Supplement to “Forecasting with a panel Tobit model”

(*Quantitative Economics*, Vol. 14, No. 1, January 2023, 117–159)

Laura Liu
Department of Economics, Indiana University

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

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

This Online Appendix consists of the following sections:

Appendix A: Theorem 2.1

A.1 *Proof of Theorem 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:

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

Sampling in a Bayesian framework involves drawing parameters from the appropriate distribution and generating data conditional on these parameters. According to As-

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

Available at http://qeconomics.org. https://doi.org/10.3982/QE1505

sumption (i), the future observations are sampled from the predictive density. This sampling can be implemented as follows: let $\tilde{\vartheta}_N$ be a draw from the posterior $p(\vartheta|Y_{1:N,0:T})$ and sample the future observations from $p(Y_{1:N,T+h}|Y_{1:N,0:T}, \tilde{\vartheta}_N)$.

We start with the bound

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

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

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

Analysis of term $B_1(\cdot)$ Note that $0 \le B_1(\cdot) < 1$. We write

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

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 $b_{i0,N} = 0$. 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

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

Then

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

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 (18) with (A.4) and noting that $0\le F_{ik,N}(\vartheta)\le 1$, we obtain

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

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

According to Assumption (ii), 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

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

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 $\epsilon > 0$,

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

The convergence statement in the last line follows from Assumption (ii) and the dominated convergence theorem:

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

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

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

$$+\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K_i} M_{ik,N}\big(\mathcal{N}_N(\bar{\vartheta}_N)\big)\right)\int_{\mathcal{N}_N(\bar{\vartheta}_N)} \|\bar{\vartheta}_N-\vartheta\|\, p(\vartheta|Y_{1:N,0:T})\,d\vartheta$$
$$\leq \left(\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K_i} M_{ik,N}\big(\mathcal{N}_N(\bar{\vartheta}_N)\big)\right)2\delta_N. \tag{A.6}$$

The last inequality follows from the definition of the neighborhood $\mathcal{N}_N(\bar{\vartheta}_N)$. Using Assumptions (ii) and (iv), 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:

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

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 $z_{it}=[1,y_{it-1}]'$. The distribution of $y_{i1}|y_{i0},\vartheta$ after integrating out λ_i is

$$y_{i1}|(y_{i0},\vartheta)\sim \text{i.i.d.}\, N\big(z_{i1}'\vartheta,2\big), \quad i=1,\dots,N.$$

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

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

Assumption (ii) This leads to the likelihood function

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

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

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

The overall posterior distribution is given by

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

We can define the shrinking neighborhood as the set

$$\mathcal{N}_N(\bar{\vartheta}_N) = \left\{\vartheta \mid (\vartheta - \bar{\vartheta}_N)'\bar{V}_N^{-1}(\vartheta - \bar{\vartheta}_N)' \le 2N^{-\eta}\right\}, \quad 0 < \eta < 1. \tag{A.11}$$

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

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

or

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

The argument can be completed by showing that

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

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

Assumption (iii) 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

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

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

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

and

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

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

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

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 (iii) is satisfied.

Assumption (iv) 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

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

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

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

We now analyze the two indicator functions separately. First,

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

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

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

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

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

This delivers the desired result.

Appendix 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, that is, let $y^{\perp}_{it}$, $t = 1, \dots, T-1$, 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 $M_0 = 1000$ draws.

B.2 *Set forecasts*

To simplify the notation, we drop $X_{1:N,-1:T}$ from the conditioning set in the remainder of this section. 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)

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

Based on posterior draws $(\lambda_i^{(j)}, \sigma_i^{2(j)}, y_{iT}^{*(j)}, \theta^{(j)})$, we can compute the conditional mean and variances $\mu^{(j)}_{iT+h|T}$, and $\sigma^{2(j)}_{iT+h|T}$, 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

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

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 $(\mu_{iT+h|T}, \sigma^2_{iT+h|T})$ under the posterior distribution. We do so using the Monte Carlo averages

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

$$
\pi_i(y) = \frac{1}{M}\sum_{j=1}^{M} p_N\big(y|\mu^{(j)}_{iT+h|T}, \sigma^{2(j)}_{iT+h|T}\big) \tag{A.15}
$$

such that

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

We also define the weights

$$
W^{(j)}_i = 1 - \Phi_N\big(-\mu^{(j)}_{iT+h|T}/\sigma^{(j)}_{iT+h|T}\big), \tag{A.17}
$$

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

Algorithm for $1-\alpha$ *set forecasts targeting pointwise coverage probability*

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

1. For $j = 1, \dots, M$: 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.
2. Evaluate the weights $\{W^{(j)}_i\}_{j=1}^M$ 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 $\{y^{(j)}_{iT+h}\}_{j=1}^M$ 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 $\{(y^{(j)}_{iT+h}, W^{(j)}_i)\}_{j=1}^M$.
   (b) Sort $\{(y^{(j)}_{iT+h}, W^{(j)}_i)\}_{j=1}^M$ in ascending order based on $y^{(j)}_{iT+h}$.
   (c) For $j = 1, \dots, M$: compute $\pi^{(j)}_i = \pi_i(y^{(j)}_{iT+h}) \approx p(y^{(j)}_{iT+h}|Y_{1:N,0:T})$ 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:

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

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

   (b) Evaluate the weights $\{W_i^{(j)}\}_{j=1}^M$ 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

$$\frac{1}{NM} \sum_{i=1}^N \sum_{j=1}^M W_i^{(j)} = 1 - \frac{1}{N} \sum_{i=1}^N \pi_{i0} = 1 - \pi_0.$$

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 $i = 1, \dots, N$:

      i. Draw $\{y_{iT+h}^{(j)}\}_{j=1}^{M}$ 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 $\{(y_{iT+h}^{(j)}, W_i^{(j)})\}_{j=1}^{M}$.

