).

Second, rather than generating h-step ahead forecasts iteratively, in practice forecasters often engage in direct estimation of an h-step-ahead prediction function. In our framework, this approach amounts to estimating a model of the form

$$y^*_{it} = \lambda_i + \rho y^*_{it-h} + \beta' x_{it-h} + u_{it}$$

with the understanding that the serial correlation in u_{it} implied by our original model (1) is ignored. A discussion of the disadvantages and advantages of multistep estimation in the context of VARs can be found in Schorfheide (2005).

4. Monte Carlo experiment

We conduct a Monte Carlo experiment to illustrate the performance of the set and density forecasts from the dynamic panel Tobit model in (1) under ideal conditions. We also discuss the estimation of the heterogeneous coefficients. We simplify the model by omitting the additional predictors x_{it} and using the RE specification. We 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 $N = 1000$ 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").[18] The measures of forecast accuracy discussed in Sections 2.3 and 2.4 are first computed for the cross-section $i = 1, \ldots, N = 1000$ and we then average the performance statistics over the $n_{\text{sim}} = 100$ Monte Carlo repetitions.

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:
$p(\lambda_i|y^*_{i0}) = \frac{1}{9}p_N(\lambda_i|\frac{5}{2}, \frac{1}{2}) + \frac{8}{9}p_N(\lambda_i|\frac{1}{4}, \frac{1}{2})$
$p(\ln\sigma_i^2|y^*_{i0}) = \frac{1}{9}p_N(\ln\sigma_i^2 - c|\frac{5}{2}, \frac{1}{2}) + \frac{8}{9}p_N(\ln\sigma_i^2 - c|\frac{1}{4}, \frac{1}{2})$, c is chosen such that $\mathbb{E}[\sigma_i^2] = 1$
Sample size: $N = 1000$, $T = 10$
Number of Monte Carlo repetitions: $N_{\text{sim}} = 100$
Fraction of zeros: 45%, Fraction of all zeros: 15%

[18] In the Online Appendix, we report additional results for Monte Carlo designs with 60% and 75% zeros, respectively. The overall message from the baseline Monte Carlo design is preserved under the alternative specifications.

Model specifications and predictors We 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.

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 (iv) normal RE and homoskedasticity. Versions (ii)–(iv) 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 (and finally censors the forecasts at 0).

Density and set forecasts To assess the density forecasts, we compute LPS and CRPS; see Section 2.3. The larger LPS and the smaller CRPS the better the forecast. The accuracy statistics are reported in columns 2 and 3 of Table 3. 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. The two "pooled" models that ignore the intercept heterogeneity perform the worst.

We consider two types of set forecasts; see Section 2.4. 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. Results are presented in columns 4 to 7 of Table 3. The "average" sets constructed from the heteroskedastic specification have good frequentist coverage properties. They attain coverage frequencies of 91.0% and 90.8%, respectively. 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.

Parameter estimates The last two columns of Table 3 summarize the bias and standard deviation of the posterior mean estimator of the homogeneous parameter ρ. Under the correctly specified "Flexible and Heterosk." model the bias is close to zero and the standard deviation is small. Replacing the flexible RE specification by a normal specification raises the bias by a factor of three. Replacing heteroskedasticity by homoskedasticity approximately increases the standard deviation by 50% because of a loss of efficiency.

TABLE 2. Summary of prior distributions.

Specification	λ	$p(\lambda\|\xi)$	σ^2	$p(\sigma^2\|\xi)$
Flexible RE and Heterosk.	$\lambda \sim N(\phi_{\lambda,k}, \Sigma_{\lambda,k})$ w.p. $\pi_{\lambda,k}$	$(\phi_{\lambda,k}, \Sigma_{\lambda,k}) \sim \text{NIG}(0, 5, 3, 2)$ $\pi_{\lambda,k} \sim \text{TSB}(1, \alpha_\lambda, K)$ $\alpha_\lambda \sim G(2, 2)$	$\ln \sigma^2 \sim N(\psi_k, \omega_k)$ w.p. $\pi_{\sigma,k}$	$(\psi_k, \omega_k) \sim \text{NIG}(\ln V^* - \ln(2)/2, 1, 3, 2\ln 2)$ $\pi_{\sigma,k} \sim \text{TSB}(1, \alpha_\sigma, K)$ $\alpha_\sigma \sim G(2, 2)$
Normal RE and Heterosk.	$\lambda \sim N(\phi_\lambda, \Sigma_\lambda)$	$(\phi_\lambda, \Sigma_\lambda) \sim \text{NIG}(0, 5, 3, 2)$	$\ln \sigma^2 \sim N(\psi, \omega)$	$(\psi, \omega) \sim \text{NIG}(\ln V^* - \ln(2)/2, 1, 3, 2\ln 2)$
Flexible RE and Homosk.	$\lambda \sim N(\phi_{\lambda,k}, \Sigma_{\lambda,k})$ w.p. $\pi_{\lambda,k}$	$(\phi_{\lambda,k}, \Sigma_{\lambda,k}) \sim \text{NIG}(0, 5, 3, 2)$ $\pi_{\lambda,k} \sim \text{TSB}(1, \alpha_\lambda, K)$ $\alpha_\lambda \sim \text{IG}(2, 2)$	$\sigma^2 \sim \text{IG}(3, 2V^*)$	N/A
Normal RE and Homosk.	$\lambda \sim N(\phi_\lambda, \Sigma_\lambda)$	$(\phi_\lambda, \Sigma_\lambda) \sim \text{NIG}(0, 5, 3, 2)$	$\sigma^2 \sim \text{IG}(3, 2V^*)$	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}^*	$y_{i0}^* \sim N(\phi_y, \Sigma_y)$, $(\phi_y, \Sigma_y) \sim \text{NIG}(0, 5, 3, 2)$			

Note: We set $V^* = \frac{1}{N}\sum_{i=1}^N \widehat{\mathbb{V}}_i(Y_{it})$, the cross-sectional average of the time-series variances of y_{it}.

TABLE 3. Monte Carlo experiment: forecast performance and parameter estimates.

	Density Forecast		Set Forecast "Average"		Set Forecast "Pointwise"		Estimates	
	LPS	CRPS	Cov.	Length	Cov.	Length	Bias($\hat{\rho}$)	StdD($\hat{\rho}$)
Flexible and Heterosk.	−0.757	0.277	0.910	1.260	0.933	1.503	−0.002	0.005
Normal and Heterosk.	−0.758	0.277	0.908	1.248	0.932	1.498	−0.006	0.005
Flexible and Homosk.	−0.902	0.294	0.929	1.506	0.942	1.698	0.007	0.008
Normal and Homosk.	−0.903	0.294	0.929	1.501	0.942	1.699	0.001	0.007
Pooled Tobit	−0.935	0.313	0.935	1.705	0.947	1.911	0.252	0.004
Pooled Linear	−1.243	0.357	0.923	1.925	0.933	1.951	0.229	0.005

Note: The Monte Carlo design is summarized in Table 1. The true values for ρ is 0.8. "Cov." is coverage frequency and "Length" is an average across i.

Imposing intercept homogeneity (pooled Tobit and pooled linear specification) leads to a substantial increase in the bias.

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 larger discrepancies between the true RE density and the draws from the posterior distribution. Because of the shrinkage effect of the prior distribution, we generally expect the cross-sectional distribution of $\mathbb{E}[\lambda_i|Y_{1:N,0:T}]$ (histograms) to be less dispersed than the distribution of λ_i (density plots). Moreover, if we observe sequences of all zeros for multiple units i, posterior inference of the corresponding λ_is should be the same.

