Forecasting with a Panel Tobit Model

Laura Liu
Indiana University

Hyungsik Roger Moon
University of Southern California and Yonsei

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

This Version: October 25, 2021

Abstract

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.

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

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

*Correspondence: L. Liu: Department of Economics, Indiana University, 100 S. Woodlawn Ave, Bloomington, IN 47405. Email: lauraliu@iu.edu. H.R. Moon: Department of Economics, University of Southern California, KAP 300, Los Angeles, CA 90089. E-mail: moonr@usc.edu. F. Schorfheide: Department of Economics, 133 S. 36th Street, University of Pennsylvania, Philadelphia, PA 19104-6297. Email: 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.

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 substantially delinquent after a period of time. The prediction of charge-off rates is interesting to banks, regulators, and investors because, from an economic perspective, charge-offs are losses on loan portfolios. If these charge-offs are large, the bank may be entering a period of distress and require additional capital. Due to mergers and acquisitions, changing business models, and changes in regulatory environments the time series dimension that is useful for forecasting is often short. The general methods developed in this paper are not tied to the charge-off rate application and can be used in any setting in which a researcher would like to analyze a panel of censored data with a large cross-sectional and a short time-series dimension.

In a panel data setting, cross-sectional heterogeneity in the data is modeled through unit-specific parameters. The more precisely these heterogeneous coefficients are estimated, the more accurate the forecasts are. The challenge in forecasting panels with a short time dimension is that they do not contain a lot of information about unit-specific parameters. A natural way of adding information to the estimation of these parameters is the use of prior distributions. For each time series, the prior information can be combined with the unit-specific likelihood function to form a posterior. Using Bayesian decision theory, the posterior distribution can be used to derive a forecast that minimizes posterior expected loss. From a frequentist perspective one obtains a forecast that will have some bias, but the resulting reduction in sampling variance often dominates the introduction of bias in a mean-squared-error sense.

The key insight in panel data applications is that, in the absence of any meaningful subjective prior information, one can extract information from the cross section and equate the prior distribution with the cross-sectional distribution of unit-specific coefficients. An empirical Bayes implementation of this idea could create a point estimate of the cross-sectional distribution of the heterogeneous coefficients and then conditions the subsequent posterior calculations on the estimated prior distribution.[1] In fact, the classic James-Stein estimator for a vector of means can be interpreted as an estimator constructed as follows. In the first step, a prior is generated by fitting a Normal distribution to a cross-section of

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

observations. In the second step, this prior is then combined with the unit-specific likelihood function to generate a posterior estimate of the unknown mean for that unit. In a panel setting the implementation is more involved but follows the same steps.

If the model is linear in the coefficients and the forecast is evaluated under a quadratic forecast error loss function, then Tweedie's formula, which expresses the posterior mean of the heterogeneous coefficients as the maximum likelihood estimate corrected by a function of the cross-sectional density of sufficient statistics, can be used to construct a forecast without having to explicitly estimate a prior distribution for the heterogeneous coefficients. This insight has been recently used, for instance, by Brown and Greenshtein (2009), Gu and Koenker (2017a,b), and Liu, Moon, and Schorfheide (2020).

In this paper 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. In particular, the full Bayesian approach is preferable for interval and density forecasts, because it captures all sources of uncertainties, including estimation uncertainty of the underlying distribution. For a linear panel data model, a full Bayesian analysis has been implemented by Liu (2021).

The contributions of our paper are threefold. First, we extend the implementation of the full Bayesian estimation in Liu (2021) to the dynamic panel Tobit model with heteroskedastic innovations and correlated random effects. The Bayesian method we develop in the paper is quite general, and the posterior sampler can be easily modified and applied to a richer class of limited dependent panel models, as discussed in Section 3.4.

Second, in regard to interval forecasting, we construct forecasts that target average posterior coverage probability across all units in our panel instead of pointwise coverage probability for each unit. We show that it is optimal to generate these forecasts as 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 also take the form of the union of one or more intervals and the value zero, or simply the value zero (singleton). For this reason we will refer to these forecasts as set instead of interval forecasts throughout this paper. Both in the Monte Carlo study and the empirical application the proposed Bayesian set forecasts have good frequentist coverage properties in the cross-section. This basic insight is connected

to similar findings in the literature on nonparametric function estimation and dates back to Wahba (1983) and Nychka (1988).

Third, we present a novel application in which we forecast bank-level loan charge-off rates. Our empirical analysis is based on more than 100 short panel data sets with a time dimension of $T = 10$. These panel data sets include predominantly credit card (CC) and residential real estate (RRE) loans and cover various (overlapping) time periods. 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 to the forecast assessment, we conduct a variety of posterior predictive checks to document that this model is able to capture the salient features of our data.

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). For instance, for the CC loan estimation sample from 2007Q2 to 2009Q4 we find that a five-percent increase in the unemployment rate lets annualized charge-off rates on CC loans increase 5.5 to 13 percentage points (posterior means) across the banks in our sample. Nonetheless, the variation in charge-off rates generated by local economic conditions is very small compared to the variation due to the heterogeneous intercept estimates.

We use two types of visualization techniques for the forecasts. First, we generate maps

that compare the spatial distribution of predicted loan losses during and after the Great Recession. More specifically, we indicate the probability and the loan charge-off rates of banks in each county covered in our sample exceed a certain threshold. Second, we plot set forecasts for each bank, sorted by the predicted charge-off rate. In the set forecast plots we document how the sets 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.

Our paper relates to several branches of the literature. The papers most closely related are Gu and Koenker (2017a,b), Liu, Moon, and Schorfheide (2020), and Liu (2021). All four of these papers focus on the estimation of the heterogeneous coefficients in linear panel data models and the subsequent use of the estimated coefficients for the purpose of prediction. Only the full Bayesian analysis in Liu (2021) has a natural extension to nonlinear models. Liu, Moon, and Schorfheide (2020), building on Brown and Greenshtein (2009), show that an empirical Bayes implementation based on Tweedie's formula can asymptotically (as the cross-sectional dimension tends to infinity) lead to forecasts that are as accurate as the so-called oracle forecasts. Here the oracle forecast is an infeasible benchmark that assumes that all homogeneous coefficients, as well as the distribution of the heterogeneous coefficients, are known to the forecaster. Liu (2021) shows that the predictive density obtained from the full Bayesian analysis converges to the oracle's predictive density in the Wasserstein metric as the cross-section gets large.

We also build on the Bayesian literature on the estimation of censored regression models. The idea of using data augmentation and Gibbs sampling to estimate a Tobit model dates back to Chib (1992). To sample the latent uncensored observations we rely on an algorithm tailored toward dynamic Tobit models by Wei (1999). Sampling from Truncated Normal distributions is implemented with a recent algorithm of Botev (2017). Our benchmark model is most closely related to the Bayesian semiparametric panel Tobit model of Li and Zheng (2008). They used their model to study female labor supply and estimated average partial effects and average transition probabilities. We generalize Li and Zheng's model by introducing heteroskedasticity through a latent unit-specific error variance and allowing for a more flexible form of correlated random effects. As mentioned previously, the former is very important for the density and set forecast performance. A broader survey of the literature on Bayesian estimation of univariate and multivariate censored regression models can be found in the handbook chapter by Li and Tobias (2011).

We model the unknown distribution of the heterogeneous coefficients (intercepts and innovation variances) as Dirichlet process mixtures (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.[2] 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 (2021). The implementation of our Gibbs sampler relies on Ishwaran and James (2001, 2002). If we restrict the number of mixture components to be equal to one, then we obtain a Normal correlated random effects model as a special case.

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 empirical results are relegated to the Online Appendix.

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

2 Model Specification and Forecast Evaluation

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

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

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. The $n_x \times 1$ vector x_{it} comprises a set of regressors and we define the homogeneous parameter $\theta = [\rho, \beta']'$. It is assumed that conditional on the heterogenous parameters and the regressors x_{it} the observations y_{it} are cross-sectionally independent. Our specification uses the lagged latent variable y_{it-1}^* on the right-hand side because it is more plausible for our empirical application. The Bayesian computations described in Section 3.2 below can be easily adapted to the alternative model, in which the lagged censored variable y_{it-1} appears on the right-hand side.

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

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

assuming cross-sectional independence of the heterogeneous coefficients.[3] Here ξ is 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_{i0}. In this case ξ would include the parameters of the conditional mean function and the non-redundant 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. The model is completed with the specification of a prior distribution for $\big(\theta, \xi\big)$. In the remainder of this section, we discuss the posterior predictive density in Section 2.1, density forecast evaluation in Section 2.2, and the construction and evaluation of set forecasts in Section 2.3.

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

2.1 Posterior Predictive Densities

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

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

Because y^*_{it+h} depends on x_{it+h} we are implicitly assuming that the sequence $X_{1:N,T+1:T+h}$ is known once the forecast is made. In the empirical application of Section 5 we focus on $h = 1$-step-ahead forecasts (see Section 3.3 for a discussion on multi-step forecasts) and define x_{it} as house price growth and unemployment growth in period $t-1$.

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

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

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

The posterior predictive distribution for unit i is given by

$$\begin{aligned} & p(y_{iT+h}|Y_{1:N,0:T}, X_{1:N,0:T+h}) \\ & = \int\int\int p(y_{iT+h}|y^*_{iT+h}) p\big(y^*_{iT+h}|y^*_{iT}, X_{i,T+1:T+h}, \lambda_i, \sigma^2_i, \theta\big) \\ & \quad \times p\big(y^*_{iT}, \lambda_i, \sigma^2_i, \theta, \xi | Y_{1:N,0:T}, X_{1:N,0:T+h}\big) dy^*_{iT} d(\lambda_i, \sigma^2_i) d(\theta, \xi). \end{aligned} \tag{3}$$

Draws from the density $p(y^*_{iT+h}|y^*_{iT}, X_{i,T+1:T+h}, \lambda_i, \sigma^2_i, \theta)$ can be generated by forward simulation of the autoregressive law of motion for y^*_{it} in (1); see also Section 3.3. To simplify the notation, we drop $X_{1:N,0:T+h}$ from the conditioning set in the remainder of this section. We denote expectations and probabilities under the posterior predictive distribution

by $\mathbb{E}^{y_{iT+h}}_{Y_{1:N,0:T}}[\cdot]$ and $\mathbb{P}^{y_{iT+h}}_{Y_{1:N,0:T}}\{\cdot\}$, respectively. More generally, we use subscripts to 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:

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

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

2.2 Evaluating Density Forecasts

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

$$\begin{aligned} LPS_h(M) \;=\; & \frac{1}{N}\sum_{i=1}^{N} \ln\big(\mathbb{I}\{y_{iT+h}=0\}\cdot \mathbb{P}^{y_{iT+h}}_{Y_{1:N,0:T}}\{y_{iT+h}=0|M\} \\ & +\mathbb{I}\{y_{iT+h}>0\}p(y_{iT+h}|Y_{1:N,0:T})\big) \end{aligned} \tag{5}$$

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 $p(y_{iT+1}|Y_{1:N,1:T})$ and a "perfect" density forecasts which assigns probability one to the realized y_{iT+h}. Then,

$$CRPS_h(M) = \frac{1}{N}\sum_{i=1}^{N}\int_0^{\infty}\big(F^{y_{iT+h}}_{Y_{1:N,0:T}}(y|M) - \mathbb{I}\{y_{iT+h}\leq y\}\big)^2 dy. \tag{6}$$

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).

2.3 Constructing and Evaluating Set Forecasts

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

$$C_{iT+h|T}(Y_{1:N,0:T}) = \{0\}\cup\left(\bigcup_{k=1}^{K_i}[a_{ik},\, b_{ik}]\right) \tag{7}$$

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

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

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 (4).[4] 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, i.e., $C_{iT+h|T}(Y_{1:N,0:T}) = \emptyset$.

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 pre-specified 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,

$$\mathbb{P}^{y_{iT+h}}_{Y_{1:N,0:T}}\big\{y_{iT+h} \in C_{iT+h|T}(Y_{1:N,0:T})\big\} \geq 1-\alpha \quad \text{for all } i, \tag{8}$$

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

$$\frac{1}{N}\sum_{i=1}^{N} \mathbb{P}^{y_{iT+h}}_{Y_{1:N,0:T}}\big\{y_{iT+h} \in C_{iT+h|T}(Y_{1:N,0:T})\big\} \geq 1-\alpha. \tag{9}$$

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 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

$$C_{iT+h|T}(Y_{1:N,0:T}) = \{0\} \cup \big\{y_{iT+h} \;\big|\; p_c(y_{iT+h}|Y_{1:N,0:T})\mathbb{I}\{y_{iT+h} \geq 0\} \geq \kappa_i\big\}, \tag{10}$$

[4]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.

where the threshold κ_i is chosen such that

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

Because $p_c(y|\cdot)$ is a continuous density, the HPD set can be represented as a collection of disjoint intervals as in (7).

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 (10) 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

$$\frac{1}{N}\sum_{i=1}^{N} \mathbb{I}\big\{y_{iT+h} \in C_{iT+h|T}(Y_{1:N,0:T})\big\} \tag{11}$$

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

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K_i}(b_{ik} - a_{ik}). \tag{12}$$

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.[5]

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

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

conditions under which

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

in $\mathbb{P}^{Y_{1:N,0:T},Y_{1:N,T+h}}$ probability as $N\longrightarrow\infty$. This result echoes results on average coverage probabilities being close to Bayesian credible levels in the literature on non-parametric function estimation, dating back to Wahba (1983) and Nychka (1988). To state the proposition we define the following probability associated with the interval $[a_{ik,N},\, b_{ik,N}]$ conditional on $(Y_{i,0:T},\vartheta)$:

$$F_{ik,N}(\vartheta)=\int_{a_{ik,N}}^{b_{ik,N}}p(y^*_{iT+h}|Y_{i,0:T},\vartheta)dy^*_{iT+h}. \tag{14}$$

Proposition 2.1 *Suppose that the posterior distribution $p(\vartheta|Y_{1:N,0:T})$ has the unique mode $\bar{\vartheta}_N$. Moreover, suppose the following assumptions hold:*

(i) There is 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

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

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

(ii) The functions $F_{ik,N}(\vartheta)$ defined in (14) are locally Lipschitz in any compact neighborhood $\mathcal{N}_N(\vartheta)$ with Lipschitz constants $M_{ik,N}(\mathcal{N}_N(\vartheta))$.

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

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

Then the empirical coverage frequency converges to the Bayesian credible level in the sense of (13).

A proof of this proposition is provided in the Online Appendix.[6] In Assumption (i) 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

[6]The statement of the proposition also holds if the model-implied data density $p(Y_{1:N,0:T})$ is replaced by an alternative data density $q(Y_{1:N,0:T})$, but the predictive density $p(Y_{1:N,T+h}|Y_{1:N,0:T})$ remains correctly specified.

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 (ii) and (iii) 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 is 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 $p(\lambda_i, y_{i0}^*, \sigma_i^2 | x_{i0}, \xi)$ and a specification of the prior distribution for (θ, ξ) in Section 3.1. Section 3.2 contains a description of the posterior sampler, Section 3.3 outlines multi-step forecasting apporaches, and Section 3.4 discusses potential generalizations of the dynamic panel Tobit model.

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_{i0}. 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 $IG(a,b)$ to denote the Beta, Gamma, and Inverse Gamma distributions, respectively. The pair (θ, σ^2) has a Normal-Inverse-Gamma distribution $NIG(m, v, a, b)$ if $\sigma^2 \sim IG(a,b)$ and $\theta|\sigma^2 \sim N(m, \sigma^2 v)$. Finally, the pair (Φ, Σ) has a matricvariate Normal-Inverse-Wishart distribution $MNIW(M, V, \nu, S)$ if $\Sigma \sim IW(\nu, S)$ has an inverse Wishart distribution and $\text{vec}(\Phi)|\Sigma \sim N(\text{vec}(M), \Sigma \otimes V)$.

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:

$$\theta \sim N(0, \tau_\theta I_{n_x+1}), \tag{15}$$

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_{i0}. Moreover, we assume that λ_i and y_{i0}^* are independent. Thus,

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

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

$$\begin{aligned} \lambda_i|\xi &\overset{iid}{\sim} N(\phi_{\lambda,k}, \Sigma_{\lambda,k}) \text{ with prob. } \pi_{\lambda,k}, \quad k = 1, \ldots, K \\ y_{i0}^*|\xi &\overset{iid}{\sim} N(\phi_y, \Sigma_y). \end{aligned} \tag{16}$$

The maximum number of mixture components K is assumed to be pre-specified.[7]

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

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

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

$$(\phi_{\lambda,k}, \Sigma_{\lambda,k}) \overset{iid}{\sim} NIG(0, \tau_\phi, 3, 2\tau_\sigma^\lambda), \quad (\phi_y, \Sigma_y) \sim NIG(0, \tau_\phi^y, 3, 2\tau_\sigma^y). \tag{17}$$

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).[8] 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 (16) and (17) 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

[7] 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 $N = 1000$.

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

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

$$\pi_{\lambda,1:K}|(\alpha_\lambda, K) \sim \begin{cases} \zeta_1, & k = 1, \\ \prod_{j=1}^{k-1}(1-\zeta_j)\,\zeta_k, & k = 2, \ldots, K-1, , \\ 1 - \sum_{j=1}^{K-1} p_j, & k = K, \end{cases} \quad \zeta_k \sim B(1, \alpha_\lambda), \quad \alpha_\lambda \sim G(2,2). \tag{18}$$

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 non-trivial 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 non-trivial components.

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

$$\sigma^2 \sim IG\big(3, 2\tau_v V^*\big). \tag{19}$$

The IG distribution is parameterized in a similar way as the IG distributions in (17). $V^* = \frac{1}{N}\sum_{i=1}^{N} \widehat{\mathbb{V}}_i(y_{it})$ 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 capture heteroskedasticity in a flexible manner, we augment the hyperparameter vector ξ and also represent the distribution of $\ln \sigma_i^2$ as a mixture of Normals:

$$\ln \sigma_i^2 | \xi \sim N\big(\psi_k, \omega_k^2\big) \text{ with prob. } \pi_{\sigma,k}, \quad k = 1, \ldots, K. \tag{20}$$

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:

$$(\psi_k, \omega_k^2) \overset{iid}{\sim} NIG\big(\ln(\tau_v V^*) - \ln(2)/2, 1, 3, 2\ln 2\big), \qquad k = 1, \ldots, K. \tag{21}$$

The parameterization 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 (19).[9] 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_{i0}. Second, we allow λ_i and y_{i0}^* to be correlated with each other conditional on x_{i0}. The CRE distribution is given by the following location and scale mixture of Normal distributions:

$$\big[\lambda_i,\, y_{i0}^*\big] \,\big|\, (x_{i0},\xi) \overset{iid}{\sim} N\big([1,\, x_{i0}']\Phi_k, \Sigma_k\big) \text{ with prob. } \pi_{\lambda,k}, \quad k=1,\ldots,K, \tag{22}$$

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 non-redundant 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 (18). The prior distribution for the coefficient matrices Φ_k and Σ_k is a multivariate generalization of the RE distribution. We assume:

$$(\Phi_k,\Sigma_k) \overset{iid}{\sim} MNIW\big(0,\tau_\phi I_{n_x+1},7,4D(\tau_\sigma)\big), \quad k=1,\ldots,K, \quad D(\tau_\sigma)=\begin{bmatrix}\tau_\sigma^\lambda & 0\\ 0 & \tau_\sigma^y\end{bmatrix}. \tag{23}$$

Under this parameterization 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.[10]

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

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

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

[9] 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 Lognormal distribution with mean $\tau_v V_*$ and variance $(\tau_v V_*)^2$.

[10] 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 parameterization of the Gamma distribution, this corresponds to an $IG(3,2D_{11}(\tau_\Sigma))=IG(3,2\tau_\sigma^\lambda)$ distribution.

