Forecasting with a Panel Tobit Model

Laura Liu[1] Hyungsik Roger Moon[2] Frank Schorfheide[3]

[1]Federal Reserve Board
[2]University of Southern California, USC Dornsife INET, and Yonsei
[3]University of Pennsylvania, CEPR, NBER, and PIER

June 2018

Goal of Research

Develop methods to:

- generate forecasts for a large number of cross-sectional units (e.g., firms, banks, households, assets)
- based on relatively short time series (e.g., due to data availability, mergers, regulatory changes, structural breaks).

Example:

$$Y_{it} = \lambda_i + U_{it}, \quad U_{it} \sim N(0,1), \quad t = 1, \dots, T, \quad i = 1, \dots, N.$$

- Forecasting Y_{iT+1} requires estimate $\hat{\lambda}_i$: $\widehat{Y}_{T+1|T} = \hat{\lambda}_i$.
- Naive (but indadmissible ...) estimate: $\hat{\lambda}_i = \frac{1}{T}\sum_{t=1}^{T} Y_{it}$.

How Can We Do Better?

- Suppose we knew that λ_i was drawn from a **prior distribution** $\pi(\cdot)$ ...
- We could construct a **posterior distribution**, using Bayes Theorem:

$$p(\lambda_i|\hat{\lambda}_i) = \frac{p(\hat{\lambda}_i|\lambda_i)\pi(\lambda_i)}{\int p(\hat{\lambda}_i|\lambda_i)\pi(\lambda_i)d\lambda_i}, \quad \hat{\lambda}_i|\lambda_i \sim N(\lambda_i, 1/T), \quad \hat{\lambda}_i = \frac{1}{T}\sum_{t=1}^{T} Y_{it}.$$

- Then minimize **posterior expected prediction loss (risk)**:

$$\hat{Y}_{iT+1} = \text{argmin}_\delta \int\int L(\lambda_i + U_{iT+1}, \delta) p(\lambda_i|\hat{\lambda}) p(U_{iT+1}) dU_{iT+1} d\lambda_i.$$

- This also minimizes integrated risk (averaging over λ_i and $\hat{\lambda}_i$).
- In practice: estimate $\pi(\cdot)$ from the cross-sectional information.

What To Do in Practice?

Option 1 – Full Bayesian analysis:

- Create a model for $\pi(\lambda_i)$, e.g. normal distribution or mixture of normal distributions with parameters ζ.
- Specify prior for ζ and estimate ζ along with the λ_i's:
  - Liu (2017) provides Bayesian implementation in linear model.
  - **Today's talk focuses on Bayesian implementation in dynamic Tobit model.**

Option 2 – Empirical Bayes

- Condition on $\pi(\lambda|\hat{\zeta})$.
- In a linear model for forecasting under quadratic loss one can use Tweedie's formula:

$$\hat{Y}_{iT+1} = \hat{\lambda}_i + \frac{1}{T}\frac{\partial}{\partial\hat{\lambda}_i} \ln p(\hat{\lambda}_i).$$

Only requires an estimate of $p(\hat{\lambda}_i)$. Implementations: Gu and Koenker (2015); Liu, Moon, and Schorfheide (2017).

Application: Modeling Loan Charge-Off Data

- Forecast loan charge-off rates for a panel of "small" banks (< 1b in assets).
- Assume banks operate in local markets and use local economic indicators (unemployment and house price) as additional predictors.
- Model – we also allow for cross-sectional heteroskedasticity:

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

  y_{it} is loan charge off-rate (for a particular type of loans) of bank i in quarter t.
- Implementation:
  - $\pi(\lambda)$: either Normal or Dirichlet process mixture of normals
  - Posterior: use Metropolis-within-Gibbs sampler with data augmentation for y_{it}^*

A Simplified Tobit Model

$$
\begin{aligned}
y_{it} &= y_{it}^* \mathbb{I}\{y_{it}^* \geq 0\}, \quad i = 1, \ldots, N, \quad t = 0, \ldots, T \\
y_{it}^* &= \lambda_i + \rho y_{it-1}^* + u_{it}, \quad u_{it} \overset{iid}{\sim} N(0, \sigma^2), \quad y_{i0}^* \sim N(\mu_{i*}, \sigma_*^2) \\
\lambda_i &\overset{iid}{\sim} \pi(\lambda).
\end{aligned}
$$

- Complications in the Tobit setup:
  - latent variables;
  - identification, esp. finite sample.
- Some Bayesian Tobit references:
  - Chib (1992): static model, use data augmentation for posterior sampling.
  - Wei (1999): dynamic model, extend posterior sampler.
  - Li and Zheng (2008): panel model, semiparametric analysis.