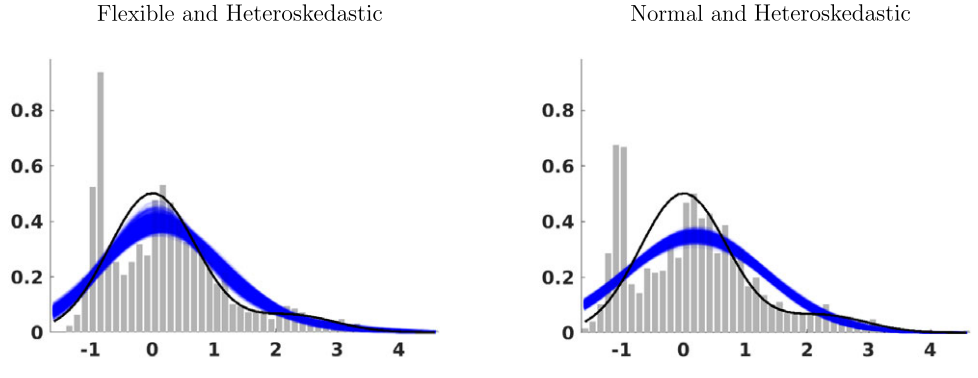

FIGURE 1. Posterior means and estimated RE distributions for λ_i. *Note*: The histograms depict $\mathbb{E}[\lambda_i|Y_{1:N,0:T}]$, $i = 1, \ldots, N$, for two different model specifications. The shaded areas are hairlines obtained by generating draws from the posterior distribution of ξ and plotting the corresponding random effects densities $p(\lambda|\xi)$. The black solid lines represent the true $p(\lambda)$.

This will create a spike in the left tail of the $\mathbb{E}[\lambda_i|Y_{1:N,0:T}]$ distribution. Both features are present in the figure.[19]

5. Empirical analysis

We 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). As mentioned in the Introduction, a charge-off occurs if a loan is deemed unlikely to be collected because the borrower has become substantially delinquent after a period of time. The prediction of charge-off rates is interesting from the perspectives of banks, regulators, and investors, because charge-offs generate losses on loan portfolios and are, in fact, a large contributor to bank losses. If these charge-off rates are large, the bank may be entering a period of distress and require additional capital.[20]

We consider a panel of "small" banks, which we define to be banks with total assets of less than one billion dollars.[21] For these banks, it is reasonable to assume that they operate in local markets. The forecasts are generated from model (1) where y_{it} are charge-off rates. As potential explanatory variables, we consider the quarter-on-quarter inflation in the house price index $\Delta \ln \mathrm{HPI}_{it-1}$, the change in the unemployment rate $\Delta \mathrm{UR}_{it-1}$, and the growth rate in personal income $\Delta \ln \mathrm{INC}_{it-1}$. Here, Δ is the temporal difference operator. The term $\beta' x_{it-1}$ therefore captures variation in regional economic conditions, which we measure at the state level. Banks located in regions with poor economic conditions may be more likely to encounter loan losses because of a higher fraction of borrowers that are unable to repay their loans. Our baseline model is based on $x_{it} = [\Delta \ln \mathrm{HPI}_{it}, \Delta \mathrm{UR}_{it}]'$, but we also consider a specification that includes personal income as a third explanatory variable and a specification without any explanatory variables.

The heterogeneous intercept λ_i can be interpreted as a bank-specific measure of the quality of the loan portfolio: the smaller λ_i, the higher the quality of the loan portfolio and the less likely a charge-off is to occur. The autoregressive component in the model captures the persistence of the composition of the loan portfolio over time, and the covariates shift the density of repayment probabilities. We consider various choices of $p(\lambda_i, y_{i0}^*, \sigma_i|x_{i,-1}, \xi)$; see Section 3.1. The data set is described in Section 5.1. Section 5.2 presents density forecast comparisons for various model specifications. Estimates of the

[19]We provide illustrative analytical examples of these effects in the Online Appendix.

[20]The accounting details are more complicated: bank balance sheets contain a contra asset account called "Allowance for Loan and Lease Losses" (ALLL). Provisions for LLL are created based on estimated credit losses and reduce the income of the bank. Charge-offs reduce the ALLL and the gross loans on the balance sheet, leaving the net amount unchanged. At this stage, the charge-offs do not lead to a further reduction of income. Whether or not a bank takes a loss provision or a charge-off is to some extent a managerial/accounting decision, although regulators require loans they classify as losses to be charged off. We abstract from strategic accounting aspects; see Moyer (1990) for a seminal paper.

[21]Monitoring potential loan losses in small banks is useful by itself. Moreover, the delinquency rates of small banks could foreshadow those rates of large banks since the small banks tend to have more subprime borrowers who are more vulnerable to minor deterioration in economic condition.

heterogeneous and homogeneous parameters are reported in Section 5.3. Posterior predictive checks are conducted in Section 5.4. Finally, Section 5.5 contains the set forecast results.

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 (CC) loans, 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 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 ($t = T$);[22] (ii) 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 4 contains some summary statistics for the two baseline samples. For the small banks in our sample, RRE loans are an important part of their loan portfolio. For approximately 45% (25%) of the banks RREs account for 20% to 50% (more than 50%) of their loan portfolio. CC 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 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

TABLE 4. Summary statistics for baseline samples.

	N	Zeros [%]	All Zeros [%]	Mean	75th	Max
RRE	2576	76	61	0.25	0.00	33.1
CC	561	43	22	3.27	4.07	260

Note: 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.

[22]There are, in general, large uncertainties during the Great Recession. Thus, accurate density and set forecasts are important.

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 *Density forecasts and model selection*

Selected samples We begin the empirical analysis by comparing the density forecast performance of several variants of (1) for the two baseline samples using $x_{it} = [\Delta \ln \mathrm{HPI}_{it}, \Delta \mathrm{UR}_{it}]'$. This comparison includes forecasts from a Tobit model and a linear model with homogeneous intercepts and homoskedastic innovation variances. Table 5 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. Second, in both RRE and CC samples, all four heteroskedastic specifications lead to very similar density forecasting performance.

All samples Figure 2 summarizes the LPS comparisons for all 111 samples. We focus on the comparison of predictive scores from the heteroskedastic specifications versus homoskedastic specifications using flexibly modeled correlated random effects. The solid line is the 45-degree line and the intersections of the dashed and dotted lines correspond to the scores associated with the baseline RRE and CC samples reported in Table 5. The figure shows that the results for the baseline samples are qualitatively representative: incorporating heteroskedasticity is important for density forecasting. We provide a figure in the Online Appendix that illustrates that LPS differentials between normal versus flexible CREs and CREs versus REs are small. In view of these results, we subsequently focus on the flexible CRE specification with heteroskedasticity.

TABLE 5. Density forecast performance.

	RRE		CC	
Specification	LPS	CRPS	LPS	CRPS
Heteroskedastic models				
Flexible CRE	−0.523	0.240	−1.921	1.957
Normal CRE	−0.521	0.240	−1.901	1.895
Flexible RE	−0.525	0.238	−1.925	1.970
Normal RE	−0.524	0.237	−1.912	1.936
Homoskedastic models				
Flexible CRE	−0.751	0.272	−2.512	2.495
Normal CRE	−0.751	0.272	−2.463	2.343
Flexible RE	−0.751	0.270	−2.630	2.613
Normal RE	−0.752	0.270	−2.535	2.391
Pooled Tobit	−0.831	0.310	−2.642	2.620
Pooled Linear	−1.594	0.374	−3.010	2.789

Note: The estimation sample ranges from 2007Q2 ($t = 0$) to 2009Q4 ($t = T = 10$). We use $x_{it} = [\Delta \ln \mathrm{HPI}_{it}, \Delta \mathrm{UR}_{it}]'$ and forecast 2010Q1 observations.