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 $Y^*_{i,0:T}$, $i = 1, \ldots, N$; see Tanner and Wong (1987) and for the Tobit model Chib (1992) and Wei (1999). The sampler for the flexible mixture representation of the CRE distribution is based on Ishwaran and James (2001, 2002). With the exception of the treatment of the latent variables $Y^*_{i,0:T}$, the computations for the Tobit model are very similar to the ones for the linear model studied in Liu (2021).

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

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

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

Our model implies that whenever $y_{it} > 0$ we can deduce that $y^*_{it} = y_{it}$. Thus, we can focus our attention on periods in which $y_{it} = 0$. In the hypothetical sample we observe

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

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

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

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

$$M_{1:s|0} = [\mu_1, \ldots, \mu_s]', \quad \mu_1 = \lambda_i + \rho y_{it_1-1} + \beta' x_{it}, \quad \mu_\tau = \lambda_i + \rho \mu_{\tau-1} + \beta' x_{it} \text{ for } \tau = 2, \ldots, s. \tag{24}$$

The covariance matrix takes the form

$$\Sigma_{1:s|0} = \sigma_i^2 \begin{bmatrix} \rho_{1,1|0} & \cdots & \rho_{1,s|0} \\ \vdots & \ddots & \vdots \\ \rho_{s,1|0} & \cdots & \rho_{s,s|0} \end{bmatrix}, \quad \rho_{i,j|0} = \rho_{j,i|0} = \rho^{j-i} \sum_{l=0}^{i-1} \rho^{2l} \text{ for } j \geq i. \tag{25}$$

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

$$\begin{aligned} M_{1:s-1|0,s} &= M_{1:s-1|0} - \Sigma_{1:s-1,s|0} \Sigma^{-1}_{ss|0} (y_{it_2+1} - \mu_s) \\ \Sigma_{1:s-1,1:s-1|0,s} &= \Sigma_{1:s-1,1:s-1|0} - \Sigma_{1:s-1,s|0} \Sigma^{-1}_{ss|0} \Sigma_{s,1:s-1|0} \end{aligned} \tag{26}$$

to deduce that

$$Y^*_{i,t_1:t_2} \sim TN_-\big(M_{1:s-1|0,s}, \Sigma_{1:s-1,1:s-1|0,s}\big). \tag{27}$$

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 (24) and (25) 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 (24) and (25). According to (22), the joint distribution of (λ_i, y^*_{i0}) is a mixture of Normals. Using the mixture component membership indicator $\gamma_{i,\lambda}$, we can express $y^*_{i0}|(\lambda_i, x_{i0}) \sim N(\mu_*(\lambda_i, x_{i0}), \sigma^2_*)$. This leads to the mean vector

$$M_{1:s} = [\mu_1, \ldots, \mu_s], \quad \mu_1 = \mu_*(\lambda_i, x_{i0}), \quad \mu_\tau = \lambda_i + \rho\mu_{\tau-1} + \beta' x_{it} \text{ for } \tau = 2, \ldots, s \tag{28}$$

and the covariance matrix

$$\Sigma_{1:s} = \sigma^2_i \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & \rho_{1,1} & \cdots & \rho_{1,s-1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \rho_{s-1,1} & \cdots & \rho_{s-1,s-1} \end{bmatrix} + \sigma^2_* \begin{bmatrix} \rho^{0+0} & \cdots & \rho^{0+(s-1)} \\ \vdots & \ddots & \vdots \\ \rho^{(s-1)+0} & \cdots & \rho^{(s-1)+(s-1)} \end{bmatrix}, \tag{29}$$

where the definition of $\rho_{i,j}$ is identical to the definition of $\rho_{i,j|0}$ in (25). One can then use the formulas in (26) to obtain the mean and covariance parameters of the Truncated Normal distribution.

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

$$y^*_{it} - \rho y^*_{it-1} - \beta' x_{it} = \lambda_i + u_{it}, \quad u_{it} \overset{iid}{\sim} N(0, \sigma^2_i), \quad t = 1, \ldots, T. \tag{30}$$

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

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

$$y^*_{it} - \rho y^*_{it-1} - \beta' x_{it} - \lambda_i = u_{it}, \quad u_{it} \overset{iid}{\sim} N(0, \sigma^2_i), \quad t = 1, \ldots, T. \tag{31}$$

However, even conditional on the mixture component membership indicator $\gamma_{i,\sigma}$, the prior for σ^2_i in (20) is not conjugate and direct sampling is not possible. Instead, we sample from this

non-standard posterior via an adaptive random walk Metropolis-Hastings (RWMH) step.[11]

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

$$y^*_{it} - \lambda_i = \rho y^*_{it-1} + \beta' x_{it} + u_{it}, \quad u_{it} \overset{iid}{\sim} N(0, \sigma^2_i), \quad i = 1, \ldots, N, \quad t = 1, \ldots, T. \tag{32}$$

The temporal and spatial independence of the u_{it}'s allows us to pool observations across i and t. Under the Normal prior in (15), the posterior distribution of $\theta = [\rho, \beta']'$ is also Normal and we can obtain draws by direct sampling.

Step 5: Drawing from $(\gamma_{i,\lambda}, \gamma_{i,\sigma})|(Y_{i,0:T}, Y^*_{i,0:T}, \lambda_i, \sigma^2_i, \theta, \xi)$**.** 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

$$\bar{\pi}_{i,\lambda,k} = \frac{\pi_{\lambda,k} p_N\big(\lambda_i | y^*_{i0}, x_{i0}, \Phi_k, \Sigma_k\big)}{\sum_{k=1}^K \pi_{\lambda,k} p_N\big(\lambda_i | y^*_{i0}, x_{i0}, \Phi_k, \Sigma_k\big)}. \tag{33}$$

Note that the conditional distribution $\lambda_i|(y^*_{i0}, x_{i0}, \Phi_k, \Sigma_k)$ is Normal, indicated by the notation $p_N(\cdot)$, and can be derived from the joint Normal distributions of the mixture components in (22). Thus,

$$\gamma_{i,\lambda}|(\Phi_{1:k}, \Sigma_{1:K}, \lambda_i) = k \text{ with prob. } \bar{\pi}_{i,\lambda,k}. \tag{34}$$

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

$$\pi_{\lambda,1:K}|(n_{\lambda,1:K}, \alpha, K) \sim TSB\left(\{1 + n_{\lambda,k}\}_{k=1}^K, \left\{\alpha_\lambda + \sum_{j=k+1}^K n_{\lambda,j}\right\}_{k=1}^K, K\right), \tag{35}$$

meaning that the ζ_k's in (18) have a $B\big(1 + n_{\lambda,k}, \alpha_\lambda + \sum_{j=k+1}^K n_{\lambda,j}\big)$ distribution. Conditional

[11] 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%.

on $\pi_{\lambda,1:K}$ the hyperparameter α_λ has a Gamma posterior distribution of the form

$$\alpha_\lambda|\pi_{\lambda,1:K} \sim G(2+K-1, 2-\ln \pi_{\lambda,K}). \tag{36}$$

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

$$p(\Phi_k, \Sigma_k|Y_{1:N,0:T}, Y^*_{1:N,0:T}, \lambda_{1:N}, \sigma^2_{1:N}, \gamma_{1:N,\lambda}, \gamma_{1:N,\sigma}, \theta) \propto p(\Phi_k, \Sigma_k) \prod_{i\in J_{\lambda,k}} p(\lambda_i, y^*_{i0}|\Phi_k, \Sigma_k). \tag{37}$$

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

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

Modifications for the simplified model specifications. If the CRE distribution is modeled parametrically instead of flexibly, then the drawing of the component 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, i.e., $\sigma^2_i = \sigma^2$ for all i, we can pool (31) in Step 3 across t and i. In combination with the prior in (19) 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 Multi-Step Forecasting

In general, there are two ways of extending one-step-ahead to multi-step-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

$$y^*_{iT+h} = \lambda_i \left(\sum_{s=0}^{h-1} \rho^s\right) + \rho^h y^*_{iT} + \beta' \left(\sum_{s=0}^{h-1} \rho^s x_{iT+h-s}\right) + \sum_{s=0}^{h-1} \rho^s u_{iT+h-s}. \tag{38}$$

If the path $x_{i,T+1:T+h}$ is given at time T, then the evaluation of the mean of y^*_{iT+h} is straightforward.[12] If the path $x_{i,T+1:T+h}$ is not known in period T but assumed to be exogenous, then the user has to specify a model for the evolution of x_{it} from which trajectories can be simulated. Suppose we have draws $(\lambda_i^{(j)}, \rho^{(j)}, \beta^{(j)}, \sigma_i^{2(j)})$ and draws $x^{(j)}_{i,T+1:T+h}$ from the posterior predictive distribution of the exogenous regressors, then we can define

$$\begin{aligned}\mu^{(j)}_{iT+h|T} &= \lambda_i^{(j)}\left(\sum_{s=0}^{h-1}(\rho^{(j)})^s\right) + (\rho^{(j)})^h y^{*(j)}_{iT} + \beta^{(j)}\left(\sum_{s=0}^{h-1}(\rho^{(j)})^s x^{(j)}_{iT+h-s}\right)\\ \sigma^{2(j)}_{iT+h|T} &= \left(\sum_{s=0}^{h-1}(\rho^{(j)})^{2s}\right)\sigma_i^{2(j)}.\end{aligned}$$

We can sample $y^{*(j)}_{iT+h}$ from a $N(\mu^{(j)}_{iT+h|T}, \sigma^{2(j)}_{iT+h|T})$ and apply the censoring to obtain a draw $y^{(j)}_{iT+h}$.

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

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

with the understanding that x_{it} is known at time $t-h$ and the serial correlation in u_{it} implied by our original model (1) is ignored. A discussion of the disadvantages and advantages of multi-step estimation in the context of VARs can be found in Schorfheide (2005).

3.4 Model Generalizations

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

Second, the panel setup can be extended to richer limited dependent variable models.

[12] In a stress-testing application of our framework the path of the exogenous variables would be set by the regulator as part of the stressed macroeconomic scenario.

Let $Y_{it} = [y_{1,it}, \ldots, y_{M,it}]'$ and $Y_{it}^* = [y_{1,it}^*, \ldots, y_{M,it}^*]'$ and consider

$$\begin{aligned} Y_{it} &= f(Y_{it}^*), \quad Y_{it}^*|(Y_{it-1}^*, x_{it}, \lambda_i, \sigma_i^2, \theta) \sim p(Y_{it}^*|Y_{it-1}^*, x_{it}, \lambda_i, \sigma_i^2, \theta), \\ &\quad (\lambda_i, Y_{i0}^*, \sigma_i^2)|\theta \sim p(\lambda_i, Y_{i0}^*, \sigma_i^2|\theta, \xi), \end{aligned} \tag{39}$$

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

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

in which the censoring of observation $y_{2,it}$ depends on the observed sign of the latent variable $y_{1,it}^*$. In order to implement richer Tobit models in our dynamic panel framework one has to modify the sampler for the conditional posterior distribution of the latent variables $Y_{i,0:T}^*|(Y_{i,0:T}, \lambda_i, \sigma_i^2, \gamma_{i,\lambda}, \gamma_{i,\sigma}, \theta, \pi)$ in Step 1 above. For instance, a posterior sampler for the (static) Type 2 Tobit model is discussed in Li and Tobias (2011). These extensions can be of interest for other panel data forecasting applications with limited dependent variables.

4 Monte Carlo Experiment

The Monte Carlo experiment is based on the dynamic panel Tobit model in (1), which we simplify by omitting the additional predictors x_{it} and using 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 $N = 1,000$ cross-sectional units and the number of time periods in the estimation sample is $T = 10$. We generate one-step-ahead forecasts for period $t = T + 1$. The fraction of zeros across all samples is 45% and for roughly 15% of the cross-sectional units the sample consists of $T = 10$ zeros ("all zeros").[13] The measures

[13] 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

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:
$p(\lambda_i \vert y_{i0}^*) = \frac{1}{9} p_N\left(\lambda_i \vert \frac{5}{2}, \frac{1}{2}\right) + \frac{8}{9} p_N\left(\lambda_i \vert \frac{1}{4}, \frac{1}{2}\right)$
$p(\ln \sigma_i^2 \vert y_{i0}^*) = \frac{1}{9} p_N\left(\ln \sigma_i^2 - c \vert \frac{5}{2}, \frac{1}{2}\right) + \frac{8}{9} p_N\left(\ln \sigma_i^2 - c \vert \frac{1}{4}, \frac{1}{2}\right)$, c is chosen such that $\mathbb{E}[\sigma_i^2] = 1$
Sample Size: $N = 1,000$, $T = 10$
Number of Monte Carlo Repetitions: $N_{sim} = 100$
Fraction of Zeros: 45%, Fraction of All-Zeros: 15%

of forecast accuracy discussed in Sections 2.2 and 2.3 are first computed for the cross section $i = 1, \ldots, N = 1,000$ and we then average the performance statistics over the $n_{sim} = 100$ Monte Carlo repetitions.[14]

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.

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

specifications.

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

Table 2: Summary of Prior Distributions

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

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

Table 3: Monte Carlo Experiment: Forecast Performance and Estimates

	Density Fcst		Set Forecast "Average"		Set Forecast "Pointwise"		Estimates	
	LPS	CRPS	Cov.	Length	Cov.	Length	Bias($\hat{\rho}$)	StdD($\hat{\rho}$)
Flexible & Heterosk.	-0.757	0.277	0.910	1.260	0.933	1.503	-0.002	0.005
Normal & Heterosk.	-0.758	0.277	0.908	1.248	0.932	1.498	-0.006	0.005
Flexible & Homosk.	-0.902	0.294	0.929	1.506	0.942	1.698	0.007	0.008
Normal & 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

Notes: 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.

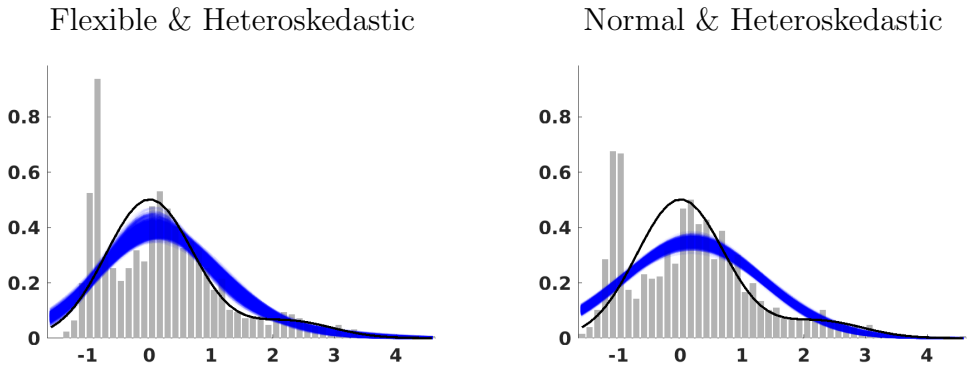
density or set forecasts that reflect parameter uncertainty, potential uncertainty about y_{iT}^*, and uncertainty about future shocks.

Density and Set Forecasts. To assess the density forecasts we compute LPS and CRPS; see Section 2.2. 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.

As discussed in Section 2.3, we consider two types of set forecasts. The first type targets the average coverage probability in the cross-section ("average"), whereas the other type targets the correct coverage probability for each unit i ("pointwise"). To assess the set forecasts we compute the coverage frequency and the average length of 90% predictive sets. 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 ho-

Figure 1: Posterior Means and Estimated RE Distributions for λ_i



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

moskedastic 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 & 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. 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.

To interpret the histograms of $\mathbb{E}[\lambda_i|Y_{1:N,0:T}]$ in view of the plotted $p(\lambda)$'s, we consider two stylized examples that capture important aspects of our setup. First, suppose that the model is static, linear, and homoskedastic, i.e., $y_{it} = \lambda_i + u_{it}$, $u_{it} \sim N(0, \sigma^2)$ and $\lambda_i \sim N(\phi_\lambda, 1)$, and

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

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

In this example, the distribution of the posterior mean estimates is less dispersed than the distribution of the λ_i's, but centered at the same mean, which is qualitatively consistent with Figure 1. Second, to understand the effect of censoring, suppose that $y_{it}^* = \lambda_i + u_{it}$ and we observe a sequence of zeros. The likelihood associated with this sequence of zeros is given by $\Phi_N^T(-\lambda_i/\sigma)$. The posterior mean for a sequence of zeros is then given by

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

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

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-offs are large, the bank may be entering a period of distress and require additional capital.[15]

[15] 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.

We consider a panel of "small" banks, which we define to be banks with total assets of less than one billion dollars. For these banks it is reasonable to assume that they operate in local markets. 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 $\beta' x_{it}$ 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 $x_{it} = [\Delta \ln \mathrm{HPI}_{it-1}, \Delta \mathrm{UR}_{it-1}]'$, 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 $p\big(\lambda_i, y_{i0}^*, \sigma_i | x_{i0}, \xi\big)$; 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 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.

Table 4: Summary Statistics for Baseline Samples

	N	Zeros [%]	All Zeros [%]	Mean	75th	Max
RRE	2,576	76	61	0.25	0.00	33.1
CC	561	43	22	3.27	4.07	260

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

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 ($t = T + 1$);[16] (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 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 $x_{it} = [\Delta \ln \text{HPI}_{it-1}, \Delta \text{UR}_{it-1}]'$. 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

[16]There are in general large uncertainties during the Great Recession, so it is challenging to conduct forecasts and necessitates more precise density and set forecasts.

Table 5: Density Forecast Performance

	RRE Charge-Offs		CC Charge-Offs	
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

Notes: The estimation sample ranges from 2007Q2 ($t = 0$) to 2009Q4 ($t = T = 10$). We use $x_{it} = [\Delta \ln \text{HPI}_{it-1}, \Delta \text{UR}_{it-1}]'$ and forecast 2010Q1 observations.

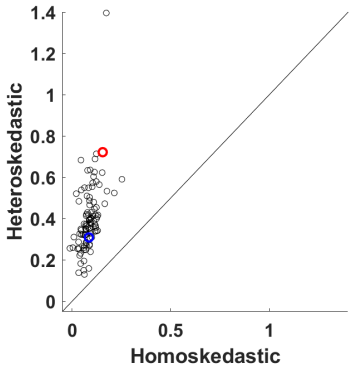
CC samples, all four heteroskedastic specifications lead to very similar density forecasting performance.

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 blue and red circles 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.

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-offs for 2010Q1 and 2018Q1, emphasizing the spatial dimension.[17] We associate each bank i with a particular county. If there are multiple banks in one county, we average the predicted probabilities.

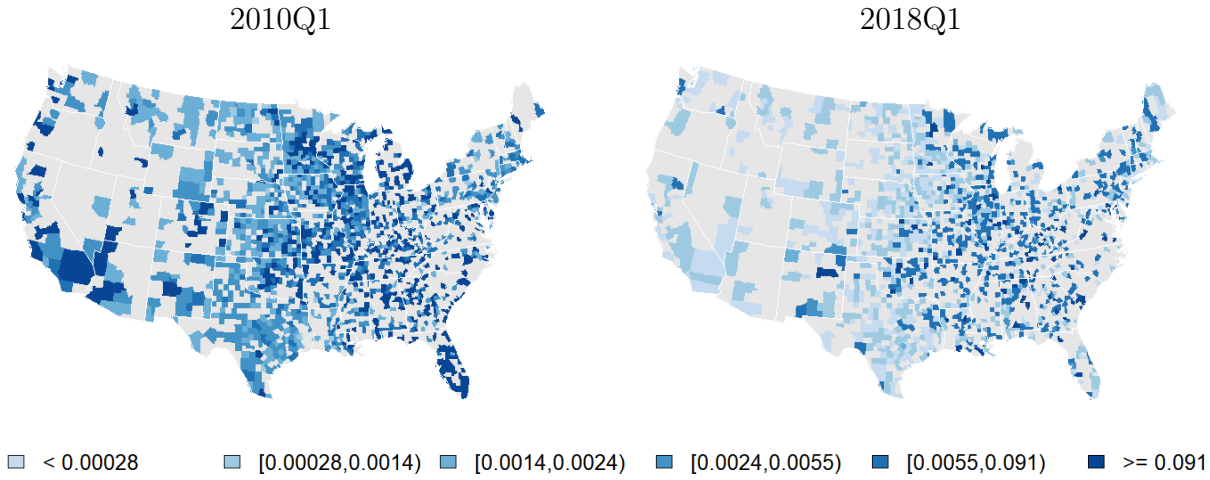
[17]Similar maps for CC charge-offs are available in the Online Appendix.