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

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

   (b) Let $\Pi = \{(\pi_i^{(j)}, y_{iT+h}^{(j)}, W_i^{(j)}) \mid i = 1, \dots, N \text{ and } j = 1, \dots, M\}$. Sort the elements in Π based on $\pi_i^{(j)}$ in descending order. Denote the sorted elements in Π by $(\pi^{(s)}, y_{T+h}^{(s)}, W^{(s)})$. 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:

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

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

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

         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 $[a_{ik}, b_{ik}]$, $k = 1, \dots, K_i$. 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

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

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

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

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

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

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

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

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

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

Case 1: $M_* = M$. Then

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

Case 2: $M_* = 0$. Then

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

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

$$\begin{aligned} &\text{CRPS}\left(\hat{F}, y^o\right) \\ &\quad = \sum_{j=2}^{M_*}\left[\hat{F}\left(y^{(j-1)}\right) - 0\right]^2\left(y^{(j)} - y^{(j-1)}\right) + \left[\hat{F}\left(y^{(M_*)}\right) - 0\right]^2\left(y^o - y^{(M_*)}\right) \end{aligned}$$

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

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

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

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

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

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

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

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

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

$$+ y^{(M)} + \big(M_*^2 - (M - M_*)^2\big) y^o \Bigg]$$

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

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

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

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

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

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

Appendix C: Supplemental information on the Monte Carlo

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

Distribution of λ_i *versus* $\mathbb{E}[\lambda_i|Y_{1:N,0:T}]$ The following two examples help to interpret the comparison of the $p(\lambda)$s and the histograms of $\mathbb{E}[\lambda_i|Y_{1:N,0:T}]$ in Figure 1 in the main paper. First, suppose that the model is static, linear, and homoskedastic, that is, $y_{it} = \lambda_i + u_{it}$, $u_{it} \sim N(0, \sigma^2)$ and $\lambda_i \sim N(\phi_\lambda, 1)$, and ϕ_λ is known (which implies $p(\lambda)$ is known). Therefore, the maximum likelihood estimator (MLE) $\hat{\lambda}_i = \lambda_i + \frac{1}{T}\sum_{t=1}^T u_{it}$ has the cross-sectional distribution $\hat{\lambda}_i \sim N(\phi_\lambda, 1 + \sigma^2/T)$ and the posterior means have the distribution

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

In this example, the distribution of the posterior mean estimates is less dispersed than the distribution of the λ_i's, but centered at the same mean, which is qualitatively consistent with Figure 1.

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

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

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

Sensitivity to fraction of zeros in sample 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

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

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

$$\text{Fraction of zeros} = 60\%: \quad p\big(\lambda_i|y_{i0}^*\big) = \frac{1}{9}p_N(\lambda_i|1.85, 0.5) + \frac{8}{9}p_N(\lambda_i|-0.4, 0.5),$$

$$\text{Fraction of zeros} = 75\%: \quad p\big(\lambda_i|y_{i0}^*\big) = \frac{1}{9}p_N(\lambda_i|1.3, 0.5) + \frac{8}{9}p_N(\lambda_i|-0.95, 0.5).$$

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.