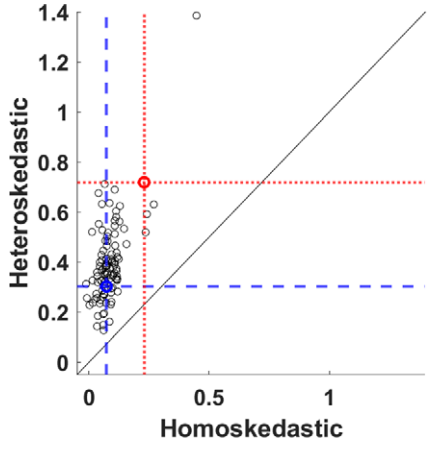

FIGURE 2. Log predictive density scores—all samples. *Note*: Flexible CRE specification. The figure illustrates pairwise comparisons of log predictive scores. We also show the 45-degree line. Log probability scores are depicted as differentials relative to pooled Tobit. The intersection of the dashed (dotted) lines corresponds to RRE (CC) baseline sample. We use $x_{it} = [\Delta \ln \text{HPI}_{it}, \Delta \text{UR}_{it}]'$.

Tail probabilities for selected samples From the density forecasts, we can compute probability forecasts for particular events. We consider the tail event $\mathbb{I}\{y_{iT+1} \geq c\}$ for $c = 1\%$ for now. Figure 3 visualizes the probabilities of the tail event for RRE charge-off rates for 2010Q1 and 2018Q1, emphasizing the spatial dimension.[23] We associate each bank i with a particular county. If there are multiple banks in one county, we average the predicted probabilities. 2010Q1 is the immediate aftermath of the Great Recession and the counties that are covered by our sample appear predominantly in dark color, indi-

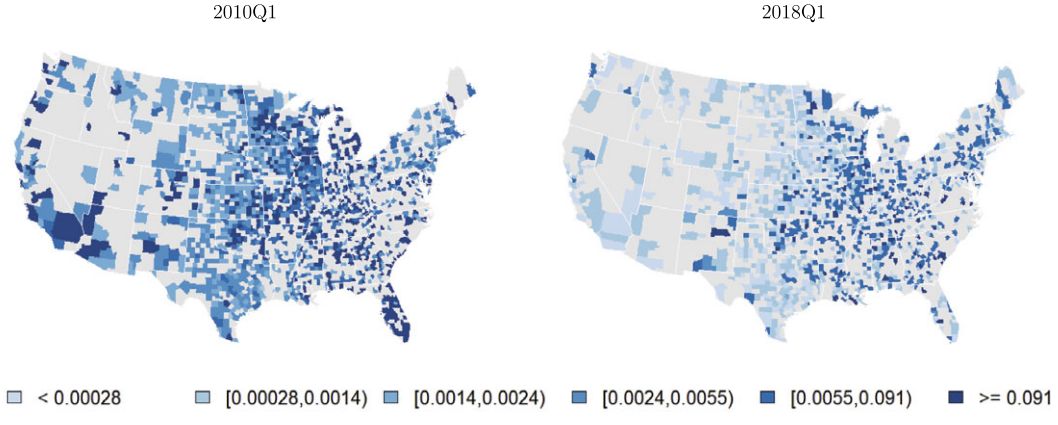

FIGURE 3. RRE charge-off rate predictive tail probabilities, spatial dimension. *Note*: Predictive tail probabilities are defined as $\mathbb{P}\{y_{iT+1} \geq c | Y_{1:N,0:T}, X_{1:N,-1:T}\}$, where $c = 1\%$. Flexible CRE specification with heteroskedasticity. The estimation samples range from 2007Q2 ($t = 0$) to 2009Q4 ($t = T = 10$) and 2015Q2 ($t = 0$) to 2017Q4 ($t = T = 10$).

[23] Similar maps for CC charge-off rates are available in the Online Appendix.

cating that predicted probabilities of the event exceed 9.1%. Banks in California, Florida, and the Midwest from Minnesota, Wisconsin, and Michigan down to Arkansas, Mississippi, and Alabama are predicted to write off a considerable fraction of their RRE loans. Eight years later, the situation has improved considerably, as the map now appears in light instead of dark color, in particular, in hard hit states such as California and Florida.

While this paper focuses on forecasting problems, the predictive densities derived from our empirical model can be embedded into more complex decision problems that more closely capture the objectives of policy makers or regulators. In this case, the predictive density is used to compute posterior expected losses associated with policy decisions. The accuracy of the loss calculation is tied to the empirical adequacy of the predictive density, which is what we are evaluating in this section.

5.3 *Parameter estimates for selected samples*

Heterogeneous parameters The distributions of posterior mean estimates of the heterogeneous coefficients for the 2007Q2 sample of RRE charge-off rates are depicted in Figure 4.[24] 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 right panel contains a scatter plot that illustrates the correlation between the posterior means of intercepts and shock standard deviations.

A notable feature of the histogram for the posterior means of $\lambda_i/(1-\rho)$ is the spike in the left tail of the distribution. Such spikes were also present in the Monte Carlo simulation; see Figure 1. The spike corresponds 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

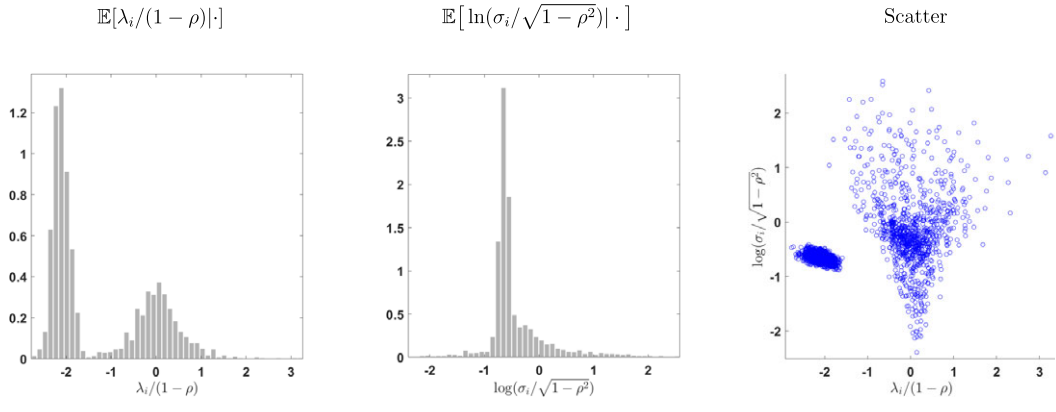

FIGURE 4. Heterogeneous coefficient estimates, RRE charge-off rates. *Note*: Heteroskedastic flexible CRE specification. The estimation sample ranges from 2007Q2 ($t=0$) to 2009Q4 ($t=T=10$). A few extreme observations are not visible in the plots. The conditioning set is $(Y_{1:N,0:T}, X_{1:N,-1:T})$.

[24]Similar plots for CC charge-off rates are available in the Online Appendix.

the histogram for the posterior means of the re-scaled log standard deviations and the right panel shows that the σ_i spike and the λ_i spike are associated with the same banks. 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. The large dispersion of σ_i estimates is consistent with the substantially better density forecast performance of the heteroskedastic models.

Because regional economic conditions have already been controlled for by $\beta' x_{it-1}$, the estimates of λ_i are more likely to be related to bank characteristics. Popular explanations for the heterogeneity in loan losses across banks, here captured by the heterogeneity of $\hat{\lambda}_i$, are attitude toward risk, that is, some banks might have a greater propensity to take risk or have better opportunities to diversify returns on their loan portfolio, and quality of credit management; see Keeton and Morris (1987) for an early contribution and Ghosh (2015, 2017) more recently.