Figure 2: Log Predictive Density Scores – All Samples



Notes: 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 blue (red) circle corresponds to RRE (CC). We use $x_{it} = [\Delta \ln \text{HPI}_{it-1}, \Delta \text{UR}_{it-1}]'$.

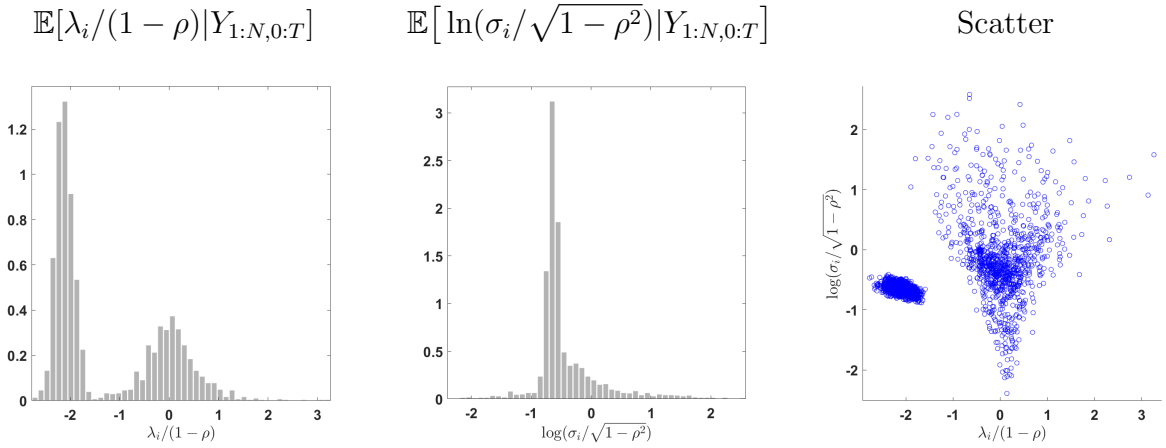
Figure 3: RRE Charge-Off Predictive Tail Probabilities, Spatial Dimension



Notes: Predictive tail probabilities are defined as $\mathbb{P}\{y_{iT+1} \geq c | Y_{1:N,0:T}\}$, 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 ($t = T = 10$).

2010Q1 is the immediate aftermath of the Great Recession and the counties that are covered by our sample appear predominantly in dark blue, indicating 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 blue instead of dark blue, in particular in hard

Figure 4: Heterogeneous Coefficient Estimates, RRE Charge-Offs



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

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.

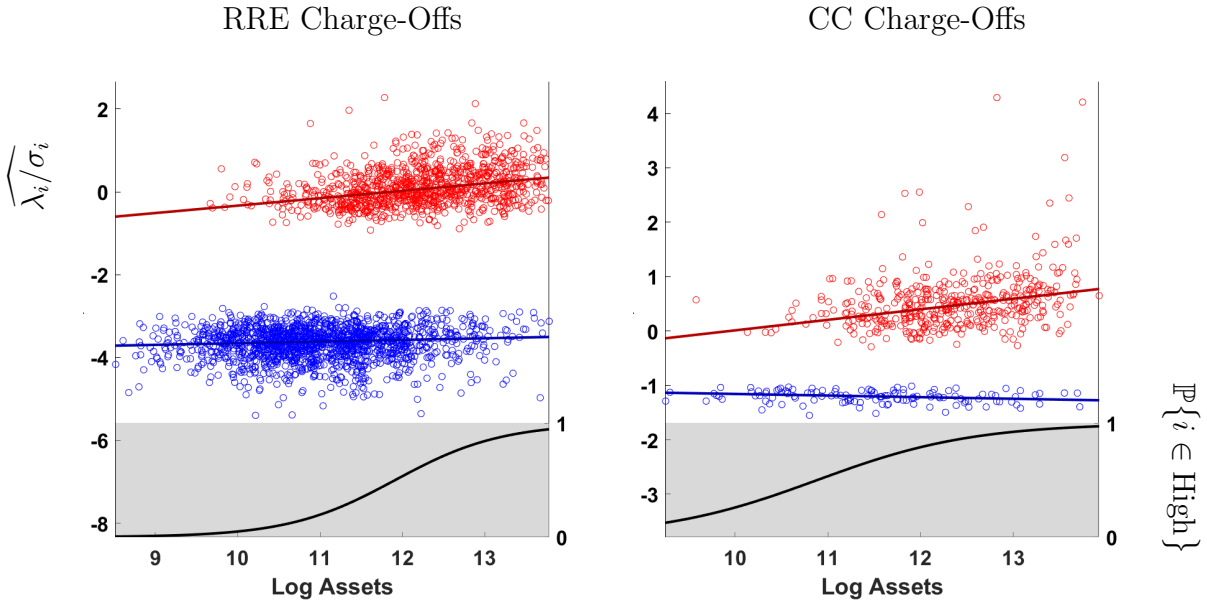
5.3 Parameter Estimates for Selected Samples

Heterogeneous Parameters. The distributions of posterior mean estimates of the heterogeneous coefficients for the 2007Q2 sample of RRE charge-offs are depicted in Figure 4.[18] 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

[18]Similar plots for CC charge-offs are available in the Online Appendix.

Figure 5: $\widehat{\lambda_i/\sigma_i}$ and $\mathbb{P}\{i \in \text{High}\}$ versus Log Assets (Bank Size)



Notes: 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 (blue) and high-λ (red) 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).

sufficiently small to explain the zero charge-off rates. In turn, the posterior mean estimate is predominantly driven by the prior. Similar spikes are visible in the 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 $\beta' x_{it}$, 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, i.e., 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 non-zero charge-offs, and bank

Table 6: Regressions of $\widehat{\lambda_i/\sigma_i}$ on Bank Characteristics

	RRE Charge-Offs						CC Charge-Offs					
	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	

Notes: 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 red and blue dots 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. $*$ indicates significance at 5% level. 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 = loglh(all)/loglh(intcpt) $- 1$ (McFadden, 1973).

size measured by the log of total assets. The top segments of the two panels contain scatter plots with group-wise 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 sample 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 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.[19] 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), 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

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

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 non-interest 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 non-performing loans, i.e., 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 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.

Table 7: Estimates of Common Parameters

	y^*_{it-1}		$\Delta \ln \text{HPI}_{it-1}$		ΔUR_{it-1}		$\Delta \ln \text{INC}_{it-1}$		LPS
	Mean	CI	Mean	CI	Mean	CI	Mean	CI	
RRE	0.22	[0.18, 0.26]	-0.05	[-0.07, -0.04]	0.85	[0.71, 0.97]			-0.5232
	0.21	[0.18, 0.25]	-0.05	[-0.06, -0.04]	0.83	[0.70, 0.96]	.001	[-.007, .010]	-0.5224
	0.29	[0.27, 0.31]							-0.5214
CC	0.41	[0.36, 0.46]	-0.16	[-0.25, -0.06]	2.60	[1.69, 3.52]			-1.9214
	0.41	[0.36, 0.45]	-0.16	[-0.26, -0.06]	2.64	[1.72, 3.58]	.016	[-.046, .079]	-1.9231
	0.48	[0.43, 0.52]							-1.9268

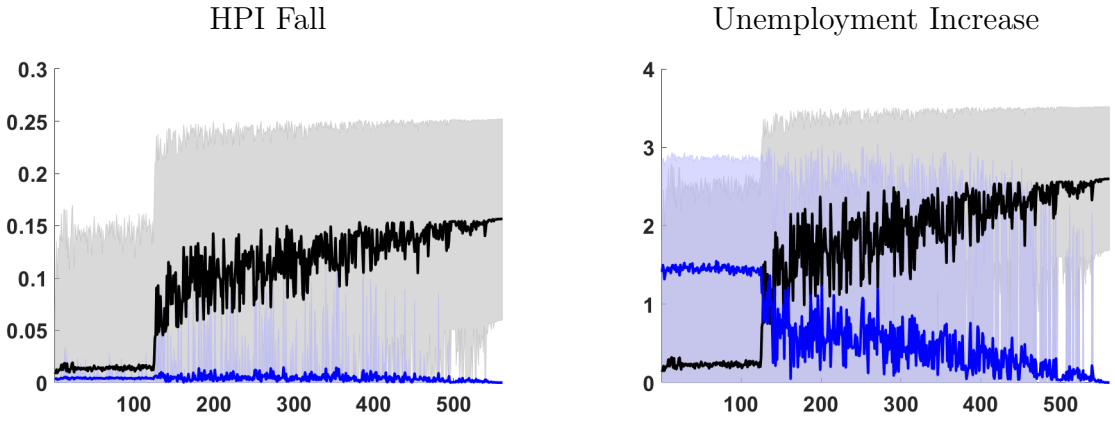
Notes: Heteroskedastic flexible CRE specification. The estimation sample ranges from 2007Q2 ($t = 0$) to 2009Q4 ($t = T = 10$). The table contains posterior means and 90% credible intervals in brackets.

We report posterior means and 90% credible intervals. For each sample we consider three specifications: (i) the baseline specification with $\Delta \ln \text{HPI}_{it-1}$ and ΔUR_{it-1}; (ii) an extended version that also includes $\Delta \ln \text{INC}_{it-1}$; (iii) and a version without regressors.

Both samples exhibit mild autocorrelation. The point estimate of ρ is 0.22 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 lead to an increase of charge-off rates by 0.05%. A 1% increase in the unemployment rate raises the charge-off rates by 0.85%. Finally, a 1% growth of personal income raises the charge-off rates by 0.001%. 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 $x_{it} = [\Delta \ln \text{HPI}_{it-1}, \Delta \text{UR}_{it-1}]'$, whereas for the RRE sample strictly speaking the model without regressors is preferred. In the Online Appendix we show scatter plots of $\hat{\lambda}_i + \beta' x_{it}$ versus $\hat{\lambda}_i$ which indicate that only a very small fraction is explained by local economic conditions. Despite the weak evidence in favor of the inclusion of local economic conditions we proceed with $x_{it} = [\Delta \ln \text{HPI}_{it-1}, \Delta \text{UR}_{it-1}]'$, whose coefficients are

Figure 6: Effects (Term I and II) of HPI and UR on CC Charge-Offs



Notes: 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. 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.

"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+1} to $\tilde{x}_{iT+1} = x_{iT+1} + \iota'\Delta x$, where the unit-length vector ι determines the direction of the perturbation of x_{iT+1} and $\Delta x > 0$ the magnitude. Accounting for censoring, we decompose the treatment effect on y_{iT+1} as follows:

$$\begin{aligned}\frac{\tilde{y}_{iT+1} - y_{iT+1}}{\Delta x} &= \beta'\iota\mathbb{I}\{\lambda_i + \rho y^*_{it-1} + \beta' x_{iT+1} + u_{iT+1} > 0\} \\ &\quad + \frac{\lambda_i + \rho y^*_{it-1} + \beta'\tilde{x}_{iT+1} + u_{iT+1}}{\Delta x}\Big(\mathbb{I}\{\lambda_i + \rho y^*_{it-1} + \beta'\tilde{x}_{iT+1} + u_{iT+1} > 0\} \\ &\qquad - \mathbb{I}\{\lambda_i + \rho y^*_{it-1} + \beta' x_{iT+1} + u_{iT+1} > 0\}\Big) \\ &= I_i + II_i.\end{aligned} \tag{40}$$

Term I_i captures the intensive margin, i.e., a bank that has non-zero charge-offs conditional on x_{iT+1} and $\tilde{x}_{iT+1}$. 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 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.[20] We sort the banks based on the posterior means $\widehat{\lambda_i/\sigma_i}$, which for $\rho = 0$ and $\beta = 0$ would determine the probability of a positive charge-off. We consider two choices for $\iota'\Delta x$: 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+1} they are unlikely to have non-zero charge-offs. For the remaining banks the posterior mean of the term I treatment effect rises under the HPI fall scenario from 0.05% to 0.16%, where the latter value is the coefficient estimate reported in Table 7. The credible intervals are fairly wide, ranging from 0% to 0.25%.

The posterior mean for component II (dark/light blue) of the treatment effect is different under the two economic scenarios. Under the house price fall term II is generally much smaller than term I. For the first 120 banks term II is small because much of $\beta'(\tilde{x}_{iT+1} - x_{iT+1})$ 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. Under the very severe unemployment scenario the low λ_i banks switch from zero to positive charge-offs which leads to a posterior mean of the average treatment effect of 1.5%. 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.

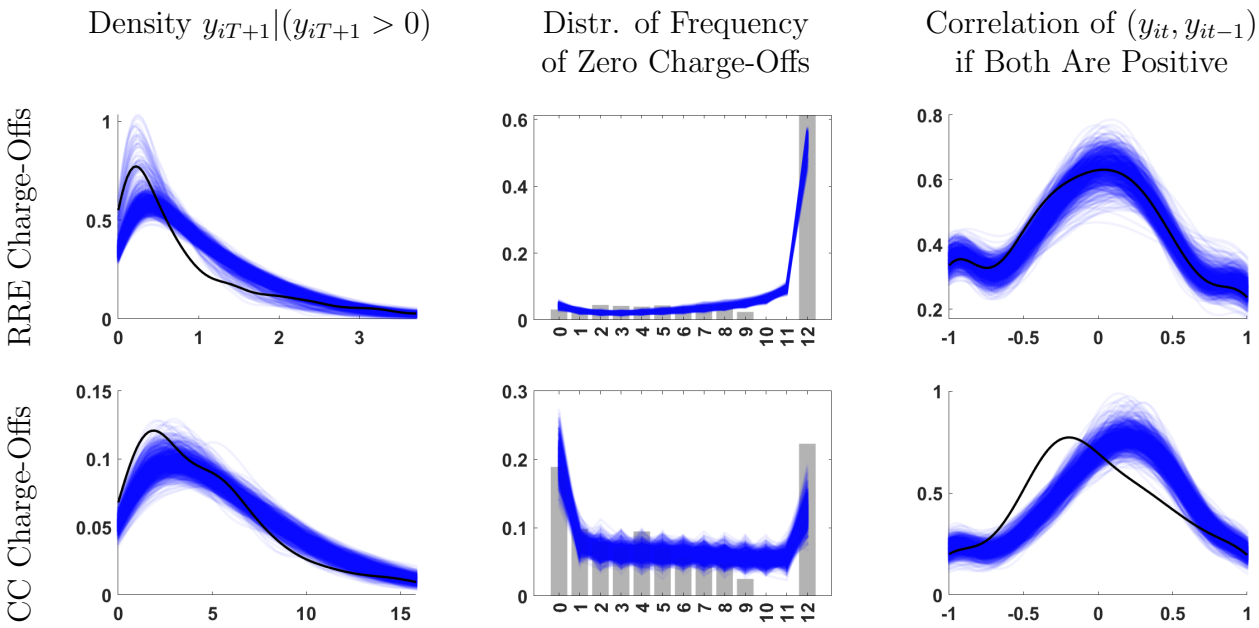
5.4 Posterior Predictive Checks for Selected Samples

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.[21] Consider the top left panel of the figure. Here, the particular characteristic, or sample statistic, under consideration is the cross-sectional density of y_{iT+1} conditional on $y_{iT+1} > 0$. The black line is computed from the actual RRE loan sample. Each blue hairline is generated as follows: (i) take a draw of (ρ, β, ξ) from the posterior distribution; (ii) conditional on these draws generate

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

[21] 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



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

$\lambda_{1:N}$, $Y^*_{1:N,0}$, 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 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 twelve quarters of the sample and roughly 5% of the banks write off RRE loans in every period. Overall, the estimated models do remarkably well in reproducing the patterns in the data. For RRE loans, the model captures the large number of all-zero samples and the fairly uniform distribution of the number of samples with zero to nine instances of $y_{it} = 0$. The only deficiency is that the model cannot explain the absence of samples with ten or eleven instances of zero charge-off rates. In the case of CC loans, the estimated model underpredicts the number of all-zero samples but generally is able to match the rest of the distribution.

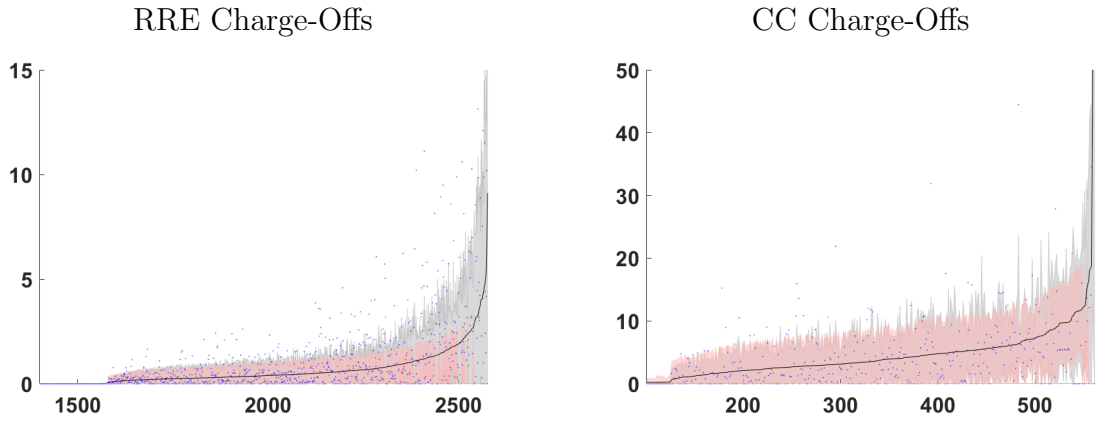
The last column of Figure 7 provides information about the models' ability to capture some of the dynamics of the charge-off data. Here the test statistic is the first-order sample autocorrelation of the $y_{i,0:T+1}$ sequence, conditional on both y_{it} and y_{it-1} being greater than zero. The panels 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 non-zero observations. With the exception of the autocorrelations in the CC charge-off sample, the two estimated models are able to reproduce the cross-sectional densities of the sample statistics.

5.5 Set Forecasts

Selected Samples. Set forecasts for 2010Q1, constructed as HPD sets from the posterior predictive distribution, are visualized in Figure 8. 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

Figure 8: Set Forecasts, Banks Sorted by $\mathbb{E}[y_{iT+1}|Y_{1:N,0:T}]$



Notes: 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.

mean forecast and the actual realization of the charge-off rate. The banks are sorted according to $\mathbb{E}[y_{iT+1}|Y_{1:N,0:T}]$. We don't show forecasts for the first 1,400 (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 length. For the RRE sample set forecasts for banks with large expected charge-offs $\mathbb{E}[y_{iT+1}|Y_{1:N,0:T}]$ become considerably shorter. In fact for $i > 2,500$ many of them become $\{0\}$. Although we plot the actual values of the charge-offs 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.[22]

We also report the frequency of the three types of set forecasts. Due to the large number

[22]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.

Table 8: Set Forecast Performance

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

Notes: 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%.