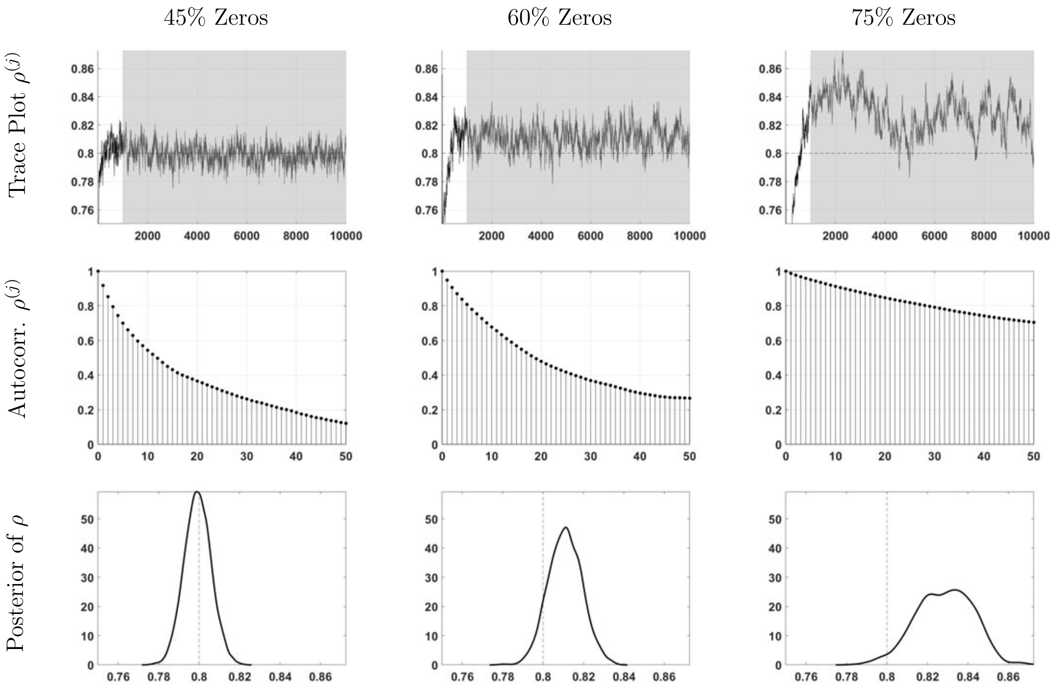

FIGURE A-1. Convergence diagnostics based on $\rho^{(j)}$ sequence. *Note*: The dashed horizontal lines in the first row and the dashed vertical lines in the last row indicate the "true" value of $\rho = 0.8$. (j) in the superscript indicates the MCMC draws. The first 1000 draws are discarded as burn-in, so the shaded regions in the first row indicate the MCMC draws kept for posterior analyses.

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 2500 nonzero observations.[1]

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

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

TABLE A-1. Monte Carlo experiment: forecast performance and parameter estimates.

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

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

units with all-zero observations assign a probability of no less than 95% to $y_{iT+1} = 0$.[2] 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. 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 nonzero 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 chance that draws of λ_i take very large negative values. This contributes to the stability of the MCMC.

[2]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.

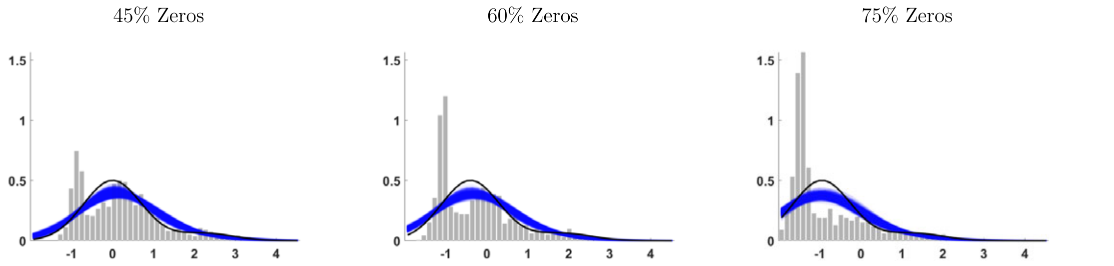

Figure A-2. Posterior means and estimated RE distributions for λ_i, Flexible and Heterosk. Specification. *Note*: The histograms depict $\mathbb{E}[\lambda_i|Y_{1:N,0:T}]$, $i = 1, \ldots, N$. The shaded areas are hairlines obtained by generating draws from the posterior distribution of ξ and plotting the corresponding random effects densities $p(\lambda|\xi)$. The this solid lines represent the true $p(\lambda)$.

Appendix D: Data set

Charge-off rates The raw data are obtained from the website of the *Federal Reserve Bank of Chicago*.[3] 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.[4] (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 and industrial;
- CLD = construction and 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;

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

[4] 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).”

- 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 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 nonmissing 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}, $t = 0, \dots, T+1$, 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	Step 1	Step 2	Step 3	Step 4	Step 5	% 0s	Mean	75%	Max
CLD	2007Q3	7903	7903	7299	3290	3146	1304	77	1.5	0.0	106.8
CLD	2007Q4	7835	7835	7219	3244	3088	1264	74	1.9	0.1	106.8
CLD	2008Q1	7692	7692	7084	3204	3032	1257	71	2.2	0.5	180.2
RRE	2007Q1	7991	7991	7393	6260	5993	2654	77	0.2	0.0	33.1
RRE	2007Q2	7993	7993	7383	6152	5894	2576	76	0.3	0.0	33.1
RRE	2007Q3	7903	7903	7299	6193	5920	2606	73	0.3	0.0	35.9
RRE	2007Q4	7835	7835	7219	6146	5859	2581	70	0.4	0.1	69.2
RRE	2008Q1	7692	7692	7084	6106	5792	2561	68	0.4	0.2	45.6
RRE	2008Q2	7701	7701	7080	6029	5721	2492	67	0.4	0.2	63.6
RRE	2008Q3	7631	7631	7008	6052	5743	2577	65	0.5	0.3	39.2
RRE	2008Q4	7559	7559	6938	6005	5679	2600	63	0.5	0.3	45.6
RRE	2009Q1	7480	7480	6849	5971	5634	2588	62	0.5	0.3	45.0
RRE	2009Q2	8103	8103	7381	5895	5564	2536	62	0.5	0.3	45.0
RRE	2009Q3	8016	8016	7302	5899	5568	2563	61	0.5	0.4	47.6

(*Continues*)