In Figure 5, we illustrate the relationship between the posterior mean estimate of $\widehat{\lambda_i/\sigma_i}$, which for $\rho = 0$ and $\beta = 0$ determines the probability of nonzero charge-offs, and bank size measured by the log of total assets. The top segments of the two panels contain scatter plots with groupwise least-absolute-deviations (LAD) regression lines (left scale). As we have seen previously in Figure 4, there are two groups of $\hat{\lambda}_i$ estimates. For simplicity, we refer to these groups as low-λ and high-λ groups, respectively. For both the RRE and CC samples, the positive relationship between bank size and riskiness of the loan portfolio $\widehat{\lambda_i/\sigma_i}$ is more pronounced for banks in the high-λ group. The slope coefficients are 0.18 and 0.19, respectively. The shaded areas at the bottom of the panels

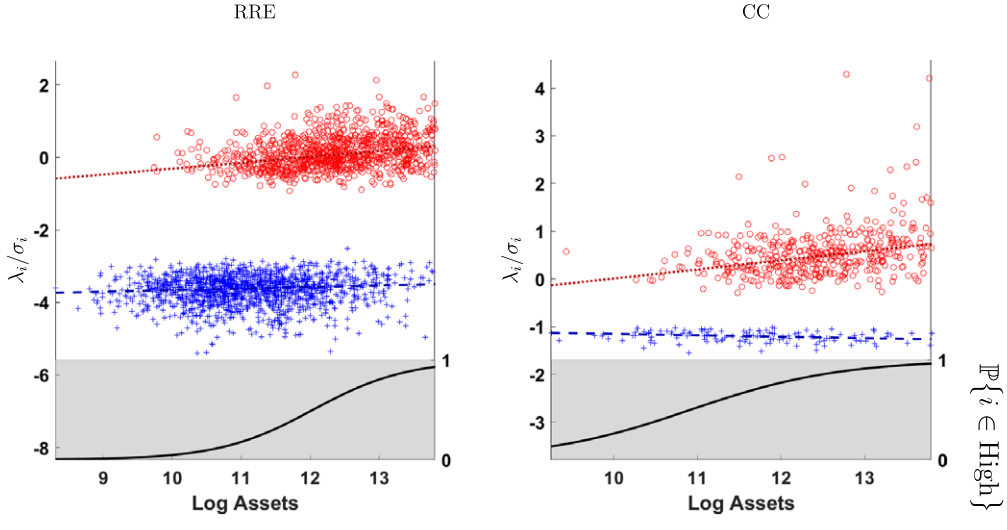

FIGURE 5. $\widehat{\lambda_i/\sigma_i}$ and $\mathbb{P}\{i \in \text{High}\}$ versus log assets (bank size). *Note*: Heteroskedastic flexible CRE specification. The estimation sample ranges from 2007Q2 ($t = 0$) to 2009Q4 ($t = T = 10$). Bank assets are measured at $t = 0$. We form a low-λ (+) and high-λ (o) group (left scale). The lines in the top segment of the plot are LAD regression lines. The black lines in the grey shaded areas are predicted probabilities from the logit models (right scale).

contain fitted probabilities (right scale) from a logit model that uses log assets as right-hand side variable. The larger the assets, the higher the probability that it belongs to the high-λ group. These results suggest that larger banks in our sample tend to hold riskier loan portfolios.

In Table 6, we present estimates from LAD regressions of $\widehat{\lambda_i/\sigma_i}$ on multiple bank characteristics (measured in period $t = 0$), separately for the low-λ and the high-λ group of banks.[25] We also report estimates for a logit model for $\mathbb{I}\{i \in \text{High}\}$. According to the logit estimates bank size (log assets), the ratio of RRE or CC loans to all loans, lending specialization (ratio of total loans to total assets), and lack of credit quality (ratio of ALLL to total loans) increase the probability that a bank belongs to the high-λ group. Capitalization (capital-to-asset ratio) and profitability (return on assets) lower the probability that a bank belongs to the high-λ group. For the group-specific regressions only a few bank variables appear to be significant. Foremost, it is bank size measured by log assets. For the RRE high-λ group, it also includes lending specialization, and for the CC high-λ group it includes diversification (share of noninterest income to total income). Operational efficiency, measured by the ratio of overhead costs to assets (OCA) is predominantly insignificant.

Ghosh (2017) studies macroeconomic and bank-level determinants of nonperforming loans, that is, loans past due 90 days or more, for the 100 largest commercial banks over the period 1992Q4 to 2016Q1. With the exception of log assets and loan fractions, we followed his study in constructing our bank-level regressors. Although our sample differs from his in several dimensions (selection of banks, measure of loan performance, and time period), we provide a brief comparison of the results for real estate loans as follows.

Ghosh (2017) finds the following significant relationships for real estate loans: log capital-to-assets (positive), log loans-to-assets (negative), log inverse credit quality (positive), log return on assets (negative). In our logit regression, the same bank characteristics have significant coefficients, but the signs of the estimates for the capital-to-asset and the loan-to-asset ratio differ. As Ghosh (2017) points out, the effect of bank capitalization on loan quality is theoretically ambiguous. On the one hand, managers in banks with low capital bases have a moral hazard incentive to engage in risky lending practices (negative relationship). On the other hand, managers in highly capitalized banks may feel confident to engage in risky lending (positive relationship). With respect to the loan-to-asset ratio, our positive estimate for RRE contradicts the notion that banks that are specialized in lending do a better job in selecting high-quality loans, and the positive relationship may reflect that these banks could have more liberal lending policies.

We also report goodness-of-fit (R^2) measures in Table 6. For the LAD regressions, we report Koenker and Machado (1999)'s quantile regression R^2. For the logit regressions, we compute McFadden (1973)'s pseudo R^2. For the RRE loans, the variation in loan quality ($\widehat{\lambda_i/\sigma_i}$) explained by bank characteristics is low. The R^2s for the group-specific LAD regressions are only 0.03 and 0.06, respectively. For the CC sample, bank characteristics

[25]Data definitions and summary statistics for the bank characteristics are provided in the Online Appendix.

TABLE 6. Regressions of $\widehat{\lambda_i/\sigma_i}$ on bank characteristics.

	RRE						CC					
	Low		High		Logit		Low		High		Logit	
Log Assets	0.05	(0.02)	0.20	(0.03)	1.55	(0.11)	0.00	(0.02)	0.21	(0.03)	1.27	(0.32)
Loan Fraction	0.07	(0.12)	0.17	(0.15)	5.07	(0.53)	10.4	(6.63)	0.61	(0.50)	1254	(179)
Capital-Asset	0.30	(0.51)	−1.36	(0.99)	−12.0	(2.83)	0.53	(0.41)	−1.42	(1.02)	−7.75	(7.53)
Loan-Asset	0.09	(0.14)	0.56	(0.23)	6.43	(0.66)	−0.20	(0.14)	−0.07	(0.21)	6.42	(2.23)
ALLL-Loan	5.82	(3.97)	12.0	(5.10)	88.1	(18.2)	−0.01	(2.23)	−1.45	(4.02)	85.8	(38.2)
Diversification	0.34	(0.35)	−0.20	(0.13)	−0.10	(0.63)	−0.05	(0.31)	1.19	(0.42)	0.82	(4.69)
Ret. on Assets	−26.1	(7.55)	−1.08	(10.06)	−122	(36.1)	28.7	(7.73)	11.1	(12.4)	−225	(130)
OCA	−19.4	(9.63)	8.25	(10.05)	47.9	(31.01)	6.38	(7.23)	16.92	(11.8)	−125	(124)
Intercept	−4.16	(0.32)	−2.86	(0.47)	−23.8	(1.60)	−1.27	(0.28)	−2.18	(0.45)	−20.4	(4.79)
Pseudo R^2	0.03		0.06		0.32		0.18		0.11		0.47	

Note: Heteroskedastic flexible CRE specification. The estimation sample ranges from 2007Q2 ($t=0$) to 2009Q4 ($t=T=10$). Bank characteristics are measured at $t=0$. Low (High) refers to small (large) $\widehat{\lambda_i/\sigma_i}$ group of banks (cutoff is approx −2 for RRE and −1 for CC); see (o) and (+) symbols in Figure 5. For banks in low and high groups, we regress $\widehat{\lambda_i/\sigma_i}$ on the variables listed in the first column using a least absolute deviations estimator. Logit refers to estimates of a logit model for $\mathbb{I}\{i \in \text{high}\}$. Standard errors are in parenthesis. Pseudo R^2 are computed as follows: LAD $= 1 - \sum|\hat{u}_i(\text{all})|/\sum|\hat{u}_i(\text{intcpt})|$ (Koenker and Machado (1999)), Logit $= 1 - \text{loglh(all)}/\text{loglh(intcpt)}$ (McFadden (1973)).