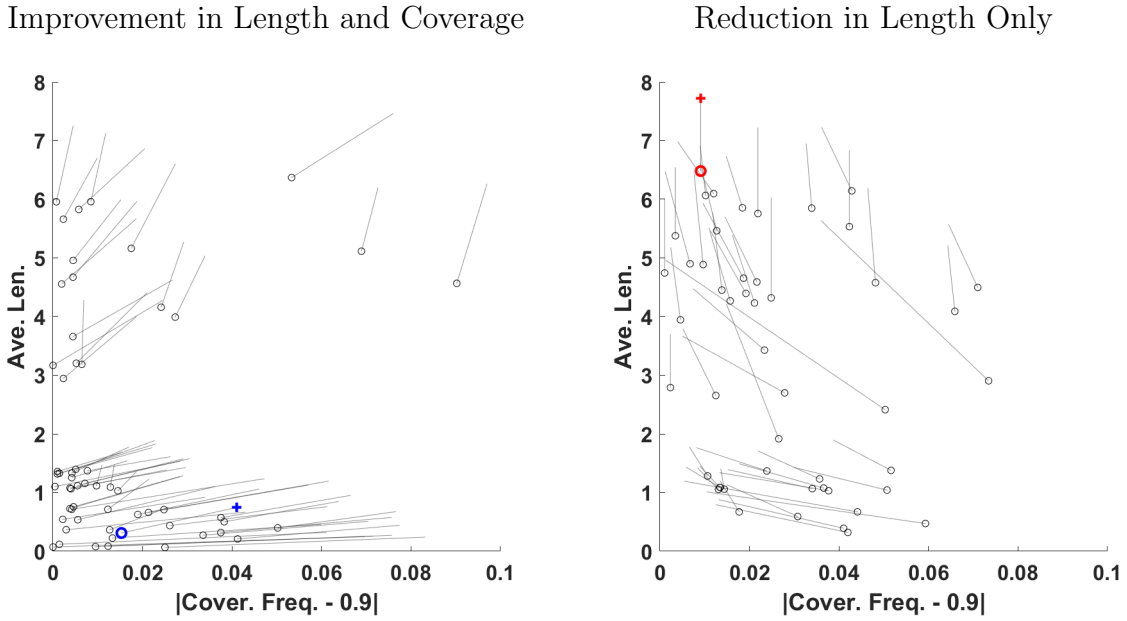
of zero observations in the RRE sample, there is a large fraction of banks, between 60% and 68%, for which the posterior predictive probability of observing $y_{iT+1} = 0$ exceeds 90%. This leads to a forecast of $\{0\}$. For the CC sample the fraction of $\{0\}$ forecasts is considerably smaller.

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 of Figure 8. Moreover, about one third of the disconnected forecasts are converted into connected forecast 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

Figure 9: Set Forecasts: Targeting Pointwise vs. Average Coverage – All Samples



Notes: Flexible CRE specification with heteroskedasticity. The blue (red) symbols correspond to the RRE (CC) baseline sample. 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.

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 frequency 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 dynamic panel versions of more general multivariate censored regression models. We can also allow for missing observations in our panel data set. Finally, even though we focused on the analysis of charge-off data, there are many other potential applications for our methods.

References

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

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

Askanazi, R., F. X. Diebold, F. Schorfheide, and M. Shin (2018): "On the Comparison of Interval Forecasts," *Journal of Time Series Analysis*, 39(6), 953–965.

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

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

Brown, L. D., and E. Greenshtein (2009): "Nonparametric Empirical Bayes and Compound Decision Approaches to Estimation of a High-dimensional Vector of Normal Means," *The Annals of Statistics*, pp. 1685–1704.

Burda, M., and M. Harding (2013): "Panel Probit with Flexible Correlated Effects: Quantifying Technology Spillovers in the Presence of Latent Heterogeneity," *Journal of Applied Econometrics*, 28(6), 956–981.

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

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

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

Del Negro, M., F. Schorfheide, F. Smets, and R. Wouters (2007): "On the Fit of New Keynesian Models," *Journal of Business and Economic Statistics*, 25(2), 123–162.

Fisher, M., and M. J. Jensen (2021): "Bayesian nonparametric learning of how skill is distributed across the mutual fund industry," *Journal of Econometrics*, forthcoming.

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

Ghosal, S., and A. van der Vaart (2017): *Fundamentals of Nonparametric Bayesian Inference*. Cambridge University Press, Cambridge.

Ghosh, A. (2015): "Banking-Industry Specific and Regional Economic Determinants of Non-performing Loans: Evidence from U.S. States," *Journal of Financial Stability*, 20, 93–104.

——— (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.

Ghosh, J. K., and R. Ramamoorthi (2003): *Bayesian Nonparametrics*. Springer Verlag, New York.

Giannone, D., M. Lenza, and G. Primiceri (2015): "Prior Selection for Vector Autoregressions," *Review of Economic and Statistics*, 97(2), 436–451.

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

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

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

Hartigan, J. (1983): *Bayes Theory Hartigan, John (1983):* Bayes Theory*, Springer Verlag, New York.* Springer Verlag, New York.

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

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

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

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

Keeton, W. R., and C. S. Morris (1987): "Why Do Banks' Loan Losses Differ," *Economic Review, FRB Kansas City*.

Koenker, R., and J. A. Machado (1999): "Goodness of Fit and Related Inverence Processes for Quantile Regression," *Journal of the American Statistical Association*, 94(448), 1296–1310.

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

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

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

Liu, L. (2021): "Density Forecasts in Panel Data Models: A Semiparametric Bayesian Perspective," *arXiv preprint arXiv:1805.04178*.

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

——— (2020): "Forecasting with Dynamic Panel Data Models," *Econometrica*, 88(1), 171–201.

McFadden, D. (1973): "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics*, ed. by P. Zarembka, pp. 105–142. Academic Press, New York.

Moyer, S. (1990): "Capital Adequacy Ratio Regulations and Accounting Choices in Commercial Banks," *Journal of Accounting and Economics*, 13, 123–154.

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

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

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

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

Schorfheide, F. (2005): "VAR Forecasting under Misspecification," *Journal of Econometrics*, 128(1), 99–136.

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

van der Vaart, A. (1998): *Asymptotic Statistics*. Cambridge University Press, Cambridge.

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

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

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

Laura Liu, Hyungsik Roger Moon, and Frank Schorfheide

This Online Appendix consists of the following sections:

A Proposition 2.1

A.1 Proof of Proposition 2.1.

Let $\vartheta = (\theta, \xi)$. The Bayes model specifies a joint distribution for the observations $(Y_{1:N,0:T}, Y_{1:N,T+h})$ and the parameters $(\vartheta, \lambda_{1:N}, \sigma^2_{1:N})$. This joint distribution can be factored into conditional distributions as follows

$$\begin{aligned} & p(Y_{1:N,0:T}, Y_{1:N,T+1}, \vartheta, \lambda_{1:N}) \\ &= p(Y_{1:N,0:T}) p(\vartheta | Y_{1:N,0:T}) \left(\prod_{i=1}^{N} p(\lambda_i, \sigma_i^2 | \vartheta, Y_{i,0:T}) p(y_{iT+h} | \lambda_i, \sigma_i^2, \vartheta, Y_{i,0:T}) \right). \end{aligned} \tag{A.1}$$

Sampling in a Bayesian framework involves drawing parameters from the appropriate distribution and generating data conditional on these parameters. Subsequently, we let $\tilde{\vartheta}_N$ be a draw from the posterior $p(\vartheta | Y_{1:N,0:T})$.

We start with the bound

$$\begin{aligned}
&\left|\frac{1}{N}\sum_{i=1}^{N}\mathbb{I}\big\{y_{iT+h}\in C_{iT+h|T}(Y_{1:N,0:T})\big\}-(1-\alpha)\right| \qquad (A.2)\\
&\leq \left|\frac{1}{N}\sum_{i=1}^{N}\mathbb{I}\big\{y_{iT+h}\in C_{iT+h|T}(Y_{1:N,0:T})\big\}-\mathbb{P}^{y_{iT+h}}_{Y_{1:N,0:T},\tilde{\vartheta}_N}\{y_{iT+h}\in C_{iT+h|T}(Y_{1:N,0:T})\}\right|\\
&\quad+\left|\frac{1}{N}\sum_{i=1}^{N}\mathbb{P}^{y_{iT+h}}_{Y_{1:N,0:T},\tilde{\vartheta}_N}\{y_{iT+h}\in C_{iT+h|T}(Y_{1:N,0:T})\}-(1-\alpha)\right|\\
&= B_1\big(Y_{1:N,0:T},Y_{1:N,T+h},\tilde{\vartheta}_N\big)+B_2\big(Y_{1:N,0:T},Y_{1:N,T+h},\tilde{\vartheta}_N\big).
\end{aligned}$$

The desired result follows if we can show that for any $\epsilon>0$

$$\lim_{N\longrightarrow\infty}\mathbb{P}^{Y_{1:N,0:T},Y_{1:N,T+h},\tilde{\vartheta}_N}\big\{B_j\big(Y_{1:N,0:T},Y_{1:N,T+h},\tilde{\vartheta}_N\big)>\epsilon\big\}=0,\quad j=1,2. \qquad (A.3)$$

Analysis of Term $B_1(\cdot)$. Note that $|B_1(\cdot)|<1$. We write

$$\begin{aligned}
&\lim_{N\longrightarrow\infty}\mathbb{P}^{Y_{1:N,0:T},Y_{1:N,T+h},\tilde{\vartheta}_N}\big\{B_1\big(Y_{1:N,0:T},Y_{1:N,T+h},\tilde{\vartheta}_N\big)>\epsilon\big\}\\
&= \lim_{N\longrightarrow\infty}\int\mathbb{P}^{Y_{1:N,T+h}}_{Y_{1:N,0:T},\tilde{\vartheta}_N}\big\{B_1\big(Y_{1:N,0:T},Y_{1:N,T+h},\tilde{\vartheta}_N\big)>\epsilon\big\}p(Y_{1:N,0:T},\tilde{\vartheta}_N)d(Y_{1:N,0:T},\tilde{\vartheta}_N)\\
&= \int\left[\lim_{N\longrightarrow\infty}\mathbb{P}^{Y_{1:N,T+h}}_{Y_{1:N,0:T},\tilde{\vartheta}_N}\big\{B_1\big(Y_{1:N,0:T},Y_{1:N,T+h},\tilde{\vartheta}_N\big)>\epsilon\big\}\right]p(Y_{1:N,0:T},\tilde{\vartheta}_N)d(Y_{1:N,0:T},\tilde{\vartheta}_N)\\
&= \int 0\cdot p(Y_{1:N,0:T},\tilde{\vartheta}_N)d(Y_{1:N,0:T},\tilde{\vartheta}_N)\\
&= 0,
\end{aligned}$$

as required. The second equality follows from the Dominated Convergence Theorem and the third equality follows from a Weak Law of Large Numbers for independently distributed random variables. Conditional on $(Y_{1:N,0:T},\tilde{\vartheta}_N)$, y_{iT+h} is sampled independently from $p(y_{iT+h}|\tilde{\vartheta}_N,Y_{1:N,0:T})$; see (A.1).

Analysis of Term $B_2(\cdot)$. To capture the probability mass at zero, define $a_{i0,N}=-\infty$ and $b_{i0,N}=\infty$. Let $\tilde{\vartheta}_N$ be a draw from the posterior $p(\vartheta|Y_{1:N,0:T})$. Recall that by construction of the set forecast

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{k=0}^{K_i}\int_{a_{ik,N}}^{b_{ik,N}}\int p(y^*_{iT+h}|Y_{i,0:T},\vartheta)p(\vartheta|Y_{1:N,0:T})d\vartheta dy^*_{iT+h}=1-\alpha.$$

Then,

$$\begin{aligned}
&\left|\frac{1}{N}\sum_{i=1}^{N}\sum_{k=0}^{K_i}\int_{a_{ik,N}}^{b_{ik,N}} p(y^*_{iT+h}|Y_{i,0:T},\tilde{\vartheta}_N)dy^*_{iT+h}-(1-\alpha)\right| \qquad (A.4)\\
&= \left|\frac{1}{N}\sum_{i=1}^{N}\sum_{k=0}^{K_i}\int_{a_{ik,N}}^{b_{ik,N}} p(y^*_{iT+h}|Y_{i,0:T},\tilde{\vartheta}_N)dy^*_{iT+1}-\int\left[\int_{a_{ik,N}}^{b_{ik,N}} p(y^*_{iT+h}|Y_{i,0:T},\vartheta)dy^*_{iT+h}\right]p(\vartheta|Y_{1:N,0:T})d\vartheta\right|\\
&= \left|\frac{1}{N}\sum_{i=1}^{N}\sum_{k=0}^{K_i}\int\left[\int_{a_{ik,N}}^{b_{ik,N}} p(y^*_{iT+1}|Y_{i,0:T},\tilde{\vartheta}_N)dy^*_{iT+1}-\int_{a_{ik,N}}^{b_{ik,N}} p(y^*_{iT+1}|Y_{i,0:T},\vartheta)dy^*_{iT+1}\right]p(\vartheta|Y_{1:N,0:T})d\vartheta\right|,
\end{aligned}$$

where we exchanged the order of integration in the second term on the right-hand side of the first equality. Combining the definition of $F_{ik,N}(\vartheta)$ in (14) with (A.4) and noting that $0 \leq F_{ik,N}(\vartheta) \leq 1$, we obtain

$$\begin{aligned}
&\left|\frac{1}{N}\sum_{i=1}^{N}\sum_{k=0}^{K_i}\int_{a_{ik,N}}^{b_{ik,N}} p(y^*_{iT+1}|Y_{i,0:T},\tilde{\vartheta}_N)dy^*_{iT+1}-(1-\alpha)\right| \qquad (A.5)\\
&= \left|\frac{1}{N}\sum_{i=1}^{N}\sum_{k=0}^{K_i}\int\left[F_{ik,N}(\tilde{\vartheta}_N)-F_{ik,N}(\vartheta)\right]p(\vartheta|Y_{1:N,0:T})d\vartheta\right|\\
&\leq \frac{1}{N}\sum_{i=1}^{N}\sum_{k=0}^{K_i}\int\left|F_{ik,N}(\tilde{\vartheta}_N)-F_{ik,N}(\vartheta)\right|p(\vartheta|Y_{1:N,0:T})d\vartheta\\
&\leq \frac{1}{N}\sum_{i=1}^{N}\sum_{k=0}^{K_i}\int_{\mathcal{N}_N(\bar{\vartheta}_N)}\left|F_{ik,N}(\tilde{\vartheta}_N)-F_{ik,N}(\vartheta)\right|p(\vartheta|Y_{1:N,0:T})d\vartheta+\int_{\mathcal{N}_N^c(\bar{\vartheta}_N)}p(\vartheta|Y_{1:N,0:T})d\vartheta\\
&= I+II,
\end{aligned}$$

say. The last inequality uses the bound $\left|F_{ik,N}(\tilde{\vartheta}_N)-F_{ik,N}(\vartheta)\right| \leq 1$ for the second term.

According to Assumption (i), we can choose a stochastic sequence of shrinking neighborhoods $\mathcal{N}_N(\bar{\vartheta}_N)$ such that

$$II \xrightarrow{p} 0$$

as $N \longrightarrow \infty$. Now consider term I. Write

$$\begin{aligned}
I &= \frac{1}{N}\sum_{i=1}^{N}\sum_{k=0}^{K_i}\mathbb{I}\{\tilde{\vartheta}_N \in \mathcal{N}_N(\bar{\vartheta}_N)\}\int_{\mathcal{N}_N(\bar{\vartheta}_N)}\left|F_{ik,N}(\tilde{\vartheta}_N)-F_{ik,N}(\vartheta)\right|p(\vartheta|Y_{1:N,0:T})d\vartheta\\
&\quad+\frac{1}{N}\sum_{i=1}^{N}\sum_{k=0}^{K_i}\mathbb{I}\{\tilde{\vartheta}_N \in \mathcal{N}_N^c(\bar{\vartheta}_N)\}\int_{\mathcal{N}_N(\bar{\vartheta}_N)}\left|F_{ik,N}(\tilde{\vartheta}_N)-F_{ik,N}(\vartheta)\right|p(\vartheta|Y_{1:N,0:T})d\vartheta\\
&= Ia+Ib,
\end{aligned}$$

say. It is straightforward to establish that term Ib converges to zero. Recall that the posterior mode is a function of $Y_{1:N,0:T}$. For any $\epsilon > 0$

$$\begin{aligned}
\mathbb{P}^{Y_{1:N,0:T},\tilde{\vartheta}_N}\{Ib > \epsilon\} &\leq \mathbb{P}^{Y_{1:N,0:T},\tilde{\vartheta}_N}\left\{\mathbb{I}\{\tilde{\vartheta}_N \in \mathcal{N}_N^c(\bar{\vartheta}_N)\}\left(\frac{1}{N}\sum_{i=1}^N K_i\right) > \epsilon\right\} \\
&= \mathbb{P}^{Y_{1:N,0:T},\tilde{\vartheta}_N}\{\tilde{\vartheta}_N \in \mathcal{N}_N^c(\bar{\vartheta}_N)\} \\
&= \int \mathbb{P}^{\tilde{\vartheta}_N}_{Y_{1:N,0:T}}\{\tilde{\vartheta}_N \in \mathcal{N}_N^c(\bar{\vartheta}_N)\}p(Y_{1:N,0:T})dY_{1:N,0:T} \\
&\longrightarrow 0.
\end{aligned}$$

The convergence statement in the last line follows from Assumption (i) and the Dominated Convergence Theorem:

$$\begin{aligned}
&\lim_{N\longrightarrow\infty} \int \mathbb{P}^{\tilde{\vartheta}_N}_{Y_{1:N,0:T}}\{\tilde{\vartheta}_N \in \mathcal{N}_N^c(\bar{\vartheta}_N)\}p(Y_{1:N,0:T})dY_{1:N,0:T} \\
&= \int \left[\lim_{N\longrightarrow\infty} \mathbb{P}^{\tilde{\vartheta}_N}_{Y_{1:N,0:T}}\{\tilde{\vartheta}_N \in \mathcal{N}_N^c(\bar{\vartheta}_N)\}\right]p(Y_{1:N,0:T})dY_{1:N,0:T} \\
&= \int 0 \cdot p(Y_{1:N,0:T})dY_{1:N,0:T}.
\end{aligned}$$

To bound term Ia we use the Lipschitz condition in Assumption (ii):

$$\begin{aligned}
Ia &\leq \frac{1}{N}\sum_{i=1}^N\sum_{k=1}^{K_i} M_{ik,N}(\mathcal{N}_N(\bar{\vartheta}_N))\mathbb{I}\{\tilde{\vartheta}_N \in \mathcal{N}_N(\bar{\vartheta}_N)\}\int_{\mathcal{N}_N(\bar{\vartheta}_N)} \|\tilde{\vartheta}_N - \vartheta\|p(\vartheta|Y_{1:N,0:T})d\vartheta \\
&\leq \frac{1}{N}\sum_{i=1}^N\sum_{k=1}^{K_i} M_{ik,N}(\mathcal{N}_N(\bar{\vartheta}_N))\mathbb{I}\{\tilde{\vartheta}_N \in \mathcal{N}_N(\bar{\vartheta}_N)\} \qquad \text{(A.6)} \\
&\quad \times \int_{\mathcal{N}_N(\bar{\vartheta}_N)} \big(\|\tilde{\vartheta}_N - \bar{\vartheta}_N\| + \|\bar{\vartheta}_N - \vartheta\|\big)p(\vartheta|Y_{1:N,0:T})d\vartheta \\
&\leq \left(\frac{1}{N}\sum_{i=1}^N\sum_{k=1}^{K_i} M_{ik,N}(\mathcal{N}_N(\bar{\vartheta}_N))\right)\mathbb{I}\{\tilde{\vartheta}_N \in \mathcal{N}_N(\bar{\vartheta}_N)\}\|\tilde{\vartheta}_N - \bar{\vartheta}_N\| \\
&\quad + \left(\frac{1}{N}\sum_{i=1}^N\sum_{k=1}^{K_i} M_{ik,N}(\mathcal{N}_N(\bar{\vartheta}_N))\right)\int_{\mathcal{N}_N(\bar{\vartheta}_N)} \|\bar{\vartheta}_N - \vartheta\|p(\vartheta|Y_{1:N,0:T})d\vartheta \\
&\leq \left(\frac{1}{N}\sum_{i=1}^N\sum_{k=1}^{K_i} M_{ik,N}(\mathcal{N}_N(\bar{\vartheta}_N))\right)2\delta_N.
\end{aligned}$$