Theory

Benchmark: oracle forecast

- Know $\pi(\lambda)$ and homogeneous parameters, but not λ_i.

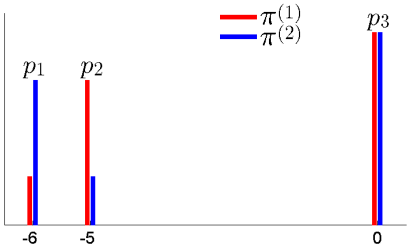
From oracle to feasible forecast

- Most importantly requires an estimate of $\pi(\lambda)$.
- Theoretical properties of feasible forecast (for linear model)
  - Liu (2017), Liu, Moon, and Schorfheide (2017)
- Identification (for Tobit model)
  - Identification in population: Hu and Shiu (2017)
  - Finite sample: left tail in $\pi(\lambda)$
  - $\Longrightarrow$ matters for posterior mean of λ_i, but not matter much for forecasts
- Estimation: parametric vs flexible treatment of $\pi(\cdot)$

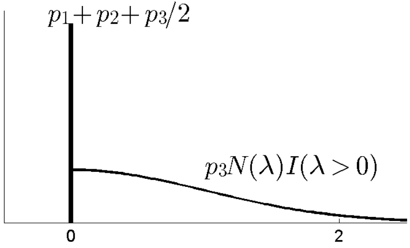
Identification Heuristics

Model: $y_{it} = \max\{0,\ \lambda_i + U_{it}\},\ U_{it} \sim N(0,1),\ T = 1.$

Forecaster j has prior $\pi^{(j)}$



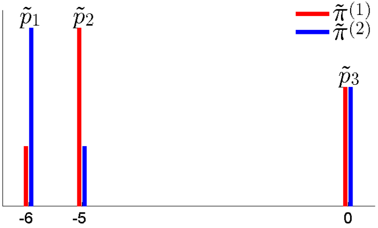
Likelihoods are (almost) identical:



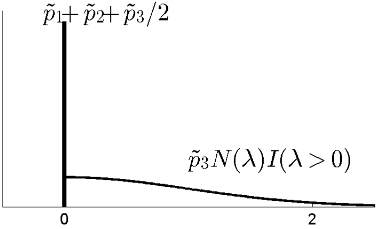
Identification Heuristics

Model: $y_{it} = \max\{0, \lambda_i + U_{it}\}$, $U_{it} \sim N(0,1)$, $T = 1$.

Posteriors after observing $y_{i1} = 0$



Forecasts after observing $y_{i1} = 0$



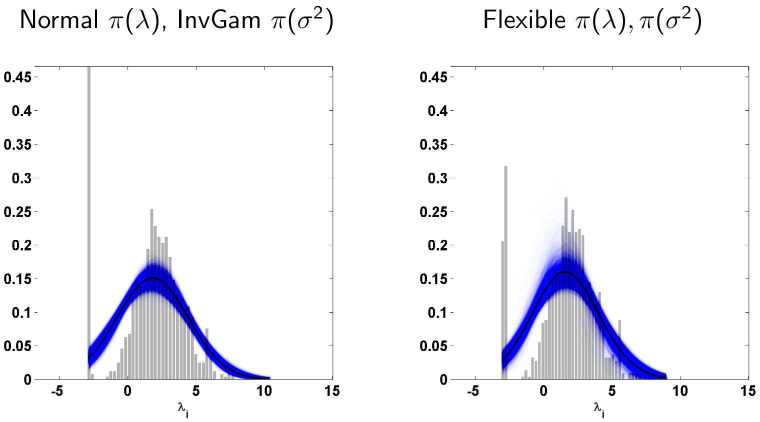
$\Longrightarrow$ Posteriors differ, but density forecasts are (almost) identical.