TABLE 7. Estimates of common parameters.

	y^*_{it-1}		$\Delta \ln \mathrm{HPI}_{it-1}$		$\Delta \mathrm{UR}_{it-1}$		$\Delta \ln \mathrm{INC}_{it-1}$		
	Mean	CI	Mean	CI	Mean	CI	Mean	CI	LPS
RRE	0.21	[0.18, 0.25]	−0.03	[−0.04, −0.02]	0.15	[0.13, 0.17]			−0.5232
	0.22		−0.03	[−0.04, −0.02]	0.15	[0.12, 0.17]	0.001	[−0.005, 0.007]	−0.5214
	0.29	[0.27, 0.31]							−0.5214
CC	0.41	[0.36, 0.46]	−0.09	[−0.15, −0.04]	0.46	[0.30, 0.62]			−1.9214
	0.41	[0.36, 0.45]	−0.10	[−0.16, −0.04]	0.46	[0.30, 0.63]	0.010	[−0.030, 0.051]	−1.9216
	0.48	[0.43, 0.52]							−1.9268

Note: 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.

are more successful in explaining variations in loan quality. The R^2 values are 0.18 and 0.11, respectively. The logit regressions attain pseudo R^2 values of 0.32 and 0.47, which indicate that the bank characteristics considered here are partly successful in determining whether a bank belongs to the low-λ or high-λ group.

Common parameters Parameter estimates of the common coefficients for the flexible CRE specification with heteroskedasticity are reported in Table 7 for the 2007Q2 samples. We report posterior means and 90% credible intervals. For each sample, we consider three specifications: (i) the baseline specification with $\Delta \ln \mathrm{HPI}_{it-1}$ and $\Delta \mathrm{UR}_{it-1}$, (ii) an extended version that also includes $\Delta \ln \mathrm{INC}_{it-1}$, and (iii) a version without regressors.

Both samples exhibit mild autocorrelation. The point estimate of ρ is 0.21 for RRE and 0.41 for CC. To report the estimates of β, we undo the standardization of the regressors. The numerical values can be interpreted as follows. For the RRE sample, under the extended specification that includes personal income growth a 1% quarter-on-quarter fall of house prices leads to an increase in charge-off rates by 0.03 percentage points. A 1% increase in the unemployment rate raises the charge-off rates by 0.15 percentage points. Finally, a 1% growth of personal income increases the charge-off rates by 0.001 percentage points. For both samples, the coefficients on persistence, house-price inflation, and unemployment rate changes are "significant," whereas the coefficient on the income growth regressor is "insignificant" in that it is small and its sign is ambiguous. Adding income growth hardly alters the coefficient estimates for house-price inflation and unemployment rate changes. The estimates for the CC sample are qualitatively similar to RRE but about three times larger in magnitude.

In the last column of Table 7, we report the LPS, now up to four decimal places, that were previously used for the comparison of density forecasts in Table 5. The values for the three configurations of x_{it} are very close. For the CC sample, the LPS criterion favors our baseline specification with $x_{it} = [\Delta \ln \mathrm{HPI}_{it}, \Delta \mathrm{UR}_{it}]'$, whereas for the RRE sample strictly speaking the model without regressors is preferred. In the Online Appendix, we show scatter plots of $\hat{\lambda}_i + \beta' x_{it-1}$ versus $\hat{\lambda}_i$, which indicate that only a very small fraction is explained by local economic conditions. Despite the quantitatively small effect

of local economic conditions on charge-offs, we proceed with $x_{it} = [\Delta \ln \mathrm{HPI}_{it}, \Delta \mathrm{UR}_{it}]'$, whose coefficients are "significant," and examine the effects of changes in house prices and unemployment more carefully.

Because the Tobit model is nonlinear, the average effect of a change in the regressors ("treatment effect") depends on λ_i. We consider a change of the regressor from its sample value x_{iT} to $\tilde{x}_{iT} = x_{iT} + \Delta x \cdot \iota$, where the unit-length vector ι determines the direction of the perturbation of x_{iT} and $\Delta x > 0$ the magnitude. Accounting for censoring, we decompose the treatment effect on y_{iT+1} as follows:

$$\begin{aligned}\frac{\tilde{y}_{iT+1} - y_{iT+1}}{\Delta x} &= \beta'\iota \mathbb{I}\{\lambda_i + \rho y^*_{iT} + \beta' x_{iT} + u_{iT+1} > 0\} \\ &\quad + \frac{\lambda_i + \rho y^*_{iT} + \beta' \tilde{x}_{iT} + u_{iT+1}}{\Delta x} \Big(\mathbb{I}\{\lambda_i + \rho y^*_{iT} + \beta' \tilde{x}_{iT} + u_{iT+1} > 0\} \\ &\qquad - \mathbb{I}\{\lambda_i + \rho y^*_{iT} + \beta' x_{iT} + u_{iT+1} > 0\} \Big) \\ &= I_i + II_i. \end{aligned} \tag{43}$$

Term I_i captures the intensive margin, that is, a bank that has nonzero charge-offs conditional on x_{iT} and $\tilde{x}_{iT}$. In this region, the Tobit model is linear and the effect is $\beta'\iota$. The second term, II_i, captures the extensive margin of banks switching between zero and positive charge-offs.

Figure 6 (see online version for colors) depicts the posterior mean and the 90% credible band of the two components of the treatment effect for the banks in the 2007Q2 CC sample.[26] We sort the banks based on the posterior means $\widehat{\lambda_i/\sigma_i}$, which for $\rho = 0$ and

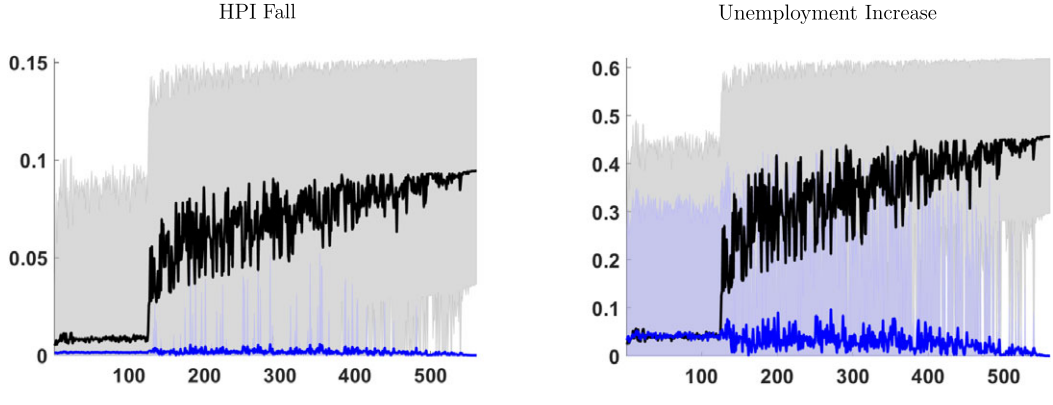

FIGURE 6. Effects (Terms I and II) of HPI and UR on CC charge-off rates. *Note*: Heteroskedastic flexible CRE specification. The estimation sample ranges from 2007Q2 ($t = 0$) to 2009Q4 ($t = T = 10$). The banks $i = 1, \ldots, N$ along the x-axis are sorted based on the posterior means $\widehat{\lambda_i/\sigma_i}$. Terms I_i are shown in black/grey and terms II_i in dark/light blue (see online version for colors). The units on the y-axis are in percent. The solid lines indicate the posterior means of the treatment effect components and the shaded areas delimit 90% credible bands.