The last inequality follows from the definition of the neighborhood $\mathcal{N}_N(\bar{\vartheta}_N)$. Using Assump-

tions (i) and (iii), we can deduce that

$$Ia \xrightarrow{p} 0, \tag{A.7}$$

in $\mathbb{P}^{Y_{1:N,0:T}}$ probability, which completes the proof. ■

A.2 A Simple Example

Consider a simple model without censoring:

$$y_{it} = \lambda_i + \theta y_{it-1} + u_{it}, \quad y_{i0} \sim N(0,1), \quad \lambda_i \sim N(\xi, 1), \quad u_{it} \sim N(0,1), \quad T = 1. \tag{A.8}$$

Define the vector of homogeneous parameters as $\vartheta = [\theta, \xi]'$. We use a prior of the form

$$p(\vartheta) \sim N(0, I).$$

In this example the predictive distribution is unimodal, which means that the HPD set constructed from the continuous part of the predictive density is a single interval. In turn, the summation of predictive interval segments over k is unnecessary. Let $x_{it} = [1, y_{it-1}]'$. The distribution of $y_{i1}|y_{i0}, \vartheta$ after integrating out λ_i is

$$y_{i1}|(y_{i0}, \vartheta) \sim iidN\big(x_{i1}'\vartheta, 2\big), \quad i = 1, \ldots, N.$$

Convergence in probability statements in Proposition 2.1 refer to the marginal distribution of the data characterized by the density

$$\begin{aligned} p(Y_{1:N,0:1}) &= (2\pi)^{-N/2-1} \left(\int \exp\left\{ -\frac{1}{2 \cdot 2} \left(\sum_{i=1}^{N} (y_{i1} - x_{i1}'\vartheta)^2 \right) - \frac{1}{2}\vartheta'\vartheta \right\} d\vartheta \right) \\ &\quad \times (2\pi)^{-N/2} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{N} y_{i0}^2 \right\}. \end{aligned}$$

Assumption (i) This leads to the likelihood function

$$p(Y_{1:N,0:1}|\vartheta) \propto \exp\left\{ -\frac{1}{2 \cdot 2} \left(\vartheta' \left(\sum_{i=1}^{N} x_{i1}x_{i1}' \right) \vartheta - 2\vartheta' \left(\sum_{i=1}^{N} x_{i1}y_{i1} \right) \right) \right\}.$$

Under the Normal prior for ϑ we obtain the following posterior mean and (scaled) variance:

$$\bar{\vartheta}_N = \left(\frac{1}{2}\sum_{i=1}^{N} x_{i1}x_{i1}' + I\right)^{-1}\left(\frac{1}{2}\sum_{i=1}^{N} x_{i1}y_{i1}\right), \quad \bar{V}_N = \left(\frac{1}{2N}\sum_{i=1}^{N} x_{i1}x_{i1}' + \frac{1}{N}I\right)^{-1}. \tag{A.9}$$

The overall posterior distribution is given by

$$\vartheta | Y_{1:N,0:1} \sim N\big(\bar{\vartheta}_N, \bar{V}_N/N\big). \tag{A.10}$$

We can define the shrinking neighborhood as the set

$$\mathcal{N}_N(\bar{\vartheta}_N) = \big\{\vartheta \;\big|\; (\vartheta - \bar{\vartheta}_N)'\bar{V}_N^{-1}(\vartheta - \bar{\vartheta}_N)' \leq 2N^{-\eta}\big\}, \quad 0 < \eta < 1. \tag{A.11}$$

Thus, for $\vartheta \in \mathcal{N}_N(\bar{\vartheta}_N)$ we have

$$\lambda_{min}(\bar{V}_N^{-1})\|\vartheta - \bar{\vartheta}_N\|^2 \leq 2N^{-\eta}$$

or

$$\|\vartheta - \bar{\vartheta}_N\| \leq \sqrt{\frac{2}{\lambda_{min}(\bar{V}_N^{-1})}}N^{-\eta/2} \equiv \delta_N.$$

The argument can be completed by showing that

$$\lambda_{min}(\bar{V}_N^{-1}) \xrightarrow{p} \epsilon_*, \quad \epsilon_* > 0$$

under $\mathbb{P}^{Y_{1:N,0}}$.

Assumption (ii) We now construct the Lipschitz constant. Consider

$$\begin{aligned}
F_{i,N}(\theta,\xi) &= \int_{a_{i,N}}^{b_{i,N}}\int_{\lambda_i} p_N(y_{i2}|\lambda_i + \theta y_{i1}, 1)p(\lambda_i|y_{i,0:1},\theta,\xi)d\lambda_i dy_{i2} \\
&= \int_{\lambda_i}\left[\int_{a_{i,N}}^{b_{i,N}} p_N(y_{i2}|\lambda_i + \theta y_{i1}, 1)dy_{i2}\right]p(\lambda_i|y_{i,0:1},\theta,\xi)d\lambda_i \\
&= \int_{\lambda_i}\Phi_N\big(g(\lambda_i + \theta y_{i1}; u_{i,N})\big)p(\lambda_i|y_{i,0:1},\theta,\xi)d\lambda_i \\
&\quad - \int_{\lambda_i}\Phi_N\big(g(\lambda_i + \theta y_{i1}; l_{i,N})\big)p(\lambda_i|y_{i,0:1},\theta,\xi)d\lambda_i,
\end{aligned}$$

where

$$g(\lambda_i + \theta y_{i1}; \zeta) = \zeta - \lambda_i - \theta y_{i1}, \quad \zeta \in \{a_{i,N}, b_{i,N}\}.$$

To find a Lipschitz constant, we construct a bound for

$$\left\| \frac{\partial}{\partial(\theta,\xi)} F_{i,N}(\theta,\xi) \right\|.$$

Define

$$F_{i,N,\zeta}(\theta,\xi) = \int_{\lambda_i} \Phi_N\big(g(\lambda_i+\theta y_{i1};\zeta)\big) p(\lambda_i|y_{i,0:1},\theta,\xi) d\lambda_i, \quad \zeta \in \{a_{i,N}, b_{i,N}\}.$$

Exchanging the order of differentiation and integration, write

$$\begin{aligned}
\frac{\partial}{\partial\theta} F_{i,N,\zeta}(\theta,\xi) &= \int_{\lambda_i} \phi_N\big(g(\lambda_i+\theta y_{i1};\zeta)\big) \left(\frac{\partial}{\partial\theta} g(\lambda_i+\theta y_{i1};\zeta)\right) p(\lambda_i|y_{i,0:1},\theta,\xi) d\lambda_i \\
&\quad + \int_{\lambda_i} \Phi_N\big(g(\lambda_i+\theta y_{i1};\zeta)\big) \left(\frac{\partial}{\partial\theta} p(\lambda_i|y_{i,0:1},\theta,\xi)\right) d\lambda_i \\
\frac{\partial}{\partial\xi} F_{i,N,\zeta}(\theta,\xi) &= \int_{\lambda_i} \phi_N\big(g(\lambda_i+\theta y_{i1};\zeta)\big) \left(\frac{\partial}{\partial\xi} g(\lambda_i+\theta y_{i1};\zeta)\right) p(\lambda_i|y_{i,0:1},\theta,\xi) d\lambda_i \\
&\quad + \int_{\lambda_i} \Phi_N\big(g(\lambda_i+\theta y_{i1};\zeta)\big) \left(\frac{\partial}{\partial\xi} p(\lambda_i|y_{i,0:1},\theta,\xi)\right) d\lambda_i.
\end{aligned}$$

Now note that

$$0 \le \phi_N(\cdot) \le (2\pi)^{-1/2}, \quad 0 \le \Phi_N(\cdot) \le 1,$$

and

$$\frac{\partial}{\partial\theta} g(\lambda_i+\theta y_{i1};\zeta) = y_{i1}, \quad \frac{\partial}{\partial\xi} g(\lambda_i+\theta y_{i1};\zeta) = 0.$$

Finally,

$$\int_{\lambda_i} \left(\frac{\partial}{\partial\theta} p(\lambda_i|y_{i,0:1},\theta,\xi)\right) d\lambda_i = \frac{\partial}{\partial\theta} \int_{\lambda_i} p(\lambda_i|y_{i,0:1},\theta,\xi) d\lambda_i = 0.$$

The same result holds for differentiation with respect to ξ. In turn, we obtain

$$\left|\frac{\partial}{\partial\theta} F_{i,N,\zeta}(\theta,\xi)\right| \le \left|\frac{y_{i1}}{\sqrt{2\pi}}\right|, \quad \left|\frac{\partial}{\partial\xi} F_{i,N,\zeta}(\theta,\xi)\right| = 0. \tag{A.12}$$

Noting that

$$F_{i,N}(\theta,\xi) = F_{i,N,u_{i,N}}(\theta,\xi) - F_{i,N,l_{i,N}}(\theta,\xi),$$

we can now define the Lipschitz constant

$$M_{i,N} = 2\left|\frac{y_{i1}}{\sqrt{2\pi}}\right| = \sqrt{\frac{2}{\pi}}|y_{i1}|,$$

which does not depend on $\mathcal{N}_N(\bar{\vartheta}_N)$. Thus, Assumption (ii) is satisfied.

Assumption (iii) Notice that in our model $\mathbb{E}[h(y_{i1})] = \mathbb{E}[h(y_{11})]$ for any i because the cross-sectional units are exchangeable. Moreover, $\mathbb{E}[h(y_{i1})|\vartheta] = \mathbb{E}[h(y_{11})|\vartheta]$ for any i. Choose M such that $M > \mathbb{E}[|y_{11}|]$. Now consider the bound

$$\begin{aligned}
&\mathbb{I}\left\{\frac{1}{N}\sum_{i=1}^{N} M_{i,N} > \sqrt{2/\pi}M\right\} \\
&= \mathbb{I}\left\{\frac{1}{N}\sum_{i=1}^{N} \sqrt{2/\pi}|y_{i1}| > \sqrt{2/\pi}M\right\} \\
&= \mathbb{I}\left\{\frac{1}{N}\sum_{i=1}^{N} \left(|y_{i1}| - \mathbb{E}\big[|y_{11}|\,\big|\,\vartheta\big] + \mathbb{E}\big[|y_{11}|\,\big|\,\vartheta\big] - \mathbb{E}[|y_{11}|]\right) > M - \mathbb{E}[|y_{11}|]\right\}
\end{aligned}$$

Let $\tilde{M} = (M - \mathbb{E}[|y_{11}|])/2$ and write

$$\begin{aligned}
&\mathbb{I}\left\{\frac{1}{N}\sum_{i=1}^{N}\sum_{k=0}^{1} M_{i,N} > \sqrt{8/\pi}M\right\} \\
&\leq \mathbb{I}\left\{\frac{1}{N}\sum_{i=1}^{N} \left(|y_{i1}| - \mathbb{E}\big[|y_{11}|\,\big|\,\vartheta\big]\right) > \tilde{M}\right\} + \mathbb{I}\left\{\left(\mathbb{E}\big[|y_{11}|\,\big|\,\vartheta\big] - \mathbb{E}[|y_{11}|]\right) > \tilde{M}\right\}
\end{aligned}$$

We now analyze the two indicator functions separately. First,

$$\begin{aligned}
&\lim_{N\longrightarrow\infty} \mathbb{P}^{Y_{1:N,0:1},\vartheta}\left\{\frac{1}{N}\sum_{i=1}^{N} \left(|y_{i1}| - \mathbb{E}\big[|y_{11}|\,\big|\,\vartheta\big]\right) > \tilde{M}\right\} \\
&= \lim_{N\longrightarrow\infty} \mathbb{E}^{\vartheta}\left[\mathbb{P}_{\vartheta}^{Y_{1:N,0:1}}\left\{\frac{1}{N}\sum_{i=1}^{N} \left(|y_{i1}| - \mathbb{E}\big[|y_{11}|\,\big|\,\vartheta\big]\right) > \tilde{M}\right\}\right] \\
&= \mathbb{E}^{\vartheta}\left[\lim_{N\longrightarrow\infty} \mathbb{P}_{\vartheta}^{Y_{1:N,0:1}}\left\{\frac{1}{N}\sum_{i=1}^{N} \left(|y_{i1}| - \mathbb{E}\big[|y_{11}|\,\big|\,\vartheta\big]\right) > \tilde{M}\right\}\right] \\
&= 0.
\end{aligned}$$

The exchange of the limit and expectation is justified by the Dominated Convergence Theorem. Conditional on ϑ the random variables $|y_{i1}|$ are independently and identically distributed and using a weak law of large numbers for $\frac{1}{N}\sum_{i=1}^{N}|y_{i1}|$ delivers the desired result.

Second, we need to control

$$\mathbb{P}^{\vartheta}\left\{\left(\mathbb{E}\big[|y_{11}|\,\big|\,\vartheta\big] - \mathbb{E}[|y_{11}|]\right) > \tilde{M}\right\}.$$

Under our prior distribution, the random variable $\mathbb{E}\big[|y_{11}| \,\big|\, \vartheta\big]$ is stochastically bounded, which means that for any $\epsilon > 0$ we can choose a $\tilde{M}$ such that

$$\mathbb{P}^{\vartheta}\left\{\big(\mathbb{E}\big[|y_{11}| \,\big|\, \vartheta\big] - \mathbb{E}[|y_{11}|]\big) > \tilde{M}\right\} < \epsilon.$$

This delivers the desired result.

B Computational Details

B.1 Gibbs Sampling

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

- $y^*_{1:N,0:T}$ with $y_{1:N,0:T}$;
- ρ with a generalized method of moments (GMM) estimator $\hat{\rho}$, such as the orthogonal differencing in Arellano and Bover (1995) (implementation details can be found in the working paper version of Liu, Moon, and Schorfheide (2018));
- λ_i with $\hat{\lambda}_i = \frac{1}{T}\sum_{t=1}^{T}(y^*_{it} - \hat{\rho} y^*_{it-1})$;
- σ_i^2 with the variance of the GMM orthogonal differencing residues for each individual i, i.e., let $y^{\perp}_{it}, t = 1, \cdots, T-1$, denote the data after orthogonal differencing transformation, then $\hat{\sigma}_i^2 = \widehat{\mathbb{V}}_i(y^{\perp}_{it} - \hat{\rho} y^{\perp}_{it-1})$, the time-series variances of $y^{\perp}_{it} - \hat{\rho} y^{\perp}_{it-1}$;
- for $z = \lambda, \sigma$, α_z with its prior mean; $\gamma_{z,i}$ with k-means clustering where $k = 10$; $\{\Phi_k, \Sigma_k, \pi_{\lambda,k}\}_{k=1}^{K}$ and $\{\psi_k, \omega_k, \pi_{\sigma,k}\}_{k=1}^{K}$ are drawn from the conditional posteriors described in Section 3.2.

The Gibbs samplers for the other dynamic panel Tobit specifications are special cases in which some of the parameter blocks drop out. The Gibbs sampler for the pooled Tobit and linear specifications are initialized via pooled OLS, which ignores the censoring. We generate a total of $M_0 + M = 10,000$ draws using the Gibbs sampler and discard the first $M_0 = 1,000$ draws.

B.2 Set Forecasts

The HPD sets generated by the algorithms presented in this subsection always include zero and be of the form

$$C_i = \{0\} \cup \left(\bigcup_{k=1}^{K_i} [a_{ik},\, b_{ik}] \right)$$

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

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

Based on posterior draws $(\lambda_i^{(j)}, \sigma_i^{2(j)}, y_{iT}^{*(j)}, \theta^{(j)})$, we can compute the conditional mean and variances $\mu_{iT+h|T}^{(j)}$, and $\sigma_{iT+h|T}^{2(j)}$, which are the primitives for the subsequent algorithms. The conditional predictive distribution of y_{iT+h} is given by a truncated Normal of the form

$$\begin{aligned} & p(y_{iT+h}|\mu_{iT+h|T}^{(j)}, \sigma_{iT+h|T}^{2(j)}) \\ &= \Phi_N\big(-\mu_{iT+h|T}^{(j)}/\sigma_{iT+h|T}^{(j)}\big)\delta_0(y_{iT+h}) + p_N(y_{iT+h}|\mu_{iT+h|T}^{(j)}, \sigma_{iT+h|T}^{2(j)})\mathbb{I}\{y_{iT+h} > 0\}, \end{aligned} \tag{A.13}$$

where $\delta_0(y)$ is the Dirac function that is 0 for $y \neq 0$, and has the properties that $\delta_0(y) \geq 0$ and $\int \delta_0(y)dy = 1$. Using a sampler for a truncated Normal distribution, it is straightforward to generate draws from the conditional predictive density.

To construct highest posterior density (HPD) sets, we need to evaluate the posterior predictive density, integrating out $(\mu_{iT+h|T}, \sigma_{iT+h|T}^2)$ under the posterior distribution. We do so using the Monte Carlo averages

$$\pi_{i0} = \frac{1}{M}\sum_{j=1}^{M} \Phi_N\big(-\mu_{iT+h|T}^{(j)}/\sigma_{iT+h|T}^{(j)}\big) \tag{A.14}$$

$$\pi_i(y) = \frac{1}{M}\sum_{j=1}^{M} p_N\left(y \,\middle|\, \mu_{iT+h|T}^{(j)},\ \sigma_{iT+h|T}^{2(j)}\right) \tag{A.15}$$

such that

$$\pi_{i0}\delta_0(y) + \pi_i(y)\mathbb{I}\{y > 0\} \approx p(y|Y_{1:N,0:T}, X_{1:N,0:T+h}). \tag{A.16}$$

We also define the weights

$$W_i^{(j)} = 1 - \Phi_N\big(-\mu_{iT+h|T}^{(j)}/\sigma_{iT+h|T}^{(j)}\big), \tag{A.17}$$

which have the property that $\frac{1}{M}\sum_{j=1}^{M} W_i^{(j)} = 1 - \pi_{i0}$.

Algorithm for $1-\alpha$ Set Forecasts Targeting Pointwise Coverage Probability:

For $i = 1, \ldots, N$:

1. For $j = 1, \ldots, M$: compute $(\mu_{iT+h|T}^{(j)}, \sigma_{iT+h|T}^{2(j)})$ based on a draw $(\lambda_i^{(j)}, \sigma_i^{2(j)}, y_{iT}^{*(j)}, \theta^{(j)})$ from the posterior distribution.

2. Evaluate the weights $\{W_i^{(j)}\}_{j=1}^M$ in (A.17) and compute π_{i0} in (A.14).

3. If $\pi_{i0} \geq 1-\alpha$, then $C_i = \{0\}$.

4. If $\pi_{i0} < 1-\alpha$, then

   (a) Draw $\{y_{iT+h}^{(j)}\}_{j=1}^M$ from the normalized continuous part of the predictive distribution $\pi_i(y)\mathbb{I}\{y>0\}/\int \pi_i(y)\mathbb{I}\{y>0\}dy$ and form the pairs $\{(y_{iT+h}^{(j)}, W_i^{(j)})\}_{j=1}^M$.

   (b) Sort $\{(y_{iT+h}^{(j)}, W_i^{(j)})\}_{j=1}^M$ in ascending order based on $y_{iT+h}^{(j)}$.

   (c) For $j = 1, \ldots, M$: compute $\pi_i^{(j)} = \pi_i(y_{iT+h}^{(j)}) \approx p(y_{iT+h}^{(j)} | Y_{1:N,0:T}, X_{1:N,0:T+h})$ based on (A.15).