Application: Bank Loan Charge-Off Rates

$$
\begin{aligned}
y_{it} &= \max\left\{y_{it}^*, 0\right\}, \quad i = 1, \dots, N, \quad t = 0, \dots, T \\
y_{it}^* &= \lambda_i + \rho y_{it-1}^* + \beta_1 \ln \text{HPI}_{it} + \beta_2 \text{UR}_{it} + u_{it}, \\
& u_{it} \overset{iid}{\sim} N(0, \sigma_i^2), \quad y_{i0}^* \sim N(\mu_{i*}, \sigma_{i*}^2),
\end{aligned}
$$

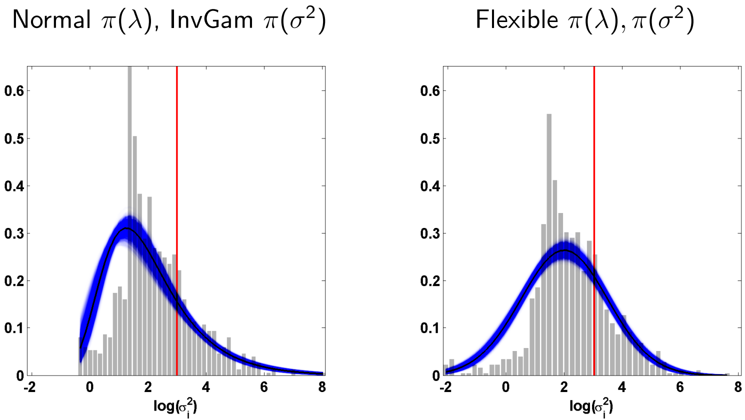
- **Background:**
  - Charge-off rates reflect bank losses.
  - Forecast loan charge-off rates for a panel of "small" ($<$ 1b in assets) banks.
  - Assume banks operate in local markets and use local economic indicators (unemployment and house price) as additional predictors.
  - y_{it}: loan charge off-rate (for a particular type of loans) of bank i in quarter t.
- **Example:**
  - Credit card charge-off rates
  - Sample: N = 875, T = 10 (2001Q2-2003Q4), fraction of 0s = 33%
  - Forecast period: 2004Q1

Posterior Means of λ_i vs. Estimated Random-Effects Distributions



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

Posterior Means of σ^2_{i*} vs. Estimated Random-Effects Distributions



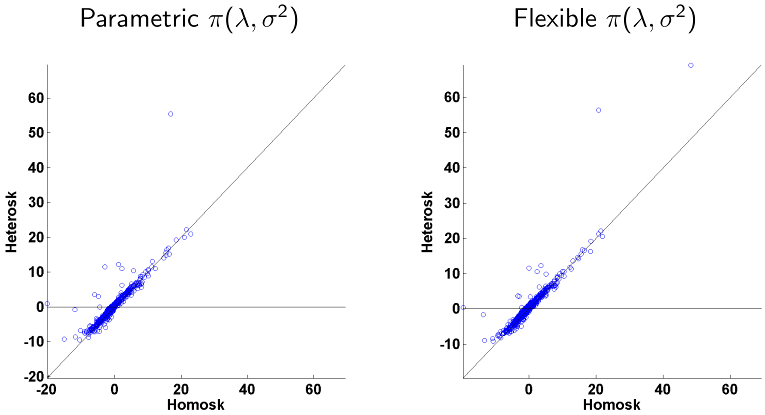
Notes: The panels depict histograms for $\ln \mathbb{E}[\sigma_i^2 | Y_{1:N,0:T}]$, $i = 1, \ldots, N$. The shaded areas are obtained by generating draws from the posterior distribution of the random effects density: $\pi(\sigma_i)|Y_{1:N,0:T}$. The estimation sample for the dynamic Tobit models ranges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