[26] A similar figure for the RRE sample is available in the Online Appendix.

$\beta = 0$ would determine the probability of a positive charge-off. We consider two choices for $\Delta x \cdot \iota$: a 5% drop in house prices (left panels) and a 5% rise in the unemployment rate within one quarter (right panels). These are severe shocks to the local economies. For the first approximately 120 banks, the posterior mean of I_i (black/grey) is close to zero. These are the banks with low values of $\hat{\lambda}_i$ that appear as a mass in the left tail of the density plot in the left panel of Figure 4. Under the baseline conditions x_{iT}, they are unlikely to have nonzero charge-offs. For the remaining banks, the posterior mean of the term I treatment effect rises under the HPI fall scenario from 0.03% to 0.1%, where the latter value is the coefficient estimate reported in Table 7. The credible intervals are fairly wide, ranging from 0% to 0.15%.

The posterior mean for component *II* (dark/light blue) of the treatment effect is qualitatively similar under the two economic scenarios. For the first 120 banks, term *II* is small because much of $\beta'(\tilde{x}_{iT} - x_{iT})$ has to compensate for the low estimate of λ_i before the latent variable y^*_{iT+1} becomes positive. For the remaining banks, the term is also small, but for a different reason: with high probability, these banks already have positive charge-offs under the baseline economic conditions. Quantitatively, the effects are larger under the very severe unemployment scenario. The switch of low λ_i banks from zero to positive charge-offs leads to a posterior mean of the average treatment effect of 0.04%. As $\widehat{\lambda_i/\sigma_i}$ increases, the expected value of term *II* decreases because it becomes more likely that the bank has positive charge-offs even under the baseline scenario.

5.4 *Posterior predictive checks for selected samples*

In order to assess the fit of the estimated panel Tobit model, we report posterior predictive checks in Figure 7. 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.[27] 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 thick solid line is computed from the actual RRE loan sample. Each 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 focus on a qualitative assessment of the model fit.

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

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

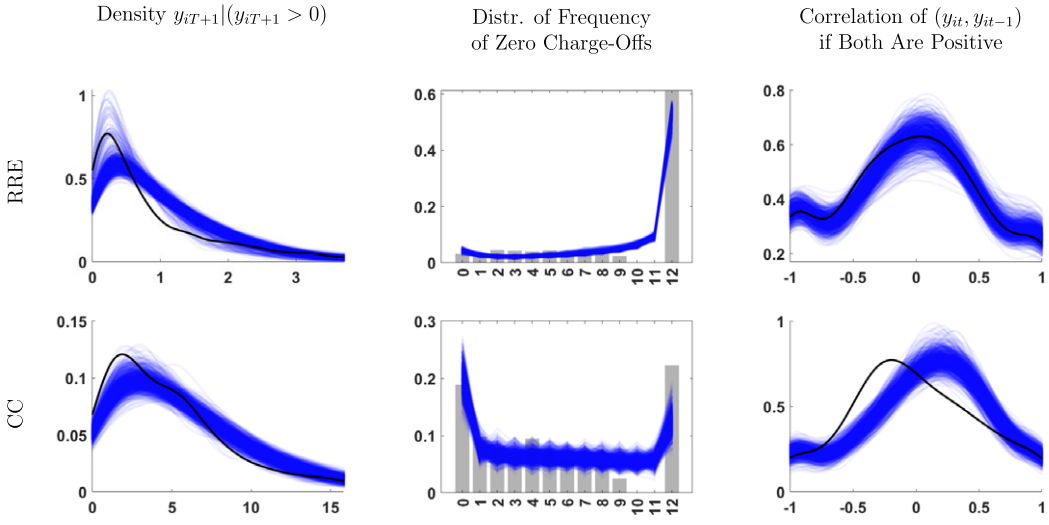

FIGURE 7. Posterior predictive checks: Cross-sectional distribution of sample statistics. *Note*: Heteroskedastic flexible CRE specification. The estimation sample ranges from 2007Q2 ($t = 0$) to 2009Q4 ($t = T = 10$). The black solid lines (left and right panels) and the histogram (center panels) are 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.

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 modes in the two actual densities. The hairlines depict the densities conditional on $y_{iT+1} > 0$. In the observed RRE sample, the fraction of $y_{iT+1} = 0$ is 0.71. The corresponding 90% interval obtained from the estimated model is [0.73, 0.79]. For CC charge-off rates, the fraction in the data is 0.43 and the corresponding 90% interval obtained from the estimated model is [0.37, 0.47].

The center panels of Figure 7 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 12 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 7 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 $y_{i,0:T+1}$ 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-8), 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 nonzero observations. With the exception of the autocorrelations in the CC sample, the two estimated models are able to reproduce the cross-sectional densities of the sample statistics.

5.5 *Set forecasts*

Selected samples Set forecasts for 2010Q1, constructed as HPD sets from the posterior predictive distribution, are visualized in Figure 8 (see online version for colors). The nominal credible level is 90%. We distinguish forecasts targeting pointwise coverage probability (grey) from forecasts targeting average coverage probability (pink). For each bank i, we plot the set forecast, the posterior mean forecast, and the actual realization of the charge-off rate. The banks are sorted according to $\mathbb{E}[y_{iT+1}|Y_{1:N,0:T}, X_{1:N,-1:T}]$. We do not show forecasts for the first 1400 (100) banks for the RRE (CC) sample because they are essentially zero.

A comparison of the grey and the pink sets in Figure 8 shows the effect of targeting average versus pointwise coverage. The upper bound as a function of i increases less under targeting average coverage probability, because the criterion allows us to shorten very wide predictive sets and lengthen narrow sets, while reducing the average

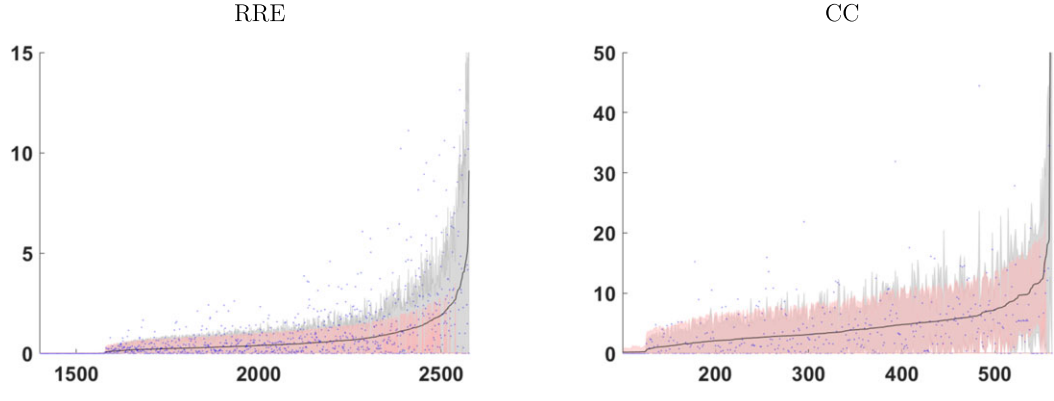

FIGURE 8. Set forecasts, banks sorted by $\mathbb{E}[y_{iT+1}|Y_{1:N,0:T}, X_{1:N,-1:T}]$. *Note*: Flexible CRE specification with heteroskedasticity. The estimation sample ranges from 2007Q2 ($t = 0$) to 2009Q4 ($t = T = 10$). We forecast 2010Q1 observations. The nominal coverage probability is 90%. Posterior mean forecasts (solid black line), actuals (blue dots), and set forecasts targeting pointwise (grey) and average (pink) coverage probability (see online version for colors).