   (d) Let $\Pi_i = \{(\pi_i^{(j)}, y_{iT+h}^{(j)}, W_i^{(j)})\}_{j=1}^M$. Sort the elements in Π_i based on $\pi_i^{(j)}$ in descending order. Denote the sorted elements in Π_i by $(\pi_i^{(s)}, y_{iT+h}^{(s)}, W_i^{(s)})$.

   (e) Note that by construction $\sum_{s=1}^M W_i^{(s)} = 1 - \pi_{i0}$. Let $\bar{\Pi}_i$ be the set of largest density values:

$$\bar{\Pi}_i = \left\{ (\pi_i^{(s)}, y_{iT+h}^{(s)}, W_i^{(s)}) \;\middle|\; s = 1, \ldots, \bar{s}, \; \sum_{s=1}^{\bar{s}} W_i^{(s)} \approx (1-\alpha-\pi_{i0})M \right\}.$$

   (f) Recall that the (j) superscript refers to draws sorted according to $y_{iT+h}^{(j)}$. For $j = 1, \ldots, M$:

      i. If (A) $j = 1$ and $(\pi_i^{(j)}, y_{iT+h}^{(j)}, W_i^{(j)}) \in \bar{\Pi}_i$, OR (B) $j > 1$, $(\pi_i^{(j-1)}, y_{iT+h}^{(j-1)}, W_i^{(j-1)}) \notin \bar{\Pi}_i$, and $(\pi_i^{(j)}, y_{iT+h}^{(j)}, W_i^{(j)}) \in \bar{\Pi}_i$, then $y_{iT+h}^{(j)}$ is the start of an interval, denoted by a_{ik}, where k is an index for the intervals.

      ii. If (A) $j = M$ and $(\pi_i^{(j)}, y_{iT+h}^{(j)}, W_i^{(j)}) \in \bar{\Pi}_i$, OR (B) $j < M$, $(\pi_i^{(j)}, y_{iT+h}^{(j)}, W_i^{(j)}) \in \bar{\Pi}_i$, and $(\pi_i^{(j+1)}, y_{iT+h}^{(j+1)}, W_i^{(j+1)}) \notin \bar{\Pi}_i$, then $y_{iT+h}^{(j)}$ is the end of an interval, denoted by b_{ik}.

   This leads to K_i intervals of the form $[a_{ik}, b_{ik}]$, $k = 1, \ldots, K_i$. If $a_{i1} = y_{iT+h}^{(1)}$, then let $a_{i1} = 0$.

(g) Delete intervals that are singletons and adjust K_i accordingly. Note that K_i may be zero for some i's.

(h) In the end, unit i's set forecast takes form

$$C_{it+h|T} = \{0\} \cup \left(\bigcup_{k=1}^{K_i} [a_{ik},\, b_{ik}] \right).$$

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

1. For $i = 1, \ldots, N$:

   (a) For $j = 1, \ldots, M$: compute $(\mu^{(j)}_{iT+h|T}, \sigma^{2(j)}_{iT+h|T})$ based on a draw $(\lambda^{(j)}_i, \sigma^{2(j)}_i, y^{*(j)}_{iT}, \theta^{(j)})$ from the posterior distribution.

   (b) Evaluate the weights $\{W^{(j)}_i\}_{j=1}^M$ in (A.17) and compute π_{i0} in (A.14).

2. Define $\pi_0 = \frac{1}{N}\sum_{i=1}^N \pi_{i0}$ (average probability of zero). Note that $\frac{1}{NM}\sum_{i=1}^N \sum_{j=1}^M W^{(j)}_i = 1 - \frac{1}{N}\sum_{i=1}^N \pi_{i0} = 1 - \pi_0$.

3. If $\pi_0 \geq 1-\alpha$ then:

   (a) Sort the units i in descending order based π_{i0}.

   (b) Assign the set $\{0\}$ to the units with the largest π_{i0} values until the desired coverage is reached. All other units i are assigned $\emptyset$.

4. Elseif $\pi_0 < 1-\alpha$, then:

   (a) For $i = 1, \ldots, N$:

      i. Draw $\{y^{(j)}_{iT+h}\}_{j=1}^M$ from the normalized continuous part of the predictive distribution normalized continuous part of the predictive distribution $\pi_i(y)\mathbb{I}\{y > 0\}/\int \pi_i(y)\mathbb{I}\{y>0\}dy$ and form the pairs $\{(y^{(j)}_{iT+h}, W^{(j)}_i)\}_{j=1}^M$.

      ii. Sort $\{(y^{(j)}_{iT+h}, W^{(j)}_i)\}_{j=1}^M$ in ascending order based on $y^{(j)}_{iT+h}$.

      iii. For $j = 1, \ldots, M$: compute $\pi^{(j)}_i = \pi_i(y^{(j)}_{iT+h}) \approx p(y^{(j)}_{iT+h}|Y_{1:N,0:T}, X_{1:N,0:T+h})$ based on (A.15).

   (b) Let $\Pi = \left\{(\pi^{(j)}_i, y^{(j)}_{iT+h}, W^{(j)}_i) \mid i = 1, \ldots, N \text{ and } j = 1, \ldots, M\right\}$. Sort the elements in Π based on $\pi^{(j)}_i$ in descending order. Denote the sorted elements in Π by $(\pi^{(s)}, y^{(s)}_{T+h}, W^{(s)})$. We dropped the i subscript from the triplet, because we are pooling across i.

(c) Let $\bar{\Pi}$ be the set of largest density values:

$$\bar{\Pi} = \left\{ \left(\pi^{(s)}, y_{T+h}^{(s)}, W^{(s)}\right) \,\middle|\, s = 1, \cdots, \bar{s}, \sum_{s=1}^{\bar{s}} W^{(s)} \approx (1 - \alpha - \pi_0)NM \right\}.$$

(d) For $i = 1, \ldots, N$:

i. For $j = 1, \ldots, M$:

A. If (A) $j = 1$ and $(\pi_i^{(j)}, y_{iT+h}^{(j)}, W_i^{(j)}) \in \bar{\Pi}$, OR (B) $j > 1$, $(\pi_i^{(j-1)}, y_{iT+h}^{(j-1)}, W_i^{(j-1)}) \notin \bar{\Pi}$, and $(\pi_i^{(j)}, y_{iT+h}^{(j)}, W_i^{(j)}) \in \bar{\Pi}$, then $y_{iT+h}^{(j)}$ is the start of an interval, denoted by a_{ik}, where k is an index for the intervals.

B. If (A) $j = M$ and $(\pi_i^{(j)}, y_{iT+h}^{(j)}, W_i^{(j)}) \in \bar{\Pi}$ OR (B) $j < M$, $(\pi_i^{(j)}, y_{iT+h}^{(j)}, W_i^{(j)}) \in \bar{\Pi}$, and $(\pi_i^{(j+1)}, y_{iT+h}^{(j+1)}, W_i^{(j+1)}) \notin \bar{\Pi}$, then $y_{iT+h}^{(j)}$ is the end of an interval, denoted by b_{ik}.

This leads to K_i intervals of the form $[a_{ik}, b_{ik}]$, $k = 1, \ldots, K_i$. If $a_{i1} = y_{iT+h}^{(1)}$, then let $a_{i1} = 0$.

ii. Delete intervals that are singletons and adjust K_i accordingly. Note that K_i may be zero for some i's.

iii. In the end, unit i's set forecast takes form

$$C_i = \{0\} \cup \left(\bigcup_{k=1}^{K_i} [a_{ik},\, b_{ik}] \right).$$

B.3 Density Forecasts

The log-predictive density can be approximated by

$$\ln p\big(y_{iT+h}|Y_{1:N,0:T}\big) \approx \begin{cases} \ln \mathbb{P}\big[y_{iT+h} = 0|Y_{1:N,0:T}\big], & \text{if } y_{iT+h} = 0, \\ \ln\left(\frac{1}{M}\sum_{j=1}^{M} p_N\big(y_{iT+h}|\mu_{iT+h|T}^{(j)}, \sigma_{iT+h|T}^{2(j)}\big)\right), & \text{otherwise.} \end{cases} \tag{A.18}$$

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

$$\hat{F}(y_{iT+h}) = \frac{1}{M}\sum_{j=1}^{M} \mathbb{I}\{y_{iT+h}^{(j)} \leq y_{iT+h}\}. \tag{A.19}$$

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

$$PIT(y_{iT+h}) \approx \hat{F}(y_{iT+h}). \tag{A.20}$$

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

$$CRPS(\hat{F}, y_{iT+h}) = \int_0^\infty \big(\hat{F}(x) - \mathbb{I}\{y_{iT+h} \leq x\}\big)^2 dx. \tag{A.21}$$

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

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

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

$$CRPS(\hat{F}, y^o) = \sum_{j=2}^{M} \big[\hat{F}(y^{(j-1)}) - 0\big]^2 (y^{(j)} - y^{(j-1)}) + \big[1 - 0\big]^2 (y^o - y^{(M)}). \tag{A.22}$$

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

$$CRPS(\hat{F}, y^o) = \big[0 - 1\big]^2 (y^{(1)} - y^o) + \sum_{j=2}^{M} \big[\hat{F}(y^{(j-1)}) - 1\big]^2 (y^{(j)} - y^{(j-1)}). \tag{A.23}$$

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

$$\begin{aligned} & CRPS(\hat{F}, y^o) \\ = & \sum_{j=2}^{M_*} \big[\hat{F}(y^{(j-1)}) - 0\big]^2 (y^{(j)} - y^{(j-1)}) + \big[\hat{F}(y^{(M_*)}) - 0\big]^2 (y^o - y^{(M_*)}) \\ & + \big[\hat{F}(y^{(M_*)}) - 1\big]^2 (y^{(M_*+1)} - y^o) + \sum_{j=M_*+2}^{M} \big[\hat{F}(y^{(j-1)}) - 1\big]^2 (y^{(j)} - y^{(j-1)}). \end{aligned} \tag{A.24}$$

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

$$CRPS(\hat{F}, y^o) = \frac{1}{M} \sum_{j=1}^{M} |y^{(j)} - y^o| - \frac{1}{M^2} \sum_{1 \leq i < j \leq M} (y^{(j)} - y^{(i)}). \tag{A.25}$$

To see their equivalence, note that (A.25) can be re-written as follows:

$$\begin{aligned}
&\frac{1}{M}\sum_{j=1}^{M}|y^{(j)}-y^o| - \frac{1}{M^2}\sum_{1\le i<j\le M}(y^{(j)}-y^{(i)}) \qquad (A.26)\\
&= \frac{1}{M}\left[\sum_{j>M_*} y^{(j)} - \sum_{j\le M_*} y^{(j)} + \big(M_* - (M-M_*)\big)y^o\right] - \frac{1}{M^2}\sum_{j=1}^{M}(2j-M-1)y^{(j)}.\\
&= \frac{1}{M^2}\left[-\sum_{j=1}^{M_*}(2j-1)y^{(j)} + \sum_{j=M_*+1}^{M}(2M-2j+1)y^{(j)}\right] + \frac{2M_*-M}{M}y^o.
\end{aligned}$$

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

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

First, let us look at the more general Case 3. After replacing $\hat{F}(y^{(j)})$, the RHS of (A.24) becomes

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

which is the same as (A.26). Similarly, for Case 1, after substituting $\hat{F}$, the RHS of (A.22)

becomes

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

which is equal to (A.26) when $M_* = M$. And for Case 2, after substituting $\hat{F}$, the RHS of (A.23) becomes

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

which is equal to (A.26) when $M_* = 0$.

C Additional Simulation Results

To examine the sensitivity of the MCMC algorithm to the fraction of zeros in the sample, we changed the design of the Monte Carlo experiment to raise the fraction of zeros. Recall from Table 1 in the main text that

$$\text{Fraction of zeros} = 45\% \quad : \quad p(\lambda_i|y_{i0}^*) = \frac{1}{9}p_N\left(\lambda_i|2.25, 0.5\right) + \frac{8}{9}p_N\left(\lambda_i|0, 0.5\right).$$

To increase the number of zeros to 60% and 75%, respectively, we consider

$$\begin{aligned}\text{Fraction of zeros} = 60\% \quad &: \quad p(\lambda_i|y_{i0}^*) = \frac{1}{9}p_N\left(\lambda_i|1.85, 0.5\right) + \frac{8}{9}p_N\left(\lambda_i|-0.4, 0.5\right)\\ \text{Fraction of zeros} = 75\% \quad &: \quad p(\lambda_i|y_{i0}^*) = \frac{1}{9}p_N\left(\lambda_i|1.3, 0.5\right) + \frac{8}{9}p_N\left(\lambda_i|-0.95, 0.5\right).\end{aligned}$$

Under the baseline configuration, the number fraction of trajectories with all zeros was 15%. Under the alternative scenarios, this fraction increases to 23% and 34%, respectively.

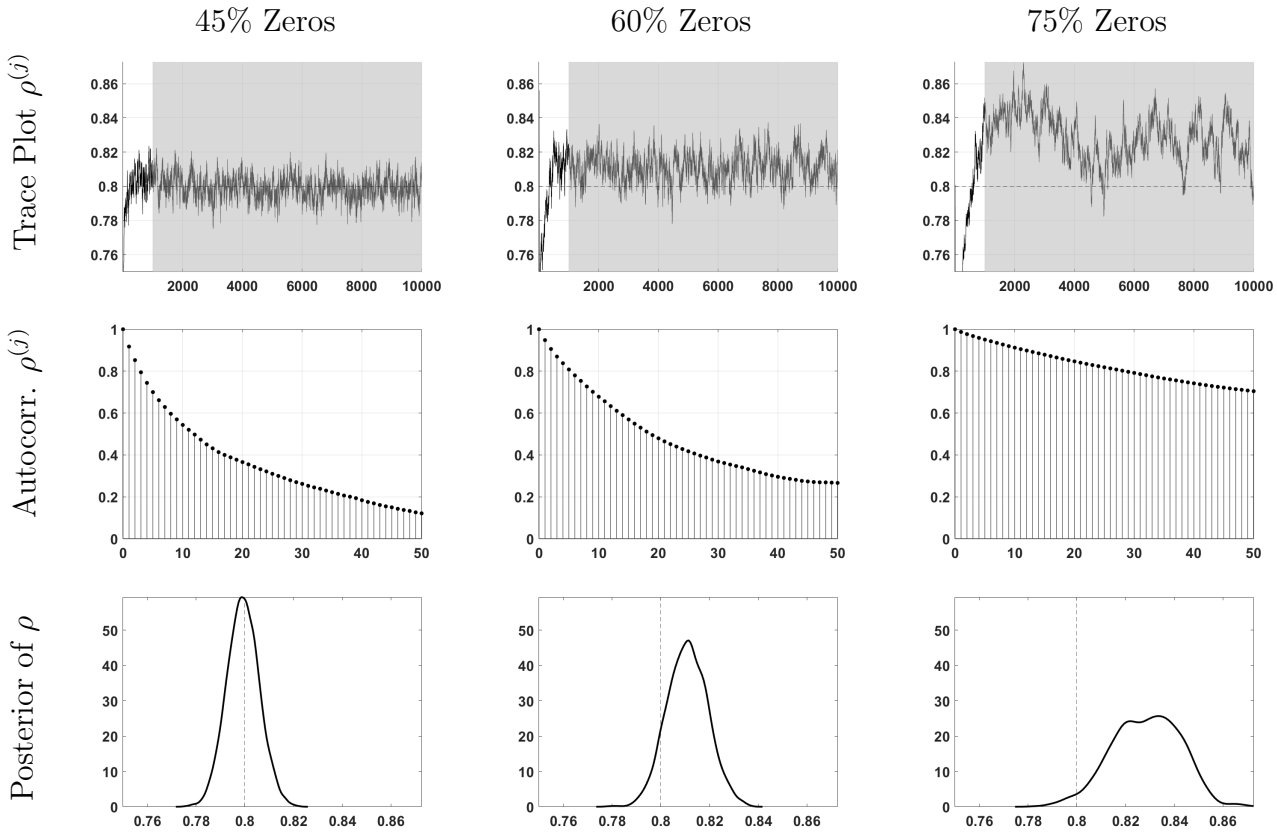
Under the alternative DGPs, the MCMC remains stable, despite the larger number of zeros in the samples. In Figure A-1 we show some convergence diagnostics based on the sequence of draws $\rho^{(j)}$. The first row contains trace plots, the second row autocorrelation functions, and the third row posterior density estimates. As the number of zeros increases, the chain becomes more persistent and the spread of the posterior increases because ρ is effectively estimated from fewer observations. Nonetheless, the algorithm remains well behaved. While 75% appears to be a large fraction, notice that the sample size is $T \cdot N = 10,000$. Thus, we still have 2,500 non-zero observations.[23]

Table A-1 reproduces and extends the results reported in Table 3 of the main text. The overall message from the baseline MC design is preserved under the alternative specifications of the DGP. The forecasts get more precise as we increase the fraction of zeros. The more zeros in the sample and the longer the zero spells, the stronger the evidence that the next observation will also be a zero. In fact, under all three designs, 100% of the units with all-zero observations assign a probability of no less than 95% to $y_{iT+1} = 0$.[24] This improves the density forecasts (lower LPS and CRPS) and shortens the predictive sets. The downside of more zeros is that the estimation of the homogeneous parameter ρ becomes more difficult.

[23] We also tried a design with 95% zeros. Not surprisingly, we experienced convergence problems for this rather extreme design.

[24] For units with all zeros, the chance of predicting zeros is large in practice, though in principle, these units still convey a slight amount of information about the common parameters and the left tail of the underlying distribution of cross-sectional heterogeneity.

Figure A-1: Convergence Diagnostics Based on $\rho^{(j)}$ Sequence



Notes: The dashed horizontal lines in the first row and the dashed vertical lines in the last row indicate the "true" value of $\rho = 0.8$. (j) in the superscript indicates the MCMC draws. The first 1,000 draws are discarded as burn-in, so the shaded regions in the first row indicate the MCMC draws kept for posterior analyses.

Both bias and standard deviation of $\hat{\rho}$ across Monte Carlo repetitions increase which is mirrored in the shape of the posterior depicted in the last row of Figure A-1. As mentioned before, this is plausible: the fewer non-zero observations, the less information about ρ is in the sample.