TABLE A-2. *Continued.*

Loan	t_0	Sample Sizes Initial	Step 1	Step 2	Step 3	Step 4	Cross-Sectional Statistics Step 5	% 0s	Mean	75%	Max
RRE	2009Q4	7940	7940	7229	5846	5508	2553	60	0.5	0.4	45.0
RRE	2010Q1	7770	7770	7077	5765	5426	2494	61	0.5	0.4	45.0
RRE	2010Q2	7770	7770	7072	5635	5308	2420	61	0.5	0.4	45.0
RRE	2010Q3	7707	7707	7013	5632	5298	2441	61	0.5	0.4	45.6
RRE	2010Q4	7608	7608	6910	5583	5255	2443	61	0.5	0.3	38.2
RRE	2011Q1	7469	7469	6784	5520	5220	2437	62	0.4	0.3	38.2
RRE	2011Q2	7472	7472	6783	5398	5110	2385	62	0.4	0.3	38.2
RRE	2011Q3	7402	7402	6716	5395	5110	2397	64	0.4	0.2	38.2
RRE	2011Q4	7334	7334	6649	5341	5059	2395	65	0.3	0.2	38.2
RRE	2012Q1	7236	7236	6546	5284	5008	2349	67	0.3	0.2	38.2
RRE	2012Q2	7234	7234	6534	5584	5283	2430	66	0.3	0.2	38.2
RRE	2012Q3	7170	7170	6465	5576	5267	2416	67	0.2	0.1	28.4
RRE	2012Q4	7073	7073	6358	5495	5197	2362	69	0.2	0.1	22.2
RRE	2013Q1	6931	6849	6212	5420	5121	2341	71	0.2	0.1	28.7
RRE	2013Q2	6934	6857	6200	5296	5008	2298	71	0.2	0.1	28.7
RRE	2013Q3	6884	6807	6144	5291	4999	2307	72	0.2	0.0	28.7
RRE	2013Q4	6803	6726	6061	5212	4932	2271	74	0.1	0.0	28.7
RRE	2014Q1	6650	6576	5913	5144	4870	2258	75	0.1	0.0	27.2
RRE	2014Q2	6650	6578	5897	5012	4746	2190	76	0.1	0.0	16.9
RRE	2014Q3	6582	6510	5821	5004	4742	2178	77	0.1	0.0	22.2
RRE	2014Q4	6502	6431	5729	4945	4691	2210	78	0.1	0.0	16.9
RRE	2015Q1	6342	6271	5564	4874	4611	2200	79	0.1	0.0	11.1
RRE	2015Q2	6348	6278	5560	4751	4500	2134	79	0.1	0.0	11.1
CC	2001Q2	9031	9031	8532	1691	1540	875	33	3.4	4.7	162.5
CC	2001Q3	8995	8995	8491	1666	1515	844	33	3.4	4.8	88.9
CC	2001Q4	8887	8887	8382	1636	1489	836	34	3.3	4.6	88.9
CC	2002Q1	8723	8723	8228	1612	1466	814	35	3.3	4.4	400.0
CC	2002Q2	8823	8823	8312	1670	1519	817	38	3.2	4.3	88.9
CC	2002Q3	8805	8805	8286	1631	1488	821	38	3.2	4.3	88.9
CC	2002Q4	8728	8728	8199	1606	1468	813	39	3.1	4.1	88.9
CC	2003Q1	8611	8611	8077	1573	1445	811	40	3.0	4.0	128.5
CC	2003Q2	8754	8754	8203	1544	1422	787	40	3.0	3.9	136.1
CC	2003Q3	8755	8755	8198	1513	1395	754	41	2.9	3.8	136.1
CC	2003Q4	8671	8671	8120	1500	1387	724	42	2.8	3.6	136.1
CC	2004Q1	8526	8526	7989	1468	1355	707	43	2.7	3.6	136.1
CC	2004Q2	8662	8662	8108	1440	1331	677	42	2.8	3.6	136.1
CC	2004Q3	8626	8626	8067	1411	1308	664	43	2.7	3.5	136.1
CC	2004Q4	8552	8552	7989	1391	1284	657	44	2.6	3.3	140.9
CC	2005Q1	8384	8384	7829	1369	1271	639	44	2.5	3.2	151.3
CC	2005Q2	8507	8507	7938	1332	1236	611	44	2.6	3.2	175.0
CC	2005Q3	8482	8482	7897	1315	1218	596	45	2.6	3.2	175.0
CC	2005Q4	8404	8404	7816	1290	1203	604	46	2.6	3.2	210.5
CC	2006Q1	8263	8263	7674	1275	1188	614	47	2.6	3.1	175.0
CC	2006Q2	8307	8307	7708	1247	1164	594	47	2.7	3.2	269.2
CC	2006Q3	8240	8240	7639	1231	1156	594	46	2.8	3.4	269.2
CC	2006Q4	8137	8137	7537	1211	1139	595	45	3.0	3.6	269.2

(*Continues*)