TABLE 8. Set forecast performance.

				Fraction of Sets of the Form		
		Coverage	Ave. Len.	$\{0\}$	$[0, b]$	$\{0\} \cup [a, b]$
RRE	Target Average Coverage	0.88	0.31	0.68	0.28	0.04
	Target Pointwise Coverage	0.94	0.75	0.61	0.36	0.03
CC	Target Average Coverage	0.91	6.48	0.02	0.81	0.17
	Target Pointwise Coverage	0.91	7.74	0.19	0.56	0.25

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

length. For the RRE sample set, forecasts for banks with large expected charge-off rates $\mathbb{E}[y_{iT+1}|Y_{1:N,0:T}, X_{1:N,-1:T}]$ become considerably shorter. In fact, for $i > 2500$, many of them become $\{0\}$. Although we plot the actual values of the charge-off rates in Figure 8, it is not possible to glean how close the empirical coverage frequency is to the nominal coverage probability. Thus, in Table 8 we report both the average length of the sets and the empirical coverage frequency. For both samples, the set forecasts that are constructed by targeting the average coverage probability have a cross-sectional coverage frequency that is close to the nominal coverage probability of 90% and they tend to be shorter than the ones obtained by targeting pointwise coverage probability.[28]

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. Roughly speaking, the forecaster 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 the RRE sample with a relatively high fraction of zeros, when targeting pointwise coverage, the average coverage probability is largely above 90%, so this mechanism manifests itself as reducing wider pointwise sets to $\{0\}$, which decreases the average coverage probability and average length at the same time. Thus, there is an increase in the fraction of $\{0\}$ forecasts; also see the right tail in the left panel of Figure 8.

For the CC sample with a relatively low fraction of zeros, when targeting pointwise coverage, the average coverage probability is already close to 90%. Switching from targeting pointwise to targeting average coverage, the majority of $\{0\}$ forecasts are converted into $[0, b]$ forecasts by adding a small continuous portion, and thereby increasing the pointwise coverage of these units to more than 90%; see the left tail in the right panel

[28]We also computed evaluation statistics for the homoskedastic specification. 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 5.

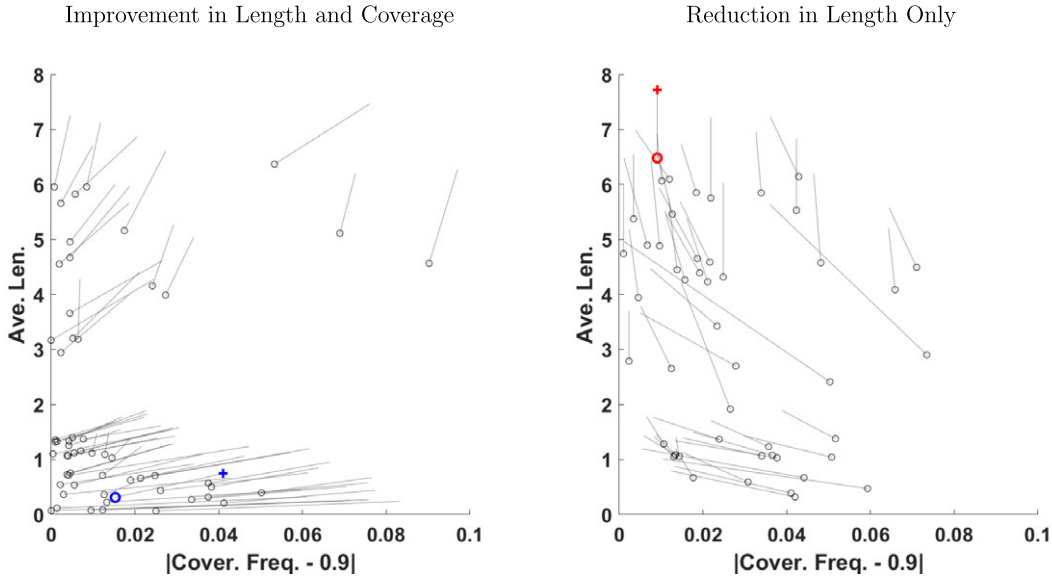

FIGURE 9. Set forecasts: Targeting pointwise versus average coverage—all samples. *Note*: Flexible CRE specification with heteroskedasticity. The bold symbols (o, +) correspond to the RRE (left panel) and CC (right panel) baseline samples. The two endpoints of each hairline indicate the coverage probability and length for a particular estimation sample. Circled endpoints correspond to targeting average coverage probability, and unmarked endpoints (or crosses for the baseline samples) represent pointwise coverage targeting. Hairlines in the left panel represent samples for which the coverage frequency gets closer to the nominal coverage probability of 90% and the length becomes shorter. The remaining samples are represented by the hairlines in the right panel.

of Figure 8. Moreover, about one-third of the disconnected forecasts are converted into connected forecasts, which is due to a lengthening of the sets for small σ_i units. In the end, the fraction of $[0, b]$ forecasts increases substantially in this case.

All samples In Figure 9, we provide information about the coverage frequency and average length size of the set forecast for all samples. We focus on a comparison between targeting pointwise versus average coverage probability in the flexible CRE specification with heteroskedasticity. Each hairline corresponds to one of the 111 different samples and the two endpoints of the hairlines indicate average length and deviation of the empirical coverage frequency from the 90% nominal credible level. The circled endpoints correspond to targeting average coverage, and unmarked endpoints (or crosses for the baseline samples) represent pointwise coverage targeting. The left panel comprises samples for which targeting average coverage brings the empirical coverage frequency closer to 90% *and* reduces the average length. Here, the hairlines point into the lower left corner of the graph. The remaining samples are represented by the hairlines in the right panel. Targeting the average coverage unambiguously reduces the average length. For 52% of the samples, it also improves the empirical coverage frequency (left panel). For the remaining 48% of the samples, the deterioration of the coverage fre-

quency is relatively small. The median improvement in coverage probability in the left panel is 0.022, whereas the median deterioration in the right panel is only 0.007. 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.

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 considered. 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 allow for stronger forms of simultaneity between the dependent variable and regressors and to account for 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

Albert, James H. and Siddhartha Chib (1993), "Bayesian analysis of binary and polychotomous response data." *Journal of the American Statistical Association*, 88, 669–679. [119, 132]

Armstrong, Timothy B., Michal Kolesár, and Mikkel Plagborg-Møller (2021), "Robust empirical Bayes confidence intervals." *Econometrica*, 90, 2567–2602. [118, 127]

Askanazi, Ross, Francis X. Diebold, Frank Schorfheide, and Minchul Shin (2018), "On the comparison of interval forecasts." *Journal of Time Series Analysis*, 39, 953–965. [127]

Atchadé, Yves F. and Jeffrey S. Rosenthal (2005), "On adaptive Markov chain Monte Carlo algorithms." *Bernoulli*, 11, 815–828. [134]

Baranchuk, Nina and Siddhartha Chib (2008), "Assessing the role of option grants to CEOs: How important is heterogeneity?" *Journal of Empirical Finance*, 15, 145–166. [118, 119, 120, 121, 132]

Botev, Zdravko I. (2017), "The normal law under linear restrictions: Simulation and estimation via minimax tilting." *Journal of the Royal Statistical Society B*, 79, 125–148. [119, 133]