Forecast Performance

Estimator	Point Fcst RMSE	Interval Fcst Cov.Freq	Interval Fcst CI Length	Density Fcst LPS
Homoskedastic Models				
Pooled Tobit	(4.54)	0.92	(8.84)	(-2.28)
Param $\pi(\lambda)$	-7%	0.92	-0.94	0.09
Flex $\pi(\lambda)$	-7%	0.92	-0.95	0.08
Flat $\pi(\lambda)$	-4%	0.92	-0.87	0.06
Heteroskedastic Models				
Param $\pi(\lambda), \pi(\sigma^2)$	5%	0.89	-1.96	0.34
Flex $\pi(\lambda), \pi(\sigma^2)$	5%	0.88	-2.00	0.32

RMSE	percentage change relative to pooled Tobit
CovFreq	nominal coverage frequency is 90%
CI Length, LPS	deviations relative to pooled Tobit

Scatter Plot of Forecast Errors



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

Interval Forecast and Density Forecast Performance

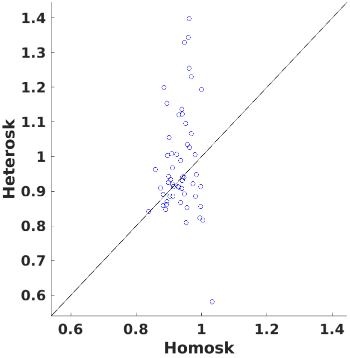
Estimator	Interval Forecast: Cover Freq	Interval Forecast: CI Len	Density Forecast: LPS	Density Forecast: CRPS
Credit Card Charge-Off Rates – Homoskedastic Models				
Normal $\pi(\lambda)$	0.92	7.90	-2.19	1.83
Flexible $\pi(\lambda)$	0.92	7.89	-2.20	1.82
Flat $\pi(\lambda)$	0.92	7.97	-2.22	1.86
Pooled Tobit	0.92	8.84	-2.28	1.98
Pooled OLS	0.95	16.98	-2.96	2.28
Credit Card Charge-Off Rates – Heteroskedastic Models				
Normal $\pi(\lambda)$, InvGam $\pi(\sigma^2)$	0.89	6.88	-1.94	1.71
Flexible $\pi(\lambda), \pi(\sigma^2)$	0.88	6.84	-1.96	1.71

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

Forecast Evaluation – Flexible $\pi(\lambda, \sigma^2)$, Other Samples

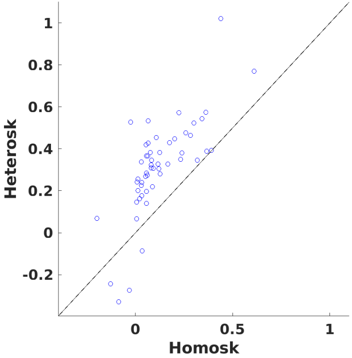
RMSE Ratios

vs. pooled linear



Log Probability Score Diffs

vs. pooled Tobit



Some Intuition

Naive vs. Pooled vs. Bayes:

- Not much cross-sectional heterogeneity: it's good to pool; naive is bad; we estimate a tight prior; imposing homogeneity works well.
- A lot of cross-sectional heterogeneity: it's bad to pool; naive is decent; we estimate a loose prior which does not generate much shrinkage.
- Intermediate cases: neither pooling nor naive is good; Bayes procedure works well.

Homoskedastic vs. Heteroskedastic:

- Estimate of σ_i determines relative weight of weight in likelihood and prior.
- Large estimate of σ_i implies lots of weight on prior which may not be good for point forecasts.

Conclusions

- Forecasting with dynamic panel data models:
  - Important to have "good" estimates of the individual effects λ_i.
  - Estimate cross-sectional distribution of λ_i.
  - Then use it as prior for Bayesian inference to sharpen inference and increase forecast accuracy.
- Complications in the Tobit setup:
  - latent variables,
  - identification, especially in finite samples.
- Bank loan charge-off rates application:
  - Bayes procedure works generally well.
  - Point forecasts: homoskedastic models
  - Interval and density forecasts: heteroskedastic models
- Future work:
  - Missing at random observations, correlated random effects...