TABLE A-2. *Continued.*

Loan	t_0	Sample Sizes					Cross-Sectional Statistics				
		Initial	Step 1	Step 2	Step 3	Step 4	Step 5	% 0s	Mean	75%	Max
CC	2007Q1	7991	7991	7393	1197	1129	574	44	3.2	3.9	269.2
CC	2007Q2	7993	7993	7383	1173	1107	561	43	3.3	4.1	269.2
CC	2007Q3	7903	7903	7299	1159	1091	544	44	3.2	4.2	175.0
CC	2007Q4	7835	7835	7219	1133	1066	534	43	3.3	4.2	175.0
CC	2008Q1	7692	7692	7084	1123	1056	527	44	3.3	4.2	175.0
CC	2008Q2	7701	7701	7080	1101	1035	512	45	3.2	4.1	158.3
CC	2008Q3	7631	7631	7008	1096	1036	509	44	3.1	4.0	158.3
CC	2008Q4	7559	7559	6938	1082	1020	506	45	3.1	3.9	149.4
CC	2009Q1	7480	7480	6849	1059	999	498	46	3.0	3.7	147.3
CC	2009Q2	8103	8103	7381	1045	989	492	45	2.8	3.7	78.5
CC	2009Q3	8016	8016	7302	1042	988	492	47	2.7	3.5	77.6
CC	2009Q4	7940	7940	7229	1032	978	479	49	2.7	3.3	400.0
CC	2010Q1	7770	7770	7077	1020	963	459	49	2.5	3.2	100.0
CC	2010Q2	7770	7770	7072	997	940	454	50	2.3	3.0	62.0
CC	2010Q3	7707	7707	7013	994	940	450	50	2.2	2.8	62.0
CC	2010Q4	7608	7608	6910	976	920	454	51	2.1	2.6	56.3
CC	2011Q1	7469	7469	6784	961	906	451	52	2.0	2.5	68.6
CC	2011Q2	7472	7472	6783	941	889	450	53	1.9	2.4	67.9
CC	2011Q3	7402	7402	6716	933	879	443	54	1.9	2.3	67.9
CC	2011Q4	7334	7334	6649	920	869	430	55	1.8	2.2	67.9
CC	2012Q1	7236	7236	6546	913	862	438	56	1.7	2.1	67.9
CC	2012Q2	7234	7234	6534	916	862	430	54	1.8	2.2	67.9
CC	2012Q3	7170	7170	6465	907	853	409	55	1.7	2.1	67.9
CON	2009Q2	8103	8103	7381	5837	5698	2600	77	0.4	0.0	77.4
CON	2009Q3	8016	8016	7302	5872	5693	2672	71	0.5	0.2	202.2
CON	2009Q4	7940	7940	7229	5814	5584	2723	65	0.5	0.5	202.2
CON	2010Q1	7770	7770	7077	5735	5461	2680	58	0.7	0.7	202.2
CON	2010Q2	7770	7770	7072	5602	5339	2600	53	0.7	0.8	202.2
CON	2010Q3	7707	7707	7013	5596	5311	2555	47	0.8	0.9	202.2
CON	2010Q4	7608	7608	6910	5545	5227	2473	42	0.9	1.0	202.2
CON	2011Q1	7469	7469	6784	5482	5133	2427	36	1.0	1.1	202.2
CON	2011Q2	7472	7472	6783	5361	5026	2328	37	1.0	1.1	202.2
CON	2011Q3	7402	7402	6716	5377	5028	2333	38	1.0	1.1	202.2
CON	2011Q4	7334	7334	6649	5324	4979	2377	38	0.9	1.0	202.2
CON	2012Q1	7236	7236	6546	5266	4932	2403	39	0.9	1.0	202.2
CON	2012Q2	7234	7234	6534	5544	5195	2530	42	0.8	1.0	76.0
CON	2012Q3	7170	7170	6465	5536	5184	2541	43	0.8	0.9	76.0
CON	2012Q4	7073	7073	6358	5457	5117	2526	43	0.8	0.9	44.7
CON	2013Q1	6931	6849	6212	5379	5042	2548	44	0.8	0.9	100.0
CON	2013Q2	6934	6857	6200	5254	4932	2465	43	0.8	0.9	100.0
CON	2013Q3	6884	6807	6144	5246	4917	2512	44	0.7	0.9	76.0
CON	2013Q4	6803	6726	6061	5165	4843	2470	44	0.7	0.9	76.0
CON	2014Q1	6650	6576	5913	5094	4767	2415	44	0.7	0.9	76.0
CON	2014Q2	6650	6578	5897	4969	4651	2326	44	0.7	0.9	35.7
CON	2014Q3	6582	6510	5821	4954	4638	2297	43	0.7	0.9	35.7
CON	2014Q4	6502	6431	5729	4894	4585	2324	43	0.7	0.9	76.0

(*Continues*)

TABLE A-2. *Continued.*

Loan	t_0	Sample Sizes Initial	Step 1	Step 2	Step 3	Step 4	Step 5	Cross-Sectional Statistics % 0s	Mean	75%	Max
CON	2015Q1	6342	6271	5564	4827	4515	2298	43	0.7	0.9	35.7
CON	2015Q2	6348	6278	5560	4704	4406	2214	43	0.7	0.9	52.9
CON	2015Q3	6271	6204	5479	4689	4402	2209	43	0.7	0.9	52.9
CON	2015Q4	6183	6117	5395	4625	4337	2222	42	0.8	0.9	113.8
CON	2016Q1	6059	5993	5256	4538	4252	2179	43	0.7	0.9	52.9

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

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

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

Housing price index (HPI_{it}) Obtained from the *Federal Housing Finance Agency* on all transactions, not seasonally adjusted. The index is available at a quarterly frequency.

Personal income (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.[5] 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.

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

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);
- 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 stock 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.

TABLE A-3. Summary statistics for bank characteristics, RRE and CC 2007Q2.