Brown, Lawrence D. and Eitan Greenshtein (2009), "Nonparametric empirical Bayes and compound decision approaches to estimation of a high-dimensional vector of normal means." *The Annals of Statistics*, 37, 1685–1704. [120]

Burda, Martin and Matthew Harding (2013), "Panel probit with flexible correlated effects: Quantifying technology spillovers in the presence of latent heterogeneity." *Journal of Applied Econometrics*, 28, 956–981. [120]

Carter, Christopher K. and Robert Kohn (1994), "On Gibbs sampling for state space models." *Biometrika*, 81, 541–553. [132]

Chib, Siddhartha (1992), "Bayes inference in the Tobit censored regression model." *Journal of Econometrics*, 51, 79–99. [118, 119, 132]

Chib, Siddhartha and Barton H. Hamilton (2002), "Semiparametric Bayes analysis of longitudinal data treatment models." *Journal of Econometrics*, 110, 67–89. [131]

Chib, Siddhartha and Ivan Jeliazkov (2006), "Inference in semiparametric dynamic models for binary longitudinal data." *Journal of the American Statistical Association*, 101, 685–700. [121]

Del Negro, Marco, Frank Schorfheide, Frank Smets, and Rafael Wouters (2007), "On the fit of new Keynesian models." *Journal of Business and Economic Statistics*, 25, 123–162. [132]

Fisher, Mark and Mark J. Jensen (2022), "Bayesian nonparametric learning of how skill is distributed across the mutual fund industry." *Journal of Econometrics*, 230, 131–153. [120]

Geweke, John (2005), *Contemporary Bayesian Econometrics and Statistics*. John Wiley & Sons, Inc. [151]

Geweke, John and Charles Whiteman (2006), "Bayesian forecasting." In *Handbook of Economic Forecasting*, Vol. 1 (Graham Elliott, Clive W. J. Granger, and Allan Timmermann, eds.), 3–80, Elsevier, New York. [124]

Ghosal, Subhashis and Aad van der Vaart (2017), *Fundamentals of Nonparametric Bayesian Inference*. Cambridge University Press, Cambridge. [128]

Ghosh, Amit (2015), "Banking-industry specific and regional economic determinants of non-performing loans: Evidence from U.S. states." *Journal of Financial Stability*, 20, 93–104. [120, 146]

Ghosh, Amit (2017), "Sector-specific analysis of non-performing loans in the U.S. banking system and their macroeconomic impact." *Journal of Economics and Business*, 93, 29–45. [120, 146, 147]

Ghosh, Jayanta K. and R. V. Ramamoorthi (2003), *Bayesian Nonparametrics*. Springer Verlag, New York. [128]

Giannone, Domenico, Michele Lenza, and Giorgio Primiceri (2015), "Prior selection for vector autoregressions." *Review of Economic and Statistics*, 97, 436–451. [132]

Gneiting, Tilmann and Adrian E. Raftery (2007), "Strictly proper scoring rules, prediction, and estimation." *Journal of the American Statistical Association*, 102, 359–378. [125]

Gu, Jiaying and Roger Koenker (2017a), "Empirical bayesball remixed: Empirical Bayes methods for longitudinal data." *Journal of Applied Economics*, 32, 575–599. [120]

Gu, Jiaying and Roger Koenker (2017b), "Unobserved heterogeneity in income dynamics: An empirical Bayes perspective." *Journal of Business & Economic Statistics*, 35, 1–16. [120]

Hartigan, John (1983), *Bayes Theory*. Springer Verlag, New York. [128]

Hirano, Keisuke (2002), "Semiparametric Bayesian inference in autoregressive panel data models." *Econometrica*, 70, 781–799. [120]

Ishwaran, Hemant and Lancelot F. James (2001), "Gibbs sampling methods for stick-breaking priors." *Journal of the American Statistical Association*, 96, 161–173. [120, 129, 132]

Ishwaran, Hemant and Lancelot F. James (2002), "Approximate drichlet process computing in finite normal mixtures." *Journal of Computational and Graphical Statistics*, 11, 508–532. [120, 130, 132]

Keane, Michael and Olena Stavrunova (2011), "A smooth mixture of Tobits model for healthcare expenditure." *Health Economics*, 20, 1126–1153. [120]

Keeton, William R. and Charles S. Morris (1987), "Why do banks' loan losses differ." Economic Review, FRB Kansas City. [146]

Koenker, Roger and Jose A. F. Machado (1999), "Goodness of fit and related inverence processes for quantile regression." *Journal of the American Statistical Association*, 94, 1296–1310. [147, 148]

Lancaster, Tony (2004), *An Introduction to Modern Bayesian Econometrics*. Blackwell Publishing. [151]

Li, Minglian and Justin Tobias (2011), "Bayesian methods in microeconometrics." In *Oxford Handbook of Bayesian Econometrics* (John Geweke, Gary Koop, and Herman van Dijk, eds.), 221–292, Oxford University Press, Oxford. [119]

Li, Tong and Xiaoyong Zheng (2008), "Semiparametric Bayesian inference for dynamic Tobit panel data models with unobserved heterogeneity." *Journal of Applied Econometrics*, 23, 699–728. [118, 119, 121, 132]

Liu, Laura (2022), "Density forecasts in panel data models: A semiparametric Bayesian perspective." *Journal of Business & Economic Statistics* (forthcoming). [118, 120, 122, 132]

Liu, Laura, Hyungsik Roger Moon, and Frank Schorfheide (2020), "Forecasting with dynamic panel data models." *Econometrica*, 88, 171–201. [120, 122]

Liu, Laura, Hyungsik Roger Moon, and Frank Schorfheide (2023), "Supplement to 'Forecasting with a panel Tobit model'." *Quantitative Economics Supplemental Material*, 14, https://doi.org/10.3982/QE1505. [120]

McFadden, Daniel (1973), "Conditional logit analysis of qualitative choice behavior." In *Frontiers in Econometrics* (Paul Zarembka, ed.), 105–142, Academic Press, New York. [147, 148]

Moyer, Susan (1990), "Capital adequacy ratio regulations and accounting choices in commercial banks." *Journal of Accounting and Economics*, 13, 123–154. [141]

Nychka, Douglas (1988), "Bayesian confidence intervals for smoothing splines." *Journal of the American Statistical Association*, 83, 1134–1143. [118, 127]

Ramey, Valerie A. (2016), "Macroeconomic shocks and their propagation." In *Handbook of Macroeconomics*, Vol. 2 (John B. Taylor and Harald Uhlig, eds.), 71–162, Elsevier, New York. [123]

Robbins, Herbert (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. [118]

Robert, Christian (1994), *The Bayesian Choice*. Springer Verlag, New York. [118, 127]

Rossi, Peter E. (2014), *Bayesian Non- and Semi-Parametric Methods and Applications*. Princeton University Press. [120]

Schorfheide, Frank (2005), "VAR forecasting under misspecification." *Journal of Econometrics*, 128, 99–136. [137]

Tanner, Martin A. and Wing Hung Wong (1987), "The calculation of posterior distributions by data augmentation." *Journal of the American Statistical Association*, 82, 528–540. [132]

van der Vaart, Aad (1998), *Asymptotic Statistics*. Cambridge University Press, Cambridge. [128]

Wahba, Grace (1983), "Bayesian "confidence intervals" for the cross-validated smoothing spline." *Journal of the Royal Statistical Society, Series B*, 45, 133–150. [118, 127]

Wei, Steven X. (1999), "A Bayesian approach to dynamic Tobit models." *Econometric Reviews*, 18, 417–439. [118, 119, 132]

Co-editor Tao Zha handled this manuscript.

Manuscript received 10 December, 2019; final version accepted 18 July, 2022; available online 28 July, 2022.