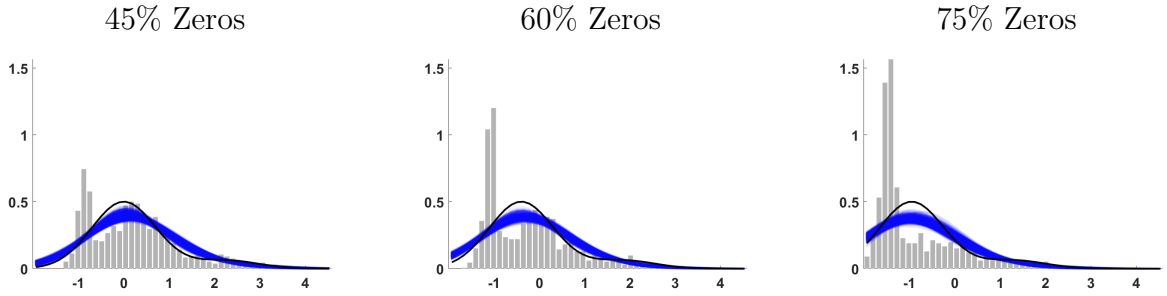
In Figure A-2 we plot the cross-sectional distribution of posterior means of λ_i as well as the estimated and "true" RE distribution. The left panel of the figure reproduces the left panel of Figure 1 in the main paper. By construction, the "true" distribution of the λ_is shifts to the left for the other two designs (center and right panel of Figure A-2). The spike in the empirical distribution of $\mathbb{E}[\lambda_i|Y_{1:N,0:T}]$ shifts to the left and increases in height because the estimated model needs to reproduce the number of zeros in the sample, which is done by lower estimates for λ_i. We are using a proper prior for the RE distribution to reduce the

Table A-1: Monte Carlo Experiment: Set Forecast and Density Forecast Performance

	Density Forecast		Set Forecast "Average"		Set Forecast "Pointwise"		Estimates	
	LPS	CRPS	Cov.	Length	Cov.	Length	Bias($\hat{\rho}$)	StdD($\hat{\rho}$)
Fraction of Zeros in Panel is 45% (From Paper)								
Flexible & Heterosk.	-0.757	0.277	0.910	1.260	0.933	1.503	-0.002	0.005
Normal & Heterosk.	-0.758	0.277	0.908	1.248	0.932	1.498	-0.006	0.005
Flexible & Homosk.	-0.902	0.294	0.929	1.506	0.942	1.698	0.007	0.008
Normal & Homosk.	-0.903	0.294	0.929	1.501	0.942	1.699	0.001	0.007
Fraction of Zeros in Panel is 60%								
Flexible & Heterosk.	-0.552	0.194	0.909	0.706	0.948	1.023	0.005	0.006
Normal & Heterosk.	-0.553	0.194	0.908	0.702	0.948	1.024	0.001	0.006
Flexible & Homosk.	-0.655	0.206	0.931	0.878	0.955	1.162	0.012	0.009
Normal & Homosk.	-0.656	0.207	0.931	0.880	0.956	1.169	0.009	0.009
Fraction of Zeros in Panel is 75%								
Flexible & Heterosk.	-0.316	0.109	0.909	0.219	0.970	0.567	0.015	0.009
Normal & Heterosk.	-0.316	0.109	0.909	0.220	0.971	0.571	0.013	0.009
Flexible & Homosk.	-0.375	0.117	0.931	0.310	0.974	0.660	0.020	0.012
Normal & Homosk.	-0.376	0.117	0.932	0.315	0.975	0.668	0.022	0.013

Notes: "Cov." is coverage frequency and "Length" is an average across i.

Figure A-2: Posterior Means and Estimated RE Distributions for λ_i, Flexible & Heterosk. Specification



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

chance that draws of λ_i take very large negative values. This contributes to the stability of the MCMC.

D Data Set

Charge-off rates. The raw data are obtained from the website of the *Federal Reserve Bank of Chicago*.[25] The raw data are available at a quarterly frequency. The charge-off rates are defined as charge-offs divided by the stock of loans and constructed in a similar manner as in Tables A-1 and A-2 of Covas, Rump, and Zakrajsek (2014). However, the construction differs in the following dimensions: (i) We focus on charge-off rates instead of net charge-off rates. (ii) We divide the charge-offs by the lagged stock of loans instead of the current stock of loans to reduce the timing issue.[26] (iii) For banks with domestic offices only (Form FFIEC 041), RIAD4645 (numerator for commercial and industrial loans) is not reported, so we switch to its domestic counterpart, RIAD4638.

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

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

We focus on "small" banks and relate the charge-off rates to local economic conditions. We include a bank in the sample if its assets are below one billion dollars. The raw data set contains missing observations and outliers that we are unable to explain with our econometric

[25] https://www.chicagofed.org/banking/financial-institution-reports/commercial-bank-data

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

model. Thus, we proceed as follows to select a subset of observations from the raw data. For each rolling sample:

1. Eliminate banks for which domestic total assets are missing for all time periods in the sample.

2. Compute average non-missing domestic total assets and eliminate banks with average assets above 1 billion dollars.

3. For each loan category, eliminate banks for which the target charge-off rate is missing for at least one period of the sample.

4. For each loan category, eliminate banks for which the target charge-off rate is negative or greater than 400% for at least one period of the sample.

5. For each loan category proceed as follows: First, for each bank, drop the two largest observations $y_{it},\ t = 0, \cdots, T+1$, and calculate the standard deviation (stdd) of the remaining observations. Then, eliminate a bank if any successive change $|y_{it} - y_{it-1}| + |y_{it+1} - y_{it}| > 10\text{stdd}$. For $t = 0$ and $t = T+1$, we only have one of the two terms and we set the other term in this selection criterion to zero.

The remaining sample sizes after each of these steps as well as some summary statistics for loan charge-off rates are reported in Table A-2.

Table A-2: Sample Sizes After Selection Steps and Summary Statistics for Charge-Off Rates

		Sample Sizes						Cross-sectional Statistics			
Loan	t_0	Initial	Step1	Step2	Step3	Step4	Step5	% 0s	Mean	75%	Max
CLD	2007Q3	7,903	7,903	7,299	3,290	3,146	1,304	77	1.5	0.0	106.8
CLD	2007Q4	7,835	7,835	7,219	3,244	3,088	1,264	74	1.9	0.1	106.8
CLD	2008Q1	7,692	7,692	7,084	3,204	3,032	1,257	71	2.2	0.5	180.2
RRE	2007Q1	7,991	7,991	7,393	6,260	5,993	2,654	77	0.2	0.0	33.1
RRE	2007Q2	7,993	7,993	7,383	6,152	5,894	2,576	76	0.3	0.0	33.1
RRE	2007Q3	7,903	7,903	7,299	6,193	5,920	2,606	73	0.3	0.0	35.9
RRE	2007Q4	7,835	7,835	7,219	6,146	5,859	2,581	70	0.4	0.1	69.2
RRE	2008Q1	7,692	7,692	7,084	6,106	5,792	2,561	68	0.4	0.2	45.6
RRE	2008Q2	7,701	7,701	7,080	6,029	5,721	2,492	67	0.4	0.2	63.6
RRE	2008Q3	7,631	7,631	7,008	6,052	5,743	2,577	65	0.5	0.3	39.2
RRE	2008Q4	7,559	7,559	6,938	6,005	5,679	2,600	63	0.5	0.3	45.6
RRE	2009Q1	7,480	7,480	6,849	5,971	5,634	2,588	62	0.5	0.3	45.0
RRE	2009Q2	8,103	8,103	7,381	5,895	5,564	2,536	62	0.5	0.3	45.0
RRE	2009Q3	8,016	8,016	7,302	5,899	5,568	2,563	61	0.5	0.4	47.6
RRE	2009Q4	7,940	7,940	7,229	5,846	5,508	2,553	60	0.5	0.4	45.0
RRE	2010Q1	7,770	7,770	7,077	5,765	5,426	2,494	61	0.5	0.4	45.0
RRE	2010Q2	7,770	7,770	7,072	5,635	5,308	2,420	61	0.5	0.4	45.0
RRE	2010Q3	7,707	7,707	7,013	5,632	5,298	2,441	61	0.5	0.4	45.6
RRE	2010Q4	7,608	7,608	6,910	5,583	5,255	2,443	61	0.5	0.3	38.2
RRE	2011Q1	7,469	7,469	6,784	5,520	5,220	2,437	62	0.4	0.3	38.2
RRE	2011Q2	7,472	7,472	6,783	5,398	5,110	2,385	62	0.4	0.3	38.2
RRE	2011Q3	7,402	7,402	6,716	5,395	5,110	2,397	64	0.4	0.2	38.2
RRE	2011Q4	7,334	7,334	6,649	5,341	5,059	2,395	65	0.3	0.2	38.2
RRE	2012Q1	7,236	7,236	6,546	5,284	5,008	2,349	67	0.3	0.2	38.2
RRE	2012Q2	7,234	7,234	6,534	5,584	5,283	2,430	66	0.3	0.2	38.2
RRE	2012Q3	7,170	7,170	6,465	5,576	5,267	2,416	67	0.2	0.1	28.4
RRE	2012Q4	7,073	7,073	6,358	5,495	5,197	2,362	69	0.2	0.1	22.2
RRE	2013Q1	6,931	6,849	6,212	5,420	5,121	2,341	71	0.2	0.1	28.7
RRE	2013Q2	6,934	6,857	6,200	5,296	5,008	2,298	71	0.2	0.1	28.7
RRE	2013Q3	6,884	6,807	6,144	5,291	4,999	2,307	72	0.2	0.0	28.7
RRE	2013Q4	6,803	6,726	6,061	5,212	4,932	2,271	74	0.1	0.0	28.7
RRE	2014Q1	6,650	6,576	5,913	5,144	4,870	2,258	75	0.1	0.0	27.2
RRE	2014Q2	6,650	6,578	5,897	5,012	4,746	2,190	76	0.1	0.0	16.9
RRE	2014Q3	6,582	6,510	5,821	5,004	4,742	2,178	77	0.1	0.0	22.2
RRE	2014Q4	6,502	6,431	5,729	4,945	4,691	2,210	78	0.1	0.0	16.9
RRE	2015Q1	6,342	6,271	5,564	4,874	4,611	2,200	79	0.1	0.0	11.1
RRE	2015Q2	6,348	6,278	5,560	4,751	4,500	2,134	79	0.1	0.0	11.1
CC	2001Q2	9,031	9,031	8,532	1,691	1,540	875	33	3.4	4.7	162.5
CC	2001Q3	8,995	8,995	8,491	1,666	1,515	844	33	3.4	4.8	88.9
CC	2001Q4	8,887	8,887	8,382	1,636	1,489	836	34	3.3	4.6	88.9

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

Table A-2: Sample Sizes After Selection Steps and Summary Statistics for Charge-Off Rates (cont.)

		Sample Sizes						Cross-sectional Statistics			
Loan	t_0	Initial	Step1	Step2	Step3	Step4	Step5	% 0s	Mean	75%	Max
CC	2002Q1	8,723	8,723	8,228	1,612	1,466	814	35	3.3	4.4	400.0
CC	2002Q2	8,823	8,823	8,312	1,670	1,519	817	38	3.2	4.3	88.9
CC	2002Q3	8,805	8,805	8,286	1,631	1,488	821	38	3.2	4.3	88.9
CC	2002Q4	8,728	8,728	8,199	1,606	1,468	813	39	3.1	4.1	88.9
CC	2003Q1	8,611	8,611	8,077	1,573	1,445	811	40	3.0	4.0	128.5
CC	2003Q2	8,754	8,754	8,203	1,544	1,422	787	40	3.0	3.9	136.1
CC	2003Q3	8,755	8,755	8,198	1,513	1,395	754	41	2.9	3.8	136.1
CC	2003Q4	8,671	8,671	8,120	1,500	1,387	724	42	2.8	3.6	136.1
CC	2004Q1	8,526	8,526	7,989	1,468	1,355	707	43	2.7	3.6	136.1
CC	2004Q2	8,662	8,662	8,108	1,440	1,331	677	42	2.8	3.6	136.1
CC	2004Q3	8,626	8,626	8,067	1,411	1,308	664	43	2.7	3.5	136.1
CC	2004Q4	8,552	8,552	7,989	1,391	1,284	657	44	2.6	3.3	140.9
CC	2005Q1	8,384	8,384	7,829	1,369	1,271	639	44	2.5	3.2	151.3
CC	2005Q2	8,507	8,507	7,938	1,332	1,236	611	44	2.6	3.2	175.0
CC	2005Q3	8,482	8,482	7,897	1,315	1,218	596	45	2.6	3.2	175.0
CC	2005Q4	8,404	8,404	7,816	1,290	1,203	604	46	2.6	3.2	210.5
CC	2006Q1	8,263	8,263	7,674	1,275	1,188	614	47	2.6	3.1	175.0
CC	2006Q2	8,307	8,307	7,708	1,247	1,164	594	47	2.7	3.2	269.2
CC	2006Q3	8,240	8,240	7,639	1,231	1,156	594	46	2.8	3.4	269.2
CC	2006Q4	8,137	8,137	7,537	1,211	1,139	595	45	3.0	3.6	269.2
CC	2007Q1	7,991	7,991	7,393	1,197	1,129	574	44	3.2	3.9	269.2
CC	2007Q2	7,993	7,993	7,383	1,173	1,107	561	43	3.3	4.1	269.2
CC	2007Q3	7,903	7,903	7,299	1,159	1,091	544	44	3.2	4.2	175.0
CC	2007Q4	7,835	7,835	7,219	1,133	1,066	534	43	3.3	4.2	175.0
CC	2008Q1	7,692	7,692	7,084	1,123	1,056	527	44	3.3	4.2	175.0
CC	2008Q2	7,701	7,701	7,080	1,101	1,035	512	45	3.2	4.1	158.3
CC	2008Q3	7,631	7,631	7,008	1,096	1,036	509	44	3.1	4.0	158.3
CC	2008Q4	7,559	7,559	6,938	1,082	1,020	506	45	3.1	3.9	149.4
CC	2009Q1	7,480	7,480	6,849	1,059	999	498	46	3.0	3.7	147.3
CC	2009Q2	8,103	8,103	7,381	1,045	989	492	45	2.8	3.7	78.5
CC	2009Q3	8,016	8,016	7,302	1,042	988	492	47	2.7	3.5	77.6
CC	2009Q4	7,940	7,940	7,229	1,032	978	479	49	2.7	3.3	400.0
CC	2010Q1	7,770	7,770	7,077	1,020	963	459	49	2.5	3.2	100.0
CC	2010Q2	7,770	7,770	7,072	997	940	454	50	2.3	3.0	62.0
CC	2010Q3	7,707	7,707	7,013	994	940	450	50	2.2	2.8	62.0
CC	2010Q4	7,608	7,608	6,910	976	920	454	51	2.1	2.6	56.3
CC	2011Q1	7,469	7,469	6,784	961	906	451	52	2.0	2.5	68.6
CC	2011Q2	7,472	7,472	6,783	941	889	450	53	1.9	2.4	67.9
CC	2011Q3	7,402	7,402	6,716	933	879	443	54	1.9	2.3	67.9
CC	2011Q4	7,334	7,334	6,649	920	869	430	55	1.8	2.2	67.9

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

Table A-2: Sample Sizes After Selection Steps and Summary Statistics for Charge-Off Rates (cont.)

		Sample Sizes						Cross-sectional Statistics			
Loan	t_0	Initial	Step1	Step2	Step3	Step4	Step5	% 0s	Mean	75%	Max
CC	2012Q1	7,236	7,236	6,546	913	862	438	56	1.7	2.1	67.9
CC	2012Q2	7,234	7,234	6,534	916	862	430	54	1.8	2.2	67.9
CC	2012Q3	7,170	7,170	6,465	907	853	409	55	1.7	2.1	67.9
CON	2009Q2	8,103	8,103	7,381	5,837	5,698	2,600	77	0.4	0.0	77.4
CON	2009Q3	8,016	8,016	7,302	5,872	5,693	2,672	71	0.5	0.2	202.2
CON	2009Q4	7,940	7,940	7,229	5,814	5,584	2,723	65	0.5	0.5	202.2
CON	2010Q1	7,770	7,770	7,077	5,735	5,461	2,680	58	0.7	0.7	202.2
CON	2010Q2	7,770	7,770	7,072	5,602	5,339	2,600	53	0.7	0.8	202.2
CON	2010Q3	7,707	7,707	7,013	5,596	5,311	2,555	47	0.8	0.9	202.2
CON	2010Q4	7,608	7,608	6,910	5,545	5,227	2,473	42	0.9	1.0	202.2
CON	2011Q1	7,469	7,469	6,784	5,482	5,133	2,427	36	1.0	1.1	202.2
CON	2011Q2	7,472	7,472	6,783	5,361	5,026	2,328	37	1.0	1.1	202.2
CON	2011Q3	7,402	7,402	6,716	5,377	5,028	2,333	38	1.0	1.1	202.2
CON	2011Q4	7,334	7,334	6,649	5,324	4,979	2,377	38	0.9	1.0	202.2
CON	2012Q1	7,236	7,236	6,546	5,266	4,932	2,403	39	0.9	1.0	202.2
CON	2012Q2	7,234	7,234	6,534	5,544	5,195	2,530	42	0.8	1.0	76.0
CON	2012Q3	7,170	7,170	6,465	5,536	5,184	2,541	43	0.8	0.9	76.0
CON	2012Q4	7,073	7,073	6,358	5,457	5,117	2,526	43	0.8	0.9	44.7
CON	2013Q1	6,931	6,849	6,212	5,379	5,042	2,548	44	0.8	0.9	100.0
CON	2013Q2	6,934	6,857	6,200	5,254	4,932	2,465	43	0.8	0.9	100.0
CON	2013Q3	6,884	6,807	6,144	5,246	4,917	2,512	44	0.7	0.9	76.0
CON	2013Q4	6,803	6,726	6,061	5,165	4,843	2,470	44	0.7	0.9	76.0
CON	2014Q1	6,650	6,576	5,913	5,094	4,767	2,415	44	0.7	0.9	76.0
CON	2014Q2	6,650	6,578	5,897	4,969	4,651	2,326	44	0.7	0.9	35.7
CON	2014Q3	6,582	6,510	5,821	4,954	4,638	2,297	43	0.7	0.9	35.7
CON	2014Q4	6,502	6,431	5,729	4,894	4,585	2,324	43	0.7	0.9	76.0
CON	2015Q1	6,342	6,271	5,564	4,827	4,515	2,298	43	0.7	0.9	35.7
CON	2015Q2	6,348	6,278	5,560	4,704	4,406	2,214	43	0.7	0.9	52.9
CON	2015Q3	6,271	6,204	5,479	4,689	4,402	2,209	43	0.7	0.9	52.9
CON	2015Q4	6,183	6,117	5,395	4,625	4,337	2,222	42	0.8	0.9	113.8
CON	2016Q1	6,059	5,993	5,256	4,538	4,252	2,179	43	0.7	0.9	52.9

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

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

Unemployment Rate ($\mathbf{UR}_{it}$). Obtained from the *Bureau of Labor Statistics*. We use seasonally adjusted monthly data, time-aggregated to quarterly frequency by simple averaging.

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

Personal Income ($\mathbf{INC}_{it}$). Raw data are obtained from the *Bureau of Labor Statistics*. All quarterly series are seasonally adjusted. We first construct quarterly state-level personal income per capita, which is only available after 2010Q1. Before 2010Q1, there is no quarterly state-level population series available. We interpolate the annual population to quarterly frequency by assuming constant population growth rate within a year, and then divide the quarterly personal income by the imputed quarterly population.[27] Then, we deflate the personal income per capita by the personal consumption expenditure price index.

Geo Coding. The annual *Summary of Deposits* data from the *Federal Deposit Insurance Corporation* also contains the state and county FIPS code associated with the headquarter location of each bank. Based on this information we can link the banks to counties and compute average forecasts for each county which are displayed in Figure 3 in the main text.

Bank Characteristics. Quarterly raw data are obtained from the website of the *Federal Reserve Bank of Chicago* (see above). We construct bank-characteristics variables as follows:

- Log Assets = log(RCON2170);
- Loan Fraction = specific loan stock / sum of all loan stocks;
- Capital-To-Asset Ratio= RCON3210/RCON2170;
- Loan-To-Asset Ratio = RCON3360/RCON2170;
- ALLL-To-Loan Ratio = RCON3123/RCON3360;
- Diversification = RIAD4079/(RIAD4079+RIAD4107);

[27] To check whether this interpolation is reasonable, we also experimented with the same interpolation after 2010Q1, and resulting time series are comparable to the available data.