	RRE						CC					
	Low		High		All		Low		High		All	
	Mean	StdD	Mean	StdD	Mean	StdD	Mean	StdD	Mean	StdD	Mean	StdD
Log Assets	11.607	0.865	12.427	0.719	12.088	0.880	12.105	0.726	12.501	0.674	12.440	0.696
Loan Fraction	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. on Assets	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*	

Note: 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 (o) and (+) symbols 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.

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

APPENDIX 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.

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-1}. Recall that the regressors are standardized to have mean zero and variance one. We set $x_{it-1} = \tilde{x}_{it-1} = \kappa[1, 1]$ and choose κ such that $\tilde{x}_{it-1}$ 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-1} 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

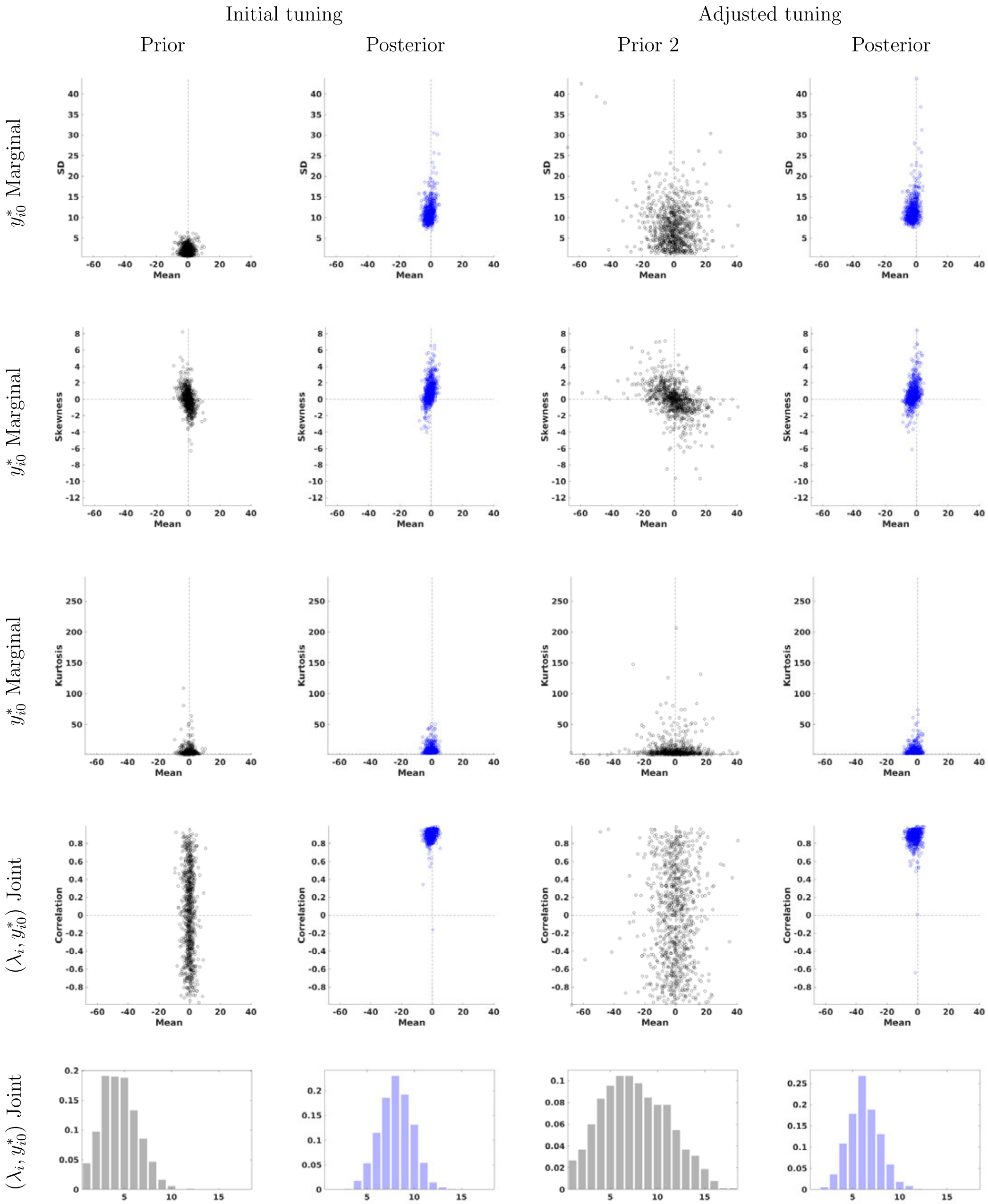

FIGURE A-3. Prior and posterior for CRE distribution, CC. *Note*: The dots in the scatter plots in rows 1 to 4 correspond to draws from the prior or posterior distribution of the hyperparameter ξ that indexes the CRE distribution. For each ξ draw, we compute the implied moments of the mixture of normals 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.

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 10. The highest probability mass is associated with 3 to 5 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.

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

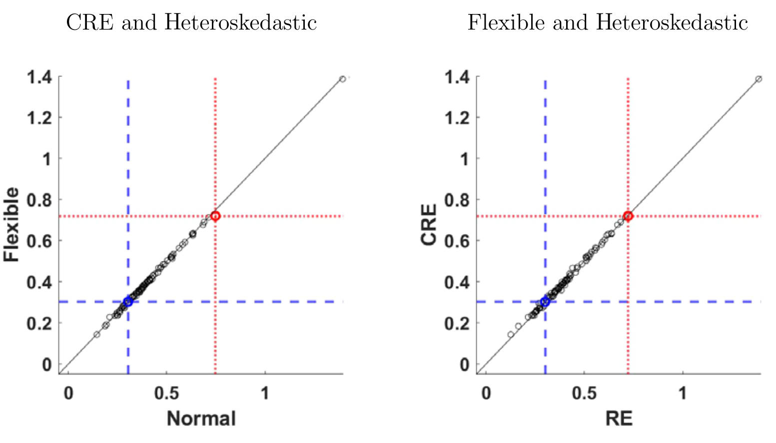

FIGURE A-4. Log predictive density scores—all samples. *Note*: The panels provide pairwise comparisons of log predictive scores. We also show the 45-degree line. Log probability scores are depicted as differentials relative to pooled Tobit. The intersection of the dashed (dotted) lines corresponds to RRE (CC) baseline sample. We use $x_{it} = [\Delta \ln \mathrm{HPI}_{it}, \Delta \mathrm{UR}_{it}]'$.

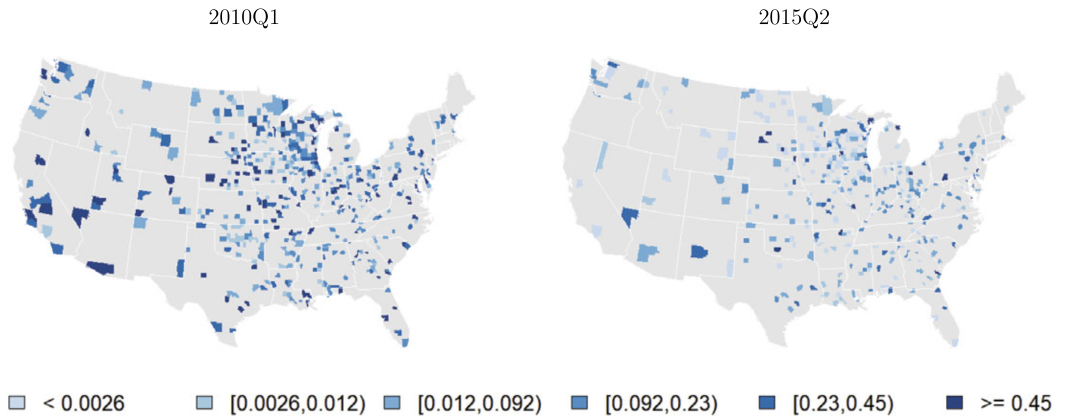

FIGURE A-5. CC charge-off rate predictive tail probabilities, spatial dimension. *Note*: Predictive tail probabilities are defined as $\mathbb{P}\{y_{iT+1} \geq c | Y_{1:N,0:T}, X_{1:N,-1:T}\}$, where $c = 5\%$. Flexible CRE specification with heteroskedasticity. The estimation samples range from 2007Q2 ($t = 0$) to 2009Q4 ($t = T = 10$) and 2012Q3 ($t = 0$) to 2015Q1 ($t = T = 10$).

E.3 *Parameter estimates*

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

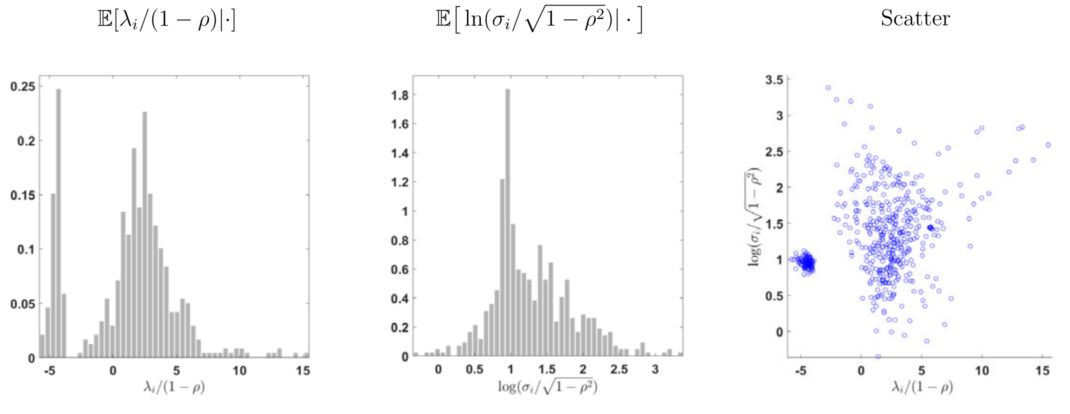

FIGURE A-6. Heterogeneous coefficient estimates, CC charge-off rates. *Note*: Heteroskedastic flexible CRE specification. The estimation sample ranges from 2007Q2 ($t = 0$) to 2009Q4 ($t = T = 10$). A few extreme observations are not visible in the plots. The conditioning set is $(Y_{1:N,0:T}, X_{1:N,-1:T})$.