Table A-3: Summary Statistics for Bank Characteristics, RRE and CC 2007Q2

	RRE Charge-Offs						CC Charge-Offs					
	Low		High		All		Low		High		All	
	Mean	StdD	Mean	StdD	Mean	StdD	Mean	StdD	Mean	StdD	Mean	StdD
LogAssets	11.607	0.865	12.427	0.719	12.088	0.880	12.105	0.726	12.501	0.674	12.440	0.696
LoanFraction	0.193	0.163	0.285	0.154	0.247	0.164	0.001	0.002	0.013	0.040	0.012	0.037
Capital-Asset	0.104	0.037	0.095	0.023	0.099	0.030	0.099	0.040	0.095	0.021	0.095	0.025
Loan-Asset	0.642	0.147	0.712	0.099	0.683	0.126	0.699	0.102	0.684	0.103	0.686	0.103
ALLL-Loan	0.013	0.005	0.012	0.005	0.013	0.005	0.012	0.007	0.013	0.006	0.013	0.006
Diversification	0.099	0.093	0.099	0.178	0.099	0.149	0.102	0.054	0.126	0.081	0.122	0.078
Ret.onAssets	0.003	0.003	0.003	0.002	0.003	0.003	0.003	0.002	0.003	0.002	0.003	0.002
OCA	0.008	0.003	0.008	0.003	0.008	0.003	0.008	0.002	0.008	0.003	0.008	0.003
Sample Size	*515*		*731*		*1246*		*61*		*333*		*394*	

Notes: Bank characteristics are the values observed at 2007Q2 ($t = 0$). 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 red and blue dots in Figure 5. The samples sizes of the "All" groups are smaller than those in Table 4 because the regression samples in Table 6 only include banks with a full set of covariates.

- Return on Assets = RIAD4340/RCON2170;
- Overhead Costs-To-Asset Ratio (OCA) = RIAD4093/RCON2170.

The unit of the balance sheet variables is thousand dollars. Except for log assets and loan fraction, the variables are similar to Ghosh (2017). The RIAD variables are year-to-date, so we take differences to obtain quarterly data. The RCON variables are stocks quantities, so we use lagged values instead of current values to overcome the timing issue in ratios. The regressions in Table 6 are based on period $t = 0$ bank characteristics, to reduce concerns about simultaneity. Summary statistics for the variables are provided in Table A-3.

E Additional Empirical Results

E.1 Tuning the CRE Prior

In order to tune the prior for the CRE distribution, we recommend visualizing certain characteristics of these distributions, such as moments and number of modes. In this subsection we consider two choices of the tuning constants, summarized in Table A-4. We refer to the first choice of τ as "initial," and the second choice as "adjusted," based on the examination of the prior and posterior distribution resulting from the "initial" choice of τ.

Table A-4: Tuning Constants for Prior Distribution, CC Sample

	τ_θ	τ_ν	τ_ϕ	τ_σ^λ	τ_σ^y
Initial	5.0	1.0	5.0	1.0	1.0
Adjusted	5.0	1.0	20.0	1.0	4.0

For each draw of the hyperparameter vector ξ from either the prior or posterior distribution, one can evaluate the moments of the CRE distribution, which is a mixture of Normals. The evaluation of a moment maps an infinite-dimensional object into a one-dimensional object whose distribution can be more easily visualized. Features of the prior for the CRE distribution for the CC sample are summarized in Figure A-3. To generate the figure, we need to choose values for the regressors x_{it}. Recall that the regressors are standardized to have mean zero and variance one. We set $x_{it} = \tilde{x}_{it} = \kappa[1, 1]$ and choose κ such that $\tilde{x}_{it}$ lies on the boundary of a 50% coverage set constructed from a bivariate Normal distribution with mean zero, variances one, and a correlation that matches the correlation of x_{it} in the sample.

The dots in the scatter plots of the first three rows of Figure A-3 represent moments of the marginal distribution of y_{i0}^*. The initial prior covers a wide range of distributions: the mean can range from -10 to 10, the standard deviation from close to 0 to 7, the distributions can be left-skewed or right-skewed, they may have a kurtosis similar to a Normal distribution or they may be very fat-tailed. The fourth row of the figure shows scatter plots of the correlation between λ_i and y_{i0}, which can range from -1 to 1.

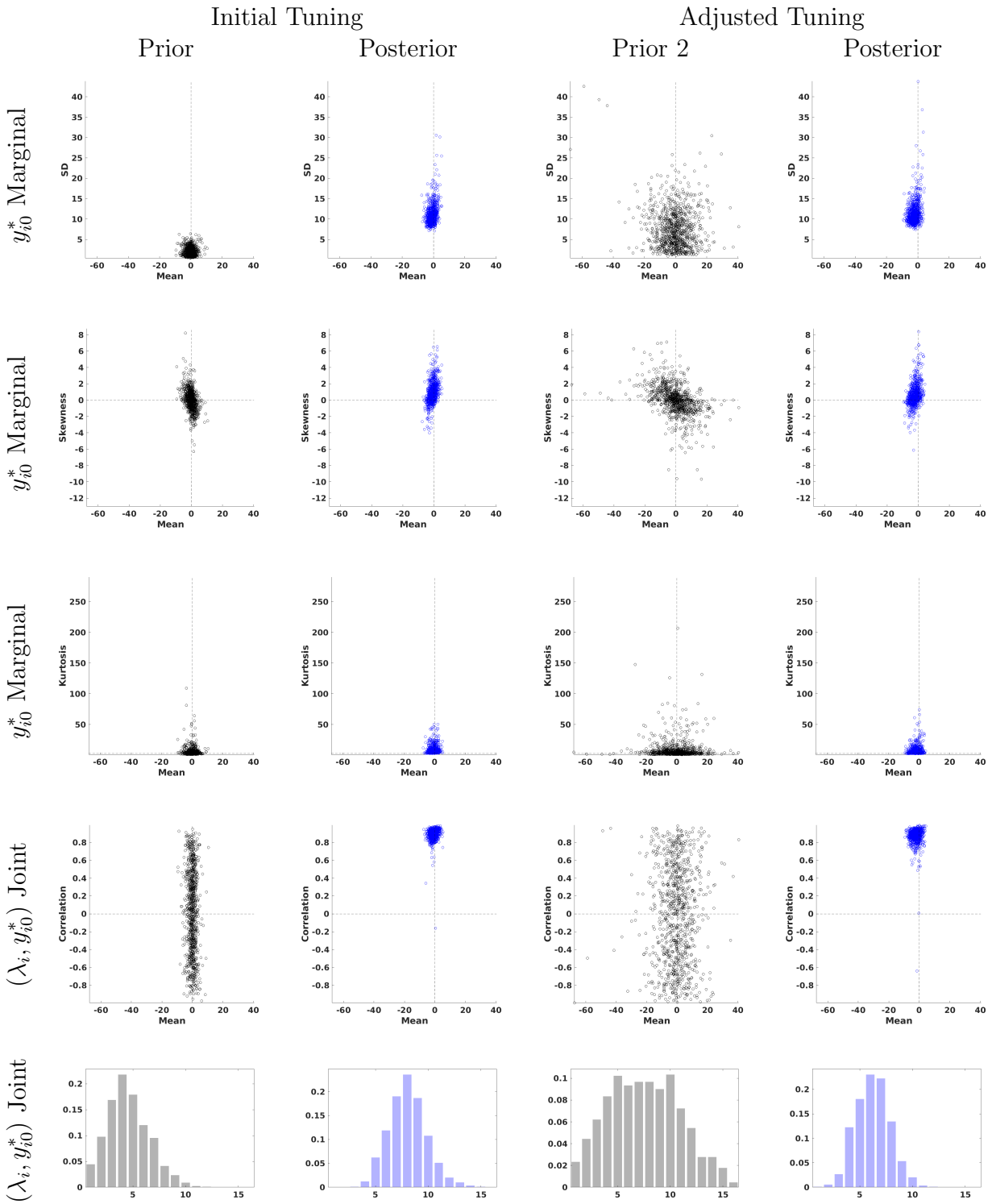
A comparison of the prior and posterior plots under the initial tuning of the prior raises two concerns. First, the posterior location and scale of the distribution of means appear to be very similar to the prior location and scale. This could mean that the likelihood function does not contain any information about the mean of the distribution of y_{i0}^*. Second, the posterior location of the distribution of standard deviations appears to be very different from the prior location. Moreover, the posterior seems to be more spread out than the prior. Thus, in this particular dimension the prior seems to assign essentially no mass in an area of the parameter space that is favored by the likelihood function which could bias the posterior estimates in a way that may not be intended by the researcher.

In view of these findings, we modify the choice of τ by raising τ_ϕ from 5 to 20 and τ_σ^y from 1 to 4. A comparison of the first and third columns of Figure A-3 indicates that the change in τ has the desired effect: the distribution of moments exhibits a larger variance.

We proceed by computing the posterior distribution for the adjusted prior. Now the prior of the means is substantially more diffuse than the posterior of the means, and the posterior of the standard deviation does no longer lie in the far tail of the prior distribution. Comparing the posterior under the initial prior to the posterior under the adjusted prior, we find that the location of the posterior distributions is quite similar. The variance of the posterior increases slightly after the adjustment of τ, but much less than the variance of the prior, so we see that the posterior is anchored by the information in the likelihood function.

The last row of Figure A-3 shows histograms of the number of modes of the CRE distribution. Recall that the CRE distribution is a mixture of Normal distribution with $K = 20$ components. This means that it could have up to 20 modes. Under the initial tuning, the prior distribution assigns probability close to one to the number of modes being between 0 to 11. The highest probability mass is associated with 4 modes. The posterior has a similar scale as the prior but is shifted to the right and peaks at 8 modes. Under the adjusted tuning, the prior distribution is more spread out which makes the posterior appear to be more concentrated relative to the prior.

Figure A-3: Prior and Posterior for CRE Distribution, CC

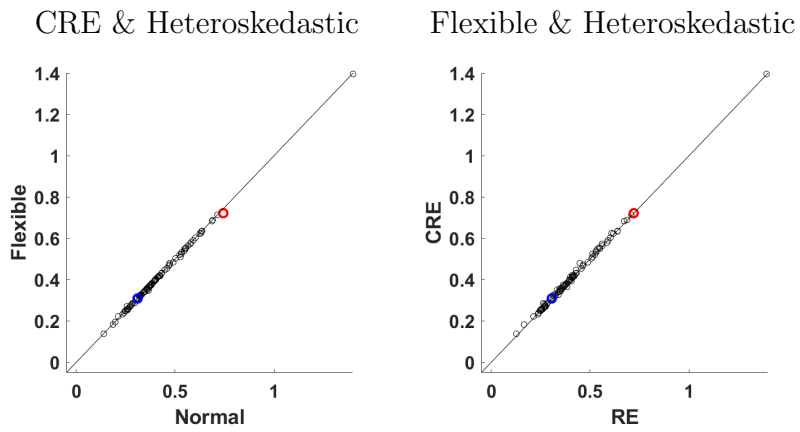


Notes: The dots in the scatter plots in rows 1 to 4 correspond to draws from the prior or posterior distribution of the hyperparameter ξ that indexes the CRE distribution. For each ξ draw we compute the implied moments of the mixture of Normals CRE distribution. SD is the standard deviation and Correlation is the correlation between λ_i and y_{i0}^*. The last row show the distribution of the number of modes.

E.2 Density Forecasts

Figure A-4 resembles Figure 2 in the main text and shows that accuracy differentials of Normal versus flexible CREs and of CREs versus REs are small.

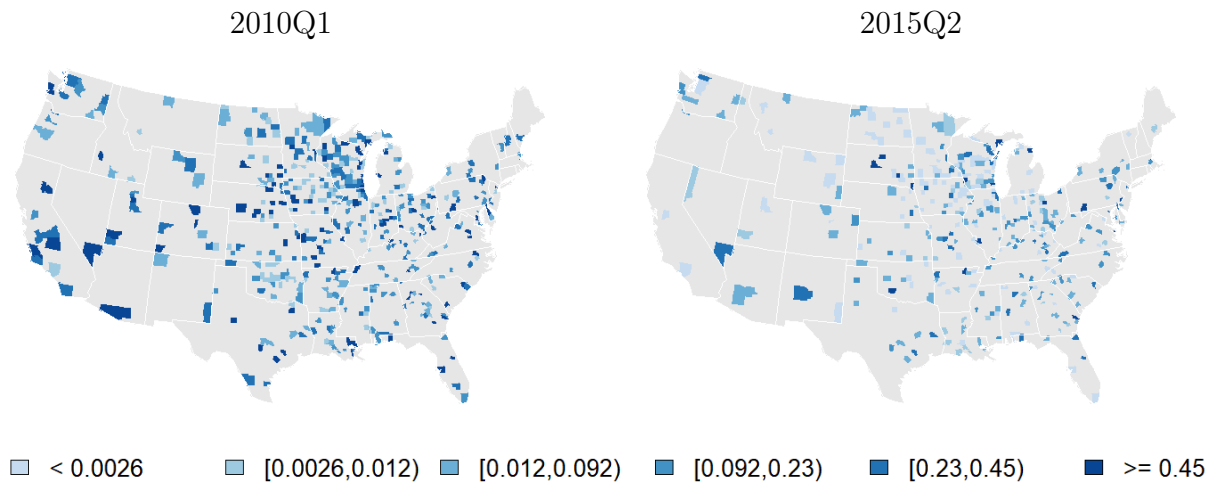
Figure A-4: Log Predictive Density Scores – All Samples



Notes: The panels provide pairwise comparisons of log predictive scores. We also show the 45-degree line. Log probability scores are depicted as differentials relative to pooled Tobit. The blue (red) circle corresponds to RRE (CC). We use $x_{it} = [\Delta \ln \text{HPI}_{it-1}, \Delta \text{UR}_{it-1}]'$.

Figure A-5 resembles Figure 3 in the main text and shows the spatial distribution of CC charge-off forecasts. Averaging across banks in each county contained in our sample we report predictive tail probabilities $\mathbb{P}\{y_{iT+1} \geq 5\% | Y_{1:N,0:T}\}$.

Figure A-5: CC Charge-Off Predictive Tail Probabilities, Spatial Dimension



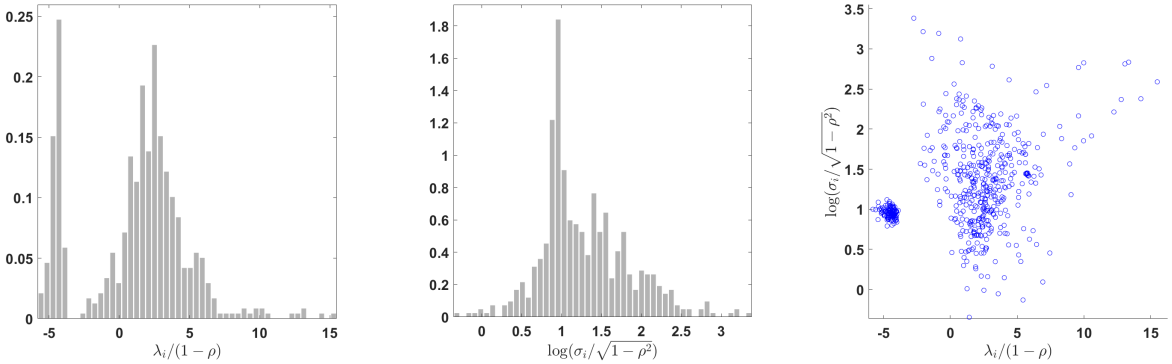
Notes: Predictive tail probabilities are defined as $\mathbb{P}\{y_{iT+1} \geq c | Y_{1:N,0:T}\}$, where $c = 5\%$. Flexible CRE specification with heteroskedasticity. The estimation samples range from 2007Q2 ($t = 0$) to 2009Q4 ($t = T = 10$) and 2012Q3 ($t = 0$) to 2015Q1 ($t = T = 10$).

E.3 Parameter Estimates

Figures A-6 (CC charge-offs) and A-7 (RRE charge-offs) resemble Figures 4 and 6 in the main text.

Figure A-6: Heterogeneous Coefficient Estimates, CC Charge-Offs

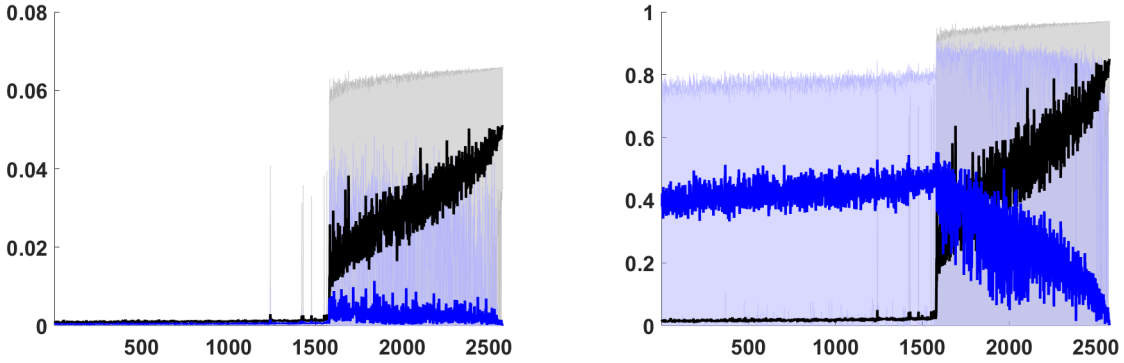
$\mathbb{E}[\lambda_i/(1-\rho)|Y_{1:N,0:T}]$ $\mathbb{E}\big[\ln(\sigma_i/\sqrt{1-\rho^2})|Y_{1:N,0:T}\big]$ Scatter



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

Figure A-7: Effects (Term I and II) of HPI and UR on RRE Charge-Offs

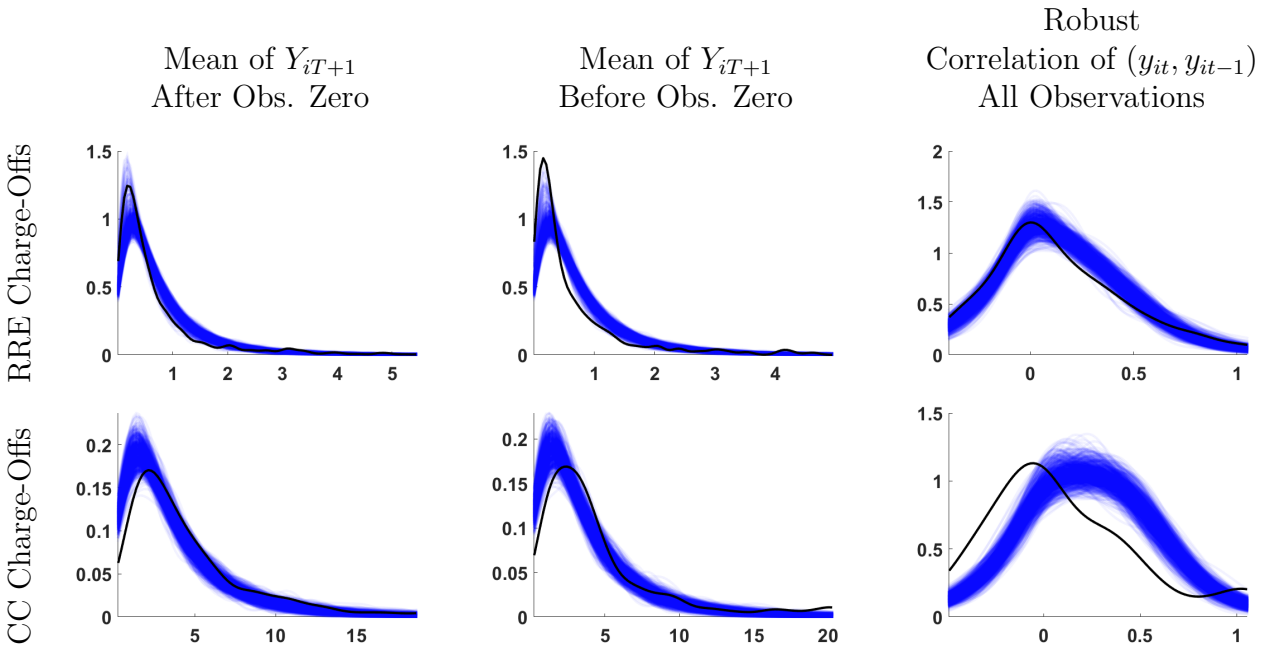
HPI Fall Unemployment Increase



Notes: 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. 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.

E.4 Predictive Checks

Figure A-8: Additional Posterior Predictive Checks: Cross-sectional Distribution of Sample Statistics



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