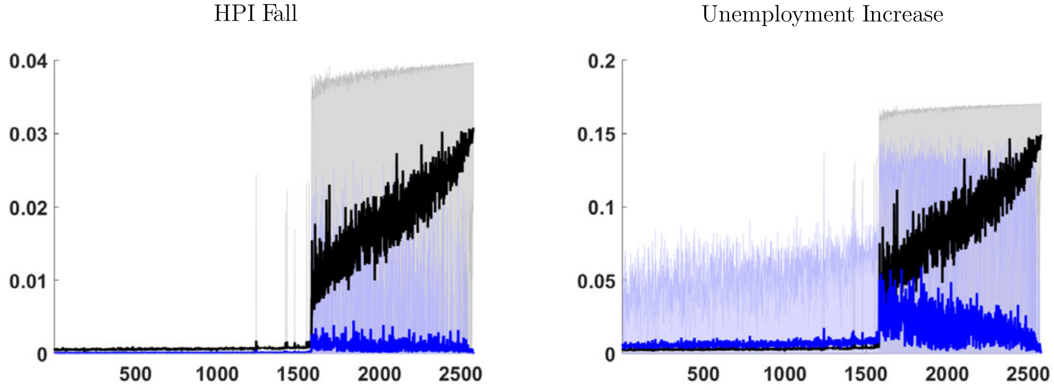

FIGURE A-7. Effects (Terms I and II) of HPI and UR on RRE charge-off rates. *Note*: Heteroskedastic flexible CRE specification. The estimation sample ranges from 2007Q2 ($t = 0$) to 2009Q4 ($t = T = 10$). The banks $i = 1, \ldots, N$ along the x-axis are sorted based on the posterior means $\widehat{\lambda_i/\sigma_i}$. Terms I_i are shown in black/grey and terms II_i in dark/light blue (see online version for colors). The units on the y-axis are in percent. The solid lines indicate the posterior means of the treatment effect components and the shaded areas delimit 90% credible bands.

E.4 *Predictive checks*

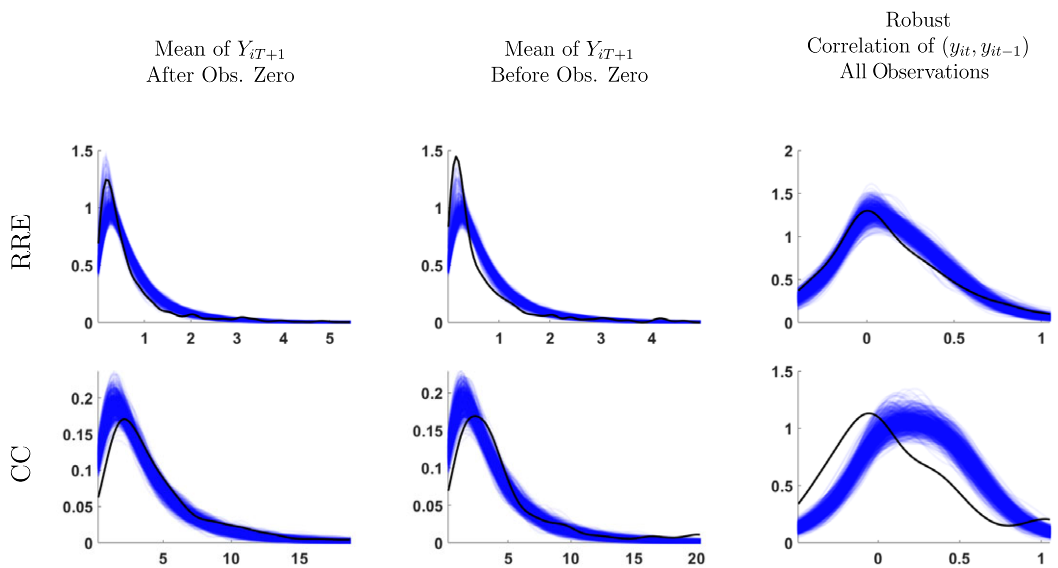

FIGURE A-8. Additional posterior predictive checks: Cross-sectional distribution of sample statistics. *Note*: Heteroskedastic flexible CRE specification. The estimation sample ranges from 2007Q2 ($t = 0$) to 2009Q4 ($t = T = 10$). The thick solid lines are computed from the actual data. Each hairline 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).

References

Arellano, Manuel and Olympia Bover (1995), "Another look at the instrumental variable estimation of error-components models." *Journal of Econometrics*, 68, 29–51. [9]

Chang, Christopher C. and Dimitris N. Politis (2016), "Robust autocorrelation estimation." *Journal of Computational and Graphical Statistics*, 25, 144–166. [29]

Covas, Francisco B., Ben Rump, and Egon Zakrajsek (2014), "Stress-testing U.S. bank holding companies: A dynamic panel quantile regression approach." *International Journal of Forecasting*, 30, 691–713. [19]

Ghosh, Amit (2017), "Sector-specific analysis of non-performing loans in the U.S. banking system and their macroeconomic impact." *Journal of Economics and Business*, 93, 29–45. [24]

Gneiting, Tilmann and Adrian E. Raftery (2007), "Strictly proper scoring rules, prediction, and estimation." *Journal of the American Statistical Association*, 102, 359–378. [14]

Liu, Laura, Hyungsik Roger Moon, and Frank Schorfheide (2018), "Forecasting with dynamic panel data models." NBER Working Paper, 25102. [9]

Co-editor Tao Zha handled this manuscript